2025 • 29 • 2

Online ISSN 2500-3259

Сетевое издание

ВАВИЛОВСКИЙ ЖУРНАЛ ГЕНЕТИКИ И СЕЛЕКЦИИ

VAVILOV JOURNAL OF GENETICS AND BREEDING

Основан в 1997 г.
Периодичность 8 выпусков в год
doi 10.18699/vjgb-25-20

2025 • 29 • 2

Online ISSN 2500-3259

Online edition

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

VAVILOV JOURNAL OF GENETICS AND BREEDING

Founded in 1997
8 issues a year
doi 10.18699/vjgb-25-20

Founders

Siberian Branch of the Russian Academy of Sciences

Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences

The Vavilov Society of Geneticists and Breeders

СОДЕРЖАНИЕ • 2025 • 29 • 2

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII • VAVILOV JOURNAL OF GENETICS AND BREEDING

CONTENTS • 2025 • 29 • 2

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):175-188

doi 10.18699/vjgb-25-21

3D cell culture models: how to obtain and characterize the main models

M.M. Abdurakhmanova [1], A.A. Leonteva [2, 1], N.S. Vasilieva [2, 1], E.V. Kuligina [1], A.A. Nushtaeva [2, 1]

[1] Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Sirius University of Science and Technology, Sirius Federal Territory, Krasnodar Region, Russia
nushtaeva.anna@gmail.com

Abstract. For many years, the gold standard in the study of malignant tumors has been the *in vitro* culture of tumor cells, *in vivo* xenografts or genetically modified animal models. Meanwhile, three-dimensional cell models (3D cultures) have been added to the arsenal of modern biomedical research. 3D cultures reproduce tissue-specific features of tissue topology. This makes them relevant tissue models in terms of cell differentiation, metabolism and the development of drug resistance. Such models are already being used by many research groups for both basic and translational research, and may substantially reduce the number of animal studies, for example in the field of oncological research. In the current literature, 3D cultures are classified according to the technique of their formation (with or without a scaffold), cultivation conditions (static or dynamic), as well as their cellular organization and function. In terms of cellular organization, 3D cultures are divided into "spheroid models", "organoids", "organs-on-a-chip" and "microtissues". Each of these models has its own unique features, which should be taken into account when using a particular model in an experiment. The simplest 3D cultures are spheroid models which are floating spherical cell aggregates. An organoid is a more complex 3D model, in which a self-organizing 3D structure is formed from stem cells (SCs) capable of self-renewal and differentiation within the model. Organ-on-a-chip models are chips of microfluidic systems that simulate dynamic physical and biological processes found in organs and tissues *in vitro*. By combining different cell types into a single structure, spheroids and organoids can act as a basis for the formation of a microtissue – a hybrid 3D model imitating a specific tissue phenotype and containing tissue-specific extracellular matrix (ECM) components. This review presents a brief history of 3D cell culture. It describes the main characteristics and perspectives of the use of "spheroid models", "organoids", "organ-on-a-chip" models and "microtissues" in immune oncology research of solid tumors.

Key words: cell aggregation; 3D cell cultures; spheroids; organoids; organ-on-a-chip; microtissue; 3D cell model culturing

For citation: Abdurakhmanova M.M., Leonteva A.A., Vasilieva N.S., Kuligina E.V., Nushtaeva A.A. 3D cell culture models: how to obtain and characterize the main models. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):175-188. doi 10.18699/vjgb-25-21

Funding. The sections **Introduction, Features of 3D tumor cell cultures: "spheroid model", "organoid", "organ-on-a-chip" and "microtissue"** and **Conclusion** (authors: A.A. Leonteva, N.S. Vasilieva, A.A. Nushtaeva) were written using funds from the project implemented within the framework of the state program of the federal territory "Sirius" "Scientific and technological development of the federal territory 'Sirius'" (Agreement No. 27-03 dated September 27, 2024); writing of the sections **Preservation of tissue-specific characteristics of cells *in vitro*** and **Cultivation of cells in 3D models** (authors M.M. Abdurakhmanova and E.V. Kuligina) as well as **Licensed access to Biorender** for preparing drawings were carried out at the expense of the grant of the Russian Science Foundation No. 24-14-00390, http://rscf.ru/project/24-14-00390/.

Acknowledgements. The authors would like to acknowledge Vladimir Richter (head of laboratory). All figures were created with BioRender.com.

Трехмерные модели культур клеток: способы получения и характеристика основных моделей

М.М. Абдурахманова [1], А.А. Леонтьева [2, 1], Н.С. Васильева [2, 1], Е.В. Кулигина [1], А.А. Нуштаева [2, 1]

[1] Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Научно-технологический университет «Сириус», федеральная территория «Сириус», Краснодарский край, Россия
nushtaeva.anna@gmail.com

Аннотация. В течение многих лет золотым стандартом в исследованиях злокачественных новообразований являлись культуры опухолевых клеток *in vitro*, ксенотрансплантаты *in vivo* или генетически модифицированные модели животных. К настоящему времени арсенал инструментов современных медико-биологических исследований пополнился трехмерными клеточными моделями (3D-культуры). 3D-культуры воспроизводят тканеспецифичные характеристики топологии ткани, что делает их релевантными тканевыми моделями с точки зрения клеточной дифференцировки, метаболизма и развития лекарственной устойчивости. Благодаря своему потенциалу такие модели уже применяются многими исследовательскими группами как для фундаментальных, так и для трансляционных исследований, и их использование позволяет значительно сократить количество экспериментов на животных, например, в области онкологии. В литературе 3D-культуры классифицируют по технике формирования (с каркасом/без каркаса), условиям культивирования (статические/динамические), а также по клеточной организации и функциям. По клеточной организации 3D-культуры разделяют на «сфероидные модели», «органоиды», «органы-на-чипе» и «микроткани». При этом каждая из моделей имеет свои характерные особенности, которые необходимо учитывать при использовании модели в эксперименте. Наиболее простые 3D-культуры – это «сфероидные модели», представляющие собой плавающие сферические агрегаты клеток. Более сложной 3D-моделью является «органоид» – самоорганизующаяся трехмерная структура, сформированная из стволовых клеток, способных к самообновлению и дифференцировке в составе модели. Микрофлюидные системы «орган-на-чипе» – это чипы, имитирующие *in vitro* основные физические и биологические процессы в органах и тканях в динамике. «Сфероиды» и «органоиды» за счет объединения различных типов клеток в единую структуру могут быть основой для формирования «микроткани» – гибридной 3D-модели, воспроизводящей специфический тканевый фенотип и содержащей тканеспецифичные компоненты внеклеточного матрикса. В данном обзоре представлена краткая история развития метода культивирования клеток *in vitro* в 3D-формате, описаны основные характеристики и перспективы применения «сфероидных моделей», «органоидов», «органов-на-чипе» и «микротканей» для исследований в области иммуноонкологии солидных опухолей.

Ключевые слова: агрегация клеток; 3D-культуры клеток; сфероиды; органоиды; орган-на-чипе; микроткань; культивирование клеточных 3D-моделей

Introduction

In the middle of the 20th century, the basic principles of *in vitro* cultivation of plant and animal cells were formed and diploid human cell lines were created (Jedrzejczak-Silicka, 2017). In the late 20th and early 21st century, 3D cell culturing methods were developed to construct cell models that more accurately reproduce the microenvironment in which cells reside in body tissues (Edmondson et al., 2014). 3D tumor cell culture techniques have been actively developing in recent decades. Compared to 2D cultures, modern 3D cell models are as close as possible to animal models and *in vivo* primary tumors in terms of the following characteristics: the apical-basal polarity of cells within the 3D model; expression level of cell genes responsible for physiological functioning of cells; heterogeneity of cellular composition; ability to secrete extracellular matrix (ECM) proteins and growth factors; drug resistance of the model and etc.

Researchers classify 3D cell cultures according to their spatial structure (Maliszewska-Olejniczak et al., 2019) and distinguish “spheroidal models”, “organoids”, “organ-on-a-chip” models and “microtissues”. In published works, the terms “spheroid”, “organoid” and “microtissue” may be mistakenly used as synonyms (Simian, Bissell, 2017). However, it should be kept in mind that all of the above models have different or only partially overlapping cell sources, construction protocols and applications and as such are not interchangeable. The reasons why the terms “spheroid model”, “organoid” and “microtissue” need to be separated are described in this review. The review also presents a brief history of the development of *in vitro* 3D cell culturing methods with a focus on the key features of 3D cellular models, which will allow researchers to determine the most physiologically relevant model for cancer immunology studies of solid tumors.

Preservation of tissue-specific characterization of cells *in vitro*

The first attempts to obtain a 3D cell model were made in 1956: Aron Arthur Moscona obtained 3D structures in the form of cell aggregates (Moscona, 1956). Moscona was the first to show that dissociated cells of different histological origin, when cultured together, are able to aggregate with each other and form a three-dimensional structure.

Radiobiologists Robert Sutherland et al. first introduced the term “spheroid” for the structures described by Aron Moscona. Sutherland and colleagues obtained multicellular spheroids from Chinese hamster lung cells (line V79). The structure of the resulting spherical cell aggregates resembled the nodules observed in animal and human carcinomas. The growth curve of cell aggregates *in vitro* was similar to the growth curve of grafts in mice. Morphological analysis of the obtained structures showed that spheroids have an outer zone containing many dividing cells, an intermediate zone, which is poorly saturated with oxygen and nutrients and contains a small number of cells in the state of mitosis, and a zone of necrotized cells. Based on the results obtained, the authors concluded that the multicellular spheroids obtained during the experiment can be used as an *in vitro* model to assess tumor growth (Sutherland, 1988).

The term “organoid” began to be used in the literature in the 1950s, but, at that time, the structures denoted by the term had nothing to do with “3D cell cultures”. For

example, William Duryee and Josephine Doherty, in their 1954 study "Nuclear and Cytoplasmic Organoids in the Living Cell", used the term "organoid" to refer to intracellular structures, namely cell organelles (Duryee, Doherty, 1954). The term "organoid" was also used to refer to tumors or abnormal cellular growths as a synonym for "teratoma" (Wolter, 1967). The development of methods for culturing organoids as 3D cellular structures dates back to 1975. James G. Reinwald and Howard Green described the first 3D model that contained normal human keratinocytes and mouse fibroblasts of the 3T3 line. In the stratified epidermis, cell division was restricted to the basal layer of growing clones, while the superficial layers consisted of terminally differentiating keratinocytes that gradually formed the keratinizing layer. Further culturing of these structures yielded "epidermal sheets" grown from small numbers of primary keratinocytes (Rheinwatd and Green, 1975). Although the term "organoid" was not used in this study, Rheinwatd and Green were the first to reconstruct a 3D tissue structure *in vitro*, and since 1980, the term "organoid" has appeared in studies on 3D cultures.

In addition, in the 1980s, the work of a group led by Mina Jahan Bissel demonstrated the important role of ECM in tumor development. Primary culture mouse mammary gland cells were cultured on a substrate of basal membrane (BM) proteins derived from Engelbreth-Holm-Swarm (EHS) mouse sarcoma. It was shown that in this conditions mammary cells formed ducts and lumen resembling secretory alveoli, and β-casein expression was detected in 90 % of the cells (Li et al., 1987). This study stimulated the development of methods to create 3D models with the consideration of the ECM. The combination of the words "3D cell culture models" was first used by Mary Helen Barcellos-Hoff et al. (Barcellos-Hoff et al., 1989) and Ole Petersen and colleagues (Petersen et al., 1992) when analyzing mammary gland cells on EHS BM substrate. Using this human mammary gland model, the group led by Barcellos-Hoff investigated alveolar morphogenesis, and the group led by Petersen was able to describe the growth pattern and differentiation of normal and malignant epithelial cells.

Until 2005, the term "organoid" was used to refer to small organ fragments consisting mainly of epithelial cells separated mechanically and/or enzymatically from stromal tissue and grown in various gels (Fata et al., 2007). However, in the last decade, the term has often been used to refer to a wider variety of 3D structures (Nikonorova et al., 2023). In 2012, The Gastrointestinal Stem Cell Consortium approved the following nomenclature for cell models of the large and small intestine: "organoid" – a 3D culture consisting of several cell types, such as cells of epithelial and mesenchymal origin; "spheroid" – a spherical 3D culture containing cells of only one cell type (Guryanov, 2016).

To clarify the nomenclature of cellular models for other tissues, the European Molecular Biology Organization organized the "Organoids" meeting in October 2016, where it was decided to apply the term "organoid" to a range of different structures, depending on the organ system (Simian, Bissell, 2017). For example, in the field of mammary gland biology, an "organoid" is a primary explant of epithelial ducts placed in ECM gels. Conversely, in intestine biology research, "organoids" may include clonal derivatives of primary epithelial stem cells (SCs) grown without mesenchyme or epithelial-mesenchymal cultures derived from embryonic stem cells (ESCs) or induced pluripotent stem cells (iPSCs) (Shamir, Ewald, 2014).

Thus, the methods of tissue fragment cultivation developed and described in the 19th and 20th centuries laid the foundation for the development of cell culture technology outside the body. The formulated principles of cell cultivation allowed to make important discoveries in the field of regenerative medicine, transplantology, biotechnology and biopharmaceutics (Simian, Bissell, 2017) (Fig. 1).

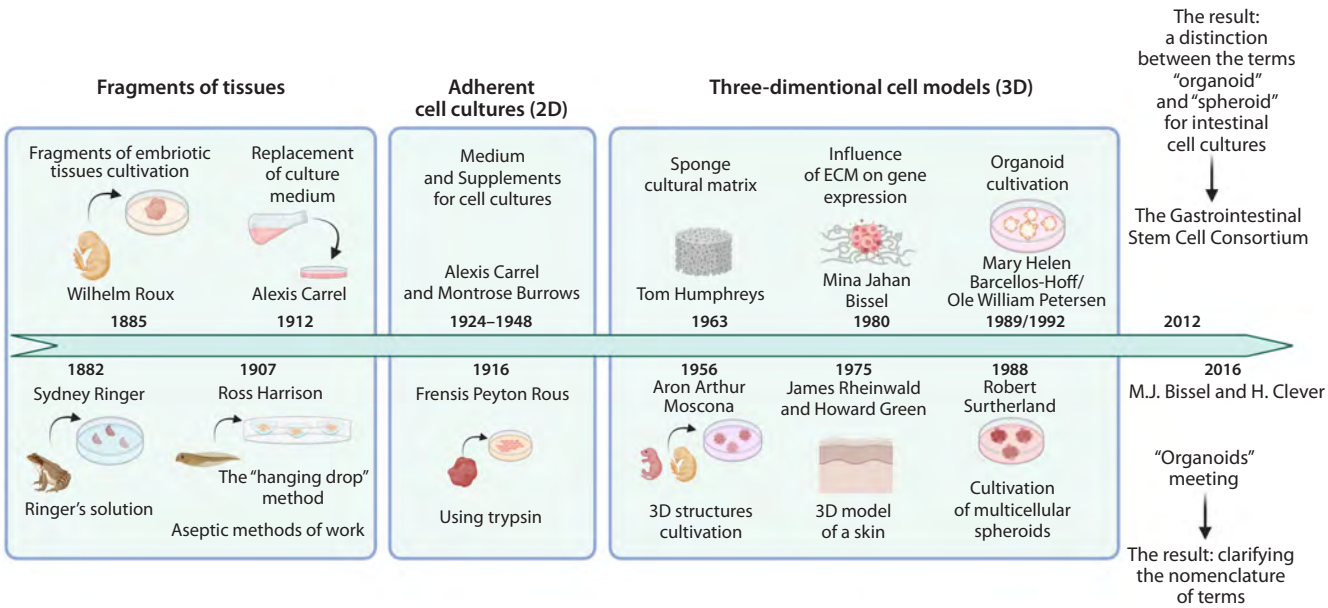


Fig. 1. Chronology of key developments in cell culturing: from tissue fragments to 2D and 3D cell models.

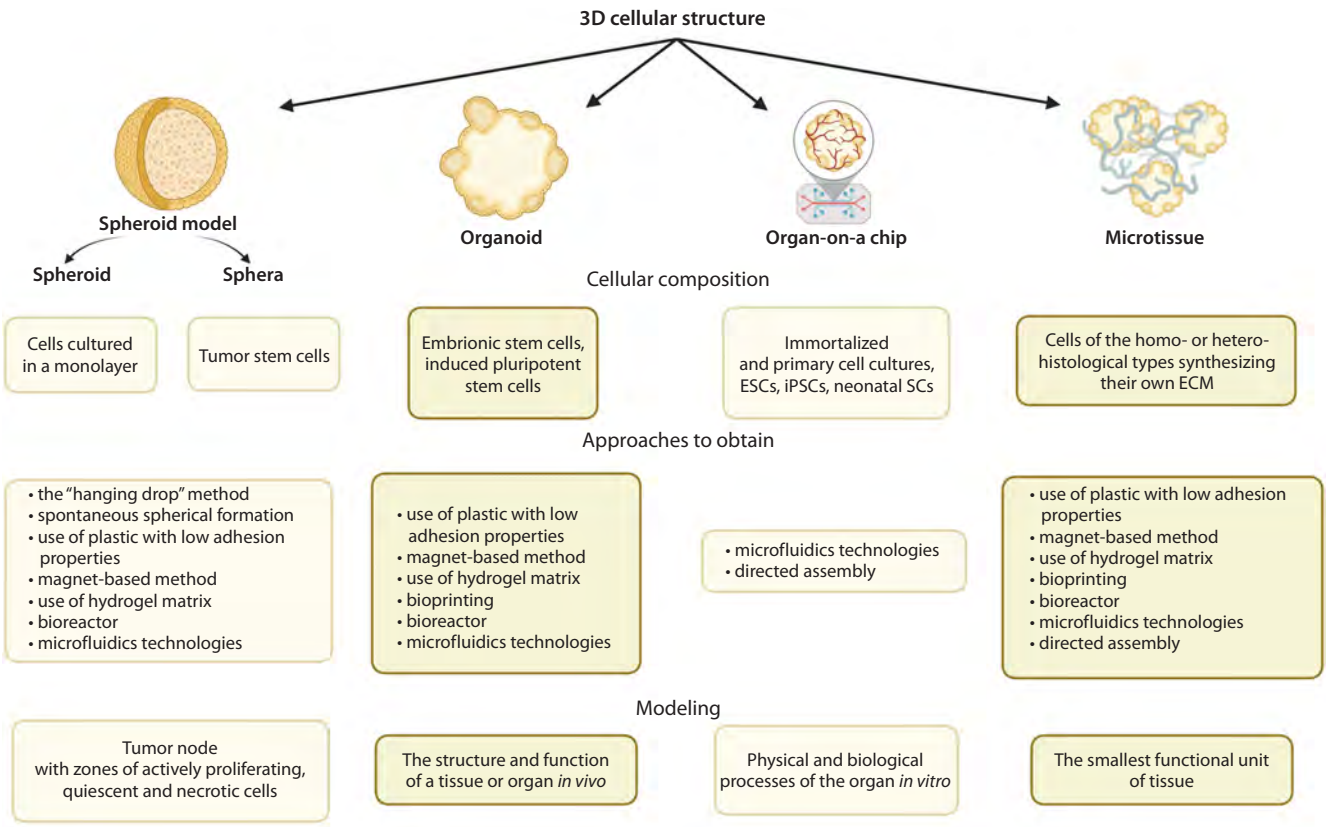


Fig. 2. Methods of obtaining and characteristic features of 3D cellular structures: "spheroid model", "organoid", "organ-on-a-chip" and "microtissue".

The specific features of 3D tumor cell cultures: "spheroid model", "organoid", "organ-on-a-chip" and "microtissue"

With the development of 3D culturing approaches, terms such as "aggregates", "spheroids", "sphere", "tumorsphere", "oncosphere", "organoid" or "organotypic spheroid" appear. They are often mistakenly used as synonyms. However, these models differ in the composition of the medium used, the cell culture surface, the cell density, the time required for formation, and the types of cells used (Rodrigues et al., 2024). That said, the ambiguity of the terminology can lead to confusion about the specific model used in a given study (Nikonorova et al., 2023). For example, Seyed Ali Karimi-fard et al. use the terms "organoid" and "mammosphere" in reference to a 3D cellular structure from MCF-7 breast adenocarcinoma tumor cells (Karimifard et al., 2024). According to the nomenclature of cellular 3D structures, "organoid" and "mammosphere" refer to different 3D models (Ponti et al., 2005; Gilazieva et al., 2020). The authors of this study refer to the publication by Sahar Moradi-Mehr et al. who describe engineered "mammospheres" as an organoid model (Moradi-Mehr et al., 2023). However, the authors of this work do not describe the model they obtained as an "organoid", but use the terms "3D MCF-7 cell culture" or "mammosphere".

We assume that the confusion in terminology is related to the novelty and speed of development of the field of 3D cell culture, as well as the desire to follow scientific trends. The importance of using appropriate terminological nomenclature was also discussed in a scientific review by V.G. Nikonorova et al. (Nikonorova et al., 2023). Despite numerous attempts to introduce nomenclature, the use of terminology is rather inconsistent among researchers; therefore, it is necessary to introduce nomenclature of cell models in the scientific community, including among Russian researchers (Kang et al., 2021; Pașca et al., 2022) (Fig. 2).

Spheroid models

Among spheroidal models, "spheres" and "spheroids" are the most common (Maliszewska-Olejniczak et al., 2019). "Spheres" include tumorospheres and tissue tumor "spheres". Tumorospheres are described as tumor cells forming 3D clusters of cell suspension growing under non-adhesive conditions. Tumor stem cells (CSC), which are associated with tumor initiation, have the potential for self-renewal and proliferation, as well as the ability to form 3D structures when cultured *in vitro* (Weiswald et al., 2015). Since sphere-forming cells are SCs, they are able to differentiate into all non-stem cell subpopulations present in the original cell culture, and thus, a tumorosphere is a mixture of CSC and differentiated cells (Maliszewska-Olejniczak et al., 2019). By contrast, tissue tumor "spheres" are derived from a patient's tumor tissue sample. The tissue sample is dissociated, allowing tumor cells to migrate from

fragments as clusters of cells and/or individual cells to form dense, compact clusters or aggregates of cells. However, this spheroid model is limited to the study of the CSC region, as it cannot reproduce the multiplicity of other cell types in a tumor, and is also poorly reproducible as some CSC remain undifferentiated (Valent et al., 2012).

"Spheroids" are aggregates of cells of a spherical shape formed in a suspension of single cells of homo- or heterogeneous cell type. The formation of such a model occurs due to homotypic intercellular adhesion, complemented by the lack of cell adhesion to the plastic of the culture vial (Sakalem et al., 2021). Such a 3D model can be formed from cells of the same lineage as well as from cells of different lineages cultured together, and allows us to assess the ability of cells to spontaneously self-organize, synthesize ECM proteins and form a specific microenvironment (Verjans et al., 2018). The spheroid resembles a non-vascularized tumor nodule – it mimics the central zone of hypoxia, the inner zone of quiescent cells and the outer zone of actively proliferating cells and is convenient as a model in the study of malignant neoplasms.

The main application area of spheroid models: in biological research as an *in vitro* tumor model, for drug testing, as a basis for tissue engineering (Daly et al., 2021; Hsu et al., 2021; Corgnac et al., 2022; George et al., 2022; Nushtaeva et al., 2022; Vasileva et al., 2022).

Advantages and disadvantages of the spheroidal model

One of the advantages of spheroid models is that they do not require an exogenous ECM (Nushtaeva et al., 2022). Such models reproduce the biochemical reactions of the original parental tumor (George et al., 2022) and intercellular interactions (Corgnac et al., 2022). In addition, spheroid models can be used as building blocks for organ-on-a-chip models and microtissues (Corgnac et al., 2022).

However, it is important to consider that depending on the method of derivation, the duration of cultivation and the size of the spheroid, the necrotic area may also increase, limiting researchers, for example in studies related to drug testing (Verjans et al., 2018). Also, not all cell lines are able to form spheroid models (Ivascu, Kubbies, 2007) and there is limited availability of cell lines derived from normal or minimally transformed tissues (Gunti et al., 2021; Han et al., 2021). In addition, a detailed selection of growth factors is required for the formation and maintenance of the spheroid model.

Prospects for the application of "spheroid models" in immunologic and cancer research

Over the last decade, immunotherapy has become a promising tool in oncotherapy (Bandara et al., 2024). Despite this, the efficacy of immunotherapy often depends on tumor histogenesis and patient characteristics. This suggests the need for improved preclinical screening models that more accurately reproduce tumor biology *in vivo*.

Spheroid models can be grown either from tumor cells alone or co-cultured with different cell types such as fibroblasts, endothelial cells, and immune cells to mimic crosstalk between different cellular compartments of patients' tumors (Abdurakhmanova et al., 2022; Heinrich et al., 2024). Although spheroids lack the vasculature and cellular heterogeneity of the primary tumor, their gene expression profiles and necrotic core formation make them similar to patients' tumors (Heinrich et al., 2024). It is currently the most used model to evaluate immunotherapeutic strategies due to its relatively low cost and high reproducibility (Boucherit et al., 2020).

Spheroid models can be used to test immunotherapy approaches, particularly to assess the efficacy of therapeutic antibodies and carry out drug screening to enhance immune cell infiltration and antitumor effects against solid tumors. For example, in the study by Melanie Grotz and colleagues, a heterotypic spheroid model of breast cancer was used to evaluate the effect of a high-affinity ligand of fibroblast activation protein on naive T-cell behavior (Grotz et al., 2024). This study showed that targeting the fibroblast activation protein is relevant for immunotherapy and effective activation of T cells in the tumor microenvironment. Spheroid models can also be used to test the efficacy of the chimeric antigen receptor (CAR) therapy approach. Veronica Bandara et al. tested their CAR-T cells targeting the non-functional purinergic receptor P2X7 and found that this approach enhanced the anti-tumor response in a spheroid model of ovarian cancer (Bandara et al., 2024). Spheroid models can also be used to investigate the role and functions of nanoscale biomolecules. In the study by Lilita Sadovska and colleagues, a 3D cellular model was developed to evaluate the effects of extracellular vesicles (EVs) in prostate cancer on human immune cells (Sadovska et al., 2018). The study showed that the majority of EVs remain bound on the surface of B cells, while a part of EVs penetrate into T cells via macropinocytosis.

In addition to generating spheroids derived from tumor cells, another approach is to develop spheroids derived from immune cells. Macrophages form spheroids and can remain viable in 3D culture for at least 16 days (Burchett et al., 2024). Y. Tanaka et al. were able to demonstrate that macrophages tend to polarize towards the anti-tumor M1 phenotype, opposing its pro-tumor M2 phenotype in the spheroid state (Tanaka et al., 2018).

However, in order to accurately mimic tumor composition and investigate the functional properties of immune cells, it is necessary to improve existing spheroid models. For example, by introducing new cell types into the spheroid in a quantitatively accurate manner. In addition, the cell ratios in the model must match what the tumor exhibits. This requires extensive study of the cellular composition of the tumor before creating the model. The most comprehensive heterotypic spheroid model was created in a study by Marcel Heinrich et al. (Heinrich et al., 2024). The authors of this study determined the number and ratio of glioblastoma tumor cells, microglia, and astrocytes to recreate a realistic brain tumor model. The inclusion of both astrocytes and

microglia in the heterotypic model significantly increased the growth of the model, and demonstrated that astrocytes play a crucial role in glioblastoma cell invasion. In addition, astrocytes and microglia contribute to a dense physical barrier that protects the tumor model from infiltration by macromolecules or immune cells.

Organoids

A significant part of 3D cell cultures is called "organoids" because, under the conditions of mimicking the 3D environment of an organism *in vitro*, cells can spontaneously self-organize, forming complex histological structures similar to the structures in the organs from which they originated. For example, mammary gland cells cultured in 3D are able to form structures similar to branched ducts (Lee et al., 2007). Currently, the term "organoid" refers to an artificial 3D structure derived from SCs and composed of organ-specific cells capable of self-organization and reflecting the structure and function of an organ *in vivo*. Such a model can be derived from ESCs, iPSCs or neonatal SCs (Sakalem et al., 2021; George et al., 2022) and provides relevant insights into tissue functionality and differentiation. Typically, "organoids" are composed of different cell types originating from different germ sheets and tend to have a higher order of self-organization compared to spheroids (Nikonorova et al., 2023).

When describing "organoids", the term "assembloids" is also used – uniting organoids formed from cells of different organs or different regions of an organ (Eke et al., 2022). Such a model should mimic the morphofunctional units of the corresponding tissues *in vivo*.

The main application area of organoids: biomedical research, drug testing, tissue engineering and transplantation therapy (Kassis et al., 2019; Hofer, Lutolf, 2021; Mesci et al., 2022; Miao et al., 2022).

Advantages and disadvantages of organoids

By altering the cell isolation procedure and varying the combination of growth factors during culturing, researchers can create organoids composed of both normal and transformed cells (Ivascu, Kubbies, 2007; Daly et al., 2021; Hsu et al., 2021; Corgnac et al., 2022), which is a powerful tool in antitumor drug screening studies. Cellular models of "organoids" can be cultured for long periods of time, genetically modified and cryopreserved, preserving their phenotypic and functional characteristics. However, it should be taken into account that the formation of a complex structure in the "organoid" model usually takes two to three months depending on the tissue type and requires a certain set of growth factors (Gunti et al., 2021).

Prospects for the application of "organoids" in immunologic and cancer research

The use of patient-derived organoids in personalized cancer immunotherapy has shown great potential. Such organoids retain the genetic and functional characteristics of the original tumors, allowing immunotherapeutic strategies to be tailored to each patient's unique cancer profile (Noorintan et al., 2024).

A study by S.D. Forsythe et al. used personalized organoid models to preclinically investigate the use of immunotherapy in the treatment of appendix cancer (Forsythe et al., 2021). Patient tumor organoids were generated using unsorted tumor cells with and without enrichment of patient immune cells derived from peripheral blood, the spleen, or lymph nodes for therapy with PD-1 (programmed cell death protein 1) inhibitors and T-cell activators. The authors demonstrated cytotoxic efficacy in a subset of immune-enhanced appendix cancer organoids from both low and high malignancy primary tumors. This study demonstrates the potential of immunotherapy for appendix cancer and the utility of immunocompetent organoids in selecting patients for clinical trials in rare cancers.

Incorporation of 3D models to predict clinical responses to screening drugs turned out to be more effective than use of traditional adherent cultures, as 3D models reproduce the features of the primary tumor to a greater extent. Z. Zhou et al. developed a standardized protocol to establish a tumor-organoid-T-cell system with breast tumor organoids and primary tumor-specific CD8+ T cells. This system facilitates high-throughput drug screening using mouse mammary tumor organoids and also allows for more accurate prediction of therapeutic responses to anticancer drugs using personalized organoids (Zhou et al., 2021). The authors showed that current epigenetic inhibitors enhance antigen presentation mediated by major histocompatibility complex class I (MHC I) on breast tumor cells. Furthermore, treatment with the histone deacetylase inhibitor BML-210 significantly sensitized breast tumor cells to the PD-1 inhibitor.

Developing co-culture systems for primary tumor epithelium that include additional cellular components without artificial addition is challenging. J.T. Neal et al. successfully created organoids derived from patient tumor epithelium that retain their own immune cells, reflecting the diversity of the tumor microenvironment (Neal et al., 2018). Populations of infiltrating CD3+ T cells expressing PD-1, cytotoxic T cells, T helper cells, T cells, B cells, NK cells and varying numbers of macrophages were observed in the personalized organoids. This method holds great promise for modeling personalized immunotherapy *in vitro* by organoids that retain their immune structure.

T.E. Schnalzger et al. developed organoids from patient-derived colon cells to study the cytotoxicity of CAR-NK cells targeting the EpCAM (cell adhesion molecule) antigen (Schnalzger et al., 2019). CAR-NK-EpCAM effectively lysed tumor cells on the first day of co-culture. The authors claim that the organoids they obtained represent a sensitive, personalized *in vitro* platform for evaluating the efficacy of CAR-based immunotherapy.

However, no matter how sophisticated organoid models are, they do not provide a physiological representation of tissue organization *in vivo*. In these models, there is no vascular system, and consequently, the diffusion of drugs, cellular products and their penetration inside the organoid is limited.

Organ-on-a-chip

Organ-on-a-chip technology has revolutionized biomedical research by providing advanced platforms for *in vitro* modeling of complex organ systems. "Organ-on-a-chip" is a technology for culturing cells in a fluid flow to mimic an artificial organ or their system, allowing the structural and functional characteristics of organs and their interactions to be reproduced. This technology is applicable to the study of disease mechanisms, responses of body systems to therapeutic agents and their toxicity profiles (Doost, Srivastava, 2024).

The organ-on-a-chip model is a small microfluidic device in the form of chips made of biocompatible materials that, through a network of microchambers, microchannels, and laminar flow, allow cells to be cultured under conditions similar to *in vivo* environments (Doost, Srivastava, 2024). Such a model can be derived from ESCs, iPSCs or neonatal SCs, as well as immortalized and primary cell cultures (Singh et al., 2022). In addition, microfluidic technologies can be combined with a "spheroid model" and/or "organoids" to form a hybrid model (Wei et al., 2023).

The main application area of organ-on-a-chip: biomedical research, drug testing, tissue engineering (Azizgolshani et al., 2021; Lohasz et al., 2021).

Advantages and disadvantages of the organ-on-a-chip model

"Organ-on-a-chip" allows full control of microfluidic systems and regulation of cellular processes in a study, mimicking dynamic human physiological processes such as respiration, peristalsis, and blood flow (Alver et al., 2024).

One of the limitations of organ-on-chip technology is the need for a material that does not affect the components of the cellular microenvironment and maintains a stable fluidic connection. Since the volume of laminar fluid is small, surface effects dominate over volume effects. In addition, laminar flow is present at the intersection of multiple fluids, and consequently the fluids may not mix properly (Danku et al., 2022).

Prospects for organ-on-a-chip application in immunologic and cancer research

Blood and lymphatic vessels play an important role in immunologic processes, moving immune cells between organs, tissues, and the lymphatic system. Microfluidic chip technology can replicate key complex and dynamic tumor characteristics such as vascularization and extravasation, improving preclinical models in the development of cancer immunotherapy (Doost, Srivastava, 2024). Most organ-on-a-chip models contain parallel channels to incorporate tumor cells into hydrogels and immune cells embedded in the hydrogel or perfused from the side channel. The specific choice of microfluidic model design is usually determined by the purpose of investigation, as throughput, dynamic characteristics (e. g., flow), and molecular sensing capabilities vary widely between models (Chernyavska et al., 2023).

Shabnam Jeibouei et al. used spheroids formed from breast cancer cells in a microfluidic chip to assess patient tumor heterogeneity and analyze migration and invasive potential (Jeibouei et al., 2024). The authors found that increased expression levels of HER2 and the macrophage marker M2a as well as the stiffness of VSMC proteins are important factors affecting tumor cell migration and invasion. M. Nguyen and colleagues reconstructed a heterotypic HER2+ breast tumor model to evaluate the effect of monoclonal antibodies. The authors cultured tumor cells, endothelial cells, blood mononuclear cells, and tumor-associated fibroblasts in a multichamber chip. This model allowed testing of monoclonal antibodies in a complex 3D system that allows perfusion of soluble molecules given the heterogeneity of the tumor (Nguyen M. et al., 2018).

Unlike adaptive immune cells, innate immune cells do not need MHC for their activation. The complexity increases significantly when adaptive immune cells have to be used in an experiment, given MHC molecules, in the presence of other MHC-mismatched cell types (Magenau et al., 2016). It is therefore crucial to develop immunocompetent organ-on-a-chip models to help us better understand how immune cells interact with organs in health and disease. Research by Irina Veith and colleagues created personalized organ-on-a-chip models of lung cancer with their autologous primary tumor, stromal, and immune cells isolated from tumor samples and measured the response to anti-PD-1 treatment (Veith et al., 2024). The microfluidic model was able to reproduce stroma-dependent mechanisms of resistance to immunotherapy, and integration of autologous immunosuppressive tumor-associated fibroblasts into the model impaired the response to anti-PD-1 therapy.

Although organ-on-a-chip models can reproduce most characteristics of individual organs and physiological flow conditions, it is unable to capture dynamic interactions between multiple organs (Kumar et al., 2024). In addition, an organ-on-a-chip still does not include all organ-specific cells and requires further refinement of the model, for example via integration of organoids into the model. Tengku Maulana created a model for infusion, recruitment and infiltration of CAR-T cells into solid tumors by integrating organ-on-a-chip approaches and patient-derived organoids. The model was used to investigate different treatment regimens with dasatinib as a pharmacologic safety switch to control CAR-T cells during therapy. The approach allowed *in vitro* evaluation of safety and efficacy in a patient-specific manner (Maulana et al., 2024).

Microtissue

A "microtissue" is a hybrid cellular 3D model that has a tissue-specific phenotype and contains tissue-specific ECM components. "Microtissues" are formed when cells in a suspension aggregate with each other and/or bind to the surrounding ECM and compactify, increasing the density of the 3D structure (Eyckmans, Chen, 2017). It is possible to form a "microtissue" by obtaining a model of "spheroids"

or "organoids" from both a single cell type and histologically different cell types (Eke et al., 2022), as well as by integrating into an organ-on-a-chip model. In this approach, "microtissues" can be spherical multicellular aggregates designed to replicate the smallest functional unit of a tissue or organ. During self-organization, cells synthesize their own ECM, re-establish cellular contacts, and thus reproduce tissue-specific functions and integrated cellular responses to environmental stimuli. Although the microtissue forms an environment that allows certain cell types to mimic their native *in vivo* behavior as closely as possible, many tissues in the body experience significant mechanical loading that alters matrix structure and cell function, which is difficult to reproduce in a 3D model (Eyckmans, Chen, 2017).

The main application area of microtissues: biomedical research, drug testing, tissue engineering and transplantation therapy (Wang Y. et al., 2020; Zhang et al., 2022).

Advantages and disadvantages of microtissue

Microtissues allow recreating complex native tissue architecture *in vivo*, including simulation of vascular network, cell-cell and cell-ECM interactions (Eke et al., 2022). Pathological processes are being modeled using microtissue for personalized screening and drug development. However, the low assembly speed for macroscale tissue simulation, building a scenario of cellular evolution in 3D dimension leading to the emergence of function rather than the formation of the final functional structure should be considered. In addition, the sources of initial cells can affect model fidelity and reproducibility (Eke et al., 2022; Schot et al., 2023; Wang O. et al., 2023).

Prospects for the application of "microtissues" in immunologic and cancer research

A microtissue is an *in vitro* biomimetic model formed from spheroids and/or organoids as biological building blocks for tissue and organ development, both through simple 3D culturing approaches and innovative engineering systems (Burdis et al., 2022). The advantage of a microtissue model is that the tissue organization can be fully engineered and the assembly of the model can be adjusted chemically or mechanically to obtain the desired tissue structure.

Claudia Martins and colleagues developed a spheroid-based heterotypic glioblastoma microtissue model to evaluate the effect of nanodrugs (Martins et al., 2023). The resulting model mimicked tumor organization, extracellular matrix production, and exhibited a cytokine signature. Macrophages within the microtissue were polarized into an M1/M2 phenotype consistent with docetaxel nanotherapy. In the study by Kazuaki Ninomiya and Tatsuhiko Taniuchi, a bio-3D printer with spheroid stacking on Kensan (microneedle matrix) was used and a microtissue was assembled by precisely stacking spheroids from normal and cancer cells. The resulting model allowed to non-invasively observe the dynamic invasion behavior of cancer cells for the first time (Ninomiya, Taniuchi, 2024). Inya Waldhauer et al. developed heterotypic 3D microtissue models to study the activity of novel IL-2-based anti-tumor immunotherapeutic drugs (Waldhauer et al., 2013). The resulting tumor cell/fibroblast/lymphocyte-based microtissue model allows us to control the penetration of antibodies and their targeting of tumor and stroma components, to study the interaction of tumor cells with immune cells in a system that more closely resembles the tumor microenvironment *in vivo*. Using bioprinting and microfluidic emulsification systems, Gyusik Hong and colleagues obtained a microtissue spheroid model with a lobular structure and realization of liver functions (Hong et al., 2021). Structured microtissue spheroids with pronounced vascularization showed improved albumin and urea secretion.

Thus, the use of the microtissue approach involves the combination of already existing 3D models to enhance the reproduction of realistic tissue features in the field of tumor immunology, and remains a promising model in the development of immunotherapy strategies.

Cell culturing in 3D models

Cultivation conditions in 3D systems should provide cells with all physical and chemical conditions necessary to mimic the *in vivo* environment. At present, there are many methods for culturing cells as part of 3D structures (Fig. 3). The following criteria should be considered when selecting a method for obtaining a 3D cell structure:

1) cell composition: a mono- (Troitskaya et al., 2021) or heterogeneous cell model (Arora et al., 2022; Nushtaeva et al., 2022);
2) method of 3D model formation: using special carrier matrices (Sulaiman et al., 2020) or without their use (Nushtaeva et al., 2022);
3) cultivation conditions: static (Arora et al., 2022) or dynamic (Coluccio et al., 2019).

Some advantages and disadvantages of methods for obtaining basic 3D models are summarized in the Table.

Conclusion

Compared to cells in adherent cultures, cells in 3D structures simulate intercellular interactions organized in space and cellular heterogeneity, which together more fully reflect tissue organization *in vivo* (Eke et al., 2022). This review discusses the nuances of terminology in 3D cell modeling, the main approaches to obtaining models, and the prospects for their use in biomedical research.

Three-dimensional "spheroid models" and "organoids" provide an opportunity to approximate the architecture and functionality of the tissue from which they originate. However, despite the advantages of these models to account for part of the microenvironment, such as stromal and immune cells, they still lack the environmental dynamics inherent to *in vivo* conditions. Organ-on-a-chip microfluidic technologies in the field of oncology combine the advantages of 3D culture in a controlled and dynamic environment. In addition, "spheroids" and "organoids" act as building blocks

Advantages and disadvantages of methods of cultivation of basic 3D models

Approach	3D model	Essence of the method	Advantages	Disadvantages	References
The "hanging drop" method	Spheroid model, microtissue	A drop of cell suspension is placed on the lid of the culture plate, the lid is inverted, causing cells to accumulate at the air-liquid interface and form a 3D structure	• The ability to work with a small number of cells without the use of expensive reagents • Obtaining a large number of 3D cultures • Model size control is possible	• The volume of the drop is limited by the need to preserve surface tension • Heterogeneity in the size of the resulting spheroids • Not suitable for long-term cultivation • Expensive when using specialized plates	Higgins et al., 2010; Nguyen O. et al., 2021
Spontaneous spheroid formation	Spheroid model	Spontaneous spheroid formation of stem-like cells when cells are cultured in 2D format	• Obtaining a 3D model without special equipment, materials and growth factors • Selective cultivation of SCs • Inexpensive method	• Heterogeneity of 3D model sizes • Lack of possibility to obtain single 3D models • No possibility to control the model size	Chen et al., 2021; Troitskaya et al., 2021
Using plastic with low adhesion properties	Spheroid model, organoid, microtissue	Forced aggregation of cells into a 3D model when cultured in plates with the bottom of the wells coated with biopolymers that prevent cell adhesion to the plastic surface	• It is possible to obtain single models • Co-culture of different cell types is possible • Model size control is possible • Generally inexpensive method	• Expensive when using specialized plates • No possibility to control the uniformity of the model	Jeong et al., 2020; Chen et al., 2021
Magnet-based methods	Spheroid model, organoid, microtissue	The cell monolayer is incubated with a suspension of magnetic nanoparticles. Cell aggregation with further formation of a 3D model occurs under the influence of magnetic force	• Rapid cell aggregation • Model size control is possible • Different cell types can be co-cultured	• Heterogeneity of cell aggregates in shape and size • Expensive method	Caleffi et al., 2021; Gaitán-Salvatella et al., 2021
Using a hydrogel matrix	Spheroid model, organoid, microtissue	The hydrogel is used as a substrate to prevent cells from adhering to the surface, or the cells are mixed with the hydrogel	• Non-toxicity of the substrate • Ease of manipulation • Possibility of long-term cultivation • Model size control is possible	• Heterogeneous composition and size of 3D models • Not suitable for cells with high invasive potential • Low stability and possible immunogenicity of the hydrogel matrix	Ravi et al., 2016; Badea et al., 2019
Bioprinting	Microtissue	The spatial organization of cells, imitating the architecture of a tissue or organ, is formed via layer-by-layer application of the material used for bioprinting. Cell printing methods: extrusion, inkjet, laser, pressurized bioprinting	• The process can be automated • Model size control is possible	• Expensive and technologically complex method	Sun et al., 2021; Eke et al., 2022
Bioreactor	Spheroid model, organoid	The cell suspension, placed in a special chamber, is subjected to continuous agitation to prevent cell adhesion to the surface. Inside the bioreactor, there is a constant circulation of nutrients and removal of cell metabolic products	• Obtaining a large number of 3D models at the industrial level	• Expensive method • No possibility to control the homogeneity of the model • Vessel rotation speed may affect physiological responses of cells • No possibility to control model size	Di Buduo et al., 2021; Khan et al., 2021

Table (end)

Approach	3D model	Essence of the method	Advantages	Disadvantages	References
Microfluidics technology	Spheroid model, organoid, organ-on-a-chip, microtissue	A chip with channels in which a constant laminar flow is maintained and transport is carried out by diffusion	• Use of a minimum number of cells and reagents • Model size control is possible • Fast model formation due to constant perfusion • Cells are minimally exposed to hypoxia due to the oxygen-permeable materials and growth factors used in the chip	• Difficulty in collecting cells for analysis • Expensive equipment is required	Bircsak et al., 2021; Nair et al., 2021
Directional assembly	Organ-on-a-chip, microtissue	Formation of the model into the desired structure occurs through chemical bonding, physical interactions, or biological adhesion between cells in spheroids or organoids	• Control of the composition and size of the model is possible • Suitable for matrix-rich tissues (bone, cartilage)	• Low reproducibility • Difficulty in reconstructing the complete tissue architecture	Kim et al., 2018; Eke et al., 2022

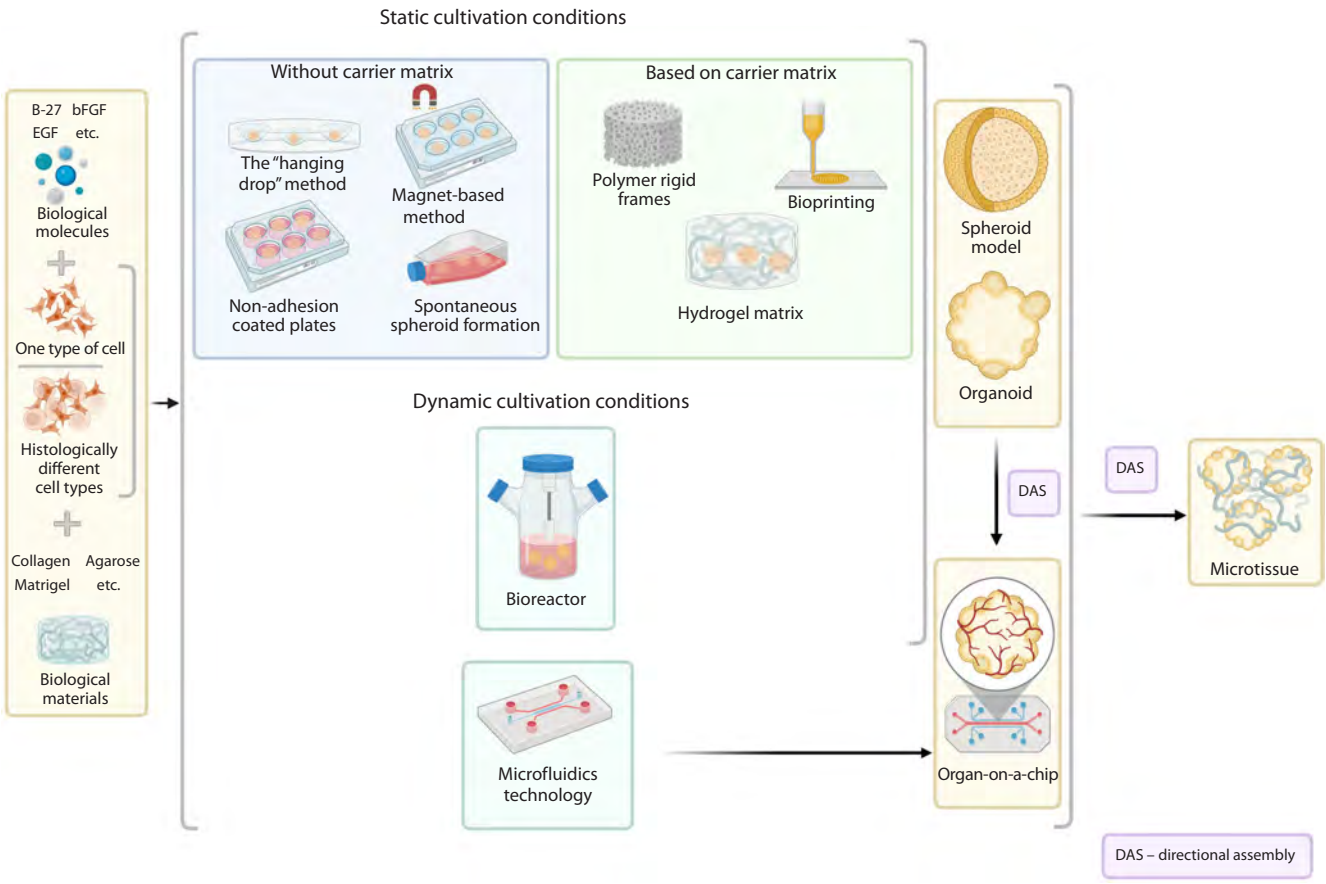


Fig. 3. Methods of obtaining 3D structures.

and form a "microtissue" that recreates the complexities of native tissue architecture *in vivo* (Eke et al., 2022).

Three-dimensional cellular models are an informative tool for investigating mechanisms of disease development and progression, as well as identifying novel biomarkers, since they are as close as possible to the primary tumor at the cellular and molecular genetic level. In addition, such models are a relevant preclinical *in vitro* platform for drug development and realization of the potential of personalized medicine.

References

Abdurakhmanova M.M., Ermakov M.S., Richter V.A., Koval O.A., Nushtaeva A.A. The optimization of methods for the establishment of heterogeneous three-dimensional cellular models of breast cancer. *Genes and Cells*. 2022;17(4):91-103. doi 10.23868/gc425244 (in Russian)

Alver C.G., Drabbe E., Agarwal A., Ishahak M. Roadblocks confronting widespread dissemination and deployment of Organs on Chips. *Nat Commun*. 2024;15:5118. doi 10.1038/s41467-024-48864-3

Arora L., Kalia M., Dasgupta S., Singh N., Verma A.K., Pal D. Development of a multicellular 3D tumor model to study cellular heterogeneity and plasticity in NSCLC tumor microenvironment. *Front Oncol*. 2022;12:881207. doi 10.3389/fonc.2022.881207

Azizgolshani H., Coppeta J.R., Vedula E.M., Marr E.E., Cain B.P., Luu R.J., Lech M.P., Kann S.H., Mulhern T.J., Tandon V., Tan K., Haroutunian N.J., Keegan P., Rogers M., Gard A.L., Baldwin K.B., de Souza J.C., Hoefler B.C., Bale S.S., Kratchman L.B., Zorn A., Patterson A., Kim E.S., Petrie T.A., Wiellette E.L., Williams C., Isenberg B.C., Charest J.L. High-throughput organ-on-chip platform with integrated programmable fluid flow and real-time sensing for complex tissue models in drug development workflows. *Lab Chip*. 2021;21(8):1454-1474. doi 10.1039/d1lc00067e

Badea M.A., Balas M., Hermenean A., Ciceu A., Herman H., Ionita D., Dinischiotu A. Influence of matrigel on single- and multiple-spheroid cultures in breast cancer research. *SLAS Discov.* 2019;24(5):563-578. doi 10.1177/2472555219834698

Bandara V., Niktaras V.M., Willett V.J., Chapman H., Lokman N.A., Macpherson A.M., Napoli S., Gundsambuu B., Foeng J., Sadlon T.J., Coombs J., McColl S.R., Barry S.C., Oehler M.K., Ricciardelli C. Engineered CAR-T cells targeting the non-functional P2X purinoceptor 7 (P2X7) receptor as a novel treatment for ovarian cancer. *Clin Transl Immunol*. 2024;13(5):e1512. doi 10.1002/cti2.1512

Barcellos-Hoff M.H., Aggeler J., Ram T.G., Bissell M.J. Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane. *Development*. 1989;105(2):223-235. doi 10.1242/dev.105.2.223

Bircsak K.M., DeBiasio R., Miedel M., Alsebahi A., Reddinger R., Saleh A., Shun T., Vernetti L.A., Gough A. A 3D microfluidic liver model for high throughput compound toxicity screening in the OrganoPlate®. *Toxicology*. 2021;450:152667. doi 10.1016/j.tox.2020.152667

Boucherit N., Gorvel L., Olive D. 3D tumor models and their use for the testing of immunotherapies. *Front Immunol.* 2020;11:603640. doi 10.3389/fimmu.2020.603640

Burchett A., Siri S., Li J., Lu X., Datta M. Novel 3-D macrophage spheroid model reveals reciprocal regulation of immunomechanical stress and mechano-immunological response. *Cell Mol Bioeng.* 2024;17(5):329-344. doi 10.1007/s12195-024-00824-z

Burdis R., Chariyev-Prinz F., Browe D.C., Freeman F.E., Nulty J., Mcdonnell E.E., Eichholz K.F., Wang B., Brama P., Kelly D.J. Spatial patterning of phenotypically distinct microtissues to engineer osteochondral grafts for biological joint resurfacing. *Biomaterials*. 2022;289:121750. doi 10.1016/j.biomaterials.2022.121750

Caleffi J.T., Aal M.C.E., Gallindo H.O.M., Caxali G.H., Crulhas B.P., Ribeiro A.O., Souza G.R., Delella F.K. Magnetic 3D cell culture: state of the art and current advances. *Life Sci.* 2021;1(286):120028. doi 10.1016/j.lfs.2021.120028

Chen Q., Cui L., Guan Y., Zhang Y. Diels-alder cross-linked, washing-free hydrogel films with ordered wrinkling patterns for multicellular spheroid generation. *Biomacromolecules.* 2021;22(8):3474-3485. doi 10.1021/acs.biomac.1c00570

Chernyavska M., Masoudnia M., Valerius T., Verdurmen W.P.R. Organ-on-a-chip models for development of cancer immunotherapies. *Cancer Immunol Immunother*. 2023;72(12):3971-3983. doi 10.1007/s00262-023-03572-7

Coluccio M.L., Perozziello G., Malara N., Parrotta E., Zhang P., Gentile F., Limongi T., Raj P.M., Cuda G., Candeloro P., Di Fabrizio E. Microfluidic platforms for cell cultures and investigations. *Microelectron Eng.* 2019;208:14-28. doi 10.1016/j.mee.2019.01.004

Corgnac S., Damei I., Gros G., Caidi A., Terry S., Chouaib S., Deloger M., Mami-Chouaib F. Cancer stem-like cells evade $CD8^+CD103^+$ tumor-resident memory T (T_{RM}) lymphocytes by initiating an epithelial-to-mesenchymal transition program in a human lung tumor model. *J Immunother Cancer*. 2022;10(4):e004527. doi 10.1136/jitc-2022-004527

Daly A.C., Davidson M.D., Burdick J.A. 3D bioprinting of high cell-density heterogeneous tissue models through spheroid fusion within self-healing hydrogels. *Nat Commun.* 2021;12(1):753. doi 10.1038/s41467-021-21029-2

Danku A.E., Dulf E., Braicu C., Jurj A., Berindan-Neagoe I. Organ-on-a-chip: a survey of technical results and problems. *Front Bioeng Biotechnol.* 2022;10:840674. doi 10.3389/fbioe.2022.840674

Di Buduo C.A., Laurent P.A., Zaninetti C., Lordier L., Soprano P.M., Ntai A., Barozzi S., La Spada A., Biunno I., Raslova H., Bussel J.B., Kaplan D.L., Balduini C.L., Pecci A., Balduini A. Miniaturized 3D bone marrow tissue model to assess response to Thrombopoietin-receptor agonists in patients. *eLife*. 2021;10:e58775. doi 10.7554/eLife.58775

Doost N.F., Srivastava S.K. A comprehensive review of organ-on-a-chip technology and its applications. *Biosensors*. 2024;14(5):225. doi 10.3390/bios14050225

Duryee W.R., Doherty J.K. Nuclear and cytoplasmic organoids in the living cell. *Ann NY Acad Sci.* 1954;58(7):1210-1231. doi 10.1111/j.1749-6632.1954.tb45904.x

Edmondson R., Broglie J.J., Adcock A.F., Yang L. Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. *Assay Drug Dev Technol.* 2014;12(4):207-218. doi 10.1089/adt.2014.573

Eke G., Vaysse L., Yao X., Escudero M., Carri A., Trevisiol E., Vieu C., Dani C., Casteilla L., Malaquin L. Cell aggregate assembly through microengineering for functional tissue emergence. *Cells*. 2022;11(9):1394. doi 10.3390/cells11091394

Eyckmans J., Chen C.S. 3D culture models of tissues under tension. *J Cell Sci*. 2017;130(1):63-70. doi 10.1242/jcs.198630

Fata J.E., Mori H., Ewald A.J., Zhang H., Yao E., Werb Z., Bissell M.J. The $MAPK^{ERK-1,2}$ pathway integrates distinct and antagonistic signals from TGFα and FGF7 in morphogenesis of mouse mammary epithelium. *Dev Biol.* 2007;306(1):193-207. doi 10.1016/j.ydbio.2007.03.013

Forsythe S.D., Erali R.A., Sasikumar S., Laney P., Shelkey E., D'Agostino R., Jr., Miller L.D., Shen P., Levine E.A., Soker S., Votanopoulos K.I. Organoid platform in preclinical investigation of personalized immunotherapy efficacy in appendiceal cancer: feasibility study. *Clin Cancer Res*. 2021;27(18):5141-5150. doi 10.1158/1078-0432.CCR-21-0982

Gaitán-Salvatella I., López-Villegas E.O., González-Alva P., Susate-Olmos F., Álvarez-Pérez M.A. Case report: formation of 3D osteoblast spheroid under magnetic levitation for bone tissue engineering. *Front Mol Biosci.* 2021;8:672518. doi 10.3389/fmolb.2021.672518

George J., Chen Y., Abdelfattah N., Yamamoto K., Gallup T.D., Adamson S.I., Rybinski B., Srivastava A., Kumar P., Lee M.G., Baskin D.S., Jiang W., Choi J.M., Flavahan W., Chuang J.H., Kim B.Y.S., Xu J., Jung S.Y., Yun K. Cancer stem cells, not bulk tumor cells, determine mechanisms of resistance to SMO inhibitors. *Cancer Res Commun.* 2022;2(6):402-416. doi 10.1158/2767-9764.CRC-22-0124

Gilazieva Z., Ponomarev A., Rutland C., Rizvanov A., Solovyeva V. Promising applications of tumor spheroids and organoids for personalized medicine. *Cancers* (*Basel*). 2020;12(10):2727. doi 10.3390/cancers12102727

Grotz M., van Gijzel L., Bitsch P., Carrara S.C., Kolmar H., Garg S. Mimicking the immunosuppressive impact of fibroblasts in a 3D multicellular spheroid model. *Front Drug Discov.* 2024;4:1427407. doi 10.3389/fddsv.2024.1427407

Gunti S., Hoke A.T.K., Vu K.P., London N.R. Organoid and spheroid tumor models: techniques and applications. *Cancers* (*Basel*). 2021; 13(4):874. doi 10.3390/cancers13040874

Guryanov M.I. Organized frequency structure of electrocardiogram during long-duration ventricular fibrillation under experimental conditions. *Sovremennye Tehnologii v Medicine = Modern Technologies in Medicine*. 2016;8(3):37-48. doi 10.17691/stm2016.8.3.04

Han S.J., Kwon S., Kim K.S. Challenges of applying multicellular tumor spheroids in preclinical phase. *Cancer Cell Int.* 2021;21(1): 152. doi 10.1186/s12935-021-01853-8

Heinrich M.A., Huynh N.T., Heinrich L., Prakash J. Understanding glioblastoma stromal barriers against NK cell attack using tri-culture 3D spheroid model. *Heliyon*. 2024;10(3):e24808. doi 10.1016/j.heliyon.2024.e24808

Higgins C.A., Richardson G.D., Ferdinando D., Westgate G.E., Jahoda C.A.B. Modelling the hair follicle dermal papilla using spheroid cell cultures. *Exp Dermatol.* 2010;19(6):546-548. doi 10.1111/j.1600-0625.2009.01007.x

Hofer M., Lutolf M.P. Engineering organoids. *Nat Rev Mater.* 2021; 6(5):402-420. doi 10.1038/s41578-021-00279-y

Hong G., Kim J., Oh H., Yun S., Kim C.M., Jeong Y., Yun W., Shim J., Jang I., Kim C., Jin S. Production of multiple cell-laden microtissue spheroids with a biomimetic hepatic-lobule-like structure. *Adv Mater*. 2021;33(36):e2102624. doi 10.1002/adma.202102624

Hsu T.W., Lu Y.J., Lin Y.J., Huang Y.T., Hsieh L.H., Wu B.H., Lin Y.C., Chen L.C., Wang H.W., Chuang J.C., Fang Y.Q., Huang C.C. Transplantation of 3D MSC/HUVEC spheroids with neuroprotective and proangiogenic potentials ameliorates ischemic stroke brain injury. *Biomaterials*. 2021;272:120765. doi 10.1016/j.biomaterials.2021.120765

Ivascu A., Kubbies M. Diversity of cell-mediated adhesions in breast cancer spheroids. *Int J Oncol.* 2007;31(6):1403-1413. doi 10.3892/ijo.31.6.1403

Jedrzejczak-Silicka M. History of cell culture. In: New Insights into Cell Culture Technology. InTech, 2017. doi 10.5772/66905

Jeibouei S., Khazraie A., Hojat A., Reza A. Human-derived Tumor-On-Chip model to study the heterogeneity of breast cancer tissue. *Biomater Adv.* 2024;162:213915. doi 10.1016/j.bioadv.2024.213915

Jeong H.S., Park C.Y., Kim J.H., Joo H.J., Choi S.C., Choi J.H., Lim I.R., Park J.H., Hong S.J., Lim D.S. Cardioprotective effects of genetically engineered cardiac stem cells by spheroid formation on ischemic cardiomyocytes. *Mol Med.* 2020;26(1):15. doi 10.1186/s10020-019-0128-8

Kang S.M., Kim D., Lee J.H., Takayama S., Park J.Y. Engineered microsystems for spheroid and organoid studies. *Adv Healthc Mater.* 2021;10(2):e2001284. doi 10.1002/adhm.202001284

Karimifard S.A., Salehzadeh-Yazdi A., Taghizadeh-Tabarsi R., Akbari-Birgani S. Mechanical effects modulate drug resistance in MCF-7-derived organoids: insights into the wnt/β-catenin pathway. *Biochem Biophys Res Commun.* 2024;695:149420. doi 10.1016/j.bbrc.2023.149420

Kassis T., Hernandez-Gordillo V., Langer R., Griffith L.G. OrgaQuant: human intestinal organoid localization and quantification using deep convolutional neural networks. *Sci Rep.* 2019;9(1):12479. doi 10.1038/s41598-019-48874-y

Khan I., Prabhakar A., Delepine C., Tsang H., Pham V., Sur M. A low-cost 3D printed microfluidic bioreactor and imaging chamber for live-organoid imaging. *Biomicrofluidics*. 2021;15(2):024105. doi 10.1063/5.0041027

Kim T.Y., Kofron C.M., King M.E., Markes A.R., Okundaye A.O., Qu Z., Mende U., Choi B.R. Directed fusion of cardiac spheroids into larger heterocellular microtissues enables investigation of cardiac action potential propagation via cardiac fibroblasts. *PLoS One*. 2018;13(5):e0196714. doi 10.1371/journal.pone.0196714

Kumar S., Wei G., Aggarwal N. Organ-on-chip technology: opportunities and challenges. *Biotechnol Notes*. 2024;5:8-12. doi 10.1016/j.biotno.2024.01.001

Lee G.Y., Kenny P.A., Lee E.H., Bissell M.J. Three-dimensional culture models of normal and malignant breast epithelial cells. *Nat Methods*. 2007;4(4):359-365. doi 10.1038/nmeth1015

Li M.L., Aggeler J., Farson D.A., Hatier C., Hassell J., Bissell M.J. Influence of a reconstituted basement membrane and its components on casein gene expression and secretion in mouse mammary epithelial cells. *Proc Natl Acad Sci USA.* 1987;84(1):136-140. doi 10.1073/pnas.84.1.136

Lohasz C., Loretan J., Sterker D., Görlach E., Renggli K., Argast P., Frey O., Wiesmann M., Wartmann M., Rausch M., Hierlemann A. A microphysiological cell-culturing system for pharmacokinetic drug exposure and high-resolution imaging of arrays of 3D microtissues. *Front Pharmacol.* 2021;12:785851. doi 10.3389/fphar.2021. 785851

Magenau J., Runaas L., Reddy P. Advances in understanding the pathogenesis of graft-versus-host disease. *Br J Haematol*. 2016;173(2): 190-205. doi 10.1111/bjh.13959

Maliszewska-Olejniczak K., Brodaczewska K.K., Bielecka Z.F., Solarek W., Kornakiewicz A., Szczylik C., Porta C., Czarnecka A.M. Development of extracellular matrix supported 3D culture of renal cancer cells and renal cancer stem cells. *Cytotechnology*. 2019; 71(1):149-163. doi 10.1007/s10616-018-0273-x

Martins C., Pacheco C., Moreira-Barbosa C., Marques-Magalhães Â., Dias S., Araújo M., Oliveira M.J., Sarmento B. Glioblastoma immuno-endothelial multicellular microtissue as a 3D *in vitro* evaluation tool of anti-cancer nano-therapeutics. *J Control Release*. 2023; 353:77-95. doi 10.1016/j.jconrel.2022.11.024

Maulana T.I., Teufel C., Cipriano M., Roosz J., Lazarevski L., van den Hil F.E., Scheller L., Orlova V., Koch A., Hudecek M., Alb M., Loskill P. Breast cancer-on-chip for patient-specific efficacy and safety testing of CAR-T cells. *Cell Stem Cell*. 2024;31(7):989-1002.e9. doi 10.1016/j.stem.2024.04.018. Epub 2024 May15. PMID 38754430

Mesci P., de Souza J.S., Martin-Sancho L., Macia A., Saleh A., Yin X., Snethlage C., Adams J.W., Avansini S.H., Herai R.H., Almenar-Queralt A., Pu Y., Szeto R.A., Goldberg G., Bruck P.T., Papes F., Chanda S.K., Muotri A.R. SARS-CoV-2 infects human brain organoids causing cell death and loss of synapses that can be rescued by treatment with Sofosbuvir. *PLoS Biol*. 2022;20(11):e3001845. doi 10.1371/journal.pbio.3001845

Miao X., Wang C., Chai C., Tang H., Hu J., Zhao Z., Luo W., Zhang H., Zhu K., Zhou W., Xu H. Establishment of gastric cancer organoid and its application in individualized therapy. *Oncol Lett*. 2022; 24(6):447. doi 10.3892/ol.2022.13567

Moradi-Mehr S., Khademy M., Akbari-Birgani S., Kafian H., Lalenejad M., Abdollahpour D., Moghimi M. Comparative evaluation of the therapeutic strategies using a minimal model of luminal-A breast cancer. *Biochem Biophys Res Commun.* 2023;666:107-114. doi 10.1016/j.bbrc.2023.05.028

Moscona A. Development of heterotypic combinations of dissociated embryonic chick cells. *Proc Soc Exp Biol Med.* 1956;92(2):410-416. doi 10.3181/00379727-92-22495

Nair A.L., Mesch L., Schulz I., Becker H., Raible J., Kiessling H., Werner S., Rothbauer U., Schmees C., Busche M., Trennheuser S., Fricker G., Stelzle M. Parallelizable microfluidic platform to model and assess in vitro cellular barriers: technology and application to study the interaction of 3D tumor spheroids with cellular barriers. *Biosensors*. 2021;11(9):314. doi 10.3390/bios11090314

Neal J.T., Li X., Zhu J., Giangarra V., Grzeskowiak C.L., Ju J., Liu I.H., Chiou S.H., Salahudeen A.A., Smith A.R., Deutsch B.C., Liao L.,

Zemek A.J., Zhao F., Karlsson K., Schultz L.M., Metzner T.J., Nadauld L.D., Tseng Y.Y., Alkhairy S., Oh C., Keskula P., Mendoza-Villanueva D., De La Vega F.M., Kunz P.L., Liao J.C., Leppert J.T., Sunwoo J.B., Sabatti C., Boehm J.S., Hahn W.C., Zheng G.X.Y., Davis M.M., Kuo C.J. Organoid modeling of the tumor immune resource organoid modeling of the tumor immune microenvironment. *Cell*. 2018;175(7):1972-1988.e16. doi 10.1016/j.cell.2018.11.021

Nguyen M., De Ninno A., Mencattini A., Mermet-Meillon F., Fornabaio G., Evans S.S., Cossutta M., Khira Y., Han W., Sirven P., Pelon F., Di Giuseppe D., Bertani F.R., Gerardino A., Yamada A., Descroix S., Soumelis V., Mechta-Grigoriou F., Zalcman G., Camonis J., Martinelli E., Businaro L., Parrini M.C. Dissecting effects of anti-cancer drugs and cancer-associated fibroblasts by on-chip reconstitution of immunocompetent tumor microenvironments. *Cell Rep*. 2018;25(13):3884-3893. doi 10.1016/j.celrep.2018.12.015

Nguyen O.P., Misun P.M., Lohasz C., Lee J., Wang W., Schroeder T., Hierlemann A. An immunocompetent microphysiological system to simultaneously investigate effects of anti-tumor natural killer cells on tumor and cardiac microtissues. *Front Immunol*. 2021;12:781337. doi 10.3389/fimmu.2021.781337

Nikonorova V.G., Chrishtop V.V., Mironov V.A., Prilepskii A.Y. Advantages and potential benefits of using organoids in nanotoxicology. *Cells*. 2023;12(4):610. doi 10.3390/cells12040610

Ninomiya K., Taniuchi T. Assembly of a tumor microtissue by stacking normal and cancer spheroids on Kenzan using a bio-3D printer to monitor dynamic cancer cell invasion in the microtissue. *Biochem Eng J*. 2024;212:109536. doi 10.1016/j.bej.2024.109536

Noorintan S.T., Angelius C., Torizal F.G. Organoid models in cancer immunotherapy: bioengineering approach for personalized treatment. *Immuno*. 2024;4(4):312-324. doi 10.3390/immuno4040020

Nushtaeva A.A., Savinkova M.M., Ermakov M.S., Varlamov M.E., Novak D.D., Richter V.A., Koval O.A. Breast cancer cells in 3D model alters their sensitivity to hormonal and growth factors. *Cell Tissue Biol*. 2022;16(6):555-567. doi 10.1134/S1990519X22060050

Paşca S.P., Arlotta P., Bateup H.S., Camp J.G., Cappello S., Gage F.H., Knoblich J.A., Kriegstein A.R., Lancaster M.A., Ming G.L., Muotri A.R., Park I.H., Reiner O., Song H., Studer L., Temple S., Testa G., Treutlein B., Vaccarino F.M. A nomenclature consensus for nervous system organoids and assembloids. *Nature*. 2022; 609(7929):907-910. doi 10.1038/s41586-022-05219-6

Petersen O.W., Rønnov-Jessen L., Howlett A.R., Bissell M.J. Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells. *Proc Natl Acad Sci USA*. 1992;89(19):9064-9068. doi 10.1073/pnas.89.19.9064

Ponti D., Costa A., Zaffaroni N., Pratesi G., Petrangolini G., Coradini D., Pilotti S., Pierotti M.A., Daidone M.G. Isolation and *in vitro* propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. *Cancer Res*. 2005;65(13):5506-5511. doi 10.1158/0008-5472.CAN-05-0626

Ravi M., Kaviya S.R., Paramesh V. Culture phases, cytotoxicity and protein expressions of agarose hydrogel induced Sp2/0, A549, MCF-7 cell line 3D cultures. *Cytotechnology*. 2016;68(3):429-441. doi 10.1007/s10616-014-9795-z

Rheinwatd J.G., Green H. Seria cultivation of strains of human epidemal keratinocytes: the formation keratinizin colonies from single cell is. *Cell*. 1975;6(3):331-343. doi 10.1016/S0092-8674(75)80001-8

Rodrigues D.B., Reis R.L., Pirraco R.P. Modelling the complex nature of the tumor microenvironment: 3D tumor spheroids as an evolving tool. *J Biomed Sci*. 2024;31(1):13. doi 10.1186/s12929-024-00997-9

Sadovska L., Zandberga E., Sagini K., Jēkabsons K., Riekstiņa U., Kalniņa Z., Llorente A., Linē A. A novel 3D heterotypic spheroid model for studying extracellular vesicle-mediated tumour and immune cell communication. *Biochem Biophys Res Commun*. 2018; 495(2):1930-1935. doi 10.1016/j.bbrc.2017.12.072

Sakalem M.E., De Sibio M.T., da Costa F.A.D.S., de Oliveira M. Historical evolution of spheroids and organoids, and possibilities of use in life sciences and medicine. *Biotechnol J*. 2021;16(5):e2000463. doi 10.1002/biot.202000463

Schnalzger T.E., De Groot M.H.P., Zhang C., Mosa M.H., Michels B.E., Röder J., Darvishi T., Wels W.S. 3D model for CAR-mediated cytotoxicity using patient-derived colorectal cancer organoids. *EMBO J*. 2019;38(12):e100928. doi 10.15252/embj.2018100928

Schot M., Araújo-Gomes N., Van Loo B., Kamperman T., Leijten J. Scalable fabrication, compartmentalization and applications of living microtissues. *Bioact Mater*. 2023;19:392-405. doi 10.1016/j.bioactmat.2022.04.005

Shamir E.R., Ewald A.J. Three-dimensional organotypic culture: experimental models of mammalian biology and disease. *Nat Rev Mol Cell Biol*. 2014;15(10):647-664. doi 10.1038/nrm3873

Simian M., Bissell M.J. Organoids: a historical perspective of thinking in three dimensions. *J Cell Biol*. 2017;216(1):31-40. doi 10.1083/jcb.201610056

Singh D., Mathur A., Arora S., Roy S., Mahindroo N. Applied surface science advances journey of organ on a chip technology and its role in future healthcare scenario. *Appl Surf Sci Adv*. 2022;9:100246. doi 10.1016/j.apsadv.2022.100246

Sulaiman S., Chowdhury S.R., Fauzi M.B., Rani R.A., Mohamadyahaya N.H., Tabata Y., Hiraoka Y., Idrus R.B.H., Hwei N.M. 3D culture of MSCS on a gelatin microsphere in a dynamic culture system enhances chondrogenesis. *Int J Mol Sci*. 2020;21(8):2688. doi 10.3390/ijms21082688

Sun M., Liu A., Yang X., Gong J., Yu M., Yao X., Wang H., He Y. 3D cell culture – can it be as popular as 2D cell culture? *Adv Nano-Biomed Res*. 2021;1(5):2000066. doi 10.1002/anbr.202000066

Sutherland R.M. Cell and environment interactions in tumor microregions: the multicell spheroid model. *Science*. 1988;240(4849): 177-184. doi 10.1126/science.2451290

Tanaka Y., Nishikawa M., Mizukami Y., Kusamori K., Ogino Y., Nishimura S., Shimizu K., Konishi S., Takahashi Y., Takakura Y. Control of polarization and tumoricidal activity of macrophages by multicellular spheroid formation. *J Control Release*. 2018;270:177-183. doi 10.1016/j.jconrel.2017.12.006

Troitskaya O., Novak D., Nushtaeva A., Savinkova M., Varlamov M., Ermakov M., Richter V., Koval O. EGFR transgene stimulates spontaneous formation of MCF7 breast cancer cells spheroids with partly loss of HER3 receptor. *Int J Mol Sci*. 2021;22(23):12937. doi 10.3390/ijms222312937

Valent P., Bonnet D., De Maria R., Lapidot T., Copland M., Melo J.V., Chomienne C., Ishikawa F., Schuringa J.J., Stassi G., Huntly B., Herrmann H., Soulier J., Roesch A., Schuurhuis G.J., Wöhrer S., Arock M., Zuber J., Cerny-Reiterer S., Johnsen H.E., Andreeff M., Eaves C. Cancer stem cell definitions and terminology: the devil is in the details. *Nat Rev Cancer*. 2012;12(11):767-775. doi 10.1038/nrc3368

Vasileva N.S., Kuligina E.V., Dymova M.A., Savinovskaya Y.I., Zinchenko N.D., Ageenko A.B., Mishinov S.V., Dome A.S., Stepanov G.A., Richter V.A., Semenov D.V. Transcriptome changes in glioma cells cultivated under conditions of neurosphere formation. *Cells*. 2022;11(19):3106. doi 10.3390/cells11193106

Veith I., Nurmik M., Mencattini A., Damei I., Lansche C., Brosseau S., Gropplero G., Corgnac S., Filippi J., Poté N., Guenzi E., Chassac A., Mordant P., Tosello J., Sedlik C., Piaggio E., Girard N., Camonis J., Shirvani H., Mami-Chouaib F., Mechta-Grigoriou F., Descroix S., Martinelli E., Zalcman G., Parrini M.C. Assessing personalized responses to anti-PD-1 treatment using patient-derived lung tumor-on-chip. *Cell Rep Med*. 2024;5(5):101549. doi 10.1016/j.xcrm.2024.101549

Verjans E.T., Doijen J., Luyten W., Landuyt B., Schoofs L. Three-dimensional cell culture models for anticancer drug screening: worth

the effort? *J Cell Physiol*. 2018;233(4):2993-3003. doi 10.1002/jcp.26052

Waldhauer I., Morra L., Lehmann S., Agarkova I., Zumstein P., Kelm J.M., Umana P., Klein C., Bacac M. Abstract LB-264: development of 3D microtissue models to study the activity of novel tumor-targeted immunotherapeutics. *Cancer Res.* 2013;73(8_Supplement):LB-264. doi 10.1158/1538-7445.AM2013-LB-264

Wang O., Han L., Lin H., Tian M., Zhang S., Duan B., Chung S., Zhang C., Lian X., Wang Y., Lei Y. Fabricating 3-dimensional human brown adipose microtissues for transplantation studies. *Bioact Mater.* 2023;22:518-534. doi 10.1016/j.bioactmat.2022.10.022

Wang Y., Kankala R.K., Zhang J., Hao L., Zhu K., Wang S., Zhang Y.S., Chen A. Modeling endothelialized hepatic tumor microtissues for drug screening. *Adv Sci*. 2020;7(21):2002002. doi 10.1002/advs.202002002

Wei W., Cardes F., Hierlemann A., Modena M.M. 3D *in vitro* blood-brain-barrier model for investigating barrier insults. *Adv Sci*. 2023;10(11):e2205752. doi 10.1002/advs.202205752

Weiswald L.B., Bellet D., Dangles-Marie V. Spherical cancer models in tumor biology. *Neoplasia*. 2015;17(1):1-15. doi 10.1016/j.neo.2014.12.004

Wolter J.R. Proliferating pigment epithelium: producing a simple organoid structure in the subrentinal space of a human eye. *Arch Ophthalmol.* 1967;77(5):651-654. doi 10.1001/archopht.1967.00980020653016

Zhang J., Li C., Meng F., Guan Y., Zhang T., Yang B., Ren Z., Liu X., Li D., Zhao J., Zhao J., Wang Y., Peng J. Functional tissue-engineered microtissue formed by self-aggregation of cells for peripheral nerve regeneration. *Stem Cell Res Ther.* 2022;13(1):3. doi 10.1186/s13287-021-02676-0

Zhou Z., Van der Jeught K., Fang Y., Yu T., Li Y., Ao Z., Liu S., Zhang L., Yang Y., Eyvani H., Cox M.L., Wang X., He X., Ji G., Schneider B.P., Guo F., Wan J., Zhang X., Lu X. An organoid-based screen for epigenetic inhibitors that stimulate antigen presentation and potentiate T-cell-mediated cytotoxicity. *Nat Biomed Eng.* 2021;5(11):1320-1335. doi 10.1038/s41551-021-00805-x

Conflict of interest. The authors declare no conflict of interest.

Received November 2, 2024. Revised November 28, 2024. Accepted November 28, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):189-199

doi 10.18699/vjgb-25-22

Isogenic induced pluripotent stem cell line ICGi036-A-1 from a patient with familial hypercholesterolaemia, derived by correcting a pathogenic variant of the gene *LDLR* c.530C>T

A.S. Zueva [1, 2], A.I. Shevchenko [1], S.P. Medvedev [1], E.A. Elisaphenko [1], A.A. Sleptcov [1, 3], M.S. Nazarenko [1, 3], N.A. Tmoyan [1, 4], S.M. Zakian [1], I.S. Zakharova [1]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia
[4] National Medical Research Center of Cardiology named after academician E.I. Chazov, Moscow, Russia

zakharova@bionet.nsc.ru

Abstract. Familial hypercholesterolaemia is a common monogenic disorder characterized by high plasma cholesterol levels leading to chronic cardiovascular disease with high risk and often early manifestation due to atherosclerotic lesions of the blood vessels. The atherosclerotic lesions in familial hypercholesterolaemia are mainly caused by pathogenic variants of the low-density lipoprotein receptor (*LDLR*) gene, which plays an important role in cholesterol metabolism. Normally, cholesterol-laden low-density lipoproteins bind to the LDLR receptor on the surface of liver cells to be removed from the bloodstream by internalisation with hepatocytes. In familial hypercholesterolaemia, the function of the receptor is impaired and the uptake of low-density lipoproteins is significantly reduced. As a result, cholesterol accumulates in the subendothelial space on the inner wall of blood vessels, triggering atherogenesis, the formation of atherosclerotic plaques. At present, there are no effective and universal approaches to the diagnosis and treatment of familial hypercholesterolaemia. A relevant approach to study the molecular genetic mechanisms of the disease and to obtain systems for screening chemical compounds as potential drugs is the generation of cellular models based on patient-specific induced pluripotent stem cells. The aim of our work was to derive an isogenic genetically modified induced pluripotent stem cell line by correcting the pathogenic allelic variant c.530C of the *LDLR* gene in the original iPSC previously obtained from a compound heterozygote patient with familial hypercholesterolaemia. The resulting isogenic iPSC line differs from the original by only one corrected nucleotide substitution, allowing us to study the direct effect of this pathogenic genetic variant on physiological changes in relevant differentiated cells. CRISPR/Cas-mediated base editing was used to correct the single nucleotide substitution. The resulting genetically modified iPSC line has pluripotency traits, a normal karyotype, a set of short tandem repeats identical to that in the original line and can be used to obtain differentiated derivatives necessary for the elaboration of relevant cell models.
Key words: familial hypercholesterolaemia; *LDLR*; induced pluripotent stem cells; genome editing; isogenic cell lines

For citation: Zueva A.S., Shevchenko A.I., Medvedev S.P., Elisaphenko E.A., Sleptcov A.A., Nazarenko M.S., Tmoyan N.A., Zakian S.M., Zakharova I.S. Isogenic induced pluripotent stem cell line ICGi036-A-1 from a patient with familial hypercholesterolaemia, derived by correcting a pathogenic variant of the gene *LDLR* c.530C>T. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):189-199. doi 10.18699/vjgb-25-22

Funding. The work was supported by the Russian Science Foundation grant No. 24-15-00346, https://rscf.ru/project/24-15-00346/.

Acknowledgements. The immunofluorescent imaging was performed using resources of the Common Facilities Center of Microscopic Analysis of Biological Objects, ICG SB RAS (https://ckp.icgen.ru/ckpmabo/), supported by the State project of the Institute of Cytology and Genetics FWNR-2022–0015.

Изогенная линия индуцированных плюрипотентных стволовых клеток ICGi036-A-1 от пациента с семейной гиперхолестеринемией, созданная путем коррекции патогенного варианта гена *LDLR* c.530C>T

А.С. Зуева [1, 2], А.И. Шевченко [1], С.П. Медведев [1], Е.А. Елисафенко [1], А.А. Слепцов [1, 3], М.С. Назаренко [1, 3], Н.А. Тмоян [1, 4], С.М. Закиян [1], И.С. Захарова [1]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Научно-исследовательский институт медицинской генетики, Томский национальный исследовательский медицинский центр Российской академии наук, Томск, Россия
[4] Национальный медицинский исследовательский центр кардиологии имени академика Е.И. Чазова Министерства здравоохранения Российской Федерации, Москва, Россия
zakharova@bionet.nsc.ru

Аннотация. Семейная гиперхолестеринемия является распространенным моногенным заболеванием, которое характеризуется повышенным содержанием холестерина в плазме крови, приводящим к хроническим заболеваниям сердечно-сосудистой системы с высоким риском и ранним проявлением развития патологий, вызванных атеросклеротическими поражениями кровеносных сосудов. Образование атеросклеротических бляшек при семейной гиперхолестеринемии в основном обусловлено патогенными вариантами гена рецептора липопротеинов низкой плотности *LDLR* (low-density lipoprotein receptor), играющего важную роль в метаболизме холестерина. В норме липопротеины низкой плотности, переносящие холестерин, связываются с рецептором LDLR на поверхности клеток печени и выводятся из кровотока путем интернализации гепатоцитами. При семейной гиперхолестеринемии происходит нарушение функционирования рецептора и значительное снижение интернализации липопротеинов низкой плотности. Это приводит к их накоплению в субэндотелиальном пространстве внутренней стенки кровеносных сосудов и вызывает атерогенез – образование атеросклеротических бляшек. На сегодняшний день не существует эффективных и универсальных подходов к диагностике и лечению семейной гиперхолестеринемии. Актуальным подходом для исследования молекулярно-генетических особенностей заболевания и разработки систем скрининга химических соединений – потенциальных лекарственных препаратов – является создание клеточных моделей на основе индуцированных плюрипотентных стволовых клеток (ИПСК) пациентов. Целью нашей работы было создание изогенной генетически модифицированной линии индуцированных плюрипотентных стволовых клеток путем коррекции патогенного аллельного варианта c.530C гена *LDLR* в линии ИПСК, полученной ранее от пациента-компаундной гетерозиготы с семейной гиперхолестеринемией. Созданная изогенная клеточная линия ИПСК отличается от исходной только одной скорректированной нуклеотидной заменой, что позволяет исследовать непосредственное влияние данного патогенного генетического варианта на физиологические изменения в релевантных дифференцированных клетках. Для коррекции однонуклеотидных замен использован CRISPR/Cas9-опосредованный метод редактирования оснований. Полученная генетически модифицированная линия ИПСК обладает свойствами плюрипотентности, имеет нормальный кариотип, идентичный исходной линии набор коротких тандемных повторов и может быть использована для формирования дифференцированных производных, необходимых при разработке релевантных клеточных моделей.

Ключевые слова: семейная гиперхолестеринемия; *LDLR*; индуцированные плюрипотентные стволовые клетки; геномное редактирование; изогенные линии клеток

Introduction

Despite advances in medical technology and revolutionary discoveries in biology, cardiovascular disease, caused mainly by atherosclerotic lesions, remains the leading cause of death worldwide, according to the World Health Organisation.

The most common inherited disease leading to atherosclerosis is familial hypercholesterolaemia (FH) (Zakharova et al., 2024b). According to the European Atherosclerosis Society, this disease occurs with a high frequency, namely one in 250 people for the heterozygous form, one in 300 thousand to one in 1 million people for the homozygous form (Ezhov et al., 2019). Nevertheless, the disease can have a latent course with difficult diagnosis and manifest acute vascular catastrophes in the form of heart attacks, strokes and other ischaemic lesions, often leading to death (Hopkins et al., 2011; Talmud et al., 2014; Ference et al., 2017). Currently, up to 70 % of suspected heterozygous carriers remain undiagnosed[1] (Rau et al., 2023).

[1] Information portal for patients with familial hypercholesterolaemia https://familyheart.org/familial-hypercholesterolemia; Information portal of the biotech genetic testing company "23andMe" https://www.23andme.com/topics/health-predispositions/fh/?srsltid=AfmBOooDFqM2USz3G0j9PZg-ng-15q__dvPbcQL6qgCzJOQodQhsLil7

FH is an autosomal dominant disorder associated with elevated low-density lipoprotein (LDL) cholesterol levels and a high risk of premature cardiovascular disease (CVD) (Harada-Shiba, 2023). FH can be heterozygous or homozygous. Patients with the homozygous FH form usually show early CVD manifestations, and without serious comprehensive therapy, their life expectancy does not exceed 30 years (Hopkins et al., 2011). FH is caused by pathogenic allelic variants in genes encoding key proteins involved in LDL clearance mediated by the LDLR (low-density lipoprotein receptor) (Gu et al., 2024). In 85 % of diagnosed FH patients, pathological conditions are caused by a disruption of the *LDLR* gene, which encodes the low-density lipoprotein receptor on the surface of hepatocytes (Hendricks-Sturrup et al., 2020). FH is rarely associated with *de novo* pathogenic allelic variants (Fularski et al., 2024). In this context, cascade genetic screening of relatives and prevention of atherosclerosis in diagnosed carriers is important.

Despite the high prevalence of FH, there are no effective treatments. According to data published by the European Atherosclerosis Society in 2022, less than 3 % of patients worldwide achieve cholesterol-lowering goals with medica-

tions (Ray et al., 2022; Harada-Shiba, 2023). The lack of effective treatment for FH is linked to the lack of relevant models for both drug trials and the study of FH pathogenesis.

A promising approach to investigate the molecular genetic basis of the disease is the generation of isogenic iPSC lines from patients with FH. Previously, we obtained the ICGi036-A iPSC line from a compound heterozygous patient with familial hypercholesterolemia. This line is registered in the Human Pluripotent Stem Cell Registry (hPSCreg) with the identifier RRID:CVCL_B5EJ (Zakharova et al., 2022a). The initial iPSC line ICGi036-A contains two allelic variants of the *LDLR* gene, which are missense mutations c.530C>T (p.Ser177Leu) and c.1054T>C (p.Cys352Arg).

The c.530C>T substitution, rs121908026 (p.Ser177Leu), is located in exon 4 of the *LDLR* gene (Semenova et al., 2020; Meshkov et al., 2021). This missense mutation results in the substitution of serine for leucine at codon 177 in the highly conserved SerAspGlu sequence in the ligand-binding domain of LDLR (Südhof et al., 1985). This substitution slows the transport of LDLR protein to the cell surface, resulting in defective receptors being unable to bind cholesterol-laden LDLRs, which leads to a reduction in LDL capture to approximately 6–31 % (Thormaehlen et al., 2015). The LDLR(NM_000527.5):c.530C>T allelic variant is reported in databases as a pathogenic variant causing FH (ClinVar ID 3686; OMIM:606945.0004; UniProt variants VAR_005327; VarSome http://varso.me/1dmA).

The c.1054T>C substitution, rs879254769 (p.Cys352Arg), is located in exon 7 of the *LDLR* gene and encodes cysteine instead of arginine at codon 352 in an epidermal growth factor-like domain (Semenova et al., 2020; Meshkov et al., 2021). The allelic variant LDLR(NM_000527.5):c.1054T>A is reported in databases as pathogenic/likely pathogenic, causing FH (ClinVar ID 251618; VarSome http://varso.me/0J8J).

There is evidence that the c.530C>T and c.1054T>C *LDLR* allelic variants can independently cause FH. For example, the heterozygous c.530C>T substitution in *LDLR* is associated with FH in several countries including the Czech Republic, India, Portugal, Poland and Spain (Bourbon et al., 2008; Palacios et al., 2012; Tichý et al., 2012; Setia et al., 2016; Sharifi et al., 2016). In addition, this allelic variant in compound heterozygosity with EX7_EX10del (c.941-?_1186+?del) in the *LDLR* gene has been reported in Brazil, and together with c.55G>C (p.Asp19His) within the *ABCG8* gene in FH patients in Malaysia (Jannes et al., 2015; Mohd Nor et al., 2019).

The c.1054T>C *LDLR* allelic variant has been found in heterozygous form in FH patients in Taiwan and Russia, and as a compound heterozygote together with c.796G>A (p.Asp266Asn) in an FH patient in Western Siberia (Meshkov et al., 2021; Shakhtshneider et al., 2021; Huang et al., 2022). In addition, our previous study confirmed positioning *in trans* between the c.530C>T and c.1054T>C allelic variants in FH patient-specific ICGi036-A iPSCs (Nazarenko et al., 2023).

In the present work, we derive and perform a detailed characterisation of the isogenic genetically modified iPSC line ICGi036-A-1 by base editing correction of *LDLR* allelic variants in the original iPSC line previously obtained from a compound heterozygous FH patient carrying pathogenic c.530C>T (p.Ser177Leu) and likely pathogenic c.1054T>C (p.Cys352Arg) *LDLR* alleles. The genetically modified iPSC line with the corrected allelic variant c.530C>T (p.Ser177Leu) can be used to obtain relevant cell types for FH modelling and drug development.

Materials and methods

Cell lines. The human PSC lines used in the present work are listed below:

- iPSC line ICGi036-A (RRID: CVCL_B5EJ) from a compound heterozygous FH patient with two *LDLR* allelic variants, namely pathogenic c.530C>T (p.Ser177Leu), rs121908026, ClinVar ID 3686, OMIM:606945.0004 and likely pathogenic c.1054T>C (p.Cys352Arg), rs879254769, ClinVar ID 251618. The previously obtained iPSC line (Zakharova et al., 2022a) was used to derive an isogenic genetically modified iPSC;
- healthy donor iPSC line ICGi022-A (RRID: CVCL_ZE02) (Malakhova et al., 2020), for pluripotency marker control;
- healthy donor embryonic stem cells (ESCs) HuES9 (HVRDe009-A) (RRID: CVCL_0057) (Cowan et al., 2004), for pluripotency marker control.

IPSCs and ESCs cultivation. IPSCs and ESCs were incubated in DMEM/F12 growth medium containing 15 % KnockOut SR (Thermo Fisher Scientific), 1 mM GlutaMax (Thermo Fisher Scientific), 1 % NEAA (Thermo Fisher Scientific), 0.25 mM 2-mercaptoethanol (Thermo Fisher Scientific) and 10 ng/ml bFGF (Sci-store). Cells were grown on a layer of mitotically inactivated mouse embryonic fibroblasts. IPSCs were cultured in an incubator at 37 °C and 5 % CO_2. IPSC and ESC colonies were enzymatically dissociated using TrypLE (Thermo Fisher Scientific) and plated in fresh medium with the addition of 2 μM ROCK inhibitor thiazovivin (STEMCELL Technologies) every 5 days.

Vectors for genetic correction of iPSCs. To insert the spacer sequence for the guide RNA, we created a universal plasmid pC9-sgRNA-mCherry (Fig. 1), allowing its delivery into cells to be detected by the mCherry fluorescence signal.

Target DNA fragments were obtained by digestion with *Pci*I and *AciG*I endonucleases (SibEnzyme) and combined using phage T4 DNA ligase. One 1,244 bp fragment, isolated from plasmid pX458 (addgene #48138), contains the U6 promoter sequence and a site for spacer cloning. Another 3,778 bp fragment, derived from plasmid pTE4560 (addgene #107526), includes the kanamycin antibiotic resistance gene and the mCherry fluorescent protein sequence. The final assembly of plasmid pC9-sgRNA-mCherry was confirmed by restriction analysis and Sanger sequencing at the Genomics Collective Use Centre of the Siberian Branch of the Russian Academy of Sciences (http://www.niboch.nsc.ru/doku.php/sequest).

Oligonucleotides for guide RNAs were selected using PnB Designer (https://fgcz-shiny.uzh.ch/PnBDesigner/) (Siegner et al., 2021).

Guide RNA expression from the U6 promoter is enhanced when nucleotide G is located immediately after the 5′-CACC-3′ sequence and before the spacer sequence (Bauer et al., 2015). Based on this, we added nucleotide G to the 5′ end of the selected oligonucleotides and then generated a complementary oligonucleotide sequence. Next, the sequence 5′-CACC-3′ was added to the 5′ end of the upper DNA strand

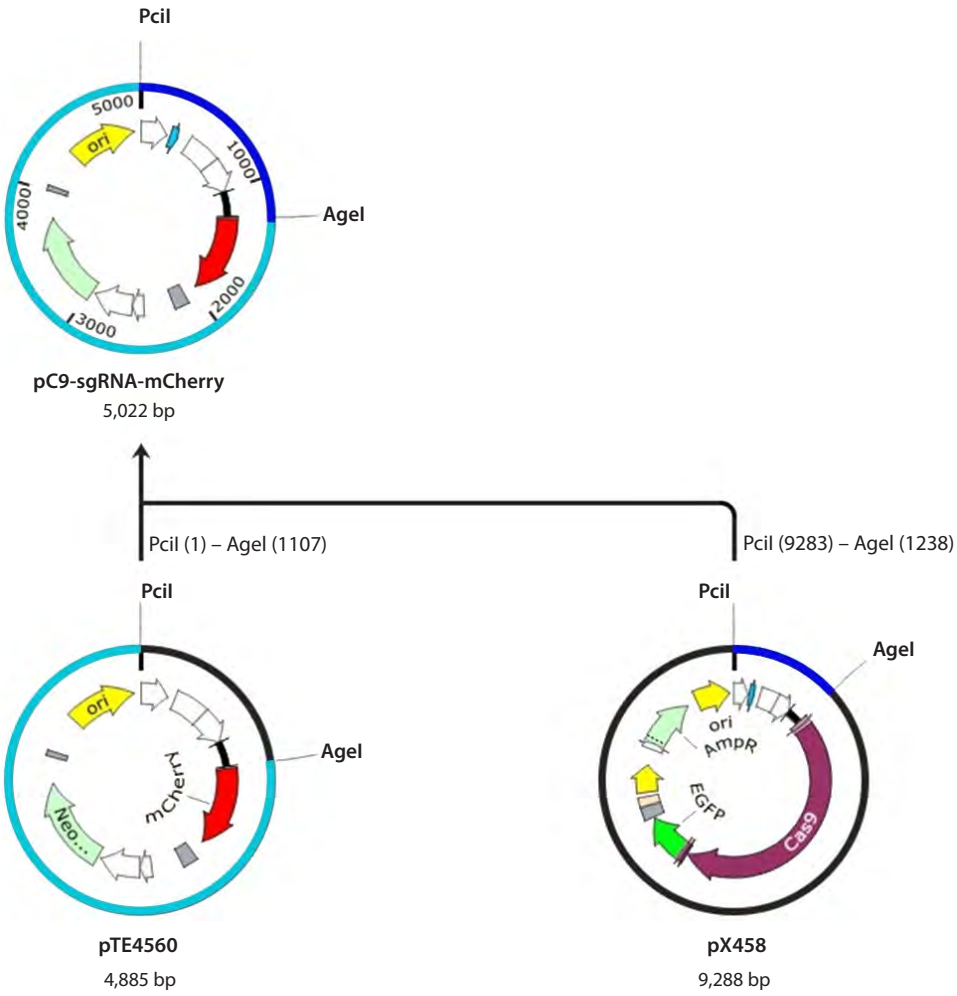


Fig. 1. Design of pC9-sgRNA-mCherry plasmid assembly from plasmids pTE4560 and pX458 in SnapGene software.

and 5′-AAAC-3′ was added to the 5′ end of the complementary oligonucleotide sequence for subsequent integration into the pC9-sgRNA-mCherry vector through *Bpi*I endonuclease restriction sites (see the Table).

Oligonucleotides were synthesised by Biosset (https://www.biosset.com/). Single-stranded oligonucleotides were phosphorylated at the 5′ end using phage T4 polynucleotide kinase (New England Biolabs) and annealed to form double-stranded molecules. Using phage T4 ligase, the resulting double-stranded molecules with sticky ends were inserted into the guide RNA spacer sites of the *Bpi*I endonuclease pre-linearised pC9-sgRNA-mCherrys plasmid. Sanger sequencing was used to confirm that the resulting plasmids contained integrated target sequences.

Two two-component plasmid systems were used for correction of targeted single-nucleotide substitutions. System 1 for correction of the c.530C>T substitution consists of the plasmid vector xCas9(3.7)-ABE(7.10) (addgene #108382),

Oligonucleotide sequences for guide RNAs modified for insertion into the plasmid vector

Number	Name	5′–3′ sequence	Length, bp	PAM
1	c.530C>T_1	cacc**G**ATCCAAGCCATCTTCGCAGT	25	CGG
2	c.530C>T_1r	aaacACTGCGAAGATGGCTTGGAT**C**	25	CGG
3	c.1054T>C_2	cacc**G**AAGATGCGAAGGTGATTTCC	25	GGG
4	c.1054T>C_2r	aaacGGAAATCACCTTCGCATCTT**C**	25	GGG

Note. Capital letters indicate oligonucleotides for guide RNAs selected using PnB Designer (https://fgcz-shiny.uzh.ch/PnBDesigner/) (Siegner et al., 2021), bold letters indicate nucleotides added to increase the expression level from the U6 promoter, lowercase letters indicate added sticky ends, and "r" indicates generated complementary oligonucleotide sequences).

encoding adenine deaminase and xCas9n nickase (adenine base editor), and the plasmid pC9-sgRNA-mCherry with an integrated guide RNA spacer sequence. System 2 for correction of the c.1054T>C substitution consists of the plasmid vector xCas9(3.7)-BE4 (addgene #108381) encoding a cytidine deaminase, xCas9n nickase (cytidine base editor), and also the plasmid pC9-sgRNA-mCherry with an integrated guide RNA spacer sequence.

Plasmid DNA isolation for subsequent lipofection was performed using the HiPure Plasmid EF Midi Kit (Magen).

Delivery of vectors to correct the c.530C>T and c.1054T>C substitutions of the *LDLR* gene sequence. Delivery of plasmids encoding base editors and guide RNAs into iPSCs was performed by lipofection using the Lipofectamine 3000 Transfection Reagent Kit (Thermo Fisher Scientific) according to the manufacturer's protocol with modifications. Twenty-four hours before lipofection, iPSCs cultured on mitotically inactivated mouse embryonic fibroblasts were plated into three wells of a 12-well plate coated with Matrigel matrix (Corning) in medium containing 15 % koSR (Thermo Fisher Scientific) and 2 μM ROCK inhibitor thiazovivine (STEMCELL Technologies). To increase the efficiency of upcoming lipofection, iPSCs were disaggregated to a single cell state during passaging using TrypLE. 4–5 hours after re-plating, when the cells had attached to the culture surface, the medium was changed to a serum-free medium containing thiazovivine and a 3-fold increased bFGF amount (30 ng/ml). 2 hours before lipofection, the medium was changed to an equivalent fresh medium without thiazovivine.

Lipofection was performed in medium without koSR and thiazovivine and with 30 ng/ml bFGF. The ratio between the amount of base-editor plasmids and plasmids with guide RNA was 3:1 in ng. Prior to lipofection, 150 μL Opti-MEM medium (Thermo Fisher Scientific) and 9 μL Lipofectamine 3000 were mixed in tube 1. In tube 2, 150 μl Opti-MEM medium, 6 μl P 3000 reagent, 125 ng each of plasmid DNA pC9-1054_2-mCherry and pC9-530-mCherry, 375 ng each of plasmid DNA xCas9(3.7)-ABE(7.10) and xCas9(3.7)-BE4 were added. The contents of tubes 1 and 2 were mixed and incubated for 15 minutes at room temperature. We added 100 μl of the mixture to 1×10^5 iPSCs growing in one well of a 12-well plate in 1 ml of medium. 26 hours after lipofection, the growth medium was removed and fresh medium containing 15 % koSR, 30 ng/ml bFGF and thiazovivin was added. The red signal of mCherry protein was detected between 24 and 48 hours after lipofection using a Nikon TiE inverted fluorescence microscope.

Selection and subcloning of the resulting iPSC clones. 48 hours after delivery of the base editing system to iPSCs, cells were disaggregated and selected by flow cytometry using the red signal of the fluorescent protein mCherry on a Sony MA900 instrument. The resulting single cell suspension of iPSCs was seeded onto culture surfaces with mitotically inactivated mouse embryonic fibroblasts. For subsequent subcloning and analysis, individual iPSC colonies grown from isolated selected cells were mechanically harvested using glass capillaries and transferred to individual tissue culture wells pre-seeded with mitotically inactivated mouse embryonic fibroblasts.

Analysis of *LDLR* gene editing results in selected iPSC clones. The editing results of the c.530C>T and c.1054T>C substitutions in the *LDLR* sequence in selected iPSC clones were analysed by PCR with subsequent Sanger sequencing. Genomic DNA was isolated from iPSCs using QuickExtract DNA Extraction Solution reagent (Lucigen) according to the manufacturer's instructions. PCR was performed using the BioMaster HS-Taq PCR-Color (2×) kit (Biolabmix) on a T100 thermal cycler amplifier (Bio-Rad). Programme parameters were as follows: 98 °C – 30 seconds; 98 °C – 15 seconds, 60 °C – 15 seconds, 72 °C – 30 seconds, 35 cycles; 72 °C – 5 minutes. Primer sequences are given in Table S1 in Supplementary Material[2].

Oligonucleotides were synthesised in Biosset (https://www.biosset.com/). Sanger sequencing reactions were performed using Big Dye Terminator V.3.1 Cycle Sequencing Kit (Applied Biosystems) and analysed at the Genomics Collective Use Centre of the Siberian Branch of the Russian Academy of Sciences (http://www.niboch.nsc.ru/doku.php/sequest) on an ABI3130XL genetic analyser.

Mycoplasma and episome detection. Testing for mycoplasma and episome contamination was performed by PCR as previously described (Choppa et al., 1998; Okita et al., 2013). Primer sequences are provided in Table S1. Program parameters for episome detection were as follows: 95 °C – 5 minutes; 95 °C – 15 seconds, 58 °C – 15 seconds, 72 °C – 20 seconds, 35 cycles; 72 °C – 5 minutes; for mycoplasma detection: 95 °C – 3 minutes; 95 °C – 15 seconds, 67 °C – 15 seconds, 72 °C – 20 seconds, 35 cycles; 72 °C – 5 minutes.

Karyotyping. Genetically modified iPSCs were karyotyped at passage 15 according to the previously described protocol using DAPI banding according to the International System of Human Cytogenetic Nomenclature (Grigor'eva et al., 2024).

STR analysis. Short tandem repeat (STR) analysis was performed with Genoanalytica (https://www.genoanalytica.ru). iPSC DNA samples were genotyped by PCR using COrDIS EXPERT 26 direct amplification reagent kit (Russia) according to the manufacturer's protocol with subsequent separation of amplicons on a 3130 Genetic Analyzer capillary electrophoresis instrument (HITACHI, Applied Biosystems Group of The Applera Corporation, Japan, USA, Registration Certificate No. FSZ 2004/1586). Electropherograms with amplicon patterns are available on request from the authors.

Quantitative RT-PCR. To analyse pluripotency gene expression, RNA was isolated from cells using Trizol reagent (Thermo Fisher Scientific) according to the manufacturer's protocol. DNAase I treatment was performed using DNA-free kit (Thermo Fisher Scientific). RNA reverse transcription was performed using M-MuLV reverse transcriptase kit (Biolabmix) and random hexamer primers (Thermo Fisher Scientific) according to the manufacturer's protocol.

Real-time PCR was used to analyse the relative expression levels of pluripotency genes (*OCT4*, *NANOG*, *SOX2*) in genetically modified and initial isogenic iPSC lines with normalisation to two housekeeping genes: *ACTB* and *B2M*. Primer sequences are given in Table S1. Reactions were performed using the BioMaster HS-qPCR SYBR Blue (2×) kit

[2] Tables S1–S4 are available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Zueva_Engl_29_2.pdf

(Biolabmix) on a T100 thermal cycler amplifier (Bio-Rad). Program parameters were as follows: 98 °C – 30 seconds; 98 °C – 15 seconds, 60 °C – 15 seconds, 72 °C – 30 seconds, 35 cycles; 72 °C – 5 minutes. For each sample, three biological and two technical replicates were analysed.

The results were evaluated with qBase+ software (CellCarta https://cellcarta.com/genomic-data-analysis/) using the generalised ΔΔCt method, taking into account the reaction efficiency calculated from the results of a six-point calibration curve.

Immunofluorescence staining. Cell preparation and antibody precipitation were performed according to the previously described protocol (Vaskova et al., 2015). Briefly, cells were fixed in 4 % formaldehyde for 10 minutes, permeabilised with 0.5 % Triton X-100 for 30 minutes (this step was omitted for surface antigens) and blocked with 1 % bovine serum albumin solution in 1X PBS. All procedures were carried out at room temperature. Incubation with primary antibodies was performed overnight at 4 °C. Secondary antibodies were incubated with cells for 1 hour in the dark at room temperature. Cell nuclei were counterstained with DAPI. Visualisation and imaging of samples were performed on a Ti-E inverted fluorescence microscope (Nikon) using NIS Advanced Research software. The list of primary and secondary antibodies is given in Table S2.

Spontaneous differentiation *in vitro*. To confirm pluripotency of genetically modified iPSCs, they were induced to differentiate into embryoid bodies within 14 days. The resulting embryoid bodies were seeded onto Matrigel (Corning)-coated Chambered Coverglass plates (Thermo Fisher Scientific) and cultured until the 21st day.

Results

In this study, we obtained a genetically modified iPSC line ICGi036-A-1 with a corrected pathogenic allelic variant c.530C>T of the *LDLR* gene, originally derived from a compound heterozygous patient associated with FH. To correct this pathogenic variant, we used the base editing technique, for which we constructed an episomal vector pC9-sgRNA-mCherry containing a universal site for cloning any guide RNA spacers of the CRISPR/Cas systems and the gene encoding mCherry red fluorescent protein. The constructed plasmid allows visualisation of transfection results within 24–48 hours and selection of cells by flow cytometry upon delivery of the editing system into the cell.

Guide RNA spacer sequences were inserted into the universal vector containing mCherry. This resulted in two plasmids: pC9-530-mCherry to correct the c.530C>T substitution and pC9-1054_2-mCherry to correct the c.1054T>C substitution in the *LDLR* gene. The resulting guide RNA plasmids, together with plasmids encoding adenine and cytidine base editor sequences, were lipofected into the ICGi036-A iPSC line (Zakharova et al., 2022a). After 24 and 48 hours, cells with red mCherry protein fluorescence were detected, indicating successful delivery and function of the plasmids (Fig. 2*a*). We selected cells with mCherry fluorescence by flow cytometry and obtained 96 clones.

We examined selected clones by Sanger sequencing and identified the subclone 130S5, in which the c.530C>T position of the pathogenic allelic variant of the *LDLR* gene was corrected to the c.530C position corresponding to the reference sequence of the gene (Fig. 2*c*). The c.1054T>C substitution was not corrected.

Short tandem repeat (STR) analysis showed that the resulting iPSC line ICGi036-A-1 matched the original isogenic line ICGi036-A and patient mononuclear cells at 26 polymorphic loci (Zakharova et al., 2022a) (Table S3).

To confirm that the resulting genetically modified ICGi036-A-1 iPSCs retained self-renewal and pluripotency properties, we examined their pluripotency markers and ability to form three germ layer derivatives. Immunofluorescence staining with antibodies against the transcription factors OCT4, NANOG, SOX2 and the surface antigen SSEA4 showed that all colonies of 130S5 iPSCs were positive for these markers (Fig. 2*d*). Analysis of the relative expression of the pluripotency genes *OCT4*, *NANOG*, *SOX2* by real-time PCR showed that their expression levels did not differ significantly from the isogenic control, the original line ICGi036-A (Fig. 2*e*). At the same time, the genetically modified ICGi036-A-1 iPSC line displayed a significantly higher expression level of pluripotency genes compared to mononuclear cells (MNCs) from which the original isogenic ICGi036-A iPSC line was derived. Analysis of spontaneous differentiation of the ICGi036-A-1 line revealed a heterogeneous population of cells, among which immunofluorescence staining revealed derivatives positive for markers attributable to ectoderm (PAX6, NF200), mesoderm (CD90, αSMA) and entoderm (CK18, AFP) (Fig. 2*f*). Thus, the genetically modified iPSC line demonstrates the ability to give rise to derivatives of three primary germ layers, which is a key property of pluripotent stem cells.

Genetically modified ICGi036-A-1 iPSCs have a normal diploid karyotype: 46,XX (Fig. 2*b*). Analysis of 23 polymorphic short tandem repeat loci validated the identity of ICGi036-A-1 iPSCs and the original isogenic ICGi036-A iPSCs. The obtained iPSCs lacked episomal vectors at passage 10 and mycoplasma contamination at passage 25 (Fig. 2*g*).

The passport of the cell line obtained is shown in Table S4.

Discussion

Advances in the technologies associated with the generation and application of induced pluripotent stem cells (iPSCs) have opened up new avenues for biological research and biomedical innovations. iPSCs are being used for human disease modelling, high-throughput drug screening and the development of advanced biomedical cell therapy products due to their available minimally invasive derivation method, unlimited proliferative potential and the ability to direct differentiation into all adult cell types (Cerneckis et al., 2024).

Patient-specific cell models derived from differentiated iPSCs help to understand the molecular genetic basis of disease and to develop more effective targeted therapies (Brooks et al., 2022). For example, our group has previously obtained iPSCs from patients with familial hypercholesterolaemia carrying pathogenic allelic variants in the *LDLR* gene (Zakharova et al., 2022a–c). Using directed iPSC differentiation, we first derived endothelial cells with LDLR pathology, modelling FH (Zakharova et al., 2024a). We found that endotheliocytes derived from FH patient-specific iPSCs, although

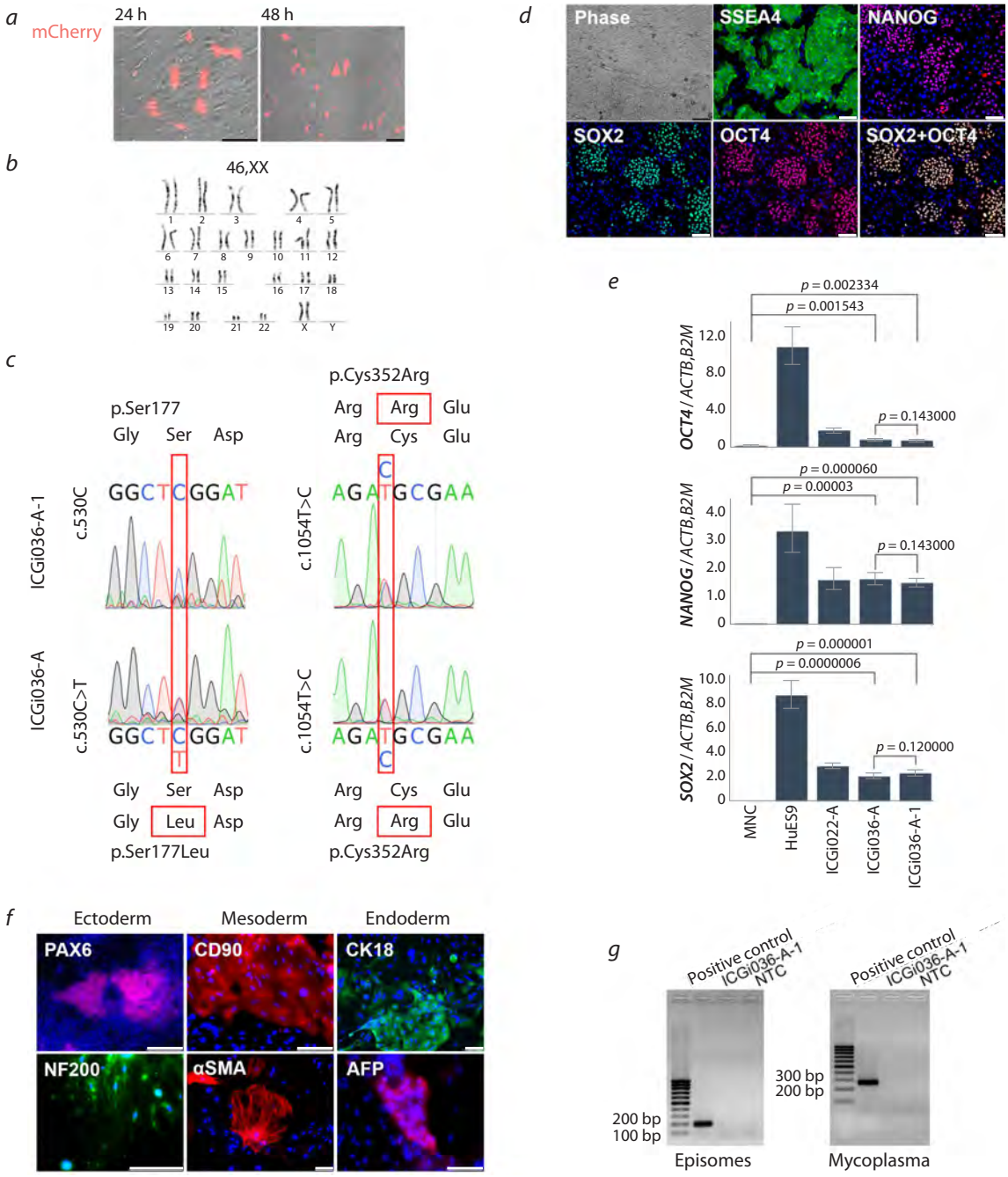


Fig. 2. Characteristics of the isogenic genetically modified iPSC line ICGi036-A-1 with the corrected pathogenic allelic variant c.530C>T.

a – mCherry protein fluorescence in iPSCs 24 and 48 h after transfection; *b* – karyotype in the resulting iPSC line; *c* – chromatograms demonstrating the corrected c.530C position in the ICGi036-A-1 line; *d* – colony morphology and staining with antibodies against pluripotency markers OCT4, NANOG, SOX2 and SSEA4 in the ICGi036-A-1 line; *e* – quantification of pluripotency gene expression of *OCT4*, *NANOG* and *SOX2* by real-time PCR; *f* – ICGi036-A-1 line retains the ability to differentiate into three germ layer derivatives; *g* – absence of episomal vectors and mycoplasma contamination in the ICGi036-A-1 line. NTC – no template control. The scale bar for all images is 100 μm.

not exposed to oxidative stress, have impaired LDLR receptor function and show signs of endothelial dysfunction. The data obtained contribute to the understanding of the molecular-genetic mechanisms underlying FH-related atherosclerosis.

The combination of iPSC technologies and genome editing methods provides isogenic cell models with similar genetic backgrounds, allowing physiological changes to be reliably studied in relevant differentiated cells (Niemitz, 2014; Omer et al., 2017; Kawatani et al., 2021; Liang et al., 2022; Wang et al., 2022; Chai et al., 2023; Bonnycastle et al., 2024). Isogenic iPSC systems can be generated either by altering the DNA sequence of healthy donor cells or by correcting a pathogenic allelic variant in patient-specific iPSCs (Nandy et al., 2023; Pavlova et al., 2023). In this work, we used CRISPR/Cas9-

mediated base editing to generate a genetically modified line ICGi036-A-1, an isogenic line of ICGi036-A iPSCs from a patient heterozygous for the pathogenic and likely pathogenic allelic variants of the *LDLR* gene with FH (Zakharova et al., 2022a). In the resulting genetically modified line ICGi036-A-1, the pathogenic allelic variant c.530C>T was corrected to the reference c.530C. The iPSC line retains pluripotency, has a normal karyotype and is identical to the original isogenic iPSC line ICGi036-A by the set of short tandem repeats.

In the resulting isogenic iPSC line ICGi036-A-1, the second position, c.1054T>C, remained uncorrected. The status of this allelic variant is currently defined as "pathogenic/likely pathogenic". The study of differentiated derivatives from heterozygous iPSCs with a corrected c.530C position will help to clarify the status of the c.1054T>C position.

We used the base editing method to correct single nucleotide substitutions in the *LDLR* gene. This method is more accurate than the classical CRISPR-Cas9 technology and allows targeted point substitutions in the DNA sequence by hydrolytic deamination, avoiding double-strand breaks (Hu et al., 2018; Porto et al., 2020). This method has already been successfully used to generate isogenic cell lines to model a number of diseases, namely Alzheimer's disease (*APOE4* gene sequence correction), sickle cell anaemia (β-globin gene), Hutchinson–Gilford progeria (lamin A gene), hereditary haemochromatosis (*HFE* gene) and some cancers (*TP53* gene) (Komor et al., 2016; Gaudelli et al., 2017; Koblan et al., 2021; Newby et al., 2021).

Verve Therapeutics is conducting the first clinical trial using base editing for FH therapy with the drug VERVE-101 starting from 2022 (ClinicalTrials.gov ID NCT05398029). *PCSK9* is the target gene for VERVE-101. Base editing of the gene disrupts PCSK9 protein synthesis, which in turn disrupts LDLR receptor degradation and leads to a reduction in plasma LDL concentrations (Rothgangl et al., 2021). However, this drug is not effective in FH patients with pathogenic allelic variants in *LDLR* that disrupt receptor synthesis or release to the cell surface, such as c.530C>T and c.1054T>C. In this context, the development of new FH cell models using safer cell genome editing systems remains a priority.

The isogenic iPSC cell lines of an FH patient we have obtained can be used to study dysfunction of relevant differentiated derivatives, such as endotheliocytes and hepatocytes, involved in FH manifestation, as well as to develop approaches for screening pharmacological compounds that are potential drugs for effective FH therapy.

There are known examples of clinical trials for some drugs selected using iPSC-based cell models, namely amyotrophic lateral sclerosis, progressive ossifying fibrodysplasia, Pendred syndrome and Alzheimer's disease (Okano, Morimoto, 2022). This approach appears to make economic sense as it allows first-line screening of compounds without the need for less relevant and more expensive animal models.

Despite the many advantages of using iPSCs for disease modelling, there are a number of limitations and challenges that need to be overcome in further research. A major drawback is that many of the differentiated derivatives obtained from iPSCs are functionally immature (Brooks et al., 2022). To address this problem, approaches are being developed to profile differentiated iPSC derivatives at the transcriptomic level at different stages and to identify robust criteria to assess the maturity of cell models (Subramanian et al., 2019; Kannan et al., 2021). Another challenge is to create reproducible, relevant models that integrate multiple iPSC-derived cell types and reflect disease pathogenesis under cell-cell interactions. Overcoming this involves using cellular organoids or assembloids (Brooks et al., 2022). However, this does not solve the problem of efficiency and reproducibility due to fluctuations caused by self-organisation within organoid systems. Bioprinting technology with a defined number of viable cells and their interaction pattern is currently being considered as a promising approach to improve the reproducibility of complex integrated cell models (Renner et al., 2020; Hofer, Lutolf, 2021; Lawlor et al., 2021).

We hope that using our cell models based on isogenic iPSC lines from patients with FH will help to better understand the mechanisms of disease progression and to develop effective drugs, increasing treatment efficacy and improving the quality and duration of patients' lives.

Conclusion

In the present study, we obtained and characterised in detail the genetically modified iPSC line ICGi036-A-1, which was derived from an isogenic line of an FH patient with compound heterozygosity for pathogenic and likely pathogenic allelic variants of the *LDLR* gene, namely c.530C>T (p.Ser177Leu) and c.1054T>C (p.Cys352Arg). The resulting line had a corrected c.530C>T position to the reference c.530C. The new iPSC line will be used to generate relevant differentiated derivatives to study FH manifestations and to develop FH-targeted therapeutic approaches.

References

Bauer D.E., Canver M.C., Orkin S.H. Generation of genomic deletions in mammalian cell lines via CRISPR/Cas9. *J Vis Exp.* 2015; 95:e52118. doi 10.3791/52118

Bonnycastle L.L., Swift A.J., Mansell E.C., Lee A., Winnicki E., Li E.S., Robertson C.C., Parsons V.A., Huynh T., Krilow C., Mohlke K.L., Erdos M.R., Narisu N., Collins F.S. Generation of human isogenic induced pluripotent stem cell lines with CRISPR prime editing. *Cris J.* 2024;7(1):53-67. doi 10.1089/crispr.2023.0066

Bourbon M., Alves A.C., Medeiros A.M., Silva S., Soutar A.K. Familial hypercholesterolaemia in Portugal. *Atherosclerosis*. 2008; 196(2):633-642. doi 10.1016/j.atherosclerosis.2007.07.019

Brooks I.R., Garrone C.M., Kerins C., Kiar C.S., Syntaka S., Xu J.Z., Spagnoli F.M., Watt F.M. Functional genomics and the future of iPSCs in disease modeling. *Stem Cell Rep*. 2022;17(5):1033-1047. doi 10.1016/j.stemcr.2022.03.019

Cerneckis J., Cai H., Shi Y. Induced pluripotent stem cells (iPSCs): molecular mechanisms of induction and applications. *Signal Transduct Target Ther.* 2024;9(1):112. doi 10.1038/s41392-024-01809-0

Chai A.C., Cui M., Chemello F., Li H., Chen K., Tan W., Atmanli A., McAnally J.R., Zhang Y., Xu L., Liu N., Bassel-Duby R., Olson E.N. Base editing correction of hypertrophic cardiomyopathy in human cardiomyocytes and humanized mice. *Nat Med.* 2023;29(2):401-411. doi 10.1038/s41591-022-02176-5

Choppa P.C., Vojdani A., Tagle C., Andrin R., Magtoto L. Multiplex PCR for the detection of *Mycoplasma fermentans, M. hominis* and *M. penetrans* in cell cultures and blood samples of patients with chronic fatigue syndrome. *Mol Cell Probes.* 1998;12(5):301-308. doi 10.1006/mcpr.1998.0186

Cowan C.A., Klimanskaya I., McMahon J., Atienza J., Witmyer J., Zucker J.P., Wang S., Morton C.C., McMahon A.P., Powers D., Melton D.A. Derivation of embryonic stem-cell lines from human blastocysts. *N Engl J Med.* 2004;350(13):1353-1356. doi 10.1056/nejmsr040330

Ezhov M.V., Bazhan S.S., Ershova A.I., Meshkov A.N., Sokolov A.A., Kukharchuk V.V., Gurevich V.S., Voevoda M.I., Sergienko I.V., Shakhtshneider E.V., Pokrovsky S.N., Konovalov G.A., Leontyeva I.V., Konstantinov V.O., Shcherbakova M.Yu., Zakharova I.N., Balakhonova T.V., Filippov A.E., Akhmedzhanov N.M., Aleksandrova O.Yu., Lipovetsky B.M. Clinical guidelines for familial hypercholesterolemia. *Ateroscleroz*. 2019;15(1):58-98 (in Russian)

Ference B.A., Ginsberg H.N., Graham I., Ray K.K., Packard C.J., Bruckert E., Hegele R.A., Krauss R.M., Raal F.J., Schunkert H., Watt G.F., Borén J., Fazio S., Horton J.D., Masana L., Nicholls S.J., Nordestgaard B.G., Van De Sluis B., Taskinen M.R., Tokgözoğlu L., Landmesser U., Laufs U., Wiklund O., Stock J.K., Chapman M.J., Catapano A.L. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J*. 2017;38(32):2459-2472. doi 10.1093/eurheartj/ehx144

Fularski P., Hajdys J., Majchrowicz G., Stabrawa M., Młynarska E., Rysz J., Franczyk B. Unveiling familial hypercholesterolemia – review, cardiovascular complications, lipid-lowering treatment and its efficacy. *Int J Mol Sci.* 2024;25(3):1637. doi 10.3390/ijms25031637

Gaudelli N.M., Komor A.C., Rees H.A., Packer M.S., Badran A.H., Bryson D.I., Liu D.R. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature.* 2017;551(7681):464-471. doi 10.1038/nature24644

Grigor'eva E.V., Malakhova A.A., Yarkova E.S., Minina J.M., Vyatkin Y.V., Nadtochy J.A., Khabarova E.A., Rzaev J.A., Medvedev S.P., Zakian S.M. Generation and characterization of two induced pluripotent stem cell lines (ICGi052-A and ICGi052-B) from a patient with frontotemporal dementia with parkinsonism-17 associated with the pathological variant c.2013T>G in the *MAPT* gene. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2024;28(7):679-687. doi 10.18699/vjgb-24-76

Gu J., Gupta R.N., Cheng H.K., Xu Y., Raal F.J. Current treatments for the management of homozygous familial hypercholesterolaemia: a systematic review and commentary. *Eur J Prev Cardiol.* 2024;31(15):1833-1849. doi 10.1093/eurjpc/zwae144

Harada-Shiba M. Impact of familial hypercholesterolemia diagnosis in real-world data. *J Atheroscler Thromb.* 2023;30(10):1303. doi 10.5551/jat.ED241

Hendricks-Sturrup R.M., Clark-Locascio J., Lu C.Y. A global review on the utility of genetic testing for familial hypercholesterolemia. *J Pers Med.* 2020;10(2):23. doi 10.3390/jpm10020023

Hofer M., Lutolf M.P. Engineering organoids. *Nat Rev Mater.* 2021;6(5):402-420. doi 10.1038/s41578-021-00279-y

Hopkins P.N., Toth P.P., Ballantyne C.M., Rader D.J. Familial hypercholesterolemias: prevalence, genetics, diagnosis and screening recommendations from the national lipid association expert panel on familial hypercholesterolemia. *J Clin Lipidol.* 2011;5(3):S9. doi 10.1016/j.jacl.2011.03.452

Hu J.H., Miller S.M., Geurts M.H., Tang W., Chen L., Sun N., Zeina C.M., Gao X., Rees H.A., Lin Z., Liu D.R. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature.* 2018;556(7699):57-63. doi 10.1038/nature26155

Huang C.C., Niu D.M., Charng M.J. Genetic analysis in a Taiwanese cohort of 750 index patients with clinically diagnosed familial hypercholesterolemia. *J Atheroscler Thromb.* 2022;29(5):639-653. doi 10.5551/jat.62773

Jannes C.E., Santos R.D., de Souza Silva P.R., Turolla L., Gagliardi A.C.M., Marsiglia J.D.C., Chacra A.P., Miname M.H., Rocha V.Z., Filho W.S., Krieger J.E., Pereira A.C. Familial hypercholesterolemia in Brazil: cascade screening program, clinical and genetic aspects. *Atherosclerosis.* 2015;238(1):101-107. doi 10.1016/j.atherosclerosis.2014.11.009

Kannan S., Farid M., Lin B.L., Miyamoto M., Kwon C. Transcriptomic entropy benchmarks stem cell-derived cardiomyocyte maturation against endogenous tissue at single cell level. *PLoS Comput Biol.* 2021;17(9):e1009305. doi 10.1371/journal.pcbi.1009305

Kawatani K., Nambara T., Nawa N., Yoshimatsu H., Kusakabe H., Hirata K., Tanave A., Sumiyama K., Banno K., Taniguchi H., Arahori H., Ozono K., Kitabatake Y. A human isogenic iPSC-derived cell line panel identifies major regulators of aberrant astrocyte proliferation in Down syndrome. *Commun Biol.* 2021;4(1):730. doi 10.1038/s42003-021-02242-7

Koblan L.W., Erdos M.R., Wilson C., Cabral W.A., Levy J.M., Xiong Z.M., Tavarez U.L., Davison L.M., Gete Y.G., Mao X., Newby G.A., Doherty S.P., Narisu N., Sheng Q., Krilow C., Lin C.Y., Gordon L.B., Cao K., Collins F.S., Brown J.D., Liu D.R. *In vivo* base editing rescues Hutchinson-Gilford progeria syndrome in mice. *Nature.* 2021;589(7843):608-614. doi 10.1038/s41586-020-03086-7

Komor A.C., Kim Y.B., Packer M.S., Zuris J.A., Liu D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature.* 2016;533(7603):420-424. doi 10.1038/nature17946

Lawlor K.T., Vanslambrouck J.M., Higgins J.W., Chambon A., Bishard K., Arndt D., Er P.X., Wilson S.B., Howden S.E., Tan K.S., Li F., Hale L.J., Shepherd B., Pentoney S., Presnell S.C., Chen A.E., Little M.H. Cellular extrusion bioprinting improves kidney organoid reproducibility and conformation. *Nat Mater.* 2021;20(2):260-271. doi 10.1038/s41563-020-00853-9

Liang Y., Sun X., Duan C., Zhou Y., Cui Z., Ding C., Gu J., Mao S., Ji S., Chan H.F., Tang S., Chen J. Generation of a gene-corrected human iPSC line (CSUASOi004-A-1) from a retinitis pigmentosa patient with heterozygous c.2699G>A mutation in the *PRPF6* gene. *Stem Cell Res.* 2022;64:103572. doi 10.1016/j.scr.2022.102911

Malakhova A.A., Grigor'eva E.V., Pavlova S.V., Malankhanova T.B., Valetdinova K.R., Vyatkin Y.V., Khabarova E.A., Rzaev J.A., Zakian S.M., Medvedev S.P. Generation of induced pluripotent stem cell lines ICGi021-A and ICGi022-A from peripheral blood mononuclear cells of two healthy individuals from Siberian population. *Stem Cell Res.* 2020;48:101952. doi 10.1016/j.scr.2020.101952

Meshkov A., Ershova A., Kiseleva A., Zotova E., Sotnikova E., Petukhova A., Zharikova A., Malyshev P., Rozhkova T., Blokhina A., Limonova A., Ramensky V., Divashuk M., Khasanova Z., Bukaeva A., Kurilova O., Skirko O., Pokrovskaya M., Mikova V., Snigir E., Akinshina A., Mitrofanov S., Kashtanova D., Makarov V., Kukharchuk V., Boytsov S., Yudin S., Drapkina O. The *LDLR*, *APOB*, and *PCSK9* variants of index patients with familial hypercholesterolemia in Russia. *Genes.* 2021;12(1):66. doi 10.3390/genes12010066

Mohd Nor N.S., Al-Khateeb A.M., Chua Y.A., Mohd Kasim N.A., Mohd Nawawi H. Heterozygous familial hypercholesterolaemia in a pair of identical twins: a case report and updated review. *BMC Pediatr.* 2019;19(1):106. doi 10.1186/S12887-019-1474-y/tables/2

Nandy K., Babu D., Rani S., Joshi G., Ijee S., George A., Palani D., Premkumar C., Rajesh P., Vijayanand S., David E., Murugesan M., Velayudhan S.R. Efficient gene editing in induced pluripotent stem cells enabled by an inducible adenine base editor with tunable expression. *Sci Rep.* 2023;13(1):21953. doi 10.1038/s41598-023-42174-2

Nazarenko M.S., Sleptcov A.A., Zarubin A.A., Salakhov R.R., Shevchenko A.I., Tmoyan N.A., Elisaphenko E.A., Zubkova E.S., Zheltysheva N.V., Ezhov M.V., Kukharchuk V.V., Parfyonova Y.V., Zakian S.M., Zakharova I.S. Calling and phasing of single-nucleotide and structural variants of the *LDLR* gene using Oxford Nano-

pore MinION. *Int J Mol Sci.* 2023;24(5):4471. doi 10.3390/ijms24054471

Newby G.A., Yen J.S., Woodard K.J., Mayuranathan T., Lazzarotto C.R., Li Y., Sheppard-Tillman H., Porter S.N., Yao Y., Mayberry K., Everette K.A., Jang Y., Podracky C.J., Thaman E., Lechauve C., Sharma A., Henderson J.M., Richter M.F., Zhao K.T., Miller S.M., Wang T., Koblan L.W., McCaffrey A.P., Tisdale J.F., Kalfa T.A., Pruett-Miller S.M., Tsai S.Q., Weiss M.J., Liu D.R. Base editing of haematopoietic stem cells rescues sickle cell disease in mice. *Nature*. 2021;595(7866):295-302. doi 10.1038/S41586-021-03609-w

Niemitz E. Isogenic iPSC-derived models of disease. *Nat Genet.* 2014;46(1):7. doi 10.1038/ng.2864

Okano H., Morimoto S. iPSC-based disease modeling and drug discovery in cardinal neurodegenerative disorders. *Cell Stem Cell*. 2022;29(2):189-208. doi 10.1016/j.stem.2022.01.007

Okita K., Yamakawa T., Matsumura Y., Sato Y., Amano N., Watanabe A., Goshima N., Yamanaka S. An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. *Stem Cells*. 2013;31(3):458-466. doi 10.1002/stem.1293

Omer L., Hudson E.A., Zheng S., Hoying J.B., Shan Y., Boyd N.L. CRISPR correction of a homozygous low-density lipoprotein receptor mutation in familial hypercholesterolemia induced pluripotent stem cells. *Hepatol Commun.* 2017;1(9):886-898. doi 10.1002/hep4.1110

Palacios L., Grandoso L., Cuevas N., Olano-Martín E., Martinez A., Tejedor D., Stef M. Molecular characterization of familial hypercholesterolemia in Spain. *Atherosclerosis*. 2012;221(1):137-142. doi 10.1016/j.atherosclerosis.2011.12.021

Pavlova S.V., Shayakhmetova L.S., Pronyaeva K.A., Shulgina A.E., Zakian S.M., Dementyeva E.V. Generation of induced pluripotent stem cell lines ICGi022-A-3, ICGi022-A-4, and ICGi022-A-5 with p.Asn515del mutation introduced in *MYBPC3* using CRISPR/Cas9. *Russ J Dev Biol.* 2023;54:96-103. doi 10.1134/S1062360423010113

Porto E.M., Komor A.C., Slaymaker I.M., Yeo G.W. Base editing: advances and therapeutic opportunities. *Nat Rev Drug Discov.* 2020;19(12):839-859. doi 10.1038/s41573-020-0084-6

Ray K.K., Ference B.A., Séverin T., Blom D., Nicholls S.J., Shiba M.H., Almahmeed W., Alonso R., Daccord M., Ezhov M., Olmo R.F., Jankowski P., Lanas F., Mehta R., Puri R., Wong N.D., Wood D., Zhao D., Gidding S.S., Virani S.S., Lloyd-Jones D., Pinto F., Perel P., Santos R.D. World Heart Federation Cholesterol Roadmap 2022. *Glob Heart*. 2022;17(1):75. doi 10.5334/gh.1154

Ray K.K., Pillas D., Hadjiphilippou S., Khunti K., Seshasai S.R.K., Vallejo-Vaz A.J., Neasham D., Addison J. Premature morbidity and mortality associated with potentially undiagnosed familial hypercholesterolemia in the general population. *Am J Prev Cardiol.* 2023;15:100580. doi 10.1016/j.ajpc.2023.100580

Renner H., Grabos M., Becker K.J., Kagermeier T.E., Wu J., Otto M., Peischard S., Zeuschner D., Tsytsyura Y., Disse P., Klingauf J., Leidel S.A., Seebohm G., Schöler H.R., Bruder J.M. A fully automated high-throughput workflow for 3D-based chemical screening in human midbrain organoids. *eLife*. 2020;9:e52904. doi 10.7554/eLife.52904

Rothgangl T., Dennis M.K., Lin P.J.C., Oka R., Witzigmann D., Villiger L., Qi W., Hruzova M., Kissling L., Lenggenhager D., Borrelli C., Egli S., Frey N., Bakker N., Walker J.A., Kadina A.P., Victorov D.V., Pacesa M., Kreutzer S., Kontarakis Z., Moor A., Jinek M., Weissman D., Stoffel M., van Boxtel R., Holden K., Pardi N., Thöny B., Häberle J., Tam Y.K., Semple S.C., Schwank G. *In vivo* adenine base editing of *PCSK9* in macaques reduces LDL cholesterol levels. *Nat Biotechnol.* 2021;39(8):949-957. doi 10.1038/s41587-021-00933-4

Semenova A.E., Sergienko I.V., García-Giustiniani D., Monserrat L., Popova A.B., Nozadze D.N., Ezhov M.V. Verification of underlying genetic cause in a cohort of Russian patients with familial hypercholesterolemia using targeted next generation sequencing. *J Cardiovasc Dev Dis.* 2020;7(2):16. doi 10.3390/jcdd7020016

Setia N., Saxena R., Arora A., Verma I.C. Spectrum of mutations in homozygous familial hypercholesterolemia in India, with four novel mutations. *Atherosclerosis*. 2016;255:31-36. doi 10.1016/j.atherosclerosis.2016.10.028

Shakhtshneider E., Ivanoshchuk D., Timoshchenko O., Orlov P., Semaev S., Valeev E., Goonko A., Ladygina N., Voevoda M. Analysis of rare variants in genes related to lipid metabolism in patients with familial hypercholesterolemia in Western Siberia (Russia). *J Pers Med.* 2021;11(11):1232. doi 10.3390/jpm11111232

Sharifi M., Walus-Miarka M., Idzior-Waluś B., Malecki M.T., Sanak M., Whittall R., Li K.W., Futema M., Humphries S.E. The genetic spectrum of familial hypercholesterolemia in south-eastern Poland. *Metabolism.* 2016;65(3):48-53. doi 10.1016/j.metabol.2015.10.018

Siegner S.M., Karasu M.E., Schröder M.S., Kontarakis Z., Corn J.E. PnB Designer: a web application to design prime and base editor guide RNAs for animals and plants. *BMC Bioinformatics*. 2021;22(1):101. doi 10.1186/s12859-021-04034-6

Subramanian A., Sidhom E.H., Emani M., Vernon K., Sahakian N., Zhou Y., Kost-Alimova M., Slyper M., Waldman J., Dionne D., Nguyen L.T., Weins A., Marshall J.L., Rosenblatt-Rosen O., Regev A., Greka A. Single cell census of human kidney organoids shows reproducibility and diminished off-target cells after transplantation. *Nat Commun.* 2019;10(1):5462. doi 10.1038/S41467-019-13382-0

Südhof T.C., Goldstein J.L., Brown M.S., Russell D.W. The LDL receptor gene: a mosaic of exons shared with different proteins. *Science*. 1985;228(4701):815-822. doi 10.1126/science.2988123

Talmud P.J., Futema M., Humphries S.E. The genetic architecture of the familial hyperlipidaemia syndromes: rare mutations and common variants in multiple genes. *Curr Opin Lipidol.* 2014;25(4):274-281. doi 10.1097/MOL.0000000000000090

Thormaehlen A.S., Schuberth C., Won H.H., Blattmann P., Joggerst-Thomalla B., Theiss S., Asselta R., Duga S., Merlini P.A., Ardissino D., Lander E.S., Gabriel S., Rader D.J., Peloso G.M., Pepperkok R., Kathiresan S., Runz H. Systematic cell-based phenotyping of missense alleles empowers rare variant association studies: a case for *LDLR* and myocardial infarction. *PLoS Genet.* 2015;11(2):e1004855. doi 10.1371/journal.pgen.1004855

Tichý L., Freiberger T., Zapletalová P., Soška V., Ravčuková B., Fajkusová L. The molecular basis of familial hypercholesterolemia in the Czech Republic: spectrum of *LDLR* mutations and genotype-phenotype correlations. *Atherosclerosis*. 2012;223(2):401-408. doi 10.1016/j.atherosclerosis.2012.05.014

Vaskova E.A., Medvedev S.P., Sorokina A.E., Nemudryy A.A., Elisaphenko E.A., Zakharova I.S., Shevchenko A.I., Kizilova E.A., Zhelezova A.I., Evshin I.S., Sharipov R.N., Minina J.M., Zhdanova N.S., Khegay I.I., Kolpakov F.A., Sukhikh G.T., Pokushalov E.A., Karaskov A.M., Vlasov V.V., Ivanova L.N., Zakian S.M. Transcriptome characteristics and X-chromosome inactivation status in cultured rat pluripotent stem cells. *Stem Cells Dev.* 2015;24(24):2912-2924. doi 10.1089/scd.2015.0204

Wang H., Luo Y., Li J., Guan J., Yang S., Wang Q. Generation of a gene corrected human isogenic iPSC line (CPGHi001-A-1) from a hearing loss patient with the TMC1 p.M418K mutation using CRISPR/Cas9. *Stem Cell Res.* 2022;60:102736. doi 10.1016/j.scr.2022.102736

Zakharova I.S., Shevchenko A.I., Tmoyan N.A., Elisaphenko E.A., Zubkova E.S., Sleptcov A.A., Nazarenko M.S., Ezhov M.V., Kukharchuk V.V., Parfyonova Y.V., Zakian S.M. Induced pluripotent stem

cell line ICGi036-A generated by reprogramming peripheral blood mononuclear cells from a patient with familial hypercholesterolemia caused due to compound heterozygous p.Ser177Leu/p.Cys352Arg mutations in *LDLR*. *Stem Cell Res.* 2022a;59:102653. doi 10.1016/j.scr.2022.102653

Zakharova I.S., Shevchenko A.I., Tmoyan N.A., Elisaphenko E.A., Kalinin A.P., Sleptcov A.A., Nazarenko M.S., Ezhov M.V., Kukharchuk V.V., Parfyonova Y.V., Zakian S.M. Induced pluripotent stem cell line ICGi037-A, obtained by reprogramming peripheral blood mononuclear cells from a patient with familial hypercholesterolemia due to heterozygous p.Trp443Arg mutations in *LDLR*. *Stem Cell Res.* 2022b;60:102703. doi 10.1016/j.scr.2022.102703

Zakharova I.S., Shevchenko A.I., Tmoyan N.A., Elisaphenko E.A., Zubkova E.S., Sleptcov A.A., Nazarenko M.S., Ezhov M.V., Kukharchuk V.V., Parfyonova Y.V., Zakian S.M. Induced pluripotent stem cell line ICGi038-A, obtained by reprogramming peripheral blood mononuclear cells from a patient with familial hypercholesterolemia due to compound heterozygous c.1246C>T/c.940+3_940+6del mutations in *LDLR*. *Stem Cell Res.* 2022c;60:102702. doi 10.1016/j.scr.2022.102702

Zakharova I.S., Shevchenko A.I., Arssan M.A., Sleptcov A.A., Nazarenko M.S., Zarubin A.A., Zheltysheva N.V., Shevchenko V.A., Tmoyan N.A., Saaya S.B., Ezhov M.V., Kukharchuk V.V., Parfyonova Y.V., Zakian S.M. IPSC-derived endothelial cells reveal LDLR dysfunction and dysregulated gene expression profiles in familial hypercholesterolemia. *Int J Mol Sci.* 2024a;25(2):689. doi 10.3390/ijms25020689

Zakharova I.S., Shevchenko A.I., Zakian S.M. Familial hypercholesterolemia: current insight and challenges in its modelling. *Pisma v Vavilovskii Zhurnal Genetiki i Selektsii = Letters to Vavilov Journal of Genetics and Breeding.* 2024b;10(1):5-14. doi 10.18699/letvjgb-2024-10-2 (in Russian)

Conflict of interest. The authors declare no conflict of interest.
Received November 25, 2024. Revised December 4, 2024. Accepted December 9, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):200-209

doi 10.18699/vjgb-25-23

Vitamin C content and profile of ascorbate metabolism gene expression in green leaves and bleached parts of the pseudostem of leek (*Allium porrum* L.) F_1 hybrids

M.A. Filyushin, T.M. Seredin, A.V. Shchennikova, E.Z. Kochieva

Federal Research Centre "Fundamentals of Biotechnology" of the Russian Academy of Sciences, Moscow, Russia
michel7753@mail.ru

Abstract. Leek (*Allium porrum* L.) is an economically important vegetable crop of the family Amaryllidaceae with a wide range of medicinal and nutritional properties, in part due to the accumulation of vitamin C (L-ascorbic acid, ascorbate). Ascorbate is an organic water-soluble compound, which performs many functions in plant cell metabolism, including as one of an effective antioxidant in plant cell responses to biotic and abiotic stress factors. Ascorbate metabolism includes biosynthesis (mainly the L-galactose pathway) and recycling (reduction of oxidized forms to ascorbate). The gene networks that determine ascorbate metabolism in leek plants are poorly understood. In this work, crosses of leek varieties/lines were carried out. Accessions of F_1 hybrids were characterized for seed germination rate, ascorbate content and expression of ascorbate biosynthesis (*PGI*, *PMI*, *PMM*, *VTC1b*, *GME1*, *GME2*, *VTC2*, *GPP*, *GalDH*, *GalLDH*) and recycling (*APX1*, *APX2*, *MDHAR1*, *MDHAR4*, *MDHAR5*, *DHAR2*, *GR*) genes in seedlings, as well as green leaves and bleached stem parts of the adult plant. A search for correlations between the level of expression of ascorbate metabolism genes and the amount of vitamin C in leeks was also carried out. It was shown that the studied hybrids are characterized by high (89–100 %) seed germination, with the exception of the hybrid from the 74 × Alligator cross (55 %). An increased level of expression of the *VTC2*, *MDHAR1*, *MDHAR4* and/or *MDHAR5* genes was detected in the seedlings and green leaves of nine F_1 hybrids, which allowed us to consider these samples promising in terms of possible stress resistance. Four hybrids that were characterized by the lowest (33 × 30, 74 × Alligator) and highest (81 × 95, 36 × 38) ascorbate content in seedlings were selected for a further detailed analysis of adult plants for the content of soluble sugars and ascorbate, gene expression and morphological characteristics (length, thickness and weight of the false stem). It was confirmed that green leaves of the 36 × 38 and 81 × 95 hybrids contain significantly more ascorbate than the 33 × 30 and 74 × Alligator hybrids. In all four hybrids, the ascorbate content was significantly lower in the bleached stems than in the green leaves. Accessions 36 × 38 and 81 × 95 were also characterized by the highest amount of soluble sugars in the bleached part of the false stem used for food. In addition, the false stem formed by the 81 × 95 hybrid was larger and heavier than the stems of the other three hybrids. A direct dependence of ascorbate content on the transcript level of ascorbate recycling genes (*APX2*, *MDHAR1*, *MDHAR4*) in green leaves was revealed, which can be used in the breeding of stress-resistant leek hybrids with a high content of vitamin C.

Key words: leek; *Allium porrum* L.; vitamin C; ascorbate biosynthesis genes; ascorbate recycling genes; soluble sugars; gene expression

For citation: Filyushin M.A., Seredin T.M., Shchennikova A.V., Kochieva E.Z. Vitamin C content and profile of ascorbate metabolism gene expression in green leaves and bleached parts of the pseudostem of leek (*Allium porrum* L.) F_1 hybrids. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):200-209. doi 10.18699/vjgb-25-23

Funding. This research was funded by the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement No. 075-15-2022-318 on April 20, 2022 on providing a grant in the form of subsidies from the Federal budget of the Russian Federation. The grant was provided as state support for the creation and development of a World-class Scientific Center "Agrotechnologies for the Future".

Содержание витамина С и профиль экспрессии генов метаболизма аскорбата в зеленых листьях и отбеленной части ложного стебля гибридов F_1 лука-порея (*Allium porrum* L.)

М.А. Филюшин, Т.М. Середин, А.В. Щенникова, Е.З. Кочиева

Федеральный исследовательский центр «Фундаментальные основы биотехнологии» Российской академии наук, Москва, Россия
michel7753@mail.ru

Аннотация. Лук-порей (*Allium porrum* L.) – экономически важная овощная культура семейства Amaryllidaceae с широким спектром лечебных и питательных качеств, в том числе благодаря накоплению витамина С (L-аскорбиновая кислота, аскорбат). Аскорбат – органическое водорастворимое соединение, которое выполняет множество функций в метаболизме клеток растения, в том числе участвует в качестве эффективного антиоксиданта в ответе растительных клеток на биотические и абиотические стрессовые факторы. Метаболизм аскорбата включает биосинтез (преимущественно L-галактозный путь) и рециклинг (восстановление окисленных форм до аскорбата). Генные сети, определяющие метаболизм аскорбата в растениях лука-порея, малоизученны. В настоящей работе гибриды F_1, полученные от 13 скрещиваний образцов лука-порея отечественной и зарубежной селекции, охарактеризованы по всхожести семян, содержанию аскорбата и экспрессии генов биосинтеза (*PGI*, *PMI*, *PMM*, *VTC1b*, *GME1*, *GME2*, *VTC2*, *GPP*, *GalDH* и *GalLDH*) и рециклинга (*APX1*, *APX2*, *MDHAR1*, *MDHAR4*, *MDHAR5*, *DHAR2*, *GR*) аскорбата в проростках, зеленых листьях и отбеленной части стебля взрослых растений. Также проведен поиск корреляций между уровнем экспрессии генов метаболизма аскорбата и количеством витамина С у лука-порея. Показано, что изучаемым гибридам присуща высокая (89–100 %) всхожесть семян, за исключением гибрида от скрещивания 74 × Аллигатор (55 %). В проростках и зеленых листьях образцов девяти гибридов F_1 выявлен повышенный уровень экспрессии генов *VTC2*, *MDHAR1*, *MDHAR4* и/или *MDHAR5*, что позволило считать эти образцы перспективными с точки зрения возможной стрессоустойчивости. Четыре гибрида, которые характеризовались наименьшим (33 × 30, 74 × Аллигатор) и наибольшим (81 × 95, 36 × 38) содержанием аскорбата в проростках, были выбраны для дальнейшего детального анализа взрослых растений на содержание растворимых сахаров и аскорбата, уровень экспрессии генов метаболизма аскорбата и морфологические характеристики (длина, толщина и масса ложного стебля). Выявлено, что зеленые листья гибридов 36 × 38 и 81 × 95 содержат существенно больше аскорбата, чем 33 × 30 и 74 × Аллигатор. В отбеленных стеблях всех четырех гибридов содержание аскорбата было значительно ниже, чем в зеленых листьях. Образцы 36 × 38 и 81 × 95 отличались также наибольшим количеством растворимых сахаров в отбеленной части ложного стебля, употребляемой в пищу. Гибрид 81 × 95 формировал ложный стебель, превышающий по размеру и массе стебель остальных трех гибридов. В зеленых листьях содержание аскорбата положительно коррелировало с уровнем экспрессии генов рециклинга аскорбата (*APX2*, *MDHAR1*, *MDHAR4*), что может быть использовано в селекции стрессоустойчивых гибридов лука-порея с повышенным содержанием витамина С.

Ключевые слова: лук-порей; *Allium porrum* L.; витамин С; гены биосинтеза аскорбата; гены рециклинга аскорбата; растворимые сахара; экспрессия генов

Introduction

Onion species, including leek (*Allium porrum* L.), have a wide range of nutritional and medicinal properties. Leek, which is also known as *A. ampeloprasum* var. *porrum* (L.) Gay, is considered an economically important vegeTable crop valued not only for its nutritional qualities but also for its antibacterial, anticancer, cardioprotective, and antioxidant properties (Celebi-Toprak, Alan, 2021).

Leek breeding is aimed at increasing the length, thickness, density and weight of the edible white (blanched) stem, improving its taste and dietary qualities, as well as increasing seed germination and resistance to stress factors and bolting (Swamy, Gowda, 2006; Celebi-Toprak, Alan, 2021). Soluble sugars (5.0–11.2 g/100 g raw weight) and vitamin C (L-ascorbic acid, ascorbate, AA) (0.9–3.6 mg/g dry weight) give leeks a delicate and sweet taste. During storage, the amount of AA in the blanched part of the false stem increases by more than 1.5 times (Lundegårdh et al., 2008; Grzelak-Błaszczyk et al., 2011; Bernaert et al., 2012; Bernaert, 2013). Both types of metabolites play significant roles in the plant's defense responses to stress factors (Yamada, Osakabe, 2018; Broad et al., 2020; Qi et al., 2020), and vitamin C is also important for human health (Hemilä, 2017).

The presence of ascorbate is also positively associated with the post-harvest shelf life of the blanched stem, since, unlike onions, the cut stem of leeks is not in a state of physiological dormancy and quickly deteriorates (Bernaert, 2013). Furthermore, ascorbate and soluble sugar-dependent signaling pathways largely determine plant ontogeny (Considine, Foyer, 2014; Yoon et al., 2021) and may therefore positively influence leek pseudostem size.

A comparison of oil extracts of leeks and another equally popular onion crop, garlic (*A. sativum* L.), showed the superiority of leeks in antioxidant activity, largely due to higher accumulations of vitamin C (Lemma et al., 2022). According to several studies, the amount of vitamin C in green leaves and the edible blanched part of the stem (false stem) of leeks can vary within 2.8–8.5 and 0.9–3.6 mg/g dry weight, respectively (Lundegårdh et al., 2008; Bernaert et al., 2012).

Vitamin C is an organic water-soluble compound that is not synthesized by humans, but is a necessary part of human diet and comes from plant foods, where the amount of ascorbate depends on the species/variety, tissue/organ, and plant growing/storage conditions (Bulley, Laing, 2016). In addition to its benefits to humans, ascorbate is involved in many aspects of plant development, including the regulation of cellular metabolism, and is also an effective antioxidant, since it is present in the cell in sufficient quantities and carries out fine regulation of the presence of various free radicals, reacting with them (Arrigoni, De Tullio, 2002).

The importance of ascorbate for the plant is emphasized by the fact that its synthesis occurs through several unique pathways, the dominant of which is the Smirnov–Wheeler L-galactose pathway, which undergoes eight stages of conversion of the initial substrate (D-fructose-6-P) into L-ascorbic acid (Bulley, Laing, 2016). The biosynthetic pathway includes reactions catalyzed sequentially by glucose-6-phosphate isomerase (PGI), mannose-6-phosphate isomerase (PMI),

phosphomannomutase (PMM), GDP-mannose pyrophosphorylase (VTC1), GDP-mannose 3′,5′-epimerase (GME), GDP-L-galactose phosphorylase (VTC2, VTC5), L-galactose-1-phosphate phosphatase (GPP), L-galactose dehydrogenase (GalDH), and L-galactono-1,4-lactone dehydrogenase (GalLDH) (Bulley, Laing, 2016).

Ascorbate recycling occurs as follows. When interacting with active forms of oxygen, as well as under the action of ascorbate peroxidases (APX) and ascorbate oxidases (AO), ascorbate is oxidized and converted into monodehydroascorbic acid (MDHA), which can be broken down into dehydroascorbic acid (DHA) and ascorbate (Bulley, Laing, 2016). Both oxidized forms (MDHA and DHA) can be reduced to ascorbate by monodehydroascorbate reductase (MDHAR) and dehydroascorbate reductase (DHAR), respectively (Bulley, Laing, 2016). Thus, the concentration of ascorbate in plant tissue is determined by the balance between the synthesis of vitamin C, its recycling and the catabolism of oxidized forms.

Gene networks of ascorbate metabolism are studied in various plant species, including cultivated species. For example, the expression level of the *VTC2* gene has been shown to be positively associated with the amount of ascorbate in plant tissue and with plant resistance to abiotic stress factors; this fact is used in breeding aimed at increasing the vitamin C content (Ali et al., 2019; Broad et al., 2020). In the model species *Arabidopsis thaliana* L., a paralog of *VTC2*, the *VTC5* gene, was found, but the rate of the L-galactose pathway is determined predominantly by the activity of *VTC2* (Dowdle et al., 2007).

Ascorbate recycling genes are studied more in terms of their role in determining plant stress tolerance. Exposure to various stress factors (both abiotic and biotic) leads to changes in the expression level of *MDHAR* genes and the activity of the enzymes they encode (Leterrier et al., 2005; Dowdle et al., 2007; Gill, Tuteja, 2010; Feng et al., 2014; Lanubile et al., 2015; Zhang et al., 2015; García et al., 2020). Overexpression of *MDHAR* genes has a positive effect on salt stress tolerance (Sultana et al., 2012; Qi et al., 2020). However, in ripe tomato fruits, this significantly reduces the ascorbate content (Haroldsen et al., 2011).

Gene networks determining ascorbate metabolism in leek plants (*A. porrum*) are poorly studied. The polymorphism and expression profile of the *VTC2* gene have been characterized, including in response to cold stress, and a correlation between the *VTC2* expression level and the ascorbate content in green leaves (positive) and the white part (negative) has been shown (Anisimova et al., 2021a, b). Three *MDHAR* genes have been identified and characterized; the level of transcripts of one of them, *MDHAR4*, positively correlates with the content of AA in the white part and green leaves of the plant (Filyushin et al., 2021). No other publications on the characterization of AA metabolism genes in leeks have been found.

The aim of this work was to obtain F_1 leek hybrids from 13 crosses of leek accessions of domestic and foreign selection and compare them by vitamin C content and the expression level of genes of AA biosynthesis (*PGI*, *PMI*, *PMM*, *VTC1b*, *GME1*, *GME2*, *VTC2*, *GPP*, *GalDH*, *GalLDH*) and recycling (*APX1*, *APX2*, *MDHAR1*, *MDHAR4*, *MDHAR5*, *DHAR2*, *GR*) in sprouts, green leaves and the blanched part of the stem of adult plants. A possible correlation between the expression level of the analyzed genes and the amount of vitamin C in leek tissues was assessed.

Materials and methods

In the study, we used seeds of F_1 hybrids obtained from 13 crosses (2022) of leek varieties/lines of domestic and foreign selection (Table 1). F_1 seeds were sown (50 pcs. from

Table 1. List of crosses of leek accessions

Cross name	Parents	F_1 seed germination, %
28 × 30	Goliaf × Vesta	100
85 × 99	Kilima × Gayia	98
94 × 113	Electro × Kazachok	99
62 × 80	k-2197 × Amerikanskiy flag	95
85 × 86	Kilima × Balder	100
12 × 13	Herbstreusen × Amarello	100
81 × 95	Winterreusen × Siegfried	100
36 × 38	Slon × Elefant	100
68 × 43	k-2120 × k-2054	89
74 × Alligator	k-2042 × Alligator	55
38 × 121	Elefant × Mateyko	100
68 × 44	k-2120 × k-2026	96
33 × 30	Letniy briz × Vesta	100

each cross) in the soil (experimental climate control facility, FRC Biotechnology RAS; day/night – 16 h/8 h, 23 °C/21 °C), germination was assessed, and the resulting plants were used in further analysis.

The aboveground part of the seedlings (30 days after germination) of the F_1 hybrids as well as green leaves and blanched false stems of adult plants at the commercial stage of development were used in further analysis.

The commercial stage of development is understood as plants before the flowering phase, the growth of which is complete, and the length and diameter of the stem have reached final dimensions. Samples of plant tissue were ground in liquid nitrogen and used to determine the content (mg/100 g raw weight) of vitamin C, sucrose, glucose and fructose using the Enzytec L-Ascorbic Acid, Enzytec™ Liquid Sucrose/D-Glucose and Enzytec™ Liquid D-Glucose/D-Fructose (R-Biopharm AG, Germany) kits, following the protocols provided by the manufacturer. Each sample type (sprout, green leaf or false stem) was ground whole and stored at –80 °C, taking the required portion for the analysis (determination of ascorbate concentration or gene expression).

Analysis of expression of the L-galactose biosynthetic pathway genes and the ascorbate recycling pathway genes was performed using quantitative real-time PCR (qRT-PCR). Total RNA was isolated from 0.2–0.5 g of ground tissue using the RNeasy Plant Mini Kit (QIAGEN, Germany). DNA impurities were removed using the RNase-free DNase set (QIAGEN, Germany), and cDNA was synthesized in the GoScript Reverse Transcription System (Promega, USA). The concentration of cDNA was determined using the Qubit® Fluorometer (Thermo Fisher Scientific, USA) and Qubit RNA HS Assay Kit (Invitrogen, USA), and 3 ng of the preparation was used in the qRT-PCR reaction. Gene-specific primers for qRT-PCR were designed based on the genomic/transcriptomic data of *A. porrum* (PRJNA310797) and *A. sativum* (PRJNA606385, PRJNA607255) available at NCBI (Table 2).

The reaction used the "2.5× Reaction mixture for qRT-PCR in the presence of SYBR Green I and ROX" kit (Synthol, Russia). qRT-PCR was performed in a CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, USA) using the program: 5 min at 95 °C, 40 cycles (15 s at 95 °C; 40 s at 60 °C). To normalize gene expression, two references were used: the glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) and *Ubiquitin* (*UBQ*) genes (Anisimova et al., 2021a). The analysis was performed in two biological and three technical replicates. The results were processed and correlation analysis was performed in GraphPad Prism v. 9.5.1 (GraphPad Software Inc., USA; https://www.graphpad.com/scientific-software/prism/).

Morphological characteristics (length, thickness and weight of the blanched part (false stem)) of leek plants were assessed in September 2023, using 10 plants of each analyzed hybrid. The plants were grown in an onion collection nursery (greenhouse coordinates 55.655182, 37.206576; Federal Scientific Vegetable Center, Moscow Region, Ph.D. in Agricultural Sciences T.M. Seredin).

Table 2. Sequences of primers used in this work for qRT-PCR analysis

Gene	Sequences (5′→3′) of primers (forward/reverse)
PGI	CCATCGGAATTGGTGGTTCTG/ GTGCAATTTGGTGGTCGATTCC
PMI	CTGCTATTTCCGGCCCTTCTA/ ACATTCCATCACCGTGAGCAG
PMM	TTCCTCGAGGTTGGGACAAGA/ TGTATGGCCGACAGTACGCT
VTC1b	TCGCTAACAAGCCAATGATCCTG/ GCTCTGTCTCCTGAGAACATG
GME1	ACACGATGATTAGTTTCAACATG/ CTTGAGGCTCAGCAGGCCA
GME2	CAGGATTAATGGGGTCAAAAGAC/ CTTGAGGCTCAGCAGGCCA
VTC2	GGTGTCAAGCGTGTGTATCTG/ TTCCCAAACAGCGGGATTGAC
GPP	GGTCACAGAAACTGATAAAGCATG/ CATCAGTAAGTAGAGCAGTGCC
GalDH	CTACCACACTTGGTGGACAC/ CCTCTGTCCGCAAGGCTTTT
GalLDH	TGGACTGCTAGGAGCAAGAGT/ GCCTTGCGTTAGGCGTCGA
APX1	GATTTATTGCGGAGAAGAACTGC/ CAACTCAGCCTTGTGCCTCA
APX2	GAAGGAAATTGAGCGGGCTC/ GAGCCATTTGGGCCACCTG
MDHAR1	CTGGCAGTAAGCGTTCTCCA/ TTTGAACCCTGGCGAGCTTG
MDHAR4	CGCAGGTTATGCAGCTCTTG/ CGCCTACGCAAGTATGAAATGC
MDHAR5	GGGGCTCGCATAGATAAGTTGA/ TCCCACGGACTTATTCAGCC
DHAR2	AGTTCTGCAAGCAAAGCCTGC/ TCCCGAGCCATCTTTGGTTAC
GR	CGGCAAAAGAAAGTTCTCAGAGG/ AGGAGACTGCTCAGTGTGGAA

Results

In this work, 13 F_1 hybrids obtained from crossing domestic and foreign leek varieties/lines were characterized (Table 1).

F_1 sprouts were characterized by the ascorbate content and the expression level of genes of the L-galactose pathway of ascorbate biosynthesis and its recycling pathway (Fig. 1, 2).

The highest amount of vitamin C was found in sprouts from crossings 36 × 38 and 81 × 95, the lowest – 33 × 30 and 74 × Alligator, while the remaining accessions showed intermediate values (5–7.5 mg/100 g raw weight) (Fig. 1).

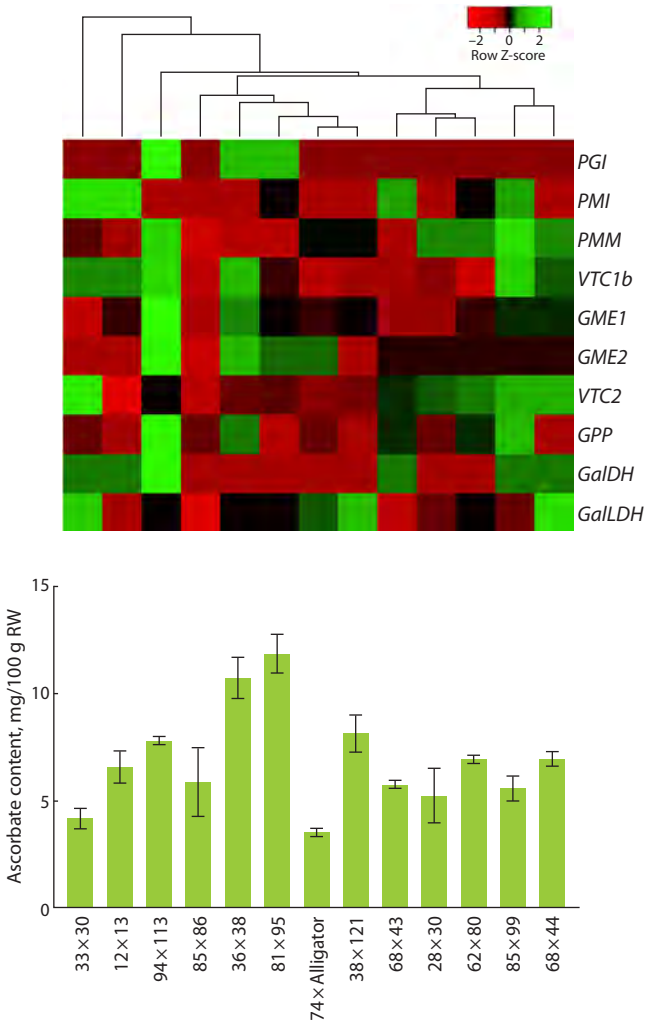


Fig. 1. Heatmap of L-galactose ascorbate biosynthetic pathway gene expression (*PGI*, *PMI*, *PMM*, *VTC1b*, *GME1*, *GME2*, *VTC2*, *GPP*, *GalDH*, *GalLDH*) and vitamin C content (mg/100 g raw weight) in seedlings of F_1 hybrids from 13 crosses of leek varieties/lines.

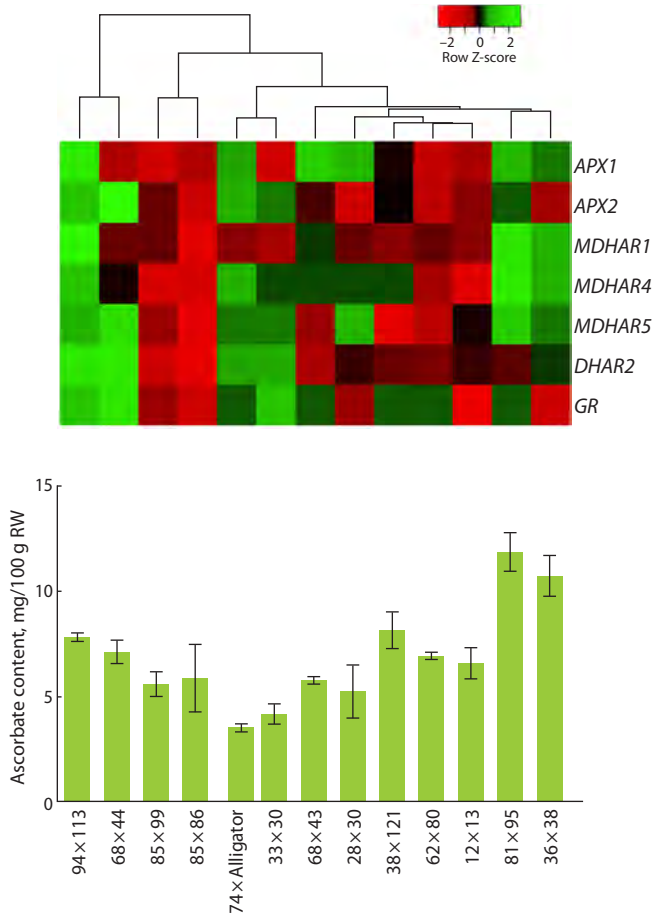


Fig. 2. Heatmap of ascorbate recycling gene expression (*APX1*, *APX2*, *MDHAR1*, *MDHAR4*, *MDHAR5*, *DHAR2*, *GR*) and vitamin C content (mg/100 g raw weight) in seedlings of F_1 hybrids from 13 crosses of leek varieties/lines.

On the heatmap of ascorbate synthesis gene expression, accessions 36 × 38 and 81 × 95 occupied adjacent positions in a cluster that also included hybrids with medium (85 × 86, 38 × 121) and the lowest (74 × Alligator) ascorbate content. Five accessions (68 × 43, 28 × 30, 62 × 80, 85 × 99, 68 × 44), characterized by similar average ascorbate content, formed a separate cluster (Fig. 1).

On the heatmap of ascorbate recycling gene expression, pairs of hybrids with the highest (36 × 38, 81 × 95) and lowest (33 × 30, 74 × Alligator) vitamin C content formed two separate clusters. We also note two clusters (94 × 113/68 × 44 and 85 × 99/85 × 86), the accessions in which were highly similar in gene expression and ascorbate content (Fig. 2).

For further analysis, we selected crosses contrasting in the amount of vitamin C in seedlings: 36 × 38 and 81 × 95 (the highest); 33 × 30 and 74 × Alligator (the lowest). In September 2023, 10 adult F_1 plants from these crosses were characterized by ascorbate content and the expression level of genes of the L-galactose pathway of ascorbate biosynthesis and ascorbate recycling genes in green leaves and the blanched false stem.

As a result, it was found that the green leaves of the 36 × 38 and 81 × 95 plants contained a similar (~35 mg/100 g raw weight) amount of ascorbate, which was expected to be significantly higher than that of the 33 × 30 and 74 × Alligator plants (~25 and ~17 mg/100 g raw weight, respectively). In the blanched false stems of all four types of plants, the ascorbate content was significantly lower than in the green leaves and did not exceed 6.5 mg/100 g raw weight. At the same time, the 33 × 30 and 74 × Alligator plants showed similar average values, and the remaining two hybrids were characterized by the lowest (81 × 95) and highest (36 × 38) amount of vitamin C (Fig. 3).

Gene expression analysis showed that in all four hybrids, the expression level of ascorbate biosynthesis genes in the blanched part of the false stem was predominantly higher than in the green leaves, with the exception of the *PMI*, *PMM*, *VTC2* and *GalLDH* genes (Fig. 4). No dependence of hybrids clustering on the expression heatmap on the ascorbate content (Fig. 3) was observed (Fig. 4).

The highest transcript level of ascorbate recycling genes was found in green leaves of the analyzed plants, except for the

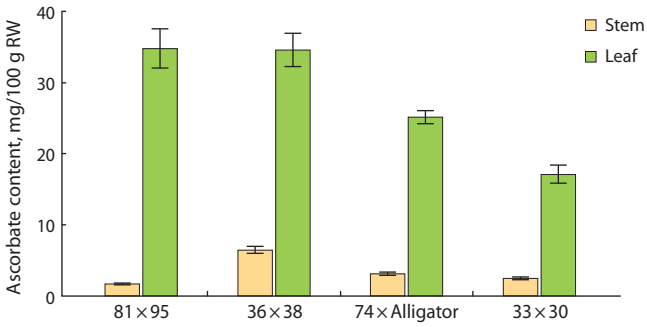


Fig. 3. Ascorbate content in the blanched false stem and green leaves of adult plants of F_1 leek hybrids from the crosses 81×95, 36×38, 74×Alligator and 33×30.

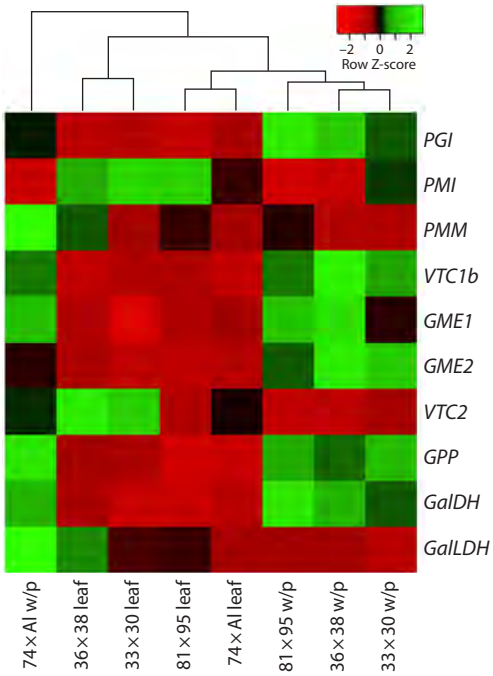


Fig. 4. Heatmap of the expression of genes of the L-galactose pathway of ascorbate biosynthesis in the white part (w/p) and leaf tissue of F_1 hybrid leek plants from the crosses 81×95, 36×38, 74×Alligator (Al) and 33×30.

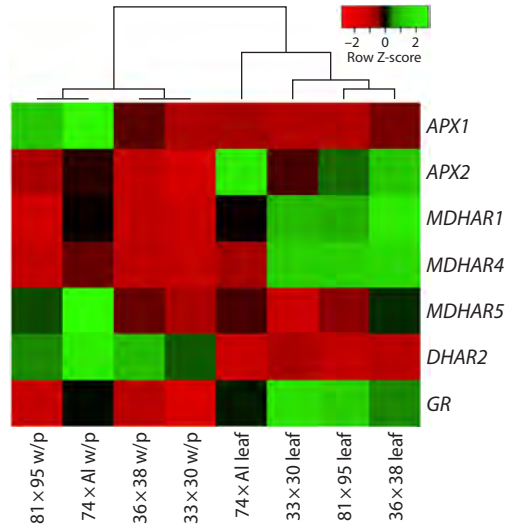


Fig. 5. Heatmap of the expression of genes of the ascorbate recycling pathway in the white part (w/p) and green leaves of F_1 hybrid leek plants from the crosses 81×95, 36×38, 74×Alligator (Al) and 33×30.

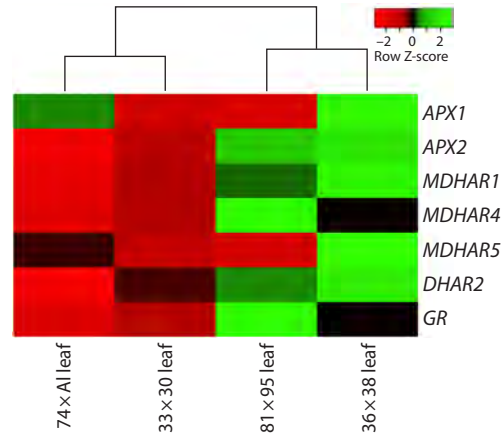


Fig. 6. Heatmap of the expression of genes of the ascorbate recycling pathway in the green leaves of F_1 hybrid leek plants from the crosses 74×Alligator (Al), 33×30, 81×95 and 36×38.

APX1, *MDHAR5*, and *DHAR2* genes with higher expression in the false stem (Fig. 5). At the same time, the heatmap clearly demonstrated the formation of two clusters, combining false stems and green leaves, respectively (Fig. 5). The "leaf" cluster included hybrids with an increased content of vitamin C in the leaves (36×38 and 81×95), and when clustering only by the level of gene expression in the leaves, the hybrids were strictly divided into groups with high (36×38 and 81×95) and low (74×Alligator, 33×30) vitamin C content (Fig. 6). Moreover, the leaves of the 36×38 and 81×95 plants (in comparison with the 74×Alligator and 33×30 hybrids) were distinguished by a significantly higher expression of ascorbate recycling genes, with rare exceptions (*APX1* and *MDHAR5* in the 81×95 plants) (Fig. 6).

In the case of the "false stem" cluster, no dependence of hybrids grouping on the amount of ascorbate was observed. For example, the 74×Alligator and 33×30 plants were assigned to separate groups (Fig. 5), despite similar vitamin C content (Fig. 3).

Next, using seedling and adult plant data, we searched for possible correlations between ascorbate content and gene expression levels, and revealed a high correlation for the ascorbate recycling genes *APX2*, *MDHAR1*, and *MDHAR4* (r = 0.94, 0.95, and 0.74, respectively) in leaves (Fig. 7).

The four analyzed hybrid lines were characterized by their morphological features. It was shown that the 81×95 hybrid forms a strong false stem 25–30 cm long, 3.5–5 cm thick, weighing 250–350 g with dense leaf arrangement. The 36×38

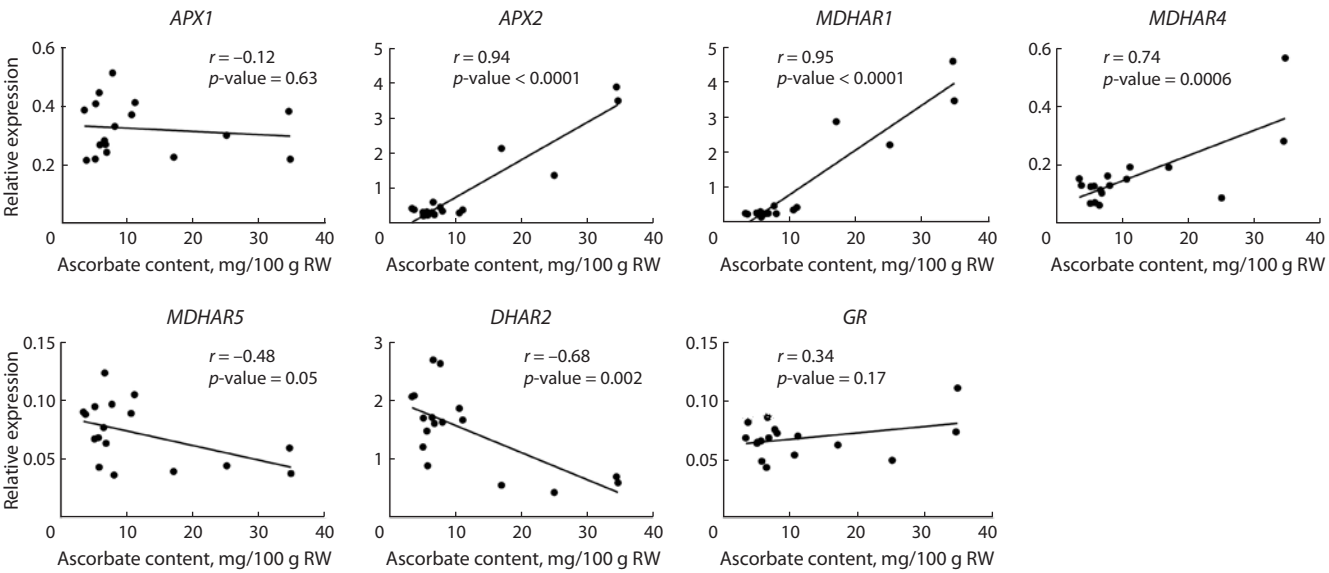


Fig. 7. Linear regression of vitamin C content and expression levels of ascorbate recycling pathway genes in green leaves of F_1 leek hybrids.
r is the Pearson correlation coefficient (correlation at $p < 0.01$).

Fig. 8. The commercial part of F_1 hybrids from the crosses 81 × 95, 36 × 38, 74 × Alligator and 33 × 30.
Scale = 10 cm.

hybrid has a strong false stem 20–25 cm long, 3–4.5 cm thick, weighing 200–300 g with dense leaf arrangement (Fig. 8). Both hybrids are characterized by 100 % seed germination (Table 1).

The false stem of the 74 × Alligator hybrid is long (20–25 cm) and thin (1.5–2.5 cm), weighing 100–150 g; the bulb is pronounced; the plants are significantly slower in growth than the other hybrids studied. The 33 × 30 hybrid is characterized by a powerful, loose false stem 20–25 cm long, 2.5–3.5 cm thick, and weighing 150–250 g; the bulb is pronounced (Fig. 8). The hybrids are characterized by 55 % (74 × Alligator) and 100 % (33 × 0) seed germination (Table 1).

The analyzed hybrids were also additionally characterized by the content of soluble sugars in green leaves and the blanched part of the false stem. It was shown that, compared to the leaves, the false stem is enriched with sucrose and contains ~1.5–3 times less fructose (Fig. 9). The amount of glucose turned out to be more variable: similar between leaves and

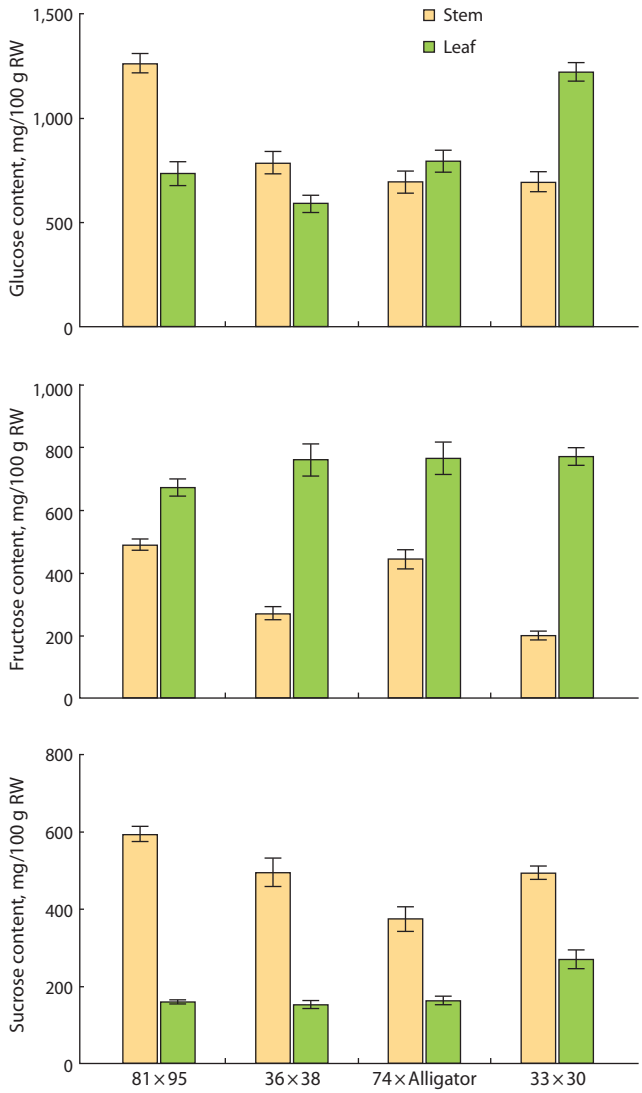


Fig. 9. Content of glucose, fructose and sucrose in the false stem and leaves of adult plants of F_1 leek hybrids from the crosses 81×95, 36×38, 74×Alligator and 33×30.

stem (36×38, 74×Alligator); the highest in the stem (81×95); the highest in the leaves (33×30) (Fig. 9). Considering the blanched part of the stem used for food, it can be seen that the highest amounts of all three sugars are contained in the F_1 hybrid from the cross 81×95; the least amount of fructose – in 33×30, and that of sucrose – in 74×Alligator; the glucose content is similar in three hybrids (36×38, 33×30, 74×Alligator) (Fig. 9).

Discussion

In this study, 13 F_1 leek hybrids with seed germination above 50 % (Table 1) were characterized for ascorbate content and expression of genes for ascorbate biosynthesis and recycling in seedlings (Fig. 1, 2). Variations in vitamin C content within 30 % between hybrids (Fig. 1) were consistent with previously demonstrated differences in ascorbate concentration in green leaves and false stems of leek (Lundegårdh et al., 2008; Bernaert et al., 2012).

The resulting selected accessions of two F_1 hybrids (81×95, 36×38), promising due to the highest vitamin C content, were further characterized at the commercial stage of development in comparison with two F_1 hybrids (74×Alligator, 33×30) with the lowest vitamin C content. Characterization included analysis of ascorbate/soluble sugar content and expression of ascorbate metabolism genes in green leaves and the blanched part of the pseudostem, as well as morphological description (Fig. 3–9).

It was shown that the results obtained from the analysis of ascorbate content in sprouts could be used to evaluate this trait only in the case of green leaves of adult plants, while the blanched part of the false stem shows variable data (Fig. 3). One of the promising hybrids (36×38) showed the highest content of vitamin C in the false stem, while the second hybrid (81×95), on the contrary, contained the lowest amount of ascorbate among the four accessions (Fig. 3).

Comparison of the ascorbate content data and clustering of the studied hybrids by the expression profile of the vitamin C synthesis genes (Fig. 4) showed that the expression level of any gene in the pathway could not be used to draw conclusions about the amount of ascorbate in the leaves or stems of leeks. In particular, the expression level of the *VTC2* gene (Fig. 1, 4), for which a relationship with ascorbate content was previously proposed (Ali et al., 2019; Broad et al., 2020), including in leeks (Anisimova et al., 2021a), is not suitable for prediction.

Nevertheless, the level of *VTC2* gene transcripts that we determined in hybrids can help to predict the degree of stress resistance of the accessions, since the level of *VTC2* expression is positively associated with plant resistance to abiotic stress factors (Ali et al., 2019; Broad et al., 2020). Based on our results (expression in seedlings or green leaves (Fig. 1, 4)), six promising hybrids can be identified, among which are two of the four selected for analysis (the studied 33×30 and 36×38, as well as 28×30, 62×80, 85×99, 68×44).

Comparison of biochemical data and clustering of hybrids by the expression of ascorbate recycling genes (Fig. 5, 6) allowed us to suggest a dependence of the level of ascorbate accumulation in leek leaves or stems on the expression profile of the pathway genes. In addition, since increased activity of *MDHAR* genes and the enzymes they encode is associated with plant stress resistance (e. g., Zhang et al., 2015; Qi et al., 2020), based on our data on three *MDHAR* genes (Fig. 2, 5, 6), we can identify seven promising hybrids, among which are all four accessions selected for analysis (studied 81×95, 36×38, 33×30, 74×Alligator, as well as 94×113, 68×44, 28×30).

Our assessment of possible statistically significant correlations between ascorbate content and the expression level of ascorbate metabolism genes revealed correlations only in green plant tissue (seedlings, green leaves) and only for three ascorbate recycling genes: *APX2*, *MDHAR1* and *MDHAR4* (Fig. 7). This confirms our previously identified positive relationship between the *MDHAR4* transcript level and the vitamin C content in leek plants (Filyushin et al., 2021).

The characteristics of the content of soluble sugars in green leaves and the blanched part of the false stem of the studied hybrids showed the absence of any dependence on the concentration of ascorbate (Fig. 9). Nevertheless, taking into ac-

count the obtained data, it can be assumed that the blanched part of the false stem in the F_1 81 × 95 hybrid, which is also characterized by the highest amount of ascorbate, has a greater nutritional value.

The morphological characteristics of the analyzed F_1 hybrids showed that the largest amounts of sugars (Fig. 9) correspond to the largest size and weight of the false stem in the hybrid from the 81 × 95 cross. The other three hybrids, taking into account the average data for all three types of sugars in the blanched part (a total of 1,500–1,600 mg/100 g raw weight (Fig. 9)), correspond to a similar length of the false stem (20–25 cm *vs.* 25–30 cm in 81 × 95). However, the variations in other stem parameters (thickness, weight) in these three hybrids do not agree in any way with either the amount of sugars or the content of each individual type of sugar.

Conclusion

Thus, the performed characterization of F_1 leek hybrids from 13 crosses made it possible to select nine accessions promising in terms of stress resistance (81 × 95, 36 × 38, 33 × 30, 74 × Alligator, 94 × 113, 28 × 30, 62 × 80, 85 × 99, 68 × 44). Of these, eight hybrids showed 95–100 % seed germination (except for 74 × Alligator, 55 %). Two hybrids (81 × 95, 36 × 38) were characterized by the highest ascorbate content in the green tissue (sprouts, green leaves) of the plant and one (36 × 38) – in the blanched part of the false stem, used as food. Hybrid 81 × 95 also accumulated the highest amount of soluble sugars in the blanched part. The found direct dependence of ascorbate content on the activity of ascorbate recycling genes (*APX2*, *MDHAR1*, *MDHAR4*) in green leaves can be used in the breeding of stress-resistant hybrids with increased vitamin C content. Further studies are needed on the possible relationship between the expression level of the *APX2*, *MDHAR1*, *MDHAR4* and *VTC2* genes and plant resistance to various stress factors, the results of which can be used in the breeding of stress-resistant leek hybrids.

References

Ali B., Pantha S., Acharya R., Ueda Y., Wu L.-B., Ashrafuzzaman M., Ishizaki T., Wissuwa M., Bulley S., Frei M. Enhanced ascorbate level improves multi-stress tolerance in a widely grown indica rice variety without compromising its agronomic characteristics. *J Plant Physiol.* 2019;240:152998. doi 10.1016/j.jplph.2019.152998

Anisimova O.K., Seredin T.M., Shchennikova A.V., Kochieva E.Z., Filyushin M.A. Estimation of the vitamin C content and GDP-L-galactose phosphorylase gene (*VTC2*) expression level in leek (*Allium porrum* L.) cultivars. *Russ J Plant Physiol.* 2021a;68(1):85-93. doi 10.1134/S1021443720060023

Anisimova O.K., Shchennikova A.V., Kochieva E.Z., Filyushin M.A. Identification and variability of the GDP-L-galactose phosphosphorylase gene *ApGGP1* in leek cultivars. *Russ J Genet.* 2021b;57(3):311-318. doi 10.1134/S1021443720060023

Arrigoni O., De Tullio M.C. Ascorbic acid: much more than just an antioxidant. *Biochim Biophys Acta.* 2002;1569(1-3):1-9. doi 10.1016/s0304-4165(01)00235-5

Bernaert N. Bioactive compounds in leek (*Allium ampeloprasum* var. *porrum*): analysis as a function of the genetic diversity, harvest time and processing techniques. Doctoral dissertation. Ghent: Ghent University, 2013

Bernaert N., De Paepe D., Bouten C., De Clercq H., Stewart D., Van Bockstaele E., De Loose M., Van Droogenbroeck B. Antioxidant capacity, total phenolic and ascorbate content as a function of the genetic diversity of leek (*Allium ampeloprasum* var. *porrum*). *Food Chem.* 2012;134:669-677. doi 10.1016/j.foodchem.2012.02.159

Broad R.C., Bonneau J.P., Hellens R.P., Johnson A.A.T. Manipulation of ascorbate biosynthetic, recycling, and regulatory pathways for improved abiotic stress tolerance in plants. *Int J Mol Sci*. 2020;21:1790. doi 10.3390/ijms21051790

Bulley S., Laing W. The regulation of ascorbate biosynthesis. *Curr Opin Plant Biol*. 2016;33:15-22. doi 10.1016/j.pbi.2016.04.010

Celebi-Toprak F., Alan A.R. In vitro gynogenesis in leek (*Allium ampeloprasum* L.). *Methods Mol Biol.* 2021;2287:171-184. doi 10.1007/978-1-0716-1315-3_7

Considine M.J., Foyer C.H. Redox regulation of plant development. *Antioxid Redox Signal.* 2014;21(9):1305-1326. doi 10.1089/ars.2013.5665

Dowdle J., Ishikawa T., Gatzek S., Rolinski S., Smirnoff N. Two genes in *Arabidopsis thaliana* encoding GDP-L-galactose phosphorylase are required for ascorbate biosynthesis and seedling. *Plant J.* 2007;52:673-689. doi 10.1111/j.1365-313X.2007.03266.x

Feng H., Liu W., Zhang Q., Wang X., Wang X., Duan X., Li F., Huang L., Kang Z. *TaMDHAR4*, a monodehydroascorbate reductase gene participates in the interactions between wheat and *Puccinia striiformis* f. sp. *tritici*. *Plant Physiol Biochem*. 2014;76:7-16. doi 10.1016/j.plaphy.2013.12.015

Filyushin M.A., Anisimova O.K., Kochieva E.Z., Shchennikova A.V. Correlation of ascorbic acid content and the pattern of monodehydroascorbate reductases (MDHARs) gene expression in leek (*Allium porrum* L.). *Russ J Plant Physiol.* 2021;68(5):849-856. doi 10.1134/S1021443721050034

García G., Clemente-Moreno M.J., Díaz-Vivancos P., García M., Hernández J.A. The apoplastic and symplastic antioxidant system in onion: response to long-term salt stress. *Antioxidants* (*Basel*). 2020;12:67. doi 10.3390/antiox9010067

Gill S.S., Tuteja N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol Biochem*. 2010;48:909-930. doi 10.1016/j.plaphy.2010.08.016

Grzelak-Błaszczyk K., Kołodziejczyk K., Badełek E., Adamicki F. Changes in the contents of mono-, di- and oligosaccharides in leek plants stored in cold room. *Eur Food Res Technol*. 2011;232(6):1027-1033. doi 10.1007/s00217-011-1476-y

Haroldsen V.M., Chi-Ham C.L., Kulkarni S., Lorence A., Bennett A.B. Constitutively expressed *DHAR* and *MDHAR* influence fruit, but not foliar ascorbate levels in tomato. *Plant Physiol Biochem*. 2011;49:1244-1249. doi 10.1016/j.plaphy.2011.08.003

Hemilä H. Vitamin C and infections. *Nutrients.* 2017;9(4):339. doi 10.3390/nu9040339

Lanubile A., Maschietto V., De Leonardis S., Battilani P., Paciolla C., Marocco A. Defense responses to mycotoxin-producing fungi *Fusarium proliferatum*, *F. subglutinans*, and *Aspergillus flavus* in kernels of susceptible and resistant maize genotypes. *Mol Plant Microbe Interact.* 2015;28(5):546-557. doi 10.1094/MPMI-09-14-0269-R

Lemma E., Yusuf Z., Desta M., Seyida S., Idris M., Mengistu S., Teneshu J. Physicochemical properties and biological activities of garlic (*Allium sativum* L.) bulb and leek (*Allium ampeloprasum* L. var. *porrum*) leaf oil extracts. *Sci World J.* 2022;2022:6573754. doi 10.1155/2022/6573754

Leterrier M., Corpas F.J., Barroso J.B., Sandalio L.M., del Río L.A. Peroxisomal monodehydroascorbate reductase, genomic clone characterization and functional analysis under environmental stress conditions. *Plant Physiol.* 2005;138(4):2111-2123. doi 10.1104/pp.105.066225

Lundegårdh B., Botek P., Schulzov V., Hajslov J., Strömberg A., Andersson H.C. Impact of different green manures on the content of S-alk(en)yl-L-cysteine sulfoxides and L-ascorbic acid in leek (*Allium porrum*). *J Agric Food Chem.* 2008;56(6):2102-2111. doi 10.1021/jf071710s

Qi Q., Yanyan D., Yuanlin L., Kunzhi L., Huini X., Xudong S. Overexpression of *SlMDHAR* in transgenic tobacco increased salt stress tolerance involving S-nitrosylation regulation. *Plant Sci*. 2020;299:110609. doi 10.1016/j.plantsci.2020.110609

Sultana S., Khew C.Y., Morshed M.M., Namasivayam P., Napis S., Ho C.L. Overexpression of monodehydroascorbate reductase from a mangrove plant (*AeMDHAR*) confers salt tolerance on rice. *J Plant Physiol*. 2012;169:311-318. doi 10.1016/j.jplph.2011.09.004

Swamy K.R.M., Gowda R.V. Leek and shallot. In: Peter K.V. (Ed.) Handbook of Herbs and Spices. Vol. 3. 2006;365-389. doi 10.1533/9781845691717.3.365

Yamada K., Osakabe Y. Sugar compartmentation as an environmental stress adaptation strategy in plants. *Semin Cell Dev Biol.* 2018;83:106-114. doi 10.1016/j.semcdb.2017.12.015

Yoon J., Cho L.H., Tun W., Jeon J.S., An G. Sucrose signaling in higher plants. *Plant Sci.* 2021;302:110703. doi 10.1016/j.plantsci.2020.110703

Zhang Y., Li Z., Peng Y., Wang X., Peng D., Li Y., He X., Zhang X., Ma X., Huang L., Yan Y. Clones of *FeSOD*, *MDHAR*, *DHAR* genes from white clover and gene expression analysis of ROS-scavenging enzymes during abiotic stress and hormone treatments. *Molecules*. 2015;20:20939-20954. doi 10.3390/molecules201119741

Conflict of interest. The authors declare no conflict of interest.
Received July 1, 2024. Revised October 8, 2024. Accepted November 8, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):210-218

doi 10.18699/vjgb-25-24

Morphological variability and genetic structure of *Miscanthus sinensis* (Poaceae) cultivated in the forest-steppe of Western Siberia

O.V. Dorogina [1, 2], I.N. Kuban [1], G.A. Zueva [1], E.V. Zhmud [1], O.Yu. Vasilyeva [1]

[1] Central Siberian Botanical Garden of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
olga-dorogina@yandex.ru

Abstract. *Miscanthus sinensis* Andersson (Poaceae) grows in monsoon climate. For this reason, when cultured under the conditions of a short growing season of Western Siberia, full-fledged seeds do not have time to form. We have studied a large number of specimens of this species from Primorsky Krai in the collection of the Central Siberian Botanical Garden, SB RAS. Using these samples, it was possible for the first time to select forms that produce high-quality mature seeds in local conditions during a short growing season, possibly due to spontaneous hybridization of early flowering forms. We obtained the first and second (G1 and G2) generations from these seeds and checked for hybrids. The aim of this study is selection, biomorphological characterization of early flowering ornamental forms of *M. sinensis* and analysis of genetic polymorphism of the selected forms (S1, S2) and the obtained G1 and G2 generations using ISSR markers. Under the conditions of introduction, the selected samples of *M. sinensis* were characterized by complex resistance, high decorativeness, reached the ontogenetic state of mature generative plants and differed from other samples in the collection by early flowering and the formation of full-fledged seeds. Thus, the forms of *M. sinensis* we selected are promising for landscape design and breeding. When studying the genetic structure of G1, G2 and two generations of the sample using ISSR markers, three effective stable unique PCR fragments were identified. A study of the genetic variability of the resulting G1 generation showed complete uniformity of genotypes. In the G2 generation, variability was observed, and we found five sets of genotypes, which were also confirmed in the dendrogram. As a result, unique molecular polymorphic fragments were identified. Their length was 300–3000 bp, and the genetic formula for certification of *M. sinensis* was compiled.

Key words: quality of seeds; decorative form of *Miscanthus sinensis*; ISSR markers; S1 and S2 selected forms; G1 and G2 generations; certification

For citation: Dorogina O.V., Kuban I.N., Zueva G.A., Zhmud E.V., Vasilyeva O.Yu. Morphological variability and genetic structure of *Miscanthus sinensis* (Poaceae) cultivated in the forest-steppe of Western Siberia. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):210-218. doi 10.18699/vjgb-25-24

Funding. This study was conducted within the framework of state assignments for Central Siberian Botanical Garden SB RAS (project "Analysis of biodiversity, conservation and restoration of rare and resource plant species using experimental methods", No. AAAA-A21-121011290025-2) and with the financial support of the Ministry of Science and Higher Education of the Russian Federation No. FSUS-2024-0024.

Морфологическая изменчивость и генетическая структура *Miscanthus sinensis* (Poaceae), культивируемого в лесостепи Западной Сибири

О.В. Дорогина [1, 2], И.Н. Кубан [1], Г.А. Зуева [1], Е.В. Жмудь [1], О.Ю. Васильева [1]

[1] Центральный сибирский ботанический сад Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
olga-dorogina@yandex.ru

Аннотация. *Miscanthus sinensis* Andersson (Poaceae), как и все другие виды рода *Miscanthus*, является растением муссонного климата, поэтому в суровых условиях лесостепи Западной Сибири не успевает сформировать жизнеспособные семена. При изучении большого числа образцов этого вида из коллекционного генофонда Центрального сибирского ботанического сада СО РАН, привезенных из различных районов Приморского края, впервые удалось отобрать две формы, S1 и S2 (selected), образующие в местных условиях в течение корот-

кого вегетационного периода качественные зрелые семена, из которых были получены первое и второе, G1 и G2 (generation), репродуктивные поколения. Цель настоящего исследования – отбор, биоморфологическая характеристика раноцветущих декоративных форм *M. sinensis* и анализ генетического полиморфизма отобранных форм (S1, S2) и полученных поколений (G1, G2) с помощью маркеров ISSR. В условиях интродукции отобранные образцы *M. sinensis* характеризовались комплексной устойчивостью к болезням и вредителям, высокой декоративностью, ранними сроками цветения и образованием полноценных семян. При изучении генетического полиморфизма с помощью маркеров ISSR у родительских образцов S1 и S2 обнаружено от одного до трех уникальных ПЦР-фрагментов. Исследование генетической изменчивости в поколении G1 показало полное единообразие генотипов. В поколении G2 наблюдалась изменчивость, при этом нами найдено пять вариантов генотипов. В результате выявлены уникальные полиморфные фрагменты, длина которых варьировала от 300 до 3000 п. н.; с их помощью составлена генетическая формула для паспортизации отобранных нами форм *M. sinensis*. Эти формы характеризуются высокими репродуктивными способностями в условиях континентального климата, перспективны для селекции, а, благодаря своим биоморфологическим особенностям, – и для ландшафтного дизайна.

Ключевые слова: качество семян; декоративная форма *Miscanthus sinensis*; ISSR-маркеры; отобранные формы S1 и S2; репродуктивные поколения G1 и G2; паспортизация

Introduction

The genus *Miscanthus* Andersson (fan grass) belongs to the subtribe Saccharinae, tribe Andropogoneae, family Poaceae. Fan grasses are perennial herbaceous plants. They are characterized by the C4 type of photosynthesis, high biomass productivity and relative unpretentiousness. These features allow them to be recommended as a promising bioenergy crop (Nie et al., 2014). The issues of biology and production process of different species of the genus *Miscanthus* are increasingly covered in modern works of Russian (Dorogina et al., 2018, 2022; Gushchina et al., 2018; Berseneva et al., 2020; Kapustyanchik et al., 2020; Anisimov et al., 2021; Yakimenko et al., 2021) and foreign (Chae et al., 2014; Gifford et al., 2014) researchers. In them, *M. sacchariflorus* (Maxim.) Hack., *M. sinensis* Andersson and, especially, *Miscanthus* × *giganteus* J.M. Greef, Deuter ex Hodk., Renvoize are considered as sources of bioethanol and biocellulose. Currently, the Soranovsky variety is grown in the forest-steppe of Western Siberia. It was selected from the natural material *M. sacchariflorus* and is a valuable industrial crop (Potseluyev, Kapustyanchik, 2018).

M. sinensis and *M. purpurascens* Andersson are widely used as ornamental grasses and a set of varieties, including variegated ones, is already presented. *M. sinensis* is the most popular choice for decorating the banks of reservoirs, rock gardens, rockeries, mixborders, as a solitaire on the lawn and for creating decorative group plantings (Grechushkina-Sukhorukova, 2022). In natural conditions, *M. sinensis* and *M. purpurascens* grow in regions with a monsoon climate. Both species can be successfully grown in continental climates, according to long-term studies of growth and development rhythms, biomorphology and ontogenesis; however, *M. sinensis* has a higher adaptive potential (Dorogina et al., 2018). It is believed that in the harsh conditions of the forest-steppe of Western Siberia, all types of *Miscanthus* (fan grass) do not have time to form full-fledged seeds (Zueva, 2020).

The collection gene pool of the Central Siberian Botanical Garden of the Siberian Branch of the Russian Academy of Sciences (CSBG SB RAS, Western Siberia, Novosibirsk), is constantly replenished with new forms and varieties of *Miscanthus* from various habitats. When replenishing the collection with samples from different areas of Primorsky Krai, Russian Federation (RF), two forms (S1 and S2 – selected) of *M. sinensis* were selected by us. In the conditions of the continental climate of the forest-steppe of Western Siberia, they form viable seeds, and two generations (G1 and G2 – generation) were obtained from them. These forms were also characterized by accelerated rates of seasonal development and a more compact habitus. They form less vegetative mass and form generative organs earlier.

The purpose of this study: selection, biomorphological characterization of early flowering ornamental forms of *Miscanthus sinensis* and analysis of genetic polymorphism of the selected S1 and S2 forms and the obtained G1 and G2 generations using ISSR markers.

Materials and methods

In 2017, *M. sinensis* seeds were collected from a coenopopulation on the Gamov Peninsula, Khasansky District, Primorsky Krai (42°58′02″N; 131°20′67″E), and sown in the ornamental plant collection plot of the CSBG SB RAS. As a result, only two plants with formed seeds were found in the forest-steppe conditions of Western Siberia (54°82′15″N; 83°10′46″E). Seeds formed on the earliest flowering shoots (more than 100 pieces were collected from five shoots). These plants, adapted to local climatic conditions, were used as initial forms (S1 and S2) to obtain subsequent G1 and G2 generations.

The experiments were conducted in two variants: in laboratory conditions in Petri dishes and when sowing in soil in a greenhouse. The dynamics of germination of the obtained seeds, the detection of the presence of a dormant period, the influence of environmental factors on their germination according to the laboratory-greenhouse-soil method of cultivation were estimated (Dyuryagina, 1982). As further studies showed, the seeds of the studied *Miscanthus* species were not in a state of deep dormancy, but in a forced dormancy.

Option I. In the third ten-day period of February, *M. sinensis* seeds were placed on filter paper for germination in laboratory conditions, at room temperature (19–20 °C), in Petri dishes in 2–3 replicates, 50 seeds in each. A small number of replicates is due to the presence of an insignificant seed harvest. Such division into replicates allowed to comply

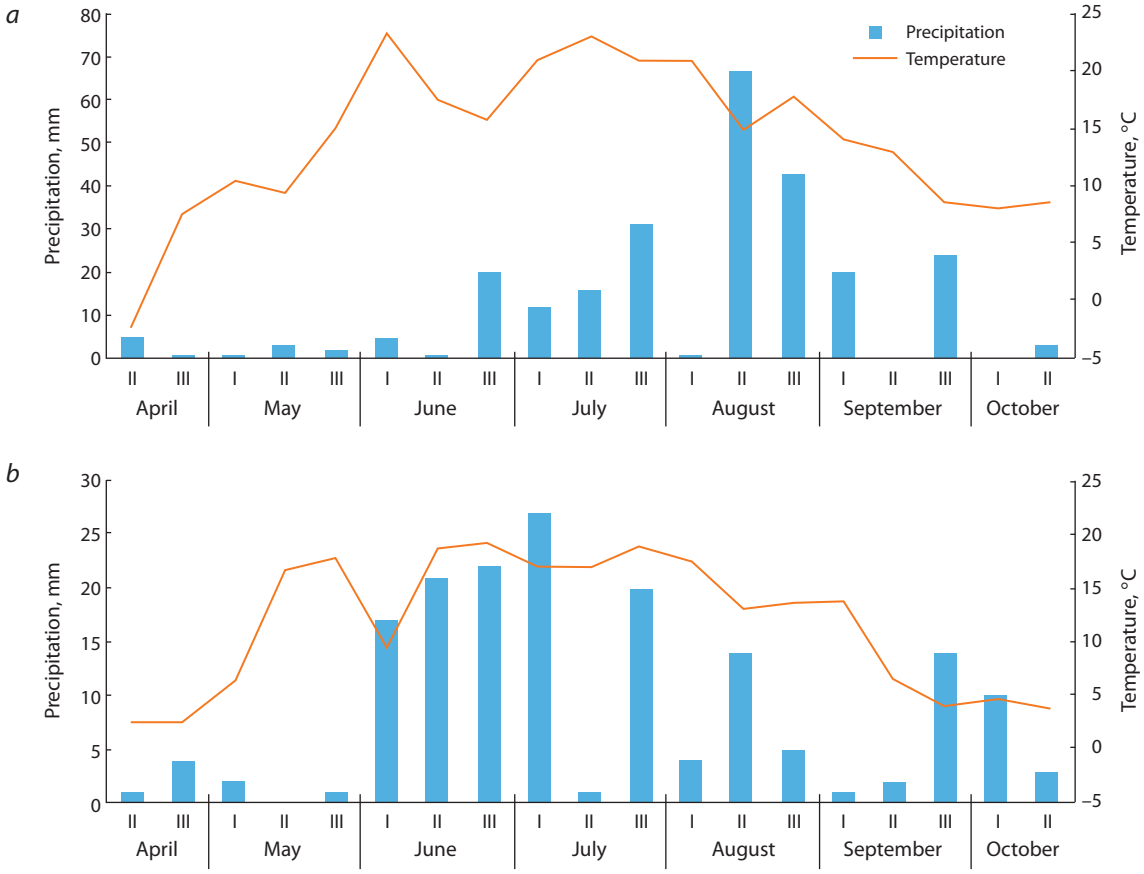


Fig. 1. Hydrothermal conditions of the 2022 (*a*) and 2023 (*b*) growing seasons.

with GOST (Methods for Testing…, 1973). The lighting was natural, filtered tap water was used as a humidifier. Regular observations of the condition of the seeds, the dynamics of germination, the nature of the growth of the primary root and primary shoot (primary leaf) were carried out.

Option II. *M. sinensis* seeds were sown in a stationary greenhouse in the second ten-day period of January in containers with a cell size of 2 × 2 cm, filled with a soil mixture of fertile soil, humus and sand. Three seeds were placed in a small depression in the soil (no more than 1 cm). They were lightly pressed down and carefully moistened by spraying, covered with glass or film, and not embedded in the soil. Germination energy and germination were recorded on the 7th and 21st day, respectively.

Hydrothermal conditions of the vegetation periods were calculated based on the data from the Ogurtsovo hydrometeorological station closest to the CSBG. The vegetation periods in 2022 and, especially, in 2023, were characterized by favorable meteorological conditions for flowering and fruiting of *M. sinensis*, similar in their dynamics to the hydrothermal indicators of the natural habitats of the species. In 2022, favorable hydrothermal conditions for plant growth and development developed from the second ten-day period of June and from the first of July. The average monthly air temperature in June and July was 17.3 and 18.9 °C, respectively (Fig. 1*a*). In 2023, the dry period in spring and the first half of summer was replaced by heavy precipitation in August (Fig. 1*b*).

To collect seeds in dry weather, panicles were cut after October 10. Based on the date of collection (ripening), flowering of this individual should have occurred in mid-August, and not in late July–early August.

The species and forms of *Miscanthus* in the collection are studied as ornamental grasses. Therefore, the following measurements and calculations were carried out by us: shoot length (cm), panicle length (cm), number of leaves on the shoot (pcs.).

To study genetic variability, we used the method of electrophoresis of intermicrosatellite regions of genomic DNA (ISSR analysis). It is known that among anonymous methods of fragment analysis, it is the most convenient, sensitive and reproducible (Nei et al., 1979; Kashin et al., 2016). DNA from laboratory-dried leaves of *M. sinensis* was isolated using the STAB method (Doyle J.J., Doyle J.L., 1987).

To study the variability between the original forms and two generations, the ISSR primers we had tested for *M. sacchariflorus* (Poaceae) were used (Dorogina et al., 2018). In this work, we used the three most informative primers, 17899A, 17898B, UBS-857, characterized by a polymorphic and reproducible pattern (Dorogina et al., 2019, 2022).

PCR was performed on a C1000 amplifier (Bio-Rad, USA). The volume of the reaction mixture was 25 µl. It consisted of the following components: 1.5 units of Taq DNA polymerase (Medigen, Russia); 2.7 mM $MgCl_2$, 0.8 mM ISSR primer (Medigen, Russia); DNA solution – 2 µl; water mQ H_2O –

2 µl. The amplification consisted of several stages: DNA denaturation for 90 s at 94 °C and 35 cycles, each of which included 40 s at 94 °C, 45 s of primer annealing and 90 s at 42–56 °C. The duration of the final stage of prolongation of the nucleotide chain was 5 min at 72 °C.

Electrophoretic separation of amplification products in 1.5 % agarose gel in 1× TAE buffer at 4 V/cm was performed. For statistical data processing, the TREECON software packages (Van de Peer, De Wachter, 1997) were used. Each ISSR marker was considered dominant, genetic distances and the polymorphism level of each primer (P, %) was calculated according to M. Nei and W.H. Li (1979).

Molecular genetic formulas for passportization of the *M. sinensis* population have been compiled according to the principle proposed by A.A. Novikova et al. (2012). Based on amplified PCR DNA fragments, genetic passports in the form of genetic formulas are presented. The genetic formula contains information about the method used, primers, and amplified DNA fragments detected in the sample under study. Statistical processing of the results using the StatSoft EXCEL 6.0 and STATISTICA v.6.0 software packages has been performed. The reliability of differences in the variability of morphometric features has been assessed using nonparametric criteria (Mann–Whitney, U-test).

Results

Morphological characteristic

The first full-fledged seeds of *M. sinensis* of local reproduction were collected in 2020. They were set during free pollination of two plants (which formed several panicles with high-quality mature seeds). Later, plants of the first (G1) and second (G2) generations of the reproduction were grown from these seeds.

The study of rhythmological and biomorphological features of the selected forms of *M. sinensis* showed the following. In the forest-steppe conditions of Western Siberia, early spring regrowth and earlier (at the end of July) flowering compared to other individuals obtained from the seeds of this sample were noted. In this regard, full-fledged seeds had time to set and ripen. Flowering of subsequent shoots has extended until October. The formation of compact bushes with shorter rhizomes is characteristic of these forms. Burgundy shoots and spreading panicles give the plants a special decorative effect (Fig. 2).

The straw is densely covered with ridges, belongs to the erect or semi-erectoid morphotype. Its height varies widely from 160 to 209 cm (Table 1). The leaf blades are elongated, linear (up to 70 cm), their width is from 0.6 to 1.2 cm. The edges of the leaf blades are very hard, the midrib is white. The panicle of this sample of *M. sinensis* is slightly drooping, characterized by 10–25 branches, 20–27 cm long (Table 1). The central axis of the panicle is shorter than the branches. Spikelets are paired. One of them has a short peduncle, and the other has a long one. The length of the spikelets reaches 4–7 mm. The spikelets at the base are pubescent with white hairs, their length can be equal to the length of the spikelet itself. These features create an additional decorative effect. We did not find any reliable differences in the variability of morphometric characteristics between G1 and G2 individuals.

The studied *M. sinensis* sample is resistant to local climatic conditions. These conditions correspond to the II–III frost

Fig. 2. *M. sinensis* in CSBG SB RAS experimental plot.
a – first generation (G1); *b* – second generation (G2).

Table 1. Morphometric indicators of *Miscanthus sinensis* shoots at the collection site of CSBG SB RAS at the end of the growing season

Samples	Length of shoots, cm		Length of panicle, cm		Number of leaves, pcs.	
S1	178.4		25.7		9.8	
S2	143.3		19.8		9.3	
	M* ± m	Cv	M ± m	Cv	M ± m	Cv
G1; *N* = 3	181.4 ± 14.3	13.7	22.4 ± 1.5	11.5	9.2 ± 0.5	10.0
G2; *N* = 7	169.4 ± 8.5	13.3	23.5 ± 1.4	16.0	7.9 ± 0.2	8.3

* M – mean value; m – error of the mean; Cv – coefficient of variation (%).

Table 2. Germination of seeds of two generations of *M. sinensis* from reproduction seeds of CSBG SB RAS in laboratory conditions and in a greenhouse

Indicators	G1	G2
Laboratory germination (28.02.2023)		
Number of seeds	150	36
Germination energy, %	84	94
Germination, %	94	100
Sowing in a greenhouse (19.01.2023)		
Number of seeds	72	72
Germination energy, %	78	72
Germination, %	68.0	72.2

resistance zones on the USDA scale (USDA Plant…, 2024). Therefore, it does not require winter shelter. During five years of introduction, *M. sinensis* was not affected by diseases and pests and was drought-resistant.

A study of the biology of germination of two generations of *M. sinensis* seeds showed that when grown in a greenhouse, the beginning of germination in both generations was observed as early as day 3. On day 7, the germination energy of G2 seeds was 10 % higher than that of G1 seeds (Table 2). At this point, mass germination ended, and after 21 days, the germination of the observed samples increased by another 10 % in G1, and all the seeds of G2 germinated.

Representatives of the genus *Miscanthus* in natural habitats begin vegetation at fairly high temperatures, therefore, for growing seedlings in a greenhouse (ontogenetic states: sprouts, juveniles and immatures), we have selected the optimal dates for sowing seeds.

When sowing seeds at the end of the second ten-day period of January, the sprouts grew and developed actively, their height on the 10th day reached 3 cm. The first true leaf in most G2 plants appeared on the 12th day after sowing. G1 sprouts lagged in development behind G2. The third true leaf in individual plants appeared on the 24th day of development. On the 50th day after sowing, the root system reached 6–8 cm in length, which exceeded the cell depth by 2–3 cm. At the same time, the leaves of the plants during the transition to the immature state, due to the small volume of cells for growth and development, noticeably turned yellow. Since there was still more than a month left before planting in open ground, the plants were transplanted into a larger container (7 × 7 cm). After transplanting, the immature plants began to actively develop and by the time of planting in open ground in the second ten-day period of May, their height varied from 9 to 31 cm.

Thus, it can be concluded that *M. sinensis* seedlings should be grown in greenhouse conditions, sowing in the second half of February. In this case, in the second half of May, the plants develop a branched fibrous root system, they enter the tillering phase, and they are ready for planting in open ground when favorable temperature conditions occur. Before planting in open ground, the plants need hardening. The seedlings were planted in a permanent place in warmed soil at a distance of 80–100 cm from each other. Thus, according to our observations, the seeds of the studied samples ripen well and do not require a period of post-harvest ripening, and laboratory conditions are more favorable for seed germination.

Marker analysis

The study of variability in generations G1 and G2 using ISSR markers in G1 showed complete uniformity of amplification

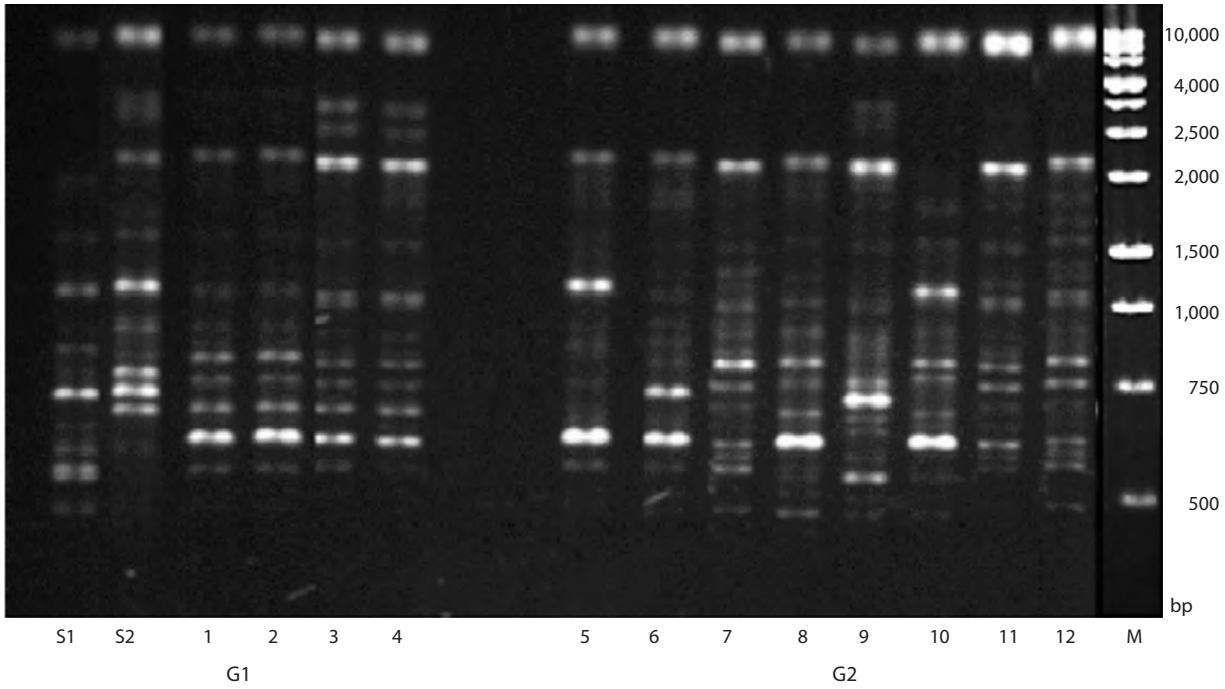


Fig. 3. Electrophoregram of PCR products obtained by DNA amplification at the representatives of *M. sinensis* and ISSR primer UBC-857.
S1, S2 – selected forms; tracks 1–4 – the first generation (G1), tracks 5–12 – the second generation (G2).

Table 3. Polymorphic fragments unique to the selected forms (S1, S2) of *M. sinensis*

Primer 5′–3′	Number of unique fragments	Length of polymorphic fragments, bp
UBS875$(AC)_8$GT	2/16 4/16	S1 510; 1,900 S2 700; 900; 2,100; 3,000
17899A$(CA)_6$AG	2/9 2/9	S1 620; 900 S2 600; 925
17898B$(CA)_6$GT	3/10 3/10	S1 350; 650; 830 S2 300; 640; 850

Note. Numerator – number of unique PCR fragments; denominator – total number of PCR fragments.

spectra. In generation G2, variability was observed, with five genotype variants revealed (Fig. 3).

On the electropherogram, in the region of less than 10,000 bp, a component present in each of the studied samples can be seen (Fig. 3). In all samples (except 10), in the region of 2,000–2,500 bp, a fragment from the S2 form is present. In the region of about 1,000 bp in samples 5 and 10 and in the region of slightly less than 750 bp in samples 6 and 9, fragments characteristic of both forms were present.

For the promising initial forms S1 and S2, selected by us, three markers were chosen for the purpose of compiling a genetic formula. They form the most polymorphic fragments (Table 3). The length of polymorphic fragments was from 300 to 3,000 bp. Using the $(Ac)_8$GT primer, 2 and 4 unique fragments were detected, two of which were characterized by the greatest length (2,100 and 3,000 bp).

We have compiled genetic formulas for the selected forms S1 and S2 of *M. sinensis* using the identified unique molecular polymorphic fragments based on the genetic formulas for *Rhododendron canadense* (Novikova et al., 2012):

S1: ISSR/(**AC)8**GT-510,1900/**(CA)6AG**-620,900/**(CA)6GT**-350,650,830

S2: ISSR/(**AC)8**GT-700,900,2100,3000/**(CA)6AG**-600,925/**(CA)6GT**-300,640,850

Discussion

In greenhouse conditions, we observed the germination of seeds of two generations of *M. sinensis* without preliminary treatment on the 3rd day, and on the 21st day, all seeds of G2 sprouted. The highest seed germination (98 and 88 %) was observed in *M. sinensis* and *M. sacchariflorus* grown in Korea at a temperature of 30 °C after scarification with 2 % NaOCl (Lee, et al., 2012). Polish scientists used sodium hypochlorite and commercial fertilizers to determine the cause of low germination of *Miscanthus* seeds and to search for methods to improve their quality (Orzeszko-Rywka, Rochalska, 2016).

The vegetation period in the geographical location of Primorsky Krai is quite long – 175–190 days, as our earlier analysis of the monsoon climate showed (Dorogina et al., 2019). Spring frosts are typical for the beginning of May, and autumn frosts, for the beginning of October. The spring-summer period is the driest. It is followed by a long summer-autumn overmoistening with up to 60 % of the annual precipitation rate.

Probably, the warm, relatively dry first half of summer in Primorsky Krai stimulates the intensive formation of generative, rather than vegetative organs of *Miscanthus*. Rains in the second half of summer and early autumn contribute to the filling of seeds and the formation of vegetative organs and buds for the next year's renewal. As mentioned above, we observed similar dynamics in Novosibirsk in 2023 (Fig. 1). That is, the conditions of this growing season in Novosibirsk turned out to be close to natural.

According to the literature data, in the steppe zone of the Stavropol Territory (RF) during the vegetation periods of 2019–2021, 13 varieties of *M. sinensis* maintained rhythmic processes similar to natural ones (Grechushkina-Sukhorukova, 2022). Thus, the beginning of the growing season was from 12.04 to 17.04, and the growth processes in the flowering phase of early flowering varieties were from 5.08 to 12.08, mid-flowering – from 16.09 to 22.09, late flowering – from 12.10 to 18.10. The author showed that the dynamic indicators of the linear growth of generative shoots correlated with the sum of effective temperatures of the growing season: in 2019 r = 0.93–0.96; in 2020 r = 0.85–0.9; in 2021, r =0.9–0.92. These data are consistent with our results obtained in the forest-steppe of Western Siberia. The forms of *M. sinensis* that we selected had the ability to vegetate for a long time while maintaining decorative properties until the onset of winter dormancy of plants and can be successfully grown from seeds in continental climate conditions.

As a result of the analysis of the obtained data for *M. sinensis* in the G2 population, we identified five genotype variants. G. Nie et al. (2014) showed that as a result of genotyping partially fertile hybrids in the hybrid population of *M. sinensis*, four genotypes were detected, and two of them were found in most plants. In Japan, tetraploid *M. sacchariflorus* and diploid *M. sinensis* are common, among which hybridization is observed (Tang et al., 2019).

However, *M. sinensis* is self-incompatible and has windborne pollen and seeds. According to our assumptions, this limits population differentiation. The degree of population differentiation using molecular markers for *M. sinensis* in individual areas of China has only been partially assessed (Chou et al., 2011; Swaminathan et al., 2012).

It is known from the literature that *M. sinensis* is a plant with a cross-pollination type (Mitros, et al., 2020). In general, inter- and intraspecific hybridization is typical for representatives of this genus, so it is characterized by rich genetic diversity and the presence of heterosis (Zhang et al., 2021). Genetic diversity is used to create *Miscanthus* hybrids. They can produce higher biomass yields and demonstrate better adaptability to various climatic conditions than their parent species (Clark et al., 2015). We did not conduct artificial hybridization. The two studied plants grew at a small distance from each other on the collection plot. G2 plants differed from the original (S1 and S2) selected forms in height, had more powerful leaves and stems. Therefore, we assumed the presence of hybrid plants as a result of spontaneous hybridization.

According to previous assumptions based on histochemical analysis of *M. sinensis* shoots, some specimens of this species can accumulate large amounts of lignin in dry, finished vegetation straws. This can complicate its industrial processing (Dorogina et al., 2019). Therefore, this species is more promising for the selection of ornamental forms.

Most of the literature sources on various aspects of the study and practical application of species, forms, hybrids and varieties of *Miscanthus* note that within the framework of the collected gene pool, serious systematic clarifications are required (Greef et al., 1997; Nishiwaki et al., 2011; Gifford et al., 2014). Analysis of genetic diversity can also provide information on the origin and composition of individual lines (Xu et al., 2013, Chen et al., 2022).

The phenotypic and genetic variability we have discovered in *M. sinensis* allows us to select forms with various economically valuable traits for further genetic improvement and development of a variety with the desired traits. For example, interspecific hybrids between *M. sacchariflorus* and *M. sinensis*, such as *Miscanthus* × *giganteus*, are promising for obtaining biomass in culture in regions with a moderate climate. Such partially fertile hybrids are interesting for improving the biomass and quality characteristics of the *Miscanthus* species (Tamura et al., 2016; Chen et al., 2022).

Thus, breeding work with *Miscanthus* in severe climatic conditions moves to a fundamentally new level. It becomes possible to both test randomly selected forms (from natural habitats) and to work with a wide range of offspring from seeds of local reproductions of different generations. This expands the possibilities of selecting forms with different traits. So, further study of phenological rhythms, biology of seed germination, morphology, the analysis of genetic diversity and the differentiation of the population using molecular markers and selection of *M. sinensis* plant forms with valuable decorative and technical (technological) characteristics is promising.

Conclusion

The forms of *M. sinensis* that we selected are highly decorative and resistant to introduction, produce viable seeds, and are promising for seed propagation and selection. They begin to bloom at the end of July and retain their decorative qualities until October. They do not require watering to maintain their decorative qualities during dry periods. *M. sinensis* grows successfully in open and shaded areas. It retains its decorative qualities in winter conditions and under snow. The offspring of the studied specimen pass all stages of ontogenesis in the conditions of Western Siberia and form viable seeds, and maintains its decorative effect from the end of July until October.

The identified polymorphic fragments in *M. sinensis* can be used for identification and taxonomy, and unique molecular polymorphic fragments, which are sequences of a certain length, are the basis for the certification of populations, forms and lines that are promising for obtaining decorative forms of *M. sinensis*. Overall, our results in developing breeding programs will help with creating *Miscanthus* varieties with elite potential.

References

Anisimov A.A., Medvedkov M.S., Skorokhodova A.N. Yield formation features in different Miscanthus species (*Miscanthus* spp.). In: From Agrarian Sciences to Agriculture. Barnaul: Altai State Agricultural University Publ., 2021;115-116 (in Russian)

Berseneva S.A., Ivleva O.E., Maslova A.O. Performance potential of species of *Miscantus* Anderss. and genus and prospects of its cultivation in Primorsky Krai. *Mežđunarodnyj Naučno-issledovatel'skij Žurnal = Int Res J*. 2020;7/2:6-10. doi 10.23670/IRJ.2020.97.7.033 (in Russian)

Chae W.B., Hong S.J., Gifford J.M., Rayburn A.L., Sacks E.J., Juvik J.A. Plant morphology, genome size, and SSR markers differentiate five distinct taxonomic groups among accessions in the genus *Miscanthus*. *GCB Bioenergy.* 2014;6:646-660. doi 10.1111/gcbb.12101

Chen Z., He Yu., Iqbal Yu., Shi Y., Huang H., Yi Z. Investigation of genetic relationships within three *Miscanthus* species using SNP markers identified by SLAF-seq. *BMC Genomics*. 2022;23(1):43. doi 10.1186/s12864-021-08277-8

Chou C.-H., Chiang Y.-C., Chiang T.-Y. Genetic variability and phytogeography of *Miscanthus sinensis* var. *condensatus*, an apomictic grass, based on RAPD fingerprints. *Can J Bot*. 2000;78(10):1262-1268. doi 10.1139/b00-102

Clark L.V., Ryan S., Nishiwaki A., Toma Yo., Zhao H., Peng J., Yoo J.H., Heo K., Yu Ch.Y., Yamada T., Sacks E.J. Genetic structure of *Miscanthus sinensis* and *Miscanthus sacchariflorus* in Japan indicates a gradient of bidirectional but asymmetric introgression January. *J Exp Bot.* 2015;66(14):4213-4225. doi 10.1093/jxb/eru511

Dorogina O.V., Vasilyeva O.Yu., Nuzhdina N.S., Buglova L.V., Gismatulina Yu.A., Zhmud E.V., Zueva G.A., Komina O.V., Tsybchenko E.A. Resource potential of some species of the genus *Miscanthus* Anderss. under conditions of continental climate of West Siberian forest-steppe. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2018;22(5):553-559. doi 10.18699/VJ18.394 (in Russian)

Dorogina O.V., Vasilyeva O.Yu., Nuzhdina N.S., Buglova I.V., Zhmud E.V., Zueva G.A., Komina O.V., Kuban I.S., Gusar A.S., Dudkin R.V. The formation and the study of a collection of the *Miscanthus* resource species gene pool in the conditions of the West Siberian forest steppe. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2019;23(7):926-932. doi 10.18699/VJ19.568 (in Russian)

Dorogina O.V., Nuzhdina N.S., Zueva G.A., Gismatulina Yu.A., Vasilyeva O.Yu. Specific shoot formation in *Miscanthus sacchariflorus* (Poaceae) under different environmental factors and DNA passportization using ISSR markers. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2022;26(1):22-29. doi 10.18699/VJGB-22-04

Doyle J.J., Doyle J.L. A rapid DNA isolation of fresh leaf tissue. *Phytochem Bull.* 1987;19:11-15

Dyuryagina G.P. On the method of introducing rare and endangered plants. *Botanicheskii Zhurnal. = Bot J.* 1982;67(5):679-687 (in Russian)

Gifford J.M., Chae W.B., Juvik J.A., Swaminathan K., Moose S.P. Mapping the genome of *Miscanthus sinensis* for QTL associated with biomass productivity. *GCB Bioenergy*. 2014;7(4):797-810. doi 10.1111/gcbb.12201

Grechushkina-Sukhorukova L.A. Dynamics of growth processes and decorative state of *Miscanthus sinensis* during introduction in the steppe zone. *Agrarnaya Nauka = Agrar Sci*. 2022;(7-8):178-182. doi 10.32634/0869-8155-2022-361-7-8-178-182 (in Russian)

Greef J.M., Deuter M., Jung C., Schondelmaier J. Genetic diversity of European *Miscanthus* species revealed by AFLP fingerprinting. *Genet Resour Crop Evol*. 1997;44(2):185-195. doi 10.1023/A:1008693214629

Gushchina V.A., Volod'kin A.A., Ostroborodova N.I., Agapkin N.D., Letuchiy A.V. Peculiarities of growth and development of introduction of *Miscanthus giganteus* in the conditions of forest-step zone in Middle Volga. *Agrarnyi Nauchnyi Zhurnal = Agrar Sci J.* 2018;1:10-13. doi 10.28983/asj.v0i1.318 (in Russian)

Kapustyanchik S.Yu., Burmakina N.V., Yakimenko V.N. Evaluation of the ecological and agrochemical state of agrocenosis with long-term growing of *Miscanthus* in Western Siberia. *Agrohimia*. 2020;9:65-73. doi 10.31857/S0002188120090082 (in Russian)

Kashin A.S., Kritskaya T.A., Schanzer I.A. Genetic polymorphism of *Tulipa gesneriana* L. evaluated on the basis of the ISSR marking data. *Russ J Genet.* 2016;52(10):1023-1033. doi 10.1134/S1022795416100045

Lee K.Y., Zhang L., Lee G.-J. Botanical and germinating characteristics of *Miscanthus* species native to Korea. *Hort Environ Biotechnol*. 2012;53(6):490-496. doi 10.1007/s13580-012-0137-9

Methods for Testing Crop and Planting Material Quality. In: Crop Seeds and Planting Material. Moscow: Izdatel'stvo Standartov, 1973;238-405 (in Russian)

Mitros T., Session A.M., James B.T., Wu G.A., Belaffif M.B., Clark L.V. Genome biology of the paleotetraploid perennial biomass crop *Miscanthus*. *Nat Commun*. 2020;11:5442. doi 10.1038/s41467-020-18923-6

Nei M., Li W.H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci USA*. 1979;76(10):5269-5273. doi 10.1073/pnas.76.10.5269

Nie G., Zhang X.-Q., Huang L.-K., Xu W.-Z., Wang J.-P., Zhang Y.-W., Ma X., Yan Y.-H., Yan H.-D. Genetic variability and population structure of the potential bioenergy crop *Miscanthus sinensis* (Poaceae) in southwest China based on SRAP markers. *Molecules*. 2014;19(8):12881-12897. doi 10.3390/molecules190812881

Nishiwaki A., Mizuguti A., Kuwabara S., Matuura H., Yamaguchi S., Toma Y., Miyashita T., Yamada T., Ishigaki G., Akashi R., Rayburn L.A., Stewart J.R. Discovery of natural *Miscanthus* (Poaceae) triploid plants in sympatric populations of *Miscanthus sacchariflorus* and *Miscanthus sinensis* in southern Japan. *Am J Bot*. 2011;98(1):154-159. doi 10.3732/ajb.1000258

Novikova A.A., Sheikina O.V., Novikov P.S., Doronina G.U. Estimation of the ISSR-markers application for systematization and genetic certification of genus *Rhododendron*. *Politematicheckij Setevoj Elektronnyj Nauchnyj Zhurnal Kubanskogo Gosudarstvennogo Agrarnogo Universiteta = Polythematic Online Scientific Journal of Kuban State Agrarian University*. 2012;82(82):916-926 (in Russian)

Orzeszko-Rywka A., Rochalska M. Possibility of seed quality improvement in *Miscanthus sinensis* (Andersson). *J Res Appl Agric Engng.* 2016;61(4):83-88

Potseluyev O.M., Kapustyanchik S.Yu. Assessment of the feasibility of growing Miscanthus under a cover of cereal crops. *Vestnik Altayskogo Gosudarstvennogo Agrarnogo Universiteta = Bulletin of Altai State Agricultural University.* 2018;10(168):55-60 (in Russian)

Swaminathan K., Chae W.B., Mitros T., Kranthi V., Xie L., Barling A., Glowacka K., Hall M., Jezowski S., Ming R., Hudson M., Juvik J.A., Rokhsar D.S. Moose S.P. A framework genetic map for *Miscanthus sinensis* from RNAseq-based markers shows recent tetraploidy. *BMC Genomics*. 2012;13:142. doi 10.1186/1471-2164-13-142

Tamura K., Uwatoko N., Yamashita H., Fujimori M., Akiyama Y., Shoji A., Sanada Y., Okumura K., Gau M. Discovery of natural interspecific hybrids between *Miscanthus sacchariflorus* and *Miscanthus sinensis* in Southern Japan: morphological characterization, genetic structure, and origin. *BioEnergy Res*. 2016;9(1):315-325. doi 10.1007/s12155-015-9683-1

Tang Y.-M., Xiao L., Igbal Y., Liao J.-F., Xiao L.-Q., Yi Z.-L., She C.-W. Molecular cytogenetic characterization and phylogenetic analysis of four *Miscantus* species (Poaceae). *Comp Cytogenet*. 2019;13(3):211-230. doi 10.3897/CompCytogen.v13i3.35346

USDA Plant Hardiness Zone Map https://planthardiness.ars.usda.gov/ (accessed: 03.2024)

Van de Peer Y.V., De Wachter R.D. Construction of evolutionary distance trees with TREECON for Windows: accounting for variation in nucleotide substitution rate among sites. *Bioinformatics*. 1997; 13(3):227-230. doi 10.1093/bioinformatics/13.3.227

Xu W.Z., Zhang X.Q., Huang L.K., Nie G., Wang J.P. Higher genetic diversity and gene flow in wild populations of *Miscanthus sinensis* in southwest China. *Biochem Syst Ecol*. 2013;48:174-181. doi 10.1016/j.bse.2012.11.024

Yakimenko V.N., Kapustyanchik S.Yu., Galitsyn G.Yu. Cultivation of Miscanthus in continental regions of Russia. *Zemledelie = Agriculture*. 2021;2:27-31. doi 10.24411/0044-3913-2021-10206 (in Russian)

Zhang G., Ge C., Xu P., Wang S., Cheng S., Han Y., Wang Y., ZhuangY., Hou X., Yu T., Xu X., Yang Y., Yin X., Wang W., Liu W., Zheng C., Sun X., Wang Z., Ming R., Dong S., Ma J., Zhang X., Chen C. The reference genome of *Miscanthus floridulus* illuminates the evolution of Saccharinae. *Nat Plants*. 2021;7:608-618. doi 10.1038/s41477-021-00908-y

Zueva G.A. Introduction of ornamental cereals and sedges in the Central Siberian Botanical Garden of the Siberian Branch of the Russian Academy of Sciences. *Vestnik of Orenburg State Pedagogical University. Electronic Scientific Journal*. 2020;3(35):30-41. doi 10.32516/2303-9922.2020.35.3 (in Russian)

Conflict of interest. The authors declare no conflict of interest.

Received May 20, 2024. Revised November 6, 2024. Accepted November 7, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):219-229

doi 10.18699/vjgb-25-25

Mitochondrial genome variation of mosquito species in the subgenus *Stegomyia* of the genus *Aedes* (Diptera: Culicidae)

A.G. Bega [1, 2], I.I. Goryacheva [1, 2], A.V. Moskaev [1, 3], B.V. Andrianov [2]

[1] Federal State University of Education, Mytishchi, Moscow Region, Russia
[2] Vavilov Institute of General Genetics of the Russian Academy of Sciences, Moscow, Russia
[3] Vernadsky Russian State University of National Economy, Balashikha, Moscow Region, Russia
andrianovb@mail.ru

Abstract. Mosquitoes in the subgenus *Stegomyia* of the genus *Aedes* are vectors of a number of vertebrate viruses, including human arboviral fevers. Of particular interest is the study of the genetic characteristics of invasive populations of species in this group. We obtained, annotated and described the mitochondrial genomes of three *Stegomyia* mosquito species of the genus *Aedes*: *Ae. albopictus*, *Ae. flavopictus* and *Ae. sibiricus*. The mitochondrial genomes of *Ae. flavopictus* and *Ae. sibiricus* were obtained from mosquitoes from synanthropic populations in the Russian Far East. The mitochondrial genome of *Ae. sibiricus* is presented for the first time. The mitochondrial genome of *Ae. albopictus* was obtained for the C6/36 cell line. We selected three primer sets, for each mosquito species, that amplify the entire mitochondrial genome except for the control region and sequenced the genomes using the Sanger method. All three new genomes have an identical gene order. We identified 13 canonical protein-coding genes, 2 ribosomal RNA genes, and 22 transport RNA genes. Protein-coding genes have canonical start and stop codons with two exceptions. The canonical stop codon "TAA" is incomplete in the *cox*1 and *cox*2 genes. The *cox*1 gene lacks the canonical start codon for methionine. Nucleotide variability is mainly represented by point nucleotide substitutions. A phylogenetic analysis of the nucleotide sequences of complete mitochondrial genomes of all known mosquitoes species in the subgenus *Stegomyia* of the genus *Aedes* was performed. The data obtained made it possible to measure the ratio of synonymous to non-synonymous substitutions (Ka/Ks) in specific protein-coding genes.
Key words: invasive species; mitochondrial genome; phylogenetic analysis; mtDNA

For citation: Bega A.G., Goryacheva I.I., Moskaev A.V., Andrianov B.V. Mitochondrial genome variation of mosquito species in the subgenus *Stegomyia* of the genus *Aedes* (Diptera: Culicidae). *Vavilovskii Zhurnal Genetiki i Selektsii* = *Vavilov J Genet Breed*. 2025;29(2):219-229. doi 10.18699/vjgb-25-25

Funding. The study was funded by the Russian Science Foundation grant (RSF) No. 24-44-10003, https://rscf.ru/en/project/24-44-10003/ "Genetic and ecological analysis of populations of the malaria mosquito *Anopheles plumbeus* as an important potential vector of vector-borne diseases in the Russian Federation and the Republic of Belarus" (the project is implemented by a scientific team together with a foreign scientific team selected by the Belarusian Republican Foundation for Basic Research (BRFBR), grant No. B23RNFM-068).

Authors' contribution. All authors contributed equally to this work.

Характеристика трех митохондриальных геномов комаров рода *Aedes* (Diptera: Culicidae) подрода *Stegomyia*

А.Г. Бега [1, 2], И.И. Горячева [1, 2], А.В. Москаев [1, 3], Б.В. Андрианов [2]

[1] Государственный университет просвещения, Мытищи, Московская область, Россия
[2] Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, Россия
[3] Российский государственный университет народного хозяйства им. В.И. Вернадского Минсельхоза Российской Федерации, Балашиха, Московская область, Россия
andrianovb@mail.ru

Аннотация. Комары рода *Aedes*, подрода *Stegomyia* являются переносчиками ряда вирусов позвоночных, в том числе возбудителей арбовирусных лихорадок человека. Особый интерес представляет изучение генетических особенностей синантропных популяций видов этой группы. Мы получили, аннотировали и описали митохондриальные геномы трех видов комаров рода *Aedes*, подрода *Stegomyia*: *Ae. albopictus*, *Ae. flavopictus* и *Ae. sibiricus*. Митохондриальные геномы *Ae. flavopictus* и *Ae. sibiricus* были получены от комаров из синантропных популяций с Дальнего Востока России. Митохондриальный геном *Ae. sibiricus* представлен нами впервые.

Митохондриальный геном *Ae. albopictus* был получен для клеточной линии C6/36. Мы подобрали три комплекта праймеров для каждого из видов комаров, которые амплифицируют весь митохондриальный геном, кроме контрольной области, и отсеквенировали геномы методом Сэнгера. Все три новых генома имеют одинаковый порядок расположения генов. Идентифицировано 13 канонических белок-кодирующих генов, 2 гена рибосомальной РНК, 22 гена транспортной РНК. Белок-кодирующие гены имеют канонические старт- и стоп-кодоны за двумя исключениями. Канонический стоп-кодон «TAA» неполный в генах *cox*1 и *cox*2. В гене *cox*1 отсутствует канонический старт-кодон для метионина. Нуклеотидная изменчивость представлена в основном точковыми нуклеотидными замещениями. Инсерции-делеции имеются в областях межгенных спейсеров. Проведен филогенетический анализ нуклеотидных последовательностей полных митохондриальных геномов всех известных видов комаров рода *Aedes*, подрода *Stegomyia*. Полученные данные позволили провести измерение соотношения синонимичных и несинонимичных замен (Ka/Ks) в конкретных белок-кодирующих генах.
Ключевые слова: инвазионный вид; митохондриальный геном; филогенетический анализ; мтДНК

Introduction

Mosquitoes of the genus *Aedes*, subgenus *Stegomyia*, are the main vectors of dengue, yellow fever and other arbovirus infections worldwide (Weetman et al., 2018). Of greatest interest and practical importance are the invasive mosquito species in this group that form dense synanthropic populations. *Aedes albopictusis* Skuse, 1894 is an invasive species native to Southeast Asia that has spread to all continents except Antarctica in the last 50 years (Medlock et al., 2012). In the Russian Federation, this species is found in the southern European part of the country. The study of its genetic variability in Russia is mainly based on the analysis of the barcode fragment of the mitochondrial gene *cox*1 (Fedorova et al., 2019; Bega et al., 2022).

The subgenus *Stegomyia* is represented in Russia by three other species of mosquitoes found in the Far East and Siberia – *Aedes flavopictus* Yamada, 1921, *Aedes sibiricus* Danilov & Filippova, 1978 and *Aedes galloisi* Yamada, 1921. These three species are considered native forest species. Comparative analyses of the genetic structure of *Ae. albopictus* and *Ae. flavopictus* populations from the Korean peninsula support this hypothesis (Shin, Jung, 2021). Previously, *Ae. flavopictus* and *Ae. sibiricus* did not form dense populations in the Far East and were only found as isolated specimens (Gutsevich et al., 1970). Recently there have been reports of sightings of these species in urban areas (Berlov, Kuberskaya, 2021; Berlov et al., 2021). We have obtained data on range expansion and the formation of dense synanthropic populations of *Ae. flavopictus* and *Ae. sibiricus* in the Russian Far East (Bega et al., 2022). This probably indicates the beginning of the formation of invasive populations of these species.

In this paper, we present the results of sequencing the mitochondrial genomes of representatives of potentially invasive populations of *Ae. flavopictus* and *Ae. sibiricus*, and the mitochondrial genome of the cell line *Ae. albopictus* C6/36, as well as the phylogenetic analysis of the obtained sequences.

The mitochondrial genome of *Ae. albopictus* is now well characterised, but some points remain controversial. The mitochondrial genomes of mosquitoes from the island of Taiwan, including the reference genome (ID NC_006817), have reading frame shifts and abnormal stop codons. This may be due to the fact that the sample was taken from an insular and presumably indigenous population. It may also be a consequence of the inclusion of nuclear copies of mitochondrial genes, or Numts, in the mitochondrial genome. Some sequences of the mitochondrial genome of *Ae. albopictus* represented in GenBank have deletions and poly(A) spacers (Battaglia et al., 2016; Ze-Ze et al., 2020). The features of the mitochondrial genome of *Ae. albopictus* cell culture have not been previously studied. C6/36 culture was obtained from mosquitoes, the place of capture of which is not precisely known (Singh, 1967). To date, the culture has been passaged in the laboratory for more than 50 years. Under cell culture conditions, with constant temperature and nutrient levels, the cells do not experience the selection factors that natural mosquito populations do. Obtaining the mitochondrial genome of a C6/36 cell culture is of interest because it shows which mitochondrial genes are under selection in natural populations. At the time of publication, only two mitochondrial genome sequences of *Ae. flavopictus*, NC_050044 and MT501510, from the southern part of the species range were available in NCBI GenBank. The genome we obtained represents a previously uncharacterised northern part of the range. The mitochondrial genome of *Ae. sibiricus* was obtained for the first time in this study. The NCBI GenBank had a mitochondrial genome for the closely related species *Ae. galloisi*. The sequences obtained in this study are of interest and can be used for further studies on the genetic characteristics of mosquitoes of the subgenus *Stegomyia*.

Materials and methods

Specimen collection and species identification. Mosquito samples were collected in the Russian Far East in the summer of 2020. We trapped *Ae. flavopictus* in Khabarovsk and *Ae. sibiricus* in Svobodny city, Amur region. *Aedes albopictus* clone C6/36 is a commercially available mosquito cell line isolated from larvae of this species (Singh, 1967). Species identification by morphological characters was carried out according to the keys in the identifiers (Gutsevich et al., 1970; Tanaka, 1979; Ree, 2003).

The taxonomic status of the mosquito we defined as *Ae. sibiricus* should be mentioned separately. Not all of the identifiers mentioned above include data on the separation of the species *Ae. sibiricus* from the previously described *Ae. galloisi* (Danilov, Filippova, 1978). We used keys to identify these species based on the colour of the legs and the struc-

ture of the hypopygium in males (Danilov, Filippova, 1978; Poltoratskaya, Mirzaeva, 2013). The species *Ae. sibiricus* is currently listed in the Mosquitoes of the World catalogue of blood-sucking mosquitoes (Wilkerson et al., 2021); however, the description of the species is only published in Russian and therefore *Ae. sibiricus* is not included in the GenBank taxonomic database.

DNA isolation and sequencing of the mitochondrial genome. Total DNA was isolated from individual adult mosquitoes. Each individual was homogenised in lysis solution. The composition of the lysis solution was as follows: 500 mM Tris-EDTA pH = 8.0, 100 μg/ml Proteinase K, 1 % Sodium N-lauroylsarcosinate, 100 mM NaCl. Lysis was performed at 50 °C for 3 hours. After lysis, the DNA was extracted with phenol. The phenol was in the upper layer. Two volumes of water were added to the resulting DNA solution, then the DNA was precipitated with isopropyl alcohol. After purification, the DNA was dissolved in deionised water.

Mitochondrial genomes were amplified using the Encyclo Plus PCR kit (Evrogen, Russia) and sequenced using the Sanger method. We selected the primers ourselves using Primer3 software (Rozen, Skaletsky, 2000) based on the *Ae. albopictus* mitochondrial genome published in the paper (Battaglia et al., 2016). PCR amplification for all primer pairs we selected was performed at an annealing temperature of 58 °C. The list of primers used is shown in Tables 1–3.

Bioinformatics analysis. Sequences were analysed using BLAST software to identify mitochondrial genes. Open reading frame start and stop codons were determined by comparison with start and stop codons of orthologous protein-coding genes in GenBank. Phylogenetic analysis was performed using the MEGA7 programme (Kumar et al., 2016). Sequences obtained from sequencing were aligned to sequences in the databases using NCBI resources (http://www.ncbi.nlm.nih.gov). We used the multiple sequence alignment algorithm Clustal W (Thompson et al., 1994). Visualisation of the mitochondrial genome ring was performed using Chloroplot software (Zheng et al., 2020). The algorithm for calculating the Ka/Ks ratio is described in the paper (Wang D. et al., 2011). We have carried out the calculation using the KaKs_Calculator software (Zhang Z. et al., 2006) using a simple substitution correction method (NG) (Nei, Gojobori, 1986). Suppose the length of the DNA sequence being compared is n and the number of substitutions between the sequences being compared is m. To calculate Ka and Ks, we need to count the number of synonymous (S) and non-synonymous (N) sites ($S + N = n$) and the number of synonymous (Sd) and non-synonymous (Nd) substitutions ($Sd + Nd = m$). Then, after correction for multiple substitutions, (Nd/N) and (Sd/S) can represent Ka and Ks, respectively. This is because the observed number of substitutions underestimates the true number of substitutions due to the divergence of sequences over time. Therefore, the calculation involved three steps: counting S and N, counting Sd and Nd, and correcting for multiple substitutions. Link to the programme distribution https://ngdc.cncb.ac.cn/biocode/tools/BT000001.

Results

Organisation of the derived mitochondrial genomes

The mitochondrial genomes of three mosquito species of the genus *Aedes*, subgenus *Stegomyia* (*Ae. albopictus*, *Ae. flavopictus* and *Ae. sibiricus*) are identical in gene order and similar in nucleotide sequence. The length of the mitochondrial genome excluding the length of control regions is as follows: *Ae. albopictus* 14,900 bp, *Ae. flavopictus* 14,893 bp, *Ae. sibiricus* 14,886 bp. Nucleotide variability is represented by point nucleotide substitutions. When comparing the mitochondrial genomes of *Ae. albopictus* and *Ae. flavopictus*, the degree of nucleotide divergence is minimal (5.74 %). The maximum degree of nucleotide divergence is observed when comparing the nucleotide sequences of the mitochondrial genomes of *Ae. albopictus* and *Ae. sibiricus* (7.51 %). The mitochondrial genomes of *Ae. flavopictus* and *Ae. sibiricus* differ by 6.62 %. All three mitochondrial genomes have a strong A + T = 78.4 % bias, which is characteristic of Diptera mitochondrial genomes.

We identified 13 canonical protein-coding genes (PCGs), 2 ribosomal RNA genes, and 22 transport RNA genes. All PCGs have canonical start and stop codons with two exceptions. The canonical stop codon "TAA" is incomplete in the *cox*1 and *cox*2 genes. It is thought that the missing base "A" is added during RNA processing. In addition, the canonical start codon for methionine is missing in the *cox*1 gene. The heavy (J) strand contains 22 genes, including 9 PCGs and 13 tRNA. The remaining 15 genes are encoded on the light strand (N-strand), including 4 PCGs, 2 rRNA and 9 tRNA. The ring genetic map of the mitochondrial genome of *Ae. sibiricus* is shown in Figure 1.

Phylogenetic analysis

The phylogenetic analysis of the nucleotide sequences of the mosquito mitochondrial genomes we obtained, using all available sequences of the mitochondrial genomes of *Ae. albopictus*, *Ae. flavopictus*, *Ae. aegypti* and *Ae. galloisi* registered in GenBank, is shown in Figure 2. The comparison region included the entire mitochondrial genome except for the control region.

Ae. albopictus, *Ae. aegypti* and *Ae. flavopictus* form independent clusters with high bootstrap support values. The mitochondrial genome of the C6/36 cell line clusters with the mitochondrial genomes of *Ae. albopictus* from the invasive part of the species' range. *Ae. sibiricus* and *Ae. galloisi* are clustered together.

PCGs variability analysis

We calculated the frequency ratio of point nucleotide substitutions leading to a change in the amino acid sequence (non-synonymous substitutions, Ka) or not leading to a change in the amino acid sequence of the protein (synonymous substitutions, Ks) Ka/Ks for PCGs in a pairwise comparison of the mitochondrial genomes obtained in this study: *Ae. albopictus* and *Ae. sibiricus*, *Ae. albopictus* and *Ae. flavopictus*, *Ae. flavopictus* and *Ae. sibiricus* (Fig. 3). The mosquito species we

Table 1. List of primers used to obtain the nucleotide sequence of the complete mitochondrial genome of C6/36 *Ae. albopictus* cell culture

Primer name	Sequence 5′→3′	Localisation on the GenBank sequence ID: OQ145430	PCR fragment length, bp
G1_18L	aatgaattgcctgataaaaagga	1–23	711
G1_18R	tgatttaatcctccaaatgc	711–692	
G4_5L	ttctataattattggggcatttgg	676–699	941
G4_5R	aaaagcatgagcagtaacaattaca	1617–1596	
G1_16L	ctggaatagtcggaacttcactaag	1505–1529	986
G1_16R	cggttaatcccccaactgta	2491–2472	
G1_15L	gccctgcacttttatgatcttt	2432–2453	919
G1_15R	tcattgatggccaataacttttt	3351–3330	
G1_14L	tggccatcaatgatattgaagtta	3339–3362	812
G1_14R	gaatcgattaggtattaatcaaaatgt	4160–4134	
G1_13L	ttggtcttttaattatcccatcaac	4111–4135	854
G1_13R	ttccccatcgtaatcctaatg	4965–4945	
G1_12L	tcgagaaggaacatttcaagg	4904–4924	895
G1_12R	ttggtaaaattaaagcaatttctacat	5799–5773	
G1_12-11L	tgtgacttccaatcacaagga	5528–5548	653
G1_12-11R	tgttgatcaagaaaaagctgcta	6181–6159	
G1_11L	tcatgaatgaaatcaaggagca	5885–5906	619
G1_11R	caagggtgaagagaatattttgg	6504–6482	
G1_10L	tttgaaactcttgcacatataatgaa	6423–6448	1027
G1_10R	tgctcctactcctgtttctgc	7450–7430	
G1_9L	gaatgaactaaagcagaaacagga	7418–7441	560
G1_9R	ttttattgaatgagaagttgtttcttt	7978–7952	
G1_9-8L	tcaccaattcgattagaaagagc	7631–7653	622
G1_9-8R	tcttcaggaagaagtcgagaattt	8253–8230	
G3_1L	aaaaattctcgacttcttcctga	8228–8250	715
G3_1R	ttgtgtatggtggttgcttttt	8943–8922	
G1_7L	agttgcctcaacatgagctt	8860–8879	858
G1_7R	gacgaaaacatcttctctgtacatt	9718–9694	
G1_6L	cagagaagatgttttcgtctagaaata	9700–9726	840
G1_6R	cccaataatgatccaaaatttca	10540–10562	
G1_5L	ttcagcctgatgaaattttgg	10530–10550	944
G1_5R	ggtcgagctccaattcatgt	11474–11455	
G1_4L	tgaattggagctcgacctgt	11458–11477	810
G1_4R	ggggtttatactgtaatagttgctgga	12268–12242	
G1_4-3L	ccttcagcaaaatcaaaagga	11994–12014	515
G1_4-3R	tcaaattcgtaaggggccta	12509–12490	
G3_2L	taggccccttacgaatttga	12490–12509	788
G3_2R	taaagggccgcagtattttg	13287–13268	
G1_2L	ctcattcaaccattcatacaagc	13204–13226	853
G1_2R	gaaaagaaatttgtgcaaatcaa	14057–14035	
G1_1L	tgatttgcacaaatttcttttca	14036–14058	679
G1_1R	ccagctaccgcggttataca	14715–14696	
G5_1L	ttgtataaccgcggtagctg	14695–14714	520
G5_1R	tgatgcttctaggaagaaatgaa	15215–15193	

Table 2. List of primers used to obtain the nucleotide sequence of the complete mitochondrial genome of *Ae. flavopictus*

Primer name	Sequence 5′→3′	Localisation on the GenBank sequence ID: OQ145431	PCR fragment length, bp
G1_18fL	aatgaaggccccgataaaaagga	1–23	710
G1_18fR	tggtttaatcctccaaatgc	710–691	
G1_17fL	ttactttctataattattggagcattt	670–696	898
G1_17fR	aaatatccctgaatgtctaagttcagt	1568–1542	
G1_16fL	ctggaatagtaggaacttctttaag	1504–1528	986
G1_16fR	cagttaatcctccaacgtta	2490–2471	
G1_15fL	gccctgctttattgtgatcttt	2431–2452	919
G1_15fR	tcattgatgcccaataaccttt	3350–3329	
G1_14fL	tgggcatcaatgatactgaagtta	3338–3361	812
G1_14fR	aaatcgattaggtattaatcagaatgt	4159–4133	
G3_14fL	ttggtcttttaattattccttcaaca	4110–4135	831
G3_3R	ggtcttcatacaatccccgt	4941–4922	
G3_3L	tcgagaaggaacatttcaagg	4903–4923	895
G1_12fR	taggtaaaattaaagcaatttctacat	5798–5772	
G1_12-11L	tgtgacttccaatcacaagga	5527–5547	656
G1_12-11R	tgttgatcaagaaaaagctgcta	6183–6161	
G1_11fL	ccatgaatgaaatcaaggagca	5884–5905	622
G1_11fR	caaggatgaagcgaatattttgg	6506–6484	
G1_10fL	tttgaaactcttggacatataatgaa	6425–6450	1027
G1_10fR	agcaccaacacctgtttctgc	7452–7432	
G1_9fL	gaatgaactaaagcagaaacaggt	7420–7443	513
G1_9fR	ttttattgaatgggaaattgtttcttt	7980–7954	
G1_9-8fL	tctacaattcgattagaaagagc	7633–7655	622
G1_9-8fR	tcttcaggaagagttcgggaattt	8255–8232	
G3_4fR	tcatatcattgacaccacaaatca	8106–8129	560
G3_4L	tctgttgctcatatgggtattgtt	8666–8643	
G4_6L	ttcgtcttcctattcgctca	8513–8532	794
G4_6R	gtttttggatttgtggtttaatttt	9307–9283	
G4_7L	aaaattaaaccacaaatccaaaaa	9283–9306	612
G4_7R	tttgggagttaatgaaaaggaa	9895–9874	
G4_8L	ttccttttcattaactcccaaag	9874–9896	817
G4_8R	tcgtaaaaatcaaccattatttacatc	10691–10665	
G1_5fL	ttcagcctgatgaaatttcgg	10532–10552	944
G1_5fR	ggtcgggctccaattcatgt	11476–11457	
G1_4fL	tgaattggagcccgacctgt	11460–11479	789
G1_4fR	ggggtttatactgtaatagttgctggg	12265–12239	
G1_4-3L	ccttcagcaaaatcaaaagga	11991–12011	515
G1_4-3fR	tcaaattcgtaaagggccaa	12506–12487	
G1_3fL	tgttccttagtaaataacttcacagca	12420–12446	809
G1_3R	tgaaggcttgtatgaatggttg	13229–13208	
G1_2L	ctcattcaaccattcatacaagc	13202–13224	829
G1_2R	gaaaagaaatttgtgcaaatcaa	14058–14036	
G1_1L	tgatttgcacaaatttcttttca	14037–14059	675
G1_1R	ccagctaccgcggttataca	14712–14693	
G5_2L	gctggcacaaattttaccaata	14708–14729	1000
G5_2R	cctatgggtcctaaatgaagaaaa	15684–15707	

Table 3. List of primers used to obtain the nucleotide sequence of the complete mitochondrial genome of *Ae. sibiricus*

Primer name	Sequence 5′→3′	Localisation on the GenBank sequence ID: OQ145432	PCR fragment length, bp
G1_18sL	aatgaattgcccgataaaaagga	1–23	706
G1_18R	tgatttaatcctccaaatgc	706–687	
G4_4L	tggagcatttggaggattaaa	683–703	599
G4_4R	caaatattttcagctttgaaggctat	1282–1257	
G3_5L	aactaatagccttcaaagctgaaa	1252–1275	417
G3_5R	tcaatttccaaatcctccaa	1669–1650	
G3_6L	ttcgaacagaacttagtcatccag	1536–1559	919
G3_6R	tcctaaagatcataaaagagcagga	2455–2431	
G1_15sL	gtcctgctctttttatgatcttt	2430–2451	919
G1_15sR	tcattgatgaccaataactttt	3349–3328	
G3_7L	tttgaacaattttaccagcaatta	3228–3251	881
G3_7R	agttgaaggaataattaaaagaccaa	4109–4134	
G3_8L	tgtatttgacccttcaactactatttttt	4050–4077	700
G3_8R	ctactaagtgaaaggggtgatttg	4750–4727	
G3_9L	gtcaacacacgcaaatcacc	4716–4735	810
G3_9R	tccttgtgattggaagtcacatatac	5546–5521	
G1_12-11L	tgtgacttccaatcacaagga	5526–5546	657
G1_12-11sR	tggtgatcaagaaaaagctgcta	6183–6161	
G1_11L	tcatgaatgaaatcaaggagca	5883–5904	614
G1_11sR	caaggatgaagagaatattttgg	6497–6475	
G3_10L	ctcttcatccttgatcaaattcc	6485–6507	959
G3_10R	cagcccctactcctgtttca	7444–7425	
G1_9sL	gaatgaactaaagctgaaacagga	7411–7434	560
G1_9sR	ttttattgaatgagaaattgtatcttt	7971–7945	
G1_9-8sL	tctccaatacgattagataaagc	7624–7646	622
G1_9-8sR	tcttcagggagaacccgagaattt	8246–8223	
G3_11L	aattctcgggttctccctga	8224–8243	897
G3_11R	ttttgaaagaagcttaattcctacatt	9121–9147	
G4_2L	ctgcttgtaaacgttcaggct	9074–9094	816
G4_2R	aactttgggagttaaagaaaaggaa	9890–9866	
G4_3L	cttccttttctttaactcccaaag	9865–9888	818
G4_3R	tcgtaaaaatcaaccattatttacatc	10683–10657	
G1_5sL	ttcagcttgatgaaattttgg	10524–10544	944
G1_5sR	ggtcgagctccaattcaggt	11468–11449	
G1_4sL	tgaattggagctcgaccagt	11452–11471	808
G1_4R	ggggtttatactgtaatagttgctgga	12260–12234	
G1_4-3sL	ccttcagcaaaatcaaaaggt	11986–12006	515
G1_4-3sR	tcaaattcggaaagggccta	12501–12482	
G1_3sL	tgttctttagtaaataacttcacagca	12415–12441	807
G1_3R	tgaaggcttgtatgaatggttg	13222–13201	
G3_12L	caaccattcatacaagccttca	13201–13222	849
G3_12R	gaaaagaaatttgtgcaaatcaa	14050–14028	
G1_1L	tgatttgcacaaatttcttttca	14029–14051	676
G1_1R	ccagctaccgcggttataca	14705–14686	
G5_3L	ttgtataaccgcggtagctg	14685–14704	358
G5_3R	ggggttatttttaataaggcaattt	15043–15019	

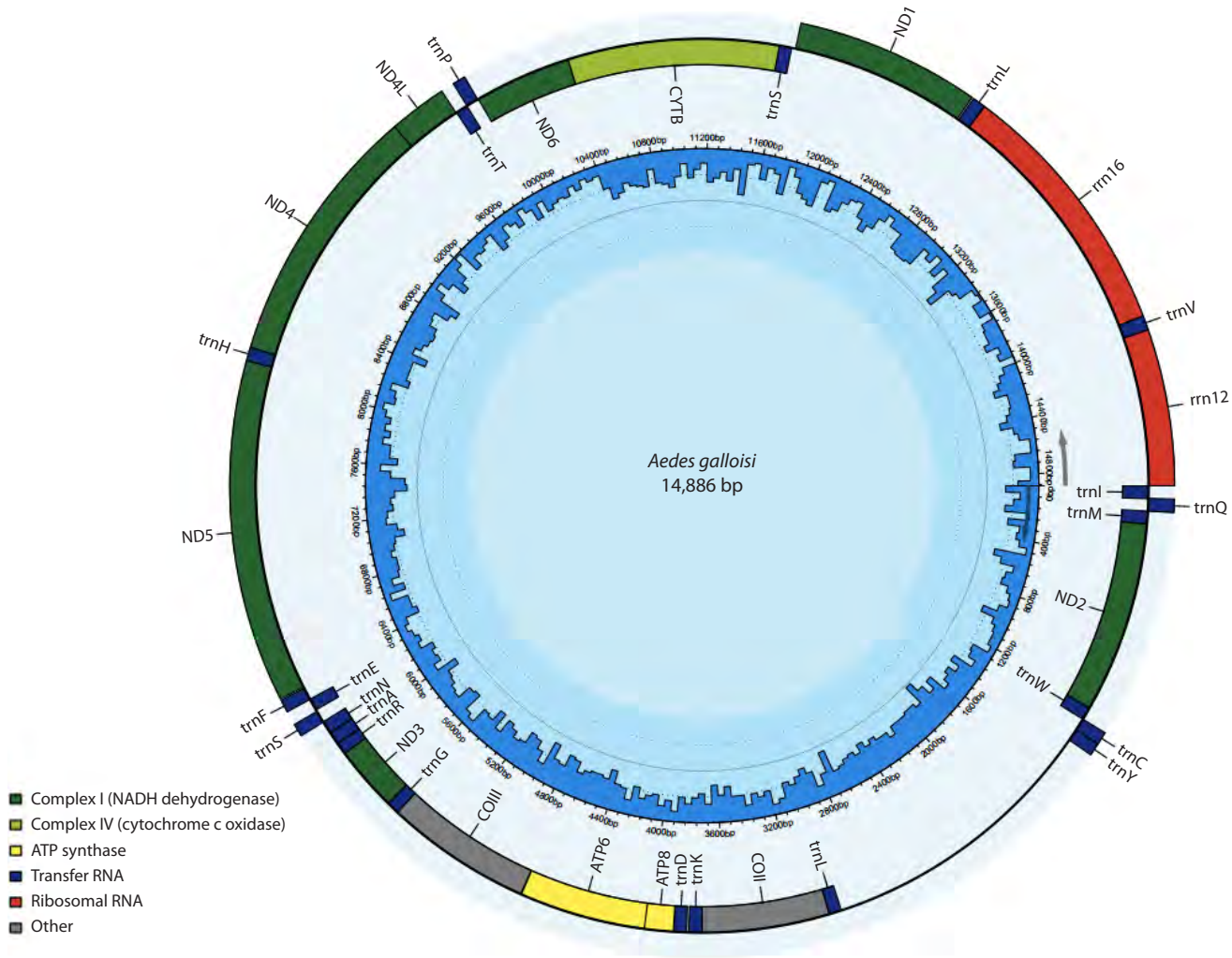


Fig. 1. Mitochondrial genome of *Ae. sibiricus* without the control site located between 12S rRNA and tRNA-Ile.

The nucleotide sequence has been deposited in the GenBank database under accession number OQ145432. The genome is registered as *Ae. galloisi* because the separation of the closely related species *Ae. galloisi* and *Ae. sibiricus* is not yet generally accepted, and the species *Ae. sibiricus* is not yet represented in the GenBank systematic database. We believe that the correct species name for the collected mosquitoes is *Ae. sibiricus*.

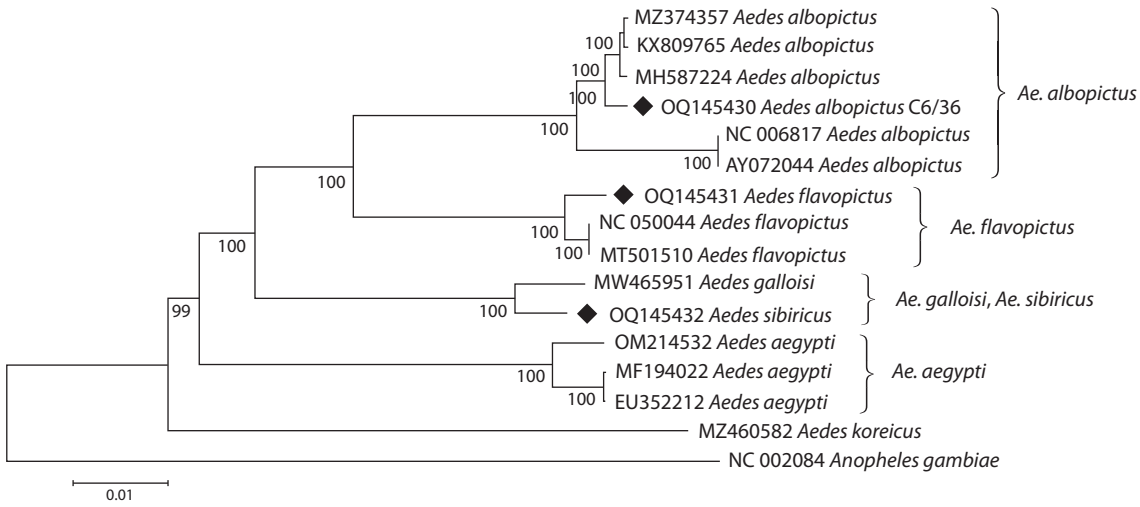


Fig. 2. NJ dendrogram of complete mitochondrial genomes.

The dendrogram is constructed using the maximum likelihood method. Branch lengths are expressed as the number of base substitutions per site. Bootstrap support values are shown next to the nodes (10,000 replicates). The complete mitochondrial genomes of *Ae. koreicus* and *Anopheles gambiae* were used as an external group. The mitochondrial genomes obtained in this study are marked with a diamond in the figure. The nucleotide sequences are registered in the GenBank database under the numbers OQ145430–OQ145432.

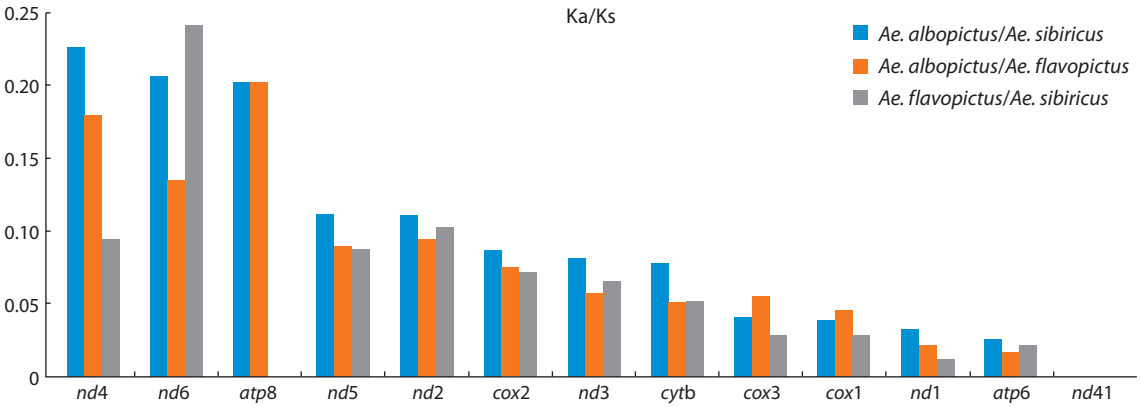


Fig. 3. Pairwise interspecies comparisons of the Ka/Ks ratio in protein-coding mitochondrial genes.

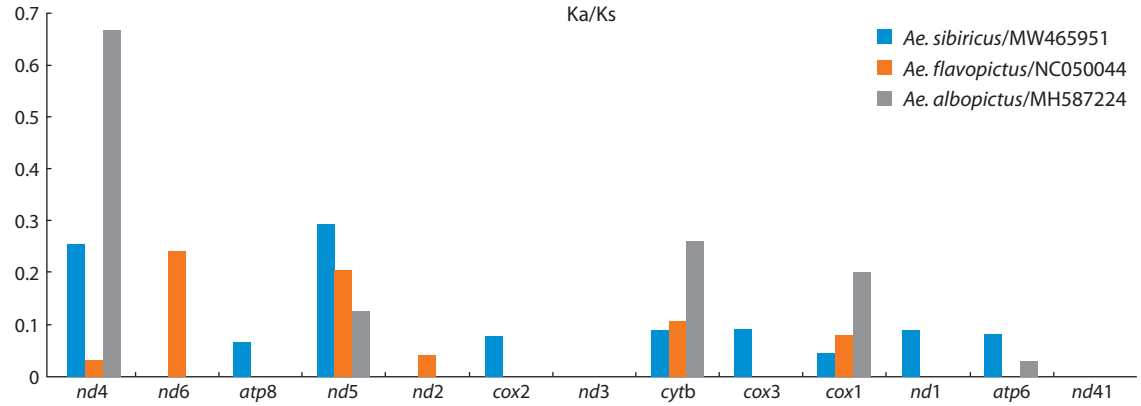


Fig. 4. Pairwise intraspecific comparisons of nucleotide variability in the magnitude of the Ka/Ks ratio of mitochondrial protein-coding genes.

studied are closely related, their habitats overlap slightly, but the centres of their ranges belong to different natural and climatic zones. *Ae. albopictus* is mainly restricted to tropical and subtropical climates, while *Ae. flavopictus* and *Ae. sibiricus* are restricted to temperate climates. At the same time, *Ae. flavopictus* predominates in zones with a monsoon climate, and *Ae. sibiricus*, in zones with a strongly continental climate. The pairwise comparison of PCGs was used to identify differences that may be adaptively relevant between the mosquito species. In Figure 3, Ka/Ks values are ranked in descending order based on the comparison of *Ae. albopictus* and *Ae. sibiricus*.

Ka/Ks ratios do not exceed 0.25 in all pairwise comparisons, indicating strong stabilising selection (Yang, Bielawski, 2000; Guo et al., 2021; Xing et al., 2022). The most variable genes between *Ae. albopictus*/*Ae. sibiricus* and *Ae. albopictus*/*Ae. flavopictus* are *nd*4, *nd*6 and *atp*8. The most conserved genes are *nd*1, *atp*6, *nd*4l and *cox*1.

In addition to interspecific comparisons, we performed intraspecific pairwise comparisons to assess the intraspecific variability of mitochondrial PCGs. The mitochondrial genome of *Ae. albopictus* cell culture C6/36 obtained in this study was compared with the genome of *Ae. albopictus* from China, GenBank ID MH587224. We compared the mitochondrial genome of *Ae. flavopictus* with the genome of *Ae. flavopictus* from Japan, GenBank ID NC050044, and the mitochondrial genome of *Ae. sibiricus* with the genome of *Ae. galloisi* from Japan, GenBank ID MW465951. The values of the Ka/Ks ratios are shown in Figure 4. The order of the genes is identical to the order of the gene rankings in Figure 3.

Within the *Ae. flavopictus* species, the highest Ka/Ks values were observed in the genes *nd*5, *nd*6, *cox*1, *cyt*b. The genes *atp*8, *cox*2, *nd*3, *cox*3, *nd*1, *atp*6, *nd*4l were conservative. When comparing the mitochondrial genomes of *Ae. albopictus* mosquitoes from the natural population and from C6/36 cell culture (Singh, 1967), the highest Ka/Ks ratio was observed in the genes *nd*4, *cyt*b, *cox*1, *nd*5. The most conserved genes were: *nd*6, *atp*8, *nd*2, *cox*2, *nd*3, *cox*3, *nd*1, *nd*4l. When comparing the mitochondrial PCGs of *Ae. sibiricus* and *Ae. galloisi*, the highest Ka/Ks values were observed in the genes *nd*4, *nd*5, *atp*8, *cox*2, *cyt*b, *cox*3, *nd*1, *cox*1, *atp*6. The following genes were conserved: *nd*6, *nd*2, *nd*3, *nd*4l.

Discussion

Organisation of the derived mitochondrial genomes

The nucleotide divergence values obtained in this study between the three closely related mosquito species are comparable and correspond to their geographical distribution in eastern Asia. *Ae. albopictus* is the most thermophilic species, characteristic of China and southern Asia. *Ae. sibiricus* is the most northerly. *Ae. flavopictus* occupies a middle position (Bega et al., 2022).

The use of molecular genetic markers to identify mosquito species is based on the use of a threshold of acceptable intraspecific variability of a given marker. This threshold is determined empirically for each marker and for each systematic group of insects (Zhang H.Z. et al., 2017). For example, for many insect groups, the threshold for intraspecific nucleotide variability of the BOLD fragment of the mitochondrial gene *cox*1 is 3 % (Hebert et al., 2003). Intraspecific variability of *Anopheles hyrcanus* s. l. mosquitoes in the Russian Far East ranged from 0.36 to 1.09 %, interspecific variability – from 2.34 to 4.50 % (Khrabrova et al., 2015). The average intraspecific variability of mosquitoes in China for the *cox*1 barcode fragment was 0.39 % (Wang G. et al., 2012). For complete mitochondrial genomes, much information has been accumulated, but there are no generally accepted quantitative generalisations.

Phylogenetic analysis

The *Ae. albopictus* mitochondrial genomes published to date can be divided into two groups. The first group was found on the island of Taiwan (presumably the native range of the species). Genomes from the second group were found in mosquitoes from the invasive part of the species range (Battaglia et al., 2016). The mitochondrial genome of the *Ae. albopictus* C6/36 cell line clustered with genomes belonging to the second group. The clustering obtained by analysing complete mitochondrial genomes is similar to that obtained in previous studies for the BOLD fragment of the *cox*1 gene (Bega et al., 2022).

PCGs variability analysis

The selection pressure on protein-coding genes can be assessed by determining their Ka/Ks ratio. We made such a comparison at the interspecific level by comparing the genomes obtained in this study. The highest Ka/Ks values for all PCGs, except *nd*6, *cox*1, *cox*3, were observed in the *Ae. albopictus*/*Ae. sibiricus* comparison. This result is in good agreement with the differences between species in terms of habitat ecology. The greater the differences in habitat between the species, the more significant the substitutions in the PCGs. The distribution of Ka/Ks values from higher to lower values within the PCGs was generally similar in all three pairwise comparisons, except for some peculiarities. For example, when comparing species from the same geographical area of the Russian Far East, *Ae. flavopictus*/*Ae. sibiricus*, the Ka/Ks ratio for the *nd*4 gene was significantly lower and no nucleotide substitutions were found at all for *atp*8. Calculating the frequency of substitutions normalised to one nucleotide, we can conclude that the *atp*8 gene in mosquitoes of the *Stegomyia* subgenus is characterised by a lower frequency of nucleotide substitutions than that in other protein-coding mitochondrial genes. A higher Ka/Ks ratio in the *atp*8 gene compared to other protein-coding mitochondrial genes was shown in a comparison of two Lepidoptera species of the genus *Gynaephora* living in different high mountain environments (Zhang B. et al., 2021), in parasitic wasp (Xing et al., 2022), and in mosquitoes of the *Anopheles* genus (Guo et al., 2021). This is probably due to the absence of strict constraints on the primary structure of the functional *atp*8 protein. In the *nd*1, *atp*6, *nd*4l and *cox*1 genes, the total frequency of nucleotide substitutions is comparable to that of other mitochondrial genes, but the Ka/Ks values are low, which confirms that these genes are under strong selective pressure.

In contrast to interspecific comparisons, the distribution of Ka/Ks between PCGs in intraspecific comparisons does not show clearly expressed general regularities, but characterises the specificity of variability accumulation for each species.

When comparing the mitochondrial genomes of *Ae. flavopictus*, the highest Ka/Ks values are observed for the *nd*5, *nd*6, *cox*1 and *cyt*b genes. This is due to the lower pressure of purifying selection. The pattern of intraspecific variability of these genes is similar to that found in the interspecific comparisons shown in Figure 3.

It is interesting to compare the mitochondrial genomes of *Ae. albopictus* mosquitoes from the natural population and from C6/36 cell culture. The highest Ka/Ks ratio is observed in the genes *nd*4, *cyt*b, *cox*1, *nd*5. The value of the Ka/Ks ratio in this case exceeds the values characteristic of both interspecific and intraspecific comparisons by a multitude, which allows us to conclude that selection in cell culture conditions is weak or absent. At the same time, the presence of fully conserved genes is observed: *nd*6, *atp*8, *nd*2, *cox*2, *nd*3, *cox*3, *nd*1, *nd*4l. This contrast in the variability of different genes of *Ae. albopictus* may be the result of the removal of a number of physiological constraints in cell culture conditions experienced by individuals in natural populations.

The variability observed when comparing the mitochondrial genomes of *Ae. sibiricus* and MW465951 mosquitoes generally corresponds to the level of interspecific variability in *Aedes* mosquitoes of the *Stegomyia* subgenus, with the exception of two abnormal genes: *nd*6 and *nd*5. The normally highly variable *nd*6 gene is monomorphic in this comparison, which may be due to the presence of stabilising selection. The *nd*5 gene, on the other hand, contains an abnormally high number of non-synonymous substitutions.

Conclusion

The study of the peculiarities of natural selection in invasive insect populations is still at the stage of accumulating material. One of the approaches used to detect the peculiarities of selection leading to the emergence of invasive populations in insects is to compare the mitochondrial genomes of native and invasive populations of the same species. Simultaneous coexistence of native and invasive populations is now known for many insect species, such as the Asian ladybird *Harmonia*

axyrid (Brown et al., 2011), the Japanese grape leafhopper *Arboridia kakogawana* (Piccinno et al., 2024), and several others. The study of the mitochondrial genomes of species that successfully synanthropise and form dense populations in urbanised areas is of interest for the discovery of mitochondrial genes involved in the genetic control of the increased viability trait characteristic of invasive insect populations.

References

Battaglia V., Gabrieli P., Brandini S., Capodiferro M.R., Javier P.A., Chen X.G., Achilli A., Semino O., Gomulski L.M., Malacrida A.R., Gasperi G., Torroni A., Olivieri A. The worldwide spread of the tiger mosquito as revealed by mitogenome haplogroup diversity. *Front Genet.* 2016;7:208. doi 10.3389/fgene.2016.00208

Bega A.G., Vu T., Goryacheva I.I., Moskaev A.V., Andrianov B.V. A barcoding and morphological identification of mosquito species of the genus *Aedes* (Diptera: Culicidae) of the Russian Far East and Northern Vietnam. *Russ J Genet.* 2022;58(3):314-325. doi 10.1134/S1022795422030024

Berlov O.E., Berlov E.Y., Artemyeva S.Yu. Findings of the tiger-mosquito *Aedes* (*Stegomyia*) *sibiricus* Danilov et Filippova, 1978 (Insecta: Diptera, Culicidae) in Irkutsk. *Baikal Zoological Journal.* 2021;2(30):118-119 (in Russian)

Berlov O.E., Kuberskaya O.V. First record of tiger mosquito *Aedes flavopictus* Yamada, 1921 (Diptera, Culicidae) in the Lower Amur area (Khabarovsk region, Russia). *Amurian Zoological Journal*. 2021;13(4):550-556. doi 10.33910/2686-9519-2021-13-4-550-556 (in Russian)

Brown P.M.J., Thomas C.E., Lombaert E., Jeffries D.L., Estoup A., Handley L.J.L. The global spread of *Harmonia axyridis* (Coleoptera: Coccinellidae): distribution, dispersal and routes of invasion. *BioControl.* 2011;56:623-641. doi 10.1007/s10526-011-9379-1

Danilov V.N., Filippova V.N. A new species of mosquito *Aedes* (*Stegomyia*) *sibiricus* sp. n. (Culicidae). *Parazitologiia = Parasitology.* 1978;12(2):170-176 (in Russian)

Fedorova M.V., Shvets O.G., Medyanik I.M., Shaikevich E.V. Genetic diversity of invasive *Aedes* (*Stegomyia*) *albopictus* (Skuse, 1895) population (Diptera, Culicidae) in Krasnodar region, Russia. *Parazitologiia = Parasitology.* 2019;53(6):518-528. doi 10.1134/S0031184719060073 (in Russian)

Guo J., Yan Z.T., Fu W.B., Yuan H., Li X.D., Chen B. Complete mitogenomes of *Anopheles peditaeniatus* and *Anopheles nitidus* and phylogenetic relationships within the genus *Anopheles* inferred from mitogenomes. *Parasit Vectors.* 2021;14(1):452. doi 10.1186/s13071-021-04963-4

Gutsevich V.A., Monchadskii A.S., Shtakel'berg A.A. Mosquitoes (Culicidae). In: Fauna of the USSR. Diptera. Vol. 3, Iss. 4. Leningrad: Nauka Publ., 1970 (in Russian)

Hebert P.D., Cywinska A., Ball S.L., de Waard J.R. Biological identifications through DNA barcodes. *Proc Biol Sci.* 2003;270(1512):313-321. doi 10.1098/rspb.2002.2218

Khrabrova N.V., Andreeva Y.V., Sibataev A.K., Alekseeva S.S., Esenbekova P.A. Mosquitoes of *Anopheles hyrcanus* (Diptera, Culicidae) group: species diagnostic and phylogenetic relationships. *Am J Trop Med Hyg.* 2015;93(3):619-622. doi 10.4269/ajtmh.14-0207

Kumar S., Stecher G., Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870-1874. doi 10.1093/molbev/msw054

Medlock J.M., Hansford K.M., Schaffner F., Versteirt V., Hendrickx G., Zeller H., Bortel W.V. A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. *Vector Borne Zoonotic Dis*. 2012;12(6):435-447. doi 10.1089/vbz.2011.0814

Nei M., Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 1986;3(5):418-426. doi 10.1093/oxfordjournals.molbev.a040410

Piccinno R., Tatti A., Avosani S., Galla G., Lazazzara V., Pedrazzoli F., Zadra N., Rodeghiero M., Seljak G., Özgen İ., Hauffe H.C., Verrastro V., Stacconi M.V.R., Mazzoni V., Rota-Stabelli O. A multidisciplinary approach to tackling invasive species: barcoding, morphology, and metataxonomy of the leafhopper *Arboridia adanae*. *Sci Rep*. 2024;14(1):2229. doi 10.1038/s41598-023-49410-9

Poltoratskaya N.V., Mirzaeva A.G. New records of the rare species *Aedes sibiricus* Danilov et Filippova, 1978 (Diptera, Culicidae) from West Siberia, Russia. *Evraziatskii Entomologicheskii Zhurnal = Euroasian Entomological Journal*. 2013;12(2):144-146. (in Russian)

Ree H.I. Taxonomic review and revised keys of the Korean mosquitoes (Diptera: Culicidae). *Entomol Res.* 2003;33(1):39-52. doi 10.1111/j.1748-5967.2003.tb00047.x

Rozen S., Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. In: Misener S., Krawetz S.A. (Eds) Bioinformatics Methods and Protocols. Methods in Molecular Biology. Vol. 132. Humana Press, Totowa, NJ, 2000. doi 10.1385/1-59259-192-2:365

Shin J., Jung J. Comparative population genetics of the invasive mosquito *Aedes albopictus* and the native mosquito *Aedes flavopictus* in the Korean peninsula. *Parasit Vectors.* 2021;14(1):377. doi 10.1186/s13071-021-04873-5

Singh K.R.P. Cell cultures derived from larvae of *Aedes albopictus* (Skuse) and *Aedes aegypti* (L.). *Curr Sci.* 1967;36(19):506-508

Tanaka K.M. A revision of the adult and larval mosquitoes of Japan (including the Ryukyu Archipelago and the Ogasawara Islands) and Korea (Diptera: Culicidae). In: Tanaka K., Mizusawa K., Saugstad E.S. (Eds) Contributions of the American Entomological Institute. APO; San Francisco; California: American Entomological Institute, 1979;987

Thompson J.D., Higgins D.G., Gibson T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22(22):4673-4680. doi 10.1093/nar/22.22.4673

Wang D., Liu F., Wang L., Huang S., Yu J. Nonsynonymous substitution rate (Ka) is a relatively consistent parameter for defining fast-evolving and slow-evolving protein-coding genes. *Biol Direct.* 2011;6:13. doi 10.1186/1745-6150-6-13

Wang G., Li C., Guo X., Xing D., Dong Y., Wang Z., Zhang Y., Liu M., Zheng Z., Zhang H., Zhu X., Wu Z., Zhao T. Identifying the main mosquito species in China based on DNA barcoding. *PLoS One.* 2012;7(10):e47051. doi 10.1371/journal.pone.0047051

Weetman D., Kamgang B., Badolo A., Moyes C.L., Shearer F.M., Coulibaly M., Pinto J., Lambrechts L., McCall P.J. *Aedes* mosquitoes and *Aedes*-borne arboviruses in Africa: current and future threats. *Int J Environ Res Public Health.* 2018;15(2):220. doi 10.3390/ijerph15020220

Wilkerson R.C., Linton Y.M., Strickman D. Mosquitoes of the World. Baltimore: Johns Hopkins University Press, 2021

Xing Z.P., Liang X., Wang X., Hu H.Y., Huang Y.X. Novel gene rearrangement pattern in mitochondrial genome of *Ooencyrtusplautus* Huang & Noyes, 1994: new gene order in *Encyrtidae* (Hymenoptera, Chalcidoidea). *ZooKeys.* 2022;1124:1-21. doi 10.3897/zookeys.1124.83811

Yang Z., Bielawski J.R. Statistical methods for detecting molecular adaptation. *Trends Ecol Evol.* 2000;15(12):496-503. doi 10.1016/s0169-5347(00)01994-7

Ze-Ze L., Borges V., Osório H.C., Machado J., Gomes J.P., Alves M.J. Mitogenome diversity of *Aedes* (*Stegomyia*) *albopictus*: detection of multiple introduction events in Portugal. *PLoS Negl Trop Dis.* 2020;14(9):e0008657. doi 10.1371/journal.pntd.0008657

Zhang B., Havird J.C., Wang E., Lv J., Xu X. Massive gene rearrangement in mitogenomes of phytoseiid mites. *Int J Biol Macromol.* 2021;186:33-39. doi 10.1016/j.ijbiomac.2021.07.011

Zhang H.G., Lv M.H., Yi W.B., Zhu W.B., Bu W.J. Species diversity can be overestimated by a fixed empirical threshold: insights from DNA barcoding of the genus *Cletus* (Hemiptera: Coreidae) and the meta-analysis of *COI* data from previous phylogeographical studies. *Mol Ecol Resour.* 2017;17(2):314-323. doi 10.1111/1755-0998.12571

Zhang Z., Li J., Yu J. Computing Ka and Ks with a consideration of unequal transitional substitutions. *BMC Evol Biol*. 2006;6:44. doi 10.1186/1471-2148-6-44

Zheng S., Poczai P., Hyvönen J., Tang J., Amiryousefi A. Chloroplot: an online program for the versatile plotting of organelle genomes. *Front Genet.* 2020;11:576124. doi 10.3389/fgene.2020.576124

Conflict of interest. The authors declare no conflict of interest.
Received May 30, 2024. Revised July 10, 2024. Accepted September 6, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):230-237

doi 10.18699/vjgb-25-26

Cytogenetics of insects in the era of chromosome-level genome assemblies

V.A. Lukhtanov [1], E.A. Pazhenkova [2]

[1] Zoological Institute of the Russian Academy of Sciences, St. Petersburg, Russia
[2] University of Ljubljana, Ljubljana, Slovenia
lukhtanov@mail.ru

Abstract. Over the past few years, a revolution has occurred in cytogenetics, driven by the emergence and spread of methods for obtaining high-quality chromosome-level genome assemblies. In fact, this has led to a new tool for studying chromosomes and chromosomal rearrangements, and this tool is thousands of times more powerful than light microscopy. This tool has revolutionized the cytogenetics of many groups of insects for which previously karyotype information, if available at all, was limited to the chromosome number. Even more impressive are the achievements of the genomic approach for studying the general patterns of chromosome organization and evolution in insects. Thus, it has been shown that rapid transformations of chromosomal numbers, which are often found in the order Lepidoptera, are most often carried out in the most parsimonious way, as a result of simple fusions and fissions of chromosomes. It has been established that these fusions and fissions are not random and occur independently in different phylogenetic lineages due to the reuse of the same ancestral chromosomal breakpoints. It has been shown that the tendency for chromosome fissions is correlated with the presence in chromosomes of the so-called interstitial telomeres, i.e. telomere-like structures located not at the ends of chromosomes, but inside them. It has been revealed that, in most insects, telomeric DNA is not just a set of short repeats, but a very long sequence consisting of $(TTAGG)_n$ (or other telomeric motifs), regularly and specifically interrupted by retrotransposons, and the telomeric motifs are diverse in terms of their length and nucleotide composition. The number of high-quality chromosome-level genome assemblies available for insects in the GenBank database is growing exponentially and now exceeds a thousand species. Therefore, the exceptional prospects for using genomic data for karyotype analysis are beyond doubt.

Key words: chromosome; karyotype; chromosomal rearrangements; telomere; meiotic drive; recombination; sex chromosomes; inversion; synteny

For citation: Lukhtanov V.A., Pazhenkova E.A. Cytogenetics of insects in the era of chromosome-level genome assemblies. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):230-237. doi 10.18699/vjgb-25-26

Funding. The work was prepared with the financial support of the Russian Science Foundation grant No. 24-14-00047, https://rscf.ru/project/24-14-00047/.

Цитогенетика насекомых в эпоху хромосомных сборок полных геномов

В.А. Лухтанов [1], Е.А. Паженкова [2]

[1] Зоологический институт Российской академии наук, Санкт-Петербург, Россия
[2] Университет Любляны, Любляна, Словения
lukhtanov@mail.ru

Аннотация. За последние несколько лет в цитогенетике произошли серьезные изменения, связанные с разработкой и распространением методов получения высококачественных хромосомных сборок полных геномов. Фактически это привело к появлению нового инструмента для изучения хромосом и хромосомных перестроек, мощность которого многократно превосходит возможности световой микроскопии. Использование этого инструмента революционизировало частную цитогенетику многих групп насекомых, для которых ранее информация о кариотипах, если она была вообще, ограничивалась элементарным подсчетом числа митотических или мейотических хромосом. Цель данного краткого обзора – обобщение достижений сравнительной и эволюционной цитогенетики насекомых, которые были получены на основании биоинформатического анализа хромосомных сборок полных геномов. С помощью этого подхода было показано, что в процессе быстрой хромосомной эволюции у чешуекрылых (отряд Lepidoptera) преобразования хромосомных чисел чаще всего осуществляются наиболее парсимониальным способом: в результате простых слияний и разделений хромосом. Установлено, что эти слияния и разделения не случайны и могут осуществляться в разных филогенетических линиях за счет повторного использования одних и тех же предковых хромосомных точек разрыва. Тенденция к разделениям хромосом скоррелирована с наличием в хромосомах так называемых интерстициальных теломер – теломеро-

подобных структур, расположенных не на концах хромосом, а внутри них. При изучении теломерных регионов выявлено, что у большинства насекомых теломерная ДНК – это не просто набор коротких повторов, а очень длинная последовательность, состоящая из $(TTAGG)_n$ (или других мотивов), регулярно и специфически прерываемая ретротранспозонами, а сами теломерные мотивы чрезвычайно разнообразны по длине и нуклеотидному составу. Число высококачественных хромосомных сборок геномов насекомых, доступных в базе данных GenBank, растет в геометрической прогрессии и уже превышает тысячу видов. Поэтому исключительные перспективы использования хромосомных сборок геномов для анализа кариотипов не вызывают сомнений.
Ключевые слова: хромосома; кариотип; хромосомные перестройки; теломера; мейотический драйв; рекомбинация; половые хромосомы; инверсии; синтении

Introduction

Progress in science is often driven by new research methods. In the field of genetics, one such fundamentally new approach, which has given a powerful impetus to the development of the entire discipline, is the DNA sequencing procedure (Heather, Chain, 2016). High-throughput sequencing methods, along with advances in bioinformatics and the development of the Hi-C DNA analysis protocol (Lieberman-Aiden et al., 2009), have recently led to a breakthrough technology for obtaining genome assemblies at the chromosome level (Dudchenko et al., 2017). This methodology has revolutionized comparative cytogenetics, stimulating the emergence of a large number of works devoted to the structure of chromosomes, patterns of chromosomal changes in evolution and role of these changes in speciation. In fact, this new technique has switched the attention of many biologists, especially bioinformaticians, from the analysis of nucleotide substitutions to the analysis of structural changes in DNA, enhancing and complementing the work previously done with a microscope.

The aim of this brief review is to analyze and summarize the advances made in comparative and evolutionary cytogenetics of insects using bioinformatic analysis of chromosome level genome assemblies.

The main stages of insect cytogenetics

Although the karyotypes of some model insect species, such as the midge *Chironomus plumosus* and the silkworm *Bombyx mori*, have been studied in great detail (Kiknadze et al., 1991, 2016; Yoshido et al., 2005), for non-model species, information on karyotypes, if available at all, is often limited to the estimation of the diploid (or haploid) number of chromosomes, an approximate description of the size characteristics of individual chromosomes and, less often, individual chromosome arms in the form of a centromeric index (Peruzzi, Eroğlu, 2013). It should be noted that obtaining the latter characteristic is, in principle, impossible for representatives of many insect orders, for example, for butterflies (Lepidoptera) and bugs (Hemiptera), since they have holocentric chromosomes, that is, they do not have a localized centromere (Mandrioli, Manicardi, 2020). Such a low average level of insect cytogenetics is largely due to the objective difficulties of studying chromosomes using a microscope: the sizes of chromosomes are often at the limit of the resolving power of light microscopy.

It is therefore not surprising that in the history of cytogenetics, beginning with its inception in the 19th century, attempts have been made to increase the resolving power of cytogenetic analysis. The first stage in the history of cytogenetics can be called the era of chromosome numbers. It arose in the second half of the 19th century, when the first descriptions and images of karyotypes containing the correct determination of the number of chromosomes appeared (e. g., Henking, 1890). The heyday of this era came in the first half of the 20th century, when the study of karyotypes became a mass phenomenon (Beliajeff, 1930; White, 1973).

Significant progress in cytogenetic research was associated with the emergence and widespread use in the second half of the 20th century of methods of differential staining of chromosomes, such as C-banding (Pardue, Gall, 1970) and G-banding (Seabright, 1971). Cytogenetics entered the era of chromosome banding. Almost simultaneously, even more powerful methods of cytogenetic analysis appeared and were developed in parallel, based on the use of the FISH method (Gall, Pardue, 1969; Langer-Safer et al., 1982) and its modifications, such as BAC-FISH (BAC Resource Consortium 2001; Yoshido et al., 2005) and chromosome painting (Schrock et al., 1996; Speicher et al., 1996). This has led to stunning advances in cytogenetics of many groups of organisms, especially vertebrates (Ferguson-Smith, Trifonov, 2007; Graphodatsky et al., 2011). As for insects, with the exception of some model species (Yoshido et al., 2005), this progress has affected them to a lesser extent. Of course, light microscopes have become much better, and the resulting images of karyotypes have become much clearer compared to what they were 100 years ago. In addition, the GISH method has made it possible to effectively detect sex chromosomes (Fukova et al., 2005; Šíchová et al., 2015). Despite this, the cytogenetics of many insect groups, for example, most families of Lepidoptera, is still at the stage of elementary counting of chromosome numbers (Pazhenkova, Lukhtanov, 2023a).

Chromosome level genome assemblies: a new tool for studying karyotypes

A revolution in the field of karyotype studies has occurred over the past six-eight years. Modern approaches to genome analysis based on obtaining long reads and using Hi-C technology (Dudchenko et al., 2017) make it possible to obtain chromosome level genome assemblies, in which all or at least most of the chromosomes are read from telomere to telomere (Miga et al., 2020; The Darwin Tree…, 2022; Zhang et al., 2023).

Currently, high-quality chromosome level genome assemblies have been obtained for representatives of most insect orders: fleas (Siphonaptera) (Driscoll et al., 2020), stoneflies (Plecoptera) (Dixon et al., 2023), dipterans (Diptera) (Zamyatin et al., 2021; Reinhardt et al., 2023), beetles (Coleoptera) (Van Dam et al., 2021; Huang et al., 2022), springtails (Collembola) (Jin et al., 2023), stick insects (Phasmatodea) (La-

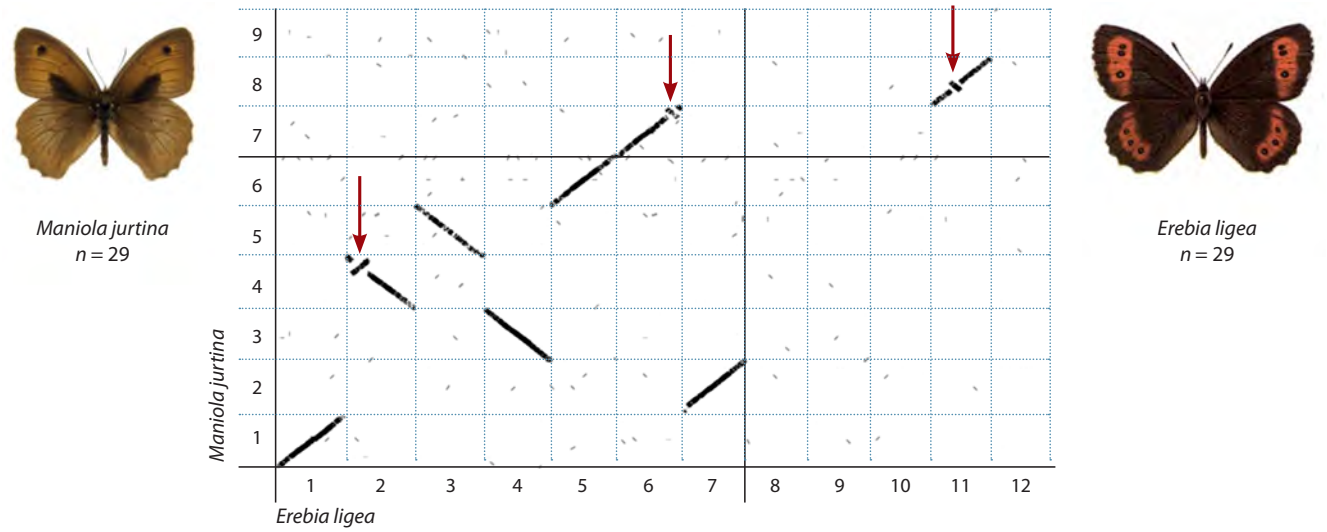


Fig. 1. Comparison of the genomes of the butterflies *Maniola jurtina* and *Erebia ligea* based on chromosome assemblies.

The butterflies *M. jurtina* and *E. ligea* have 29 chromosomes in the haploid set. The first 12 chromosomes of *E. ligea* are mapped on the abscissa axis. The first nine chromosomes of *M. jurtina* are mapped on the ordinate axis. The diagonals on the graph show the regions of macrosynteny. Inversions are marked with red arrows. It is seen that chromosome 1 of *M. jurtina* is homologous to chromosome 1 of *E. ligea*, chromosome 2 of *M. jurtina* is homologous to chromosome 7 of *E. ligea*, etc. (according to: Pazhenkova, Lukhtanov, 2023b).

vanchy et al., 2024), Hymenoptera (Sun et al., 2021), mayflies (Ephemeroptera) (Farr et al., 2023), Hemiptera (Biello et al., 2021; Mathers et al., 2021; Chen H. et al., 2022; Wang et al., 2024), Orthoptera (Li R. et al., 2024), Trichoptera (Ge et al., 2024), Psocoptera (Feng et al., 2022), Neuroptera (Wang et al., 2022), Odonata (Patterson et al., 2024), Thysanoptera (Yingning et al., 2024), Dermaptera (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_963082975.1/) and a large number of Lepidoptera species (Mackintosh et al., 2022a; Gauthier et al., 2023; Wright et al., 2024). These assemblies contain information on the haploid number of chromosomes and the size of each chromosome, measured in the number of base pairs. Almost always, there is also information on the presence and size of the sex chromosome X (Z for Lepidoptera and caddisflies, in which females are the heterogametic sex) and, less often, the sex chromosome Y (W for Lepidoptera).

The use of chromosome level genome assemblies has actually led to the emergence of a new methodology and a new tool for studying chromosomes and chromosomal rearrangements. The resolution power of this tool significantly exceeds the capabilities of light microscopy. The basis of the methodology is to obtain pairwise or multiple alignments of chromosome assemblies of different species. These alignments are usually presented in the form of circular plots (Krzywinski et al., 2009). Another analysis option is to obtain pairwise comparisons presented in the form of dot plots (Li H., 2018), in which the nucleotide sequences of individual chromosomes are plotted along the abscissa and ordinate axes, starting with the first, largest chromosome (Fig. 1). Such graphs clearly demonstrate macrosyntenic regions and identify chromosome fusions/fissions, as well as chromosomal inversions. The latter are visible on the graph as segments that are perpendicular to the main diagonals (Fig. 1).

The use of this tool has revolutionized cytogenetics of many insect groups, for which karyotype information, if any, was previously limited to elementary counts of mitotic or meiotic chromosomes. However, even more impressive are the achievements of the genomic approach to the study of general patterns of chromosome organization and karyotype evolution, including the analysis of chromosomal rearrangements and chromosomal syntenies (Biello et al., 2021; Mathers et al., 2021; Sun et al., 2021; Van Dam et al., 2021; Höök et al., 2023; Hundsdoerfer et al., 2023; Wright et al., 2024), as well as the reconstruction of ancestral karyotypes (Chen X. et al., 2023; Wright et al., 2024). Chromosome level genome assemblies have been used to study meiotic drive (Reinhardt et al., 2023; Boman et al., 2024), sex chromosome evolution (Mackintosh et al., 2022b; Berner et al., 2023; Höök et al., 2023, 2024), interspecific transfer of chromosomal inversion during interspecific hybridization (Seixas et al., 2021), the role of chromosomal rearrangements in the evolution of recombination frequency (Näsvall et al., 2023), and to identify genomic coordinates for breakpoints that give rise to chromosomal rearrangements (Zamyatin et al., 2021).

Chromosomal conservatism and rapid karyotypic evolution

In our work (Pazhenkova, Lukhtanov, 2023a), we used the analysis of chromosome level genome assemblies to solve one of the mysteries of evolutionary cytogenetics. It is known that the chromosome numbers of many insects are conservative and remain unchanged or with minimal changes for tens and hundreds of millions of years (White, 1973). For example, in the order Lepidoptera (butterflies and moths), the ancestral haploid chromosome number $n = 31$ has been preserved for 200 million years, although $n = 30$ is often found in some species along with $n = 31$. In the blue butterflies (the family Lycaenidae), the haploid number $n = 24$ predominates, although $n = 23$ is often found (Robinson, 1971). This suggests that large chromosomal rearrangements are rare in the evolution of the order Lepidoptera. At the same time, in some genera of butterflies, there are explosions of karyotypic variability,

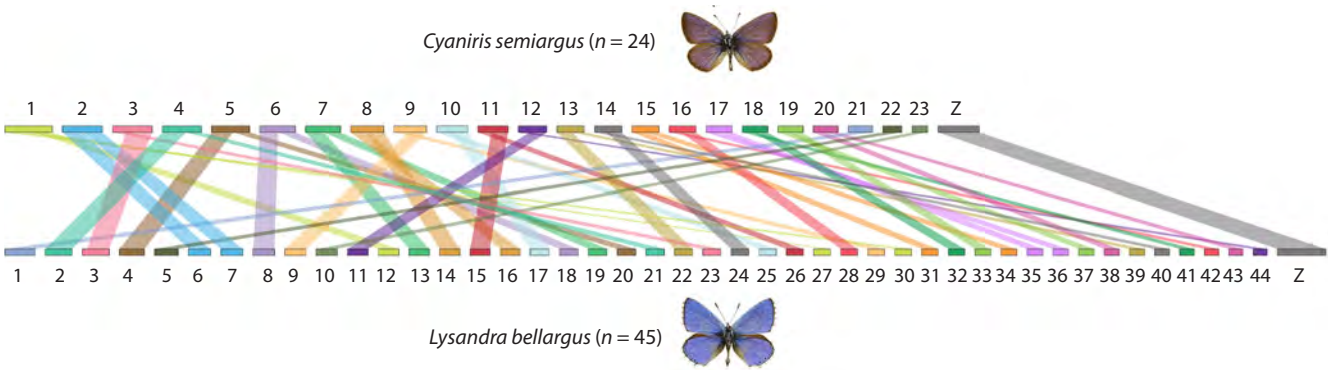


Fig. 2. Schematic representation of chromosomes and regions of macrosynteny in the karyotypes of the butterflies *Cyaniris semiargus* (*n* = 23+Z) and *Lysandra bellargus* (*n* = 44+Z).

The karyotype of *L. bellargus* differs from the karyotype of *C. semiargus* by chromosome fissions that have occurred in 21 out of 23 autosomes (according to: Pazhenkova, Lukhtanov, 2023a).

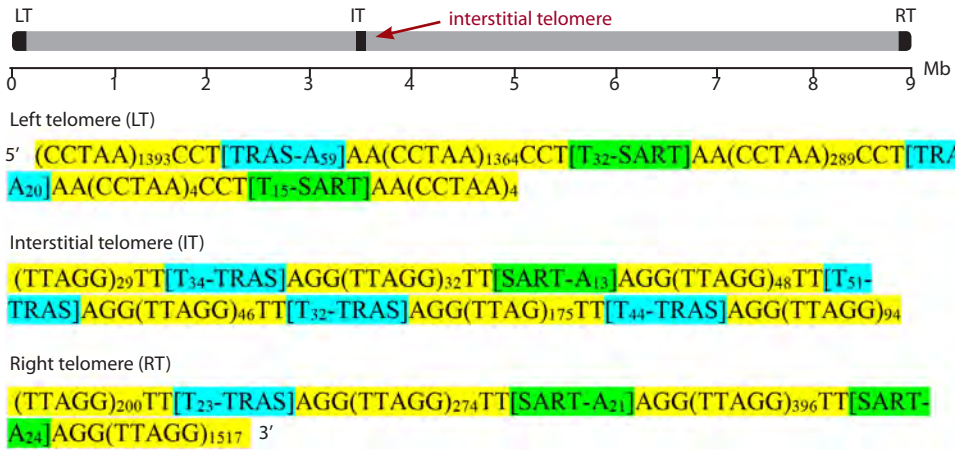


Fig. 3. Telomere and interstitial telomere structure in chromosome 38 of *Lysandra bellargus*.

Each telomere is a long array $(CCTAA)_n/(TTAGG)_n$ (yellow) interspersed with retrotransposable TRAS (blue) and SART (green) elements. The TRAS and SART elements have long A_n/T_n and T_n/A_n tails and are specifically inserted between CCT/AGG and AA/TT nucleotides of the TTAGG motif (according to: Pazhenkova, Lukhtanov, 2023a).

and chromosome numbers change dramatically in a very short time, for example, during the divergence of two closely related species (White, 1973). The chromosomal mechanisms of such rapid karyotypic evolution were unclear. In addition, it was unclear how real the phenomenon of chromosomal conservatism itself was, since the preservation of the ancestral chromosome number does not exclude intra-chromosomal rearrangements.

Analysis of chromosome level genome assemblies in multiple Lepidoptera species showed that in the evolutionary phase of chromosomal conservatism, most autosomes are indeed stable. However, this does not apply to the sex chromosome Z. Fusions of the Z chromosome with one of the autosomes, independently occurring in different evolutionary lineages, lead to multiple variants of the NeoZ chromosome and a decrease in the haploid number by one unit.

As for the explosive karyotypic evolution, the most rapid changes in chromosome numbers are carried out in a parsimonious way: as a result of simple fusions and fissions of chromosomes (Fig. 2). Moreover, these fusions and fissions are not random and can be carried out in different phylogenetic lineages due to the repeated use of the same ancestral chromosomal breakpoints (Pazhenkova, Lukhtanov, 2023a). It should also be noted that the tendency for breaks is correlated with the presence of the so-called interstitial telomeres in chromosomes, i. e. telomere-like structures located not at the ends of chromosomes, but inside them (Fig. 3).

Telomeric DNA of insects

It is believed that in most insects, telomeric DNA consists of a canonical five-letter TTAGG motif, which is repeated hundreds and thousands of times at the ends of chromosomes (Kuznetsova et al., 2020). However, the analysis of telomeric DNA in 220 insect species in our studies (Lukhtanov, 2022; Lukhtanov, Pazhenkova, 2023), as well as other works that appeared in parallel (Zhou et al., 2022; Fajkus et al., 2023), showed that in addition to the canonical TTAGG motif, insects contain a large number of other variants of telomeric repeats, the length of which varies from 1 to 11 nucleotides (see the Table).

Even more intriguing is the fact that the vast majority of insects have telomeres with a complex multilayer structure (Lukhtanov, Pazhenkova, 2023). In these telomeres, blocks of short telomeric motifs are regularly interrupted by retrotrans-

Non-canonical telomeric motifs found in insects (Lukhtanov, Pazhenkova, 2023)

1 bp motif	5 bp motif	6 bp motif	8 bp motif	10 bp motif	11 bp motif
T	TTGGG	TTAGGG	TTATTGGG	TTAGGGATGG	TTAGGTCTGGG
	TCAGG	TTCCTC		TTAGGGGTGG	TTAGTCTTGGG
		TTTGGG		TTAGGGTGGT	TTAGGTTGGGG
		TCTGGG		TTAGTTTGGG	TTAGGTTCGGG
				TTTGTTTGGG	TTAGGTTTGGG
				TTATTGAGGT	TTGGGTCTGGG
					TTGCGTCTGGG
					TTGCGTCAGGG

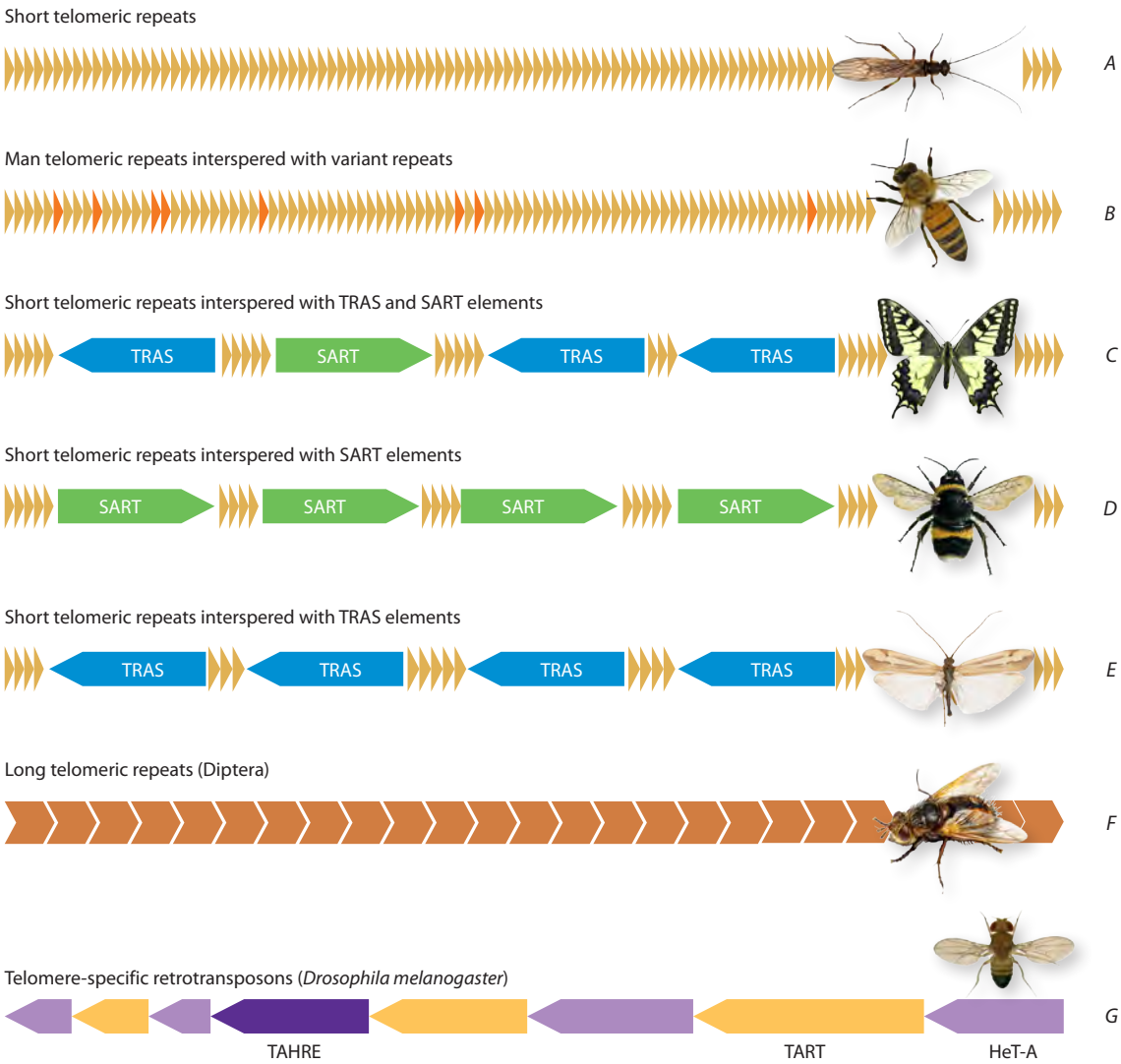


Fig. 4. Types of telomeric DNA organization in insects.

A – short repeats; *B* – main short repeats interspersed with variant short repeats, known in the honey bee *Apis mellifera* (Wallberg et al., 2019); *C* – main short repeats with insertions of telomere-specific non-LTR retrotransposons of the *TRAS* and *SART* families (found in Lepidoptera); *D* – main short repeats (5–11 bp) with insertions of telomere-specific non-LTR retrotransposons of the *SART* family (found in most studied species of Hemiptera, Coleoptera and many Hymenoptera); *E* – main short repeats with insertions of telomere-specific non-LTR retrotransposons of the *TRAS* family (found in Trichoptera); *F* – long (173–381 bp) repeats (found in Diptera); *G* – telomere-specific non-LTR retrotransposons of the *HeT-A*, *TAHRE* and *TART* families (found in *Drosophila melanogaster*) (Biessmann et al., 2000; Casacuberta, Pardue, 2003).

posons that are specifically embedded in repeats of both the canonical TTAGG motif (Fig. 3) and other non-canonical motifs. In our opinion, such a structure indirectly indicates the presence of two parallel mechanisms for maintaining telomere length during cell divisions in insects: the classical telomerase mechanism and a mechanism based on transpositions. In general, insects are characterized by a great diversity in the organization of telomeric DNA, which is summarized in Figure 4.

Prospects for the use of chromosome level genome assemblies in cytogenetics

The number of high-quality chromosome level genome assemblies available in the GenBank database is growing exponentially due to the activities of various laboratories, and primarily The Wellcome Sanger Institute in the UK (The Darwin Tree…, 2022). Currently, GenBank contains information on chromosome assemblies of genomes (including determination of the haploid number of chromosomes) for 1,118 insect species (https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=50557, dated May 15, 2024). Thus, it can already be stated that over the past three-four years, the number of new insect karyotypes, obtained using bioinformatic analysis of genomes, is comparable to or even exceeds the number of karyotypes studied using routine cytogenetic analysis. In addition, chromosomal assemblies of genomes carry several orders of magnitude more information about the obtained karyotypes.

The successes of the genomic approach discussed above do not mean that classical cytogenetics, which provides information on the real spatial configurations of chromosomes, should be discounted. Classical cytogenetics is also needed to validate chromosome level genome assemblies, in particular to confirm the number of chromosomes (Pazhenkova, Lukhtanov, 2023a) and the structure of telomeric DNA motifs (Dalla Benetta et al., 2020; Stoianova et al., 2024). Examples of such validation reinforce the conclusion that chromosome level genome assemblies are a reliable source of information on karyotypes. For instance, for butterflies and moths (order Lepidoptera), information on chromosome assemblies, including determination of the haploid number of chromosomes, is available for 452 species (https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=7088, from May 15, 2024). For more than half of them, there are data on chromosome numbers obtained using light microscopy methods. Having compared these data, we found complete agreement in the haploid chromosome number calculations made using the bioinformatics approach and light microscopy methods (Pazhenkova, Lukhtanov, 2023a).

Thus, the exceptional prospects of using chromosome level genome assemblies for karyotype analysis are beyond doubt.

References

BAC Resource Consortium. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature*. 2001;409:953-958. doi 10.1038/35057192

Beliajeff N.K. Die Chromosomenkomplexe und ihre Beziehung zur Phylogenie bei den Lepidopteren. *Z Vererbungsl*. 1930;14:369-399. doi 10.1007/BF01848966

Berner D., Ruffener S., Blattner L.A. Chromosome-level assemblies of the *Pieris mannii* butterfly genome suggest Z-origin and rapid evolution of the W chromosome. *Genome Biol Evol*. 2023;15(6):evad111. doi 10.1093/gbe/evad111

Biello R., Singh A., Godfrey C.J., Fernández F.F., Mugford S.T., Powell G., Hogenhout S.A., Mathers T.C. A chromosome-level genome assembly of the woolly apple aphid, *Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae). *Mol Ecol Resour.* 2021;21(1):316-326. doi 10.1111/1755-0998.13258

Biessmann H., Zurovcova M., Yao J.G., Lozovskaya E., Walter M.F. A telomeric satellite in *Drosophila virilis* and its sibling species. *Chromosoma.* 2000;109:372-380. doi 10.1007/s004120000094

Boman J., Wiklund C., Vila R., Backström N. Meiotic drive against chromosome fusions in butterfly hybrids. *Chromosome Res*. 2024;32(2):7. doi 10.1007/s10577-024-09752-0

Casacuberta E., Pardue M.L. Transposon telomeres are widely distributed in the *Drosophila* genus: TART elements in the *virilis* group. *Proc Natl Acad Sci USA.* 2003;100:3363-3368. doi 10.1073/pnas.0230353100

Chen H., Qiao G., Liang A. Chromosome-level genome assembly of *Callitettix versicolor* (rice spittlebug). *Genome Biol Evol.* 2022;14(9):evac130. doi 10.1093/gbe/evac130

Chen X., Wang Z., Zhang C., Hu J., Lu Y., Zhou H., Mei Y., Cong Y., Guo F., Wang Y., He K., Liu Y., Li F. Unraveling the complex evolutionary history of lepidopteran chromosomes through ancestral chromosome reconstruction and novel chromosome nomenclature. *BMC Biol.* 2023;21:265. doi 10.1186/s12915-023-01762-4

Dalla Benetta E., Antoshechkin I., Yang T., Nguyen H.Q.M., Ferree P.M., Akbari O.S. Genome elimination mediated by gene expression from a selfish chromosome. *Sci Adv*. 2020;6:eaaz9808. doi 10.1126/sciadv.aaz9808

Dixon A., Macadam C.R., Natural History Museum Genome Acquisition Lab, Darwin Tree of Life Barcoding collective, Wellcome Sanger Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium. The genome sequence of the spring stonefly, *Protonemura montana* (Kimmins, 1941) [version 1; peer review: awaiting peer review]. *Wellcome Open Res*. 2023;8:333. doi 10.12688/wellcomeopenres.19790.1

Driscoll T.P., Verhoeve V.I., Gillespie J.J., Johnston J.S., Guillotte M.L., Rennoll-Bankert K.E., Rahman M.S., Hagen D., Elsik C.G., Macaluso K.R., Azad A.F. A chromosome-level assembly of the cat flea genome uncovers rampant gene duplication and genome size plasticity. *BMC Biol.* 2020;18(1):70. doi 10.1186/s12915-020-00802-7

Dudchenko O., Batra S.S., Omer A.D., Nyquist S.K., Hoeger M., Durand N.C., Shamim M.S., Machol I., Lander E.S., Aiden A.P., Aiden E.L. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017;356(6333):92-95. doi 10.1126/science.aal3327

Fajkus P., Adámik M., Nelson A.D.L., Kilar A.M., Franek M., Bubeník M., Frydrychová R.Č., Votavová A., Sýkorová E., Fajkus J., Peška V. Telomerase RNA in Hymenoptera (Insecta) switched to plant/ciliate-like biogenesis. *Nucleic Acids Res*. 2023;51(1):420-433. doi 10.1093/nar/gkac1202

Farr A., Macadam C.R., Natural History Museum Genome Acquisition Lab; Darwin Tree of Life Barcoding collective; Wellcome Sanger Institute Tree of Life programme; Wellcome Sanger Institute Scientific Operations: DNA Pipelines collective; Tree of Life Core Informatics collective; Darwin Tree of Life Consortium. The genome sequence of the Northern Summer Mayfly, *Siphlonurus alternatus* (Say, 1824) [version 1; peer review: 2 approved, 3 approved with reservations]. *Wellcome Open Res*. 2023,8:488. doi 10.12688/wellcomeopenres.20172.1

Feng S., Opit G., Deng W., Stejskal V., Li Z. A chromosome-level genome of the booklouse, *Liposcelis brunnea*, provides insight into louse evolution and environmental stress adaptation. *GigaScience*. 2022;11:giac062. doi 10.1093/gigascience/giac062

Ferguson-Smith M.A., Trifonov V. Mammalian karyotype evolution. *Nat Rev Genet*. 2007;8(12):950-962. doi 10.1038/nrg2199

Fuková I., Nguyen P., Marec F. Codling moth cytogenetics: karyotype, chromosomal location of rDNA, and molecular differentiation of sex chromosomes. *Genome*. 2005;48(6):1083-1092. doi 10.1139/g05-063

Gall J.G., Pardue M.L. Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc Natl Acad Sci USA*. 1969;63(2):378-383. doi 10.1073/pnas.63.2.378

Gauthier J., Meier J., Legeai F., McClure M., Whibley A., Bretaudeau A., Boulain H., Parrinello H., Mugford S.T., Durbin R., Zhou C., McCarthy S., Wheat C.W., Piron-Prunier F., Monsempes C., François M.C., Jay P., Noûs C., Persyn E., Jacquin-Joly E., Meslin C., Montagné N., Lemaitre C., Elias M. First chromosome scale genomes of ithomiine butterflies (Nymphalidae: Ithomiini): comparative models for mimicry genetic studies. *Mol Ecol Resour.* 2023;23(4):872-885. doi 10.1111/1755-0998.13749

Ge X., Peng L., Deng Z., Du J., Sun C., Wang B. Chromosome-scale genome assemblies of *Himalopsyche anomala* and *Eubasilissa splendida* (Insecta: Trichoptera). *Sci Data.* 2024;11(1):267. doi 10.1038/s41597-024-03097-3

Graphodatsky A.S., Trifonov V.A., Stanyon R. The genome diversity and karyotype evolution of mammals. *Mol Cytogenet*. 2011;4:22. doi 10.1186/1755-8166-4-22

Heather J.M., Chain B. The sequence of sequencers: the history of sequencing DNA. *Genomics*. 2016;107(1):1-8. doi 10.1016/j.ygeno.2015.11.003

Henking H. Untersuchung über die erster Entwicklungsvorgänge in den Eien der Insekten. I. Das Ei von *Pieris brassicae* L. nebst Bemerkungen über Samen und Samenbildung. *Z Wiss Zool*. 1890;49:503-564

Höök L., Näsvall K., Vila R., Wiklund C., Backström N. High-density linkage maps and chromosome level genome assemblies unveil direction and frequency of extensive structural rearrangements in wood white butterflies (*Leptidea* spp.). *Chromosome Res.* 2023;31(1):2. doi 10.1007/s10577-023-09713-z

Höök L., Vila R., Wiklund C., Backström N. Temporal dynamics of faster Neo-Z evolution in butterflies. *Evolution*. 2024;78:1554-1567. doi 10.1093/evolut/qpae082

Huang Q., Sim S.B., Geib S.M., Childers A., Liu J., Wei X., Han W., Posada-Florez F., Xue A.Z., Li Z., Evans J.D. Identification of sex chromosomes and primary sex ratio in the small hive beetle, a worldwide parasite of honey bees. *GigaScience*. 2022;12:giad056. doi 10.1093/gigascience/giad056

Hundsdoerfer A.K., Schell T., Patzold F., Wright C.J., Yoshido A., Marec F., Daneck H., Winkler S., Greve C., Podsiadlowski L., Hiller M., Pippel M. High-quality haploid genomes corroborate 29 chromosomes and highly conserved synteny of genes in *Hyles* hawkmoths (Lepidoptera: Sphingidae). *BMC Genomics*. 2023;24(1):443. doi 10.1186/s12864-023-09506-y

Jin J., Zhao Y., Zhang G., Pan Z., Zhang F. The first chromosome-level genome assembly of *Entomobrya proxima* Folsom, 1924 (Collembola: Entomobryidae). *Sci Data*. 2023;10(1):541. doi 10.1038/s41597-023-02456-w

Kiknadze I.I., Shilova A.I., Kerkis I.E., Shobanov N.A., Zelentsov N.I., Grebenyuk L.P., Istomina A.G., Prasolov V.A. Karyotypes and Morphology of Larvae of the Tribe Chironomini. An Atlas. Novosibirsk: Nauka Publ., 1991 (in Russian)

Kiknadze I., Istomina A., Golygina V., Gunderina L. Karyotypes of Palearctic and Holarctic Species of the Genus *Chironomus*. Novosibirsk: Academic Publishing House "Geo", 2016

Krzywinski M., Schein J., Birol I., Connors J., Gascoyne R., Horsman D., Jones S.J., Marra M.A. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639-1645. doi 10.1101/gr.092759.109

Kuznetsova V., Grozeva S., Gokhman V. Telomere structure in insects: a review. *J Zool Syst Evol Res*. 2020;58:127-158. doi 10.1111/jzs.12332

Langer-Safer P.R., Levine M., Ward D.C. Immunological method for mapping genes on drosophila polytene chromosomes. *Proc Natl Acad Sci USA*. 1982;79(14):4381-4385. doi 10.1073/pnas.79.14.4381

Lavanchy G., Brandt A., Bastardot M., Dumas Z., Labédan M., Massy M., Toubiana W., Tran Van P., Luchetti A., Scali V., Mantovani B., Schwander T. Evolution of alternative reproductive systems in *Bacillus* stick insects. *Evolution*. 2024;78(6):1109-1120. doi 10.1093/evolut/qpae045

Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094-3100

Li R., Qin Y., Rong W., Deng W., Li X. Chromosome-level genome assembly of the pygmy grasshopper *Eucriotettix oculatus* (Orthoptera: Tetrigoidea). *Sci Data*. 2024;11:431. doi 10.1038/s41597-024-03276-2

Lieberman-Aiden E., van Berkum N.L., Williams L., Imakaev M., Ragoczy T., Telling A., Amit I., Lajoie B.R., Sabo P.J., Dorschner M.O., Sandstrom R., Bernstein B., Bender M.A., Groudine M., Gnirke A., Stamatoyannopoulos J., Mirny L.A., Lander E.S., Dekker J. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289-293. doi 10.1126/science.1181369

Lukhtanov V.A. Diversity and evolution of telomere and subtelomere DNA sequences in insects. *bioRxiv*. 2022. doi 10.1101/2022.04.08.487650

Lukhtanov V.A., Pazhenkova E.A. Diversity and evolution of telomeric motifs and telomere DNA organization in insects. *Biol J Linn Soc*. 2023;140:536-555. doi 10.1093/biolinnean/blad068

Mackintosh A., Laetsch D.R., Baril T., Ebdon S., Jay P., Vila R., Hayward A., Lohse K. The genome sequence of the scarce swallowtail, *Iphiclides podalirius*. *G3* (*Bethesda*). 2022a;12(9):jkac193. doi 10.1093/g3journal/jkac193

Mackintosh A., Laetsch D.R., Baril T., Foster R.G., Dincă V., Vila R., Hayward A., Lohse K. The genome sequence of the lesser marbled fritillary, *Brenthis ino*, and evidence for a segregating neo-Z chromosome. *G3* (*Bethesda*). 2022b;12(6):jkac069. doi 10.1093/g3journal/jkac069

Mandrioli M., Manicardi G.C. Holocentric chromosomes. *PLoS Genet*. 2020;16(7):e1008918. doi 10.1371/journal.pgen.1008918

Mathers T.C., Wouters R.H.M., Mugford S.T., Swarbreck D., van Oosterhout C., Hogenhout S.A. Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. *Mol Biol Evol*. 2021;38(3):856-875. doi 10.1093/molbev/msaa246

Miga K.H., Koren S., Rhie A., Vollger M.R., Gershman A., Bzikadze A., Brooks S., … Pevzner P.A., Gerton J.L., Sullivan B.A., Eichler E.E., Phillippy A.M. Telomere-to-telomere assembly of a complete human X chromosome. *Nature*. 2020;585(7823):79-84. doi 10.1038/s41586-020-2547-7

Näsvall K., Boman J., Höök L., Vila R., Wiklund C., Backström N. Nascent evolution of recombination rate differences as a consequence of chromosomal rearrangements. *PLoS Genet.* 2023;19(8):e1010717. doi 10.1371/journal.pgen.1010717

Pardue M.L., Gall J.G. Chromosomal localization of mouse satellite DNA. *Science*. 1970;168(3937):1356-1358. doi 10.1126/science.168.3937.1356

Patterson C.W., Bonillas-Monge E., Brennan A., Grether G.F., Mendoza-Cuenca L., Tucker R., Vega-Sánchez Y.M., Drury J. A chromosome-level genome assembly for the smoky rubyspot damselfly (*Hetaerina titia*). *J Hered*. 2024;115(1):103-111. doi 10.1093/jhered/esad070

Pazhenkova E.A., Lukhtanov V.A. Chromosomal conservatism vs chromosomal megaevolution: enigma of karyotypic evolution in Lepidoptera. *Chromosome Res*. 2023a;31:16. doi 10.1007/s10577-023-09725-9

Pazhenkova E.A., Lukhtanov V.A. Whole-genome analysis reveals the dynamic evolution of holocentric chromosomes in satyrine butterflies. *Genes*. 2023b;14(2):437. doi 10.3390/genes14020437

Peruzzi L., Eroğlu H.E. Karyotype asymmetry: again, how to measure and what to measure? *Comp Cytogenet*. 2013;7(1):1-9. doi 10.3897/CompCytogen.v7i1.4431

Reinhardt J.A., Baker R.H., Zimin A.V., Ladias C., Paczolt K.A., Werren J.H., Hayashi C.Y., Wilkinson G.S. Impacts of sex ratio meiotic drive on genome structure and function in a stalk-eyed fly. *Genome Biol Evol*. 2023;15(7):evad118. doi 10.1093/gbe/evad118

Robinson R. Lepidoptera Genetics. Oxford: Pergamon Press, 1971

Schrock E., du Manoir S., Veldman T., Schoell B., Wienberg J., Ferguson-Smith M.A., Ning Y., Ledbetter D.H., Bar-Am I., Soenksen D., Garini Y., Ried T. Multicolor spectral karyotyping of human chromosomes. *Science*. 1996;273(5274):494-497. doi 10.1126/science.273.5274.494

Seabright M. A rapid banding technique for human chromosomes. *Lancet*. 1971;2(7731):971-972. doi 10.1016/s0140-6736(71)90287-x

Seixas F.A., Edelman N.B., Mallet J. Synteny-based genome assembly for 16 species of *Heliconius* butterflies, and an assessment of structural variation across the genus. *Genome Biol Evol*. 2021;13(7):evab069. doi 10.1093/gbe/evab069

Šíchová J., Voleníková A., Dincă V., Nguyen P., Vila R., Sahara K., Marec F. Dynamic karyotype evolution and unique sex determination systems in *Leptidea* wood white butterflies. *BMC Evol Biol*. 2015;15:89. doi 10.1186/s12862-015-0375-4

Speicher M.R., Ballard S.G., Ward D.C. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nat Genet*. 1996;12:368-375. doi 10.1038/ng0496-368

Stoianova D., Grozeva S., Golub N.V., Anokhin B.A., Kuznetsova V.G. The first FISH-confirmed non-canonical telomeric motif in Heteroptera: *Cimex lectularius* Linnaeus, 1758 and *C. hemipterus* (Fabricius, 1803) (Hemiptera, Cimicidae) have a 10 bp motif $(TTAGGGATGG)_n$. *Genes*. 2024;15(8):1026. doi 10.3390/genes15081026

Sun C., Huang J., Wang Y., Zhao X., Su L., Thomas G.W.C., Zhao M., … Schaack S., Barribeau S.M., Williams P.H., Waterhouse R.M., Mueller R.L. Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits. *Mol Biol Evol*. 2021;38(2):486-501. doi 10.1093/molbev/msaa240

The Darwin Tree of Life Project Consortium. Sequence locally, think globally: The Darwin Tree of Life Project. *Proc Natl Acad Sci USA*. 2022;119:e2115642118. doi 10.1073/pnas.2115642118

Van Dam M.H., Cabras A.A., Henderson J.B., Rominger A.J., Pérez Estrada C., Omer A.D., Dudchenko O., Lieberman Aiden E., Lam A.W. The Easter Egg Weevil (*Pachyrhynchus*) genome reveals syntenic patterns in Coleoptera across 200 million years of evolution. *PLoS Genet*. 2021;17(8):e1009745. doi 10.1371/journal.pgen.1009745

Wallberg A., Bunikis I., Pettersson O.V., Mosbech M.B., Childers A.K., Evans J.D., Mikheyev A.S., Robertson H.M., Robinson G.E., Webster M.T. A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome length scaffolds. *BMC Genomics*. 2019;20:275. doi 10.1186/s12864-019-5642-0

Wang Y., Zhang R., Wang M., Zhang L., Shi C.M., Li J., Fan F., Geng S., Liu X., Yang D. The first chromosome-level genome assembly of a green lacewing *Chrysopa pallens* and its implication for biological control. *Mol Ecol Resour.* 2022;22(2):755-767. doi 10.1111/1755-0998.13503

Wang Y., Luo Y., Ge Y., Liu S., Liang W., Wu C., Wei S., Zhu J. Chromosome-level genome assembly of the predatory stink bug *Arma custos*. *Sci Data*. 2024;11:417. doi 10.1038/s41597-024-03270-8

White M.J.D. Animal Cytology and Evolution. Cambridge: Cambridge University Press, 1973

Wright C.J., Stevens L., Mackintosh A., Lawniczak M., Blaxter M. Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera. *Nat Ecol Evol.* 2024;8(4):777-790. doi 10.1038/s41559-024-02329-4

Yingning L., Shuhua W., Wenting D., Miao M., Ying W., Rong Z., Liping B. Chromosome-level genome assembly of *Odontothrips loti* Haliday (Thysanoptera: Thripidae). *Sci Data*. 2024;11(1):451. doi 10.1038/s41597-024-03289-x

Yoshido A., Bando H., Yasukochi Y., Sahara K. The *Bombyx mori* karyotype and the assignment of linkage groups. *Genetics*. 2005;170:675-685. doi 10.1534/genetics.104.040352

Zamyatin A., Avdeyev P., Liang J., Sharma A., Chen C., Lukyanchikova V., Alexeev N., Tu Z., Alekseyev M.A., Sharakhov I.V. Chromosome-level genome assemblies of the malaria vectors *Anopheles coluzzii* and *Anopheles arabiensis*. *GigaScience*. 2021;10(3):giab017. doi 10.1093/gigascience/giab017

Zhang T., Xing W., Wang A., Zhang N., Jia L., Ma S., Xia Q. Comparison of long-read methods for sequencing and assembly of lepidopteran pest genomes. *Int J Mol Sci*. 2023;24(1):649. doi 10.3390/ijms24010649

Zhou Y., Wang Y., Xiong X., Appel A.G., Zhang C., Wang X. Profiles of telomeric repeats in Insecta reveal diverse forms of telomeric motifs in Hymenopterans. *Life Sci Alliance.* 2022;5:e202101163. doi 10.26508/lsa.202101163

Conflict of interest. The authors declare no conflict of interest.

Received May 24, 2024. Revised September 11, 2024. Accepted September 11, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):238-247

doi 10.18699/vjgb-25-27

Aquaporins and their role in plant-microbial systems

T.R. Kudriashova, A.A. Kryukov, A.I. Gorenkova, A.P. Yurkov

All-Russia Research Institute for Agricultural Microbiology, Pushkin, St. Petersburg, Russia
t.kudryashova@arriam.ru

Abstract. Global losses of agricultural products from water scarcity could be greater than from all other causes combined. Water deficiency in plants can result from insufficient precipitation, elevated air temperatures, and other factors that reduce the water available in the soil. Most terrestrial plants are able to form symbiosis with arbuscular mycorrhizal fungi. Arbuscular mycorrhiza plays a key role in the mineral nutrition of many terrestrial plant species. Water transport in plants is regulated primarily by aquaporins, transmembrane proteins. Aquaporins help plants save water, which is an important component of the plant's adaptation strategy to water scarcity. Some studies suggest that arbuscular mycorrhizal fungi can decrease the expression of aquaporin genes in plants under drought conditions, which reduces water transport within host plant tissues and conserves available water. On the other hand, there is little scientific evidence of the interaction mechanisms between plants and arbuscular mycorrhizal fungi during aquaporin regulation. In addition, the information in different sources on the aquaporin functions in different plant species may be contradictory. Plant aquaporins are represented by several subfamilies; their number varies for different species. A more comprehensive study of these transporters can enhance our understanding of water transport in plants and assess how arbuscular mycorrhizal fungi can influence it. This review contains data on the history of studies of the structure, localization, phylogeny, and functions of aquaporins. Advancing the study of the symbiotic system functioning may contribute to the development of biofertilizers based on soil microorganisms for agricultural uses in the Russian Federation.

Key words: aquaporins; AQP; arbuscular mycorrhiza; drought; water transport in plants; symbiosis

For citation: Kudriashova T.R., Kryukov A.A., Gorenkova A.I., Yurkov A.P. Aquaporins and their role in plant-microbial systems. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):238-247. doi 10.18699/vjgb-25-27

Funding. The work was supported by RSF grant No. 22-16-00064.

Аквапорины и их роль в растительно-микробных системах

Т.Р. Кудряшова, А.А. Крюков, А.И. Горенкова, А.П. Юрков

Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии, Пушкин, Санкт-Петербург, Россия
t.kudryashova@arriam.ru

Аннотация. Мировые потери сельскохозяйственной продукции из-за дефицита воды, вероятно, более значительны, чем от других причин, вместе взятых. Причины дефицита воды у растений могут быть связаны с недостатком атмосферных осадков, высокой температурой воздуха и другими факторами, которые могут привести к снижению содержания доступной для растений воды в почве. Большинство наземных растений способно вступать в симбиоз с грибами арбускулярной микоризы. Такой симбиоз выполняет ключевую роль в минеральном питании многих видов наземных растений. Транспорт воды в растениях, ее использование регулируются, в первую очередь, с участием трансмембранных белков – аквапоринов. С помощью аквапоринов растение может «экономить» воду, что является важным элементом стратегии адаптации растения к условиям дефицита воды. По некоторым сведениям, грибы арбускулярной микоризы в условиях засухи способны снижать экспрессию генов аквапоринов растения, тем самым уменьшая транспорт воды внутри тканей растения-хозяина, что приводит к ее «экономии». С другой стороны, в настоящее время в научной литературе информации о механизмах взаимодействия растения и грибов арбускулярной микоризы при регуляции работы аквапоринов недостаточно. Кроме того, имеющиеся в различных источниках сведения о работе аквапоринов у разных видов растений могут противоречить друг другу. Аквапорины в растениях представлены несколькими подсемействами, и их число для разных видов варьирует. Изучение этого семейства транспортеров важно для понимания водного транспорта в растениях и оценки влияния на него со стороны грибов арбускулярной микоризы. В обзоре собраны данные об истории изучения, структуре, локализации, филогении, функциях аквапоринов. Развитие знаний о функционировании симбиотических систем будет способствовать созданию биоудобрений на основе микробной биомассы для использования в сельском хозяйстве Российской Федерации.

Ключевые слова: аквапорины; AQP; арбускулярная микориза; засуха; транспорт воды в растениях; симбиоз

Introduction

Stressful conditions during drought affect plant life in many aspects; under conditions of water deficit, the rate of nutrient uptake from the soil decreases, which has implications for biomass growth and crop yields (Ahanger, Agarwal, 2017). Proteins from the Aquaporin family (AQP) are involved in the transport of water in plants. This family is part of a larger major intrinsic proteins (MIP) family (Nielsen et al., 2002; Zhou Y., MacKinnon, 2003). This family received its name after the first water transporter was found in the lens fibers of mammals (including humans), which was later named Aquaporin 0. Aquaporins are represented by integral membrane proteins forming transmembrane pores in cells. During a genomic AQP family study across various plants (24 species including algae, mosses, lycophytes, dicotyledons, and monocotyledons), the aquaporins were divided into eight subfamilies, evolving from large intrinsic proteins (LIPs), found in diatom algae, to tonoplast intrinsic proteins (TIPs) (Hussain et al., 2020). Five out of the eight MIP subfamilies are found in seed plants (including monocotyledons and dicotyledons): plasma membrane intrinsic proteins (PIPs), tonoplast intrinsic proteins (TIPs), nodulin 26-like intrinsic proteins (NIPs), small basic intrinsic proteins (SIPs), and X intrinsic proteins (XIPs) (Danielson, Johanson, 2008). Hybrid intrinsic proteins (HIPs) and GLpF-like intrinsic proteins (GIPs) are found only in mosses (Abascal et al., 2014; Singh et al., 2020).

Arbuscular mycorrhizal fungi (AMF) enhance nutrient uptake, particularly that of phosphorus, in host plants while also regulating their water balance and transport (Schachtman et al., 1998; Huey et al., 2020). At the same time, mycorrhization changes the regulation of AQP genes in plants. In arbuscular mycorrhizal (AM) symbiosis, the root cortical cells form a periarbuscular membrane that surrounds each arbuscule, creating a separation between the fungus and the plant cytoplasm. This process establishes a plant-fungus relationship, helping the host plant obtain water and nutrients, and improving drought tolerance (Kakouridis et al., 2022). In synergy with other microorganisms, an arbuscular mycorrhiza has the potential to supersede classical chemical fertilizers that have adverse effects on ecosystems and reduce land degradation caused by drought in particular (Kuila, Ghosh, 2022; Seka et al., 2022). It is suggested that reducing aquaporin expression during water deficiency may be a way to minimize water loss (Quiroga et al., 2017). The review is aimed to study the impact of AM on plant water exchange and the AQP family's role in increasing plant drought tolerance.

Background. Aquaporin structure

Water transport in the plant is mediated by three pathways: apoplastic, symplastic, and transmembrane transport. The latter involves aquaporins present on biological membranes and forming channels (Singh et al., 2020). Such channels ensure the movement of water in two directions (Chrispeels, Agre, 1994).

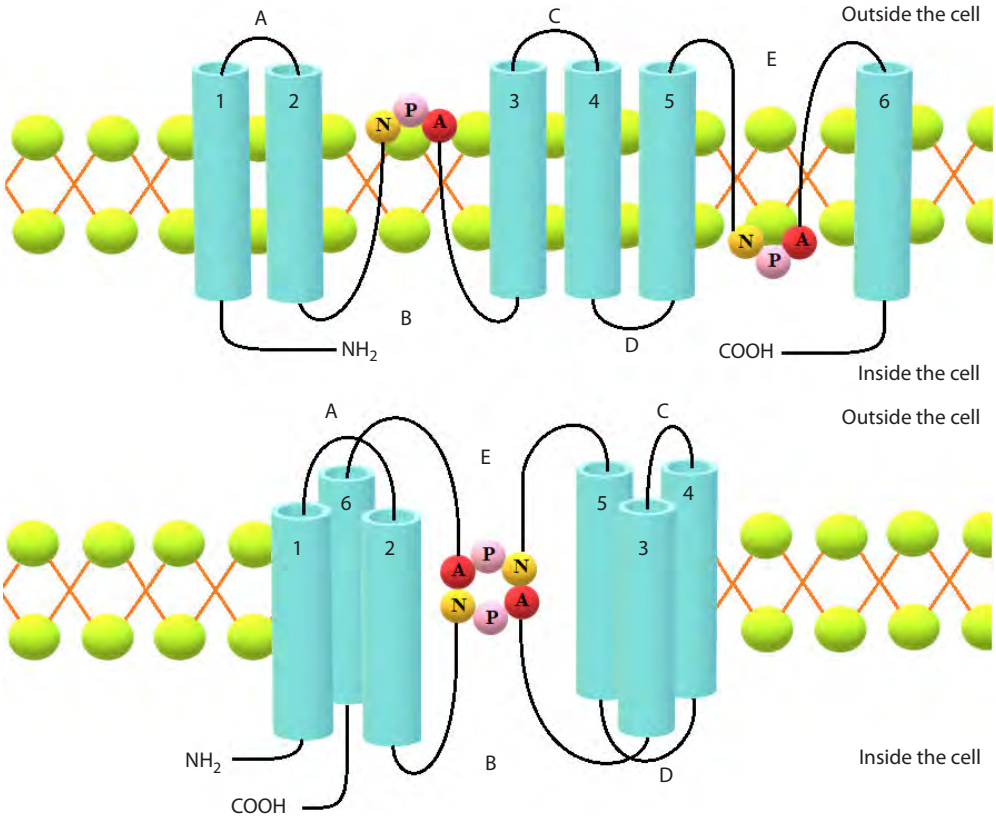
Aquaporins were discovered in the 1980s when researchers studied the mechanisms of water transport across cell membranes. Peter Agre began studying red blood cell proteins at Johns Hopkins University in 1988. He discovered a protein that was bound to antibodies targeting glycophorin. This protein was called CHIP28 (channel-forming integral protein) (Agre et al., 1993).

Although the basic concept of water movement across biological membranes had already been established in the early 1950s, G.M. Preston discovered the first CHIP28 water channel in human erythrocytes only in 1992 (Preston et al., 1992); it was subsequently named Aquaporin-1 (AQP-1). Later, a similar protein was identified in *Escherichia coli* and named aquaglyceroporin (GLP-F).

In 1992, P. Agre and colleagues cloned the *AQP1* gene to determine its structure. AQP1 is a transmembrane protein with four transmembrane domains that form a narrow channel through which water molecules can move (Agre et al., 1993). The first identified plant aquaporin was AtTIP1;1, a transmembrane protein found on the vacuolar membrane of *Arabidopsis thaliana* (Maurel et al., 1993). By 1999, MIPs included 150 representatives that had been identified on cell membranes of various organisms, from bacteria to humans (Lagrée et al., 1999). To this date, more than 7,541 MIP homologs have been discovered in 484 eukaryotic species (Irisarri et al., 2024). MIP family proteins are confirmed to localize on the cell membranes of all living organisms. Plasma and inner membranes as well as viral envelopes are key localization sites for MIPs. The Nobel Prize awarded to Peter Agre and Roderick MacKinnon for the discovery of AQPs in 2003 (discovery of the 3D molecular structure of the bacterial potassium channel and the explanation of the nature of their selectivity) brought AQPs into the spotlight (Knepper, Nielsen, 2004). This discovery demonstrated how water can rapidly and efficiently pass through cell membranes despite their hydrophobic nature.

The function of aquaporins in the water uptake and transport in mycorrhizal plants has been investigated since 1997. Initial studies revealed the expression of aquaporin genes of the TIP subfamily in Medicago (*Medicago truncatula*) and Petroselinum (*Petroselinum crispum*) inoculated with AMF. The first analysis of transport properties was performed for *MtAQP1* (Krajinski et al., 2000). *RiAQP1*, the first AMF AQP, is believed to have been found in *Rhizophagus irregularis* in 2009. *RiAQP1* was distinctively expressed during cold and drought stress in the roots of the host plant (Aroca et al., 2009). To the present day, Russian and foreign scientists continue the vigorous studies of the aquaporin functions in various tissues and organs in genetics, biotechnology, medicine, and agriculture.

All members of the aquaporin family in plants have a similar structure. An aquaporin is a tetramer composed of monomers, each with six transmembrane domains (1–6) and five connecting loops (A–E) localized on the intra- (B, D) or extracytosolic (A, C, E) side of the membrane. Loops A and D have an asparagine-proline-alanine sequence (the



Aquaporin protein structure according to (Kapilan et al., 2018), with revisions.

NPA motif) and form hydrophobic α-helixes. Each NPA sequence is oriented toward the center of the AQP pore. The sequences contribute to the contraction of the central pore and, in combination with the dipole moment of two α-helixes enveloping the membrane, prevent proton (H^+) permeation (see the Figure).

Both aquaporin terminal ends (N and C) are oriented to the cytoplasmic side of the membrane and help carry out the specific regulation of the aquaporin activity. In addition, four conserved sequences form a typical aromatic arginine constriction near the extracytosolic pore mouth that functions as the main selectivity filter (Hussain et al., 2020). Aquaporin activity is regulated by post-translational modifications (phosphorylation and methylation), pH, Ca^{2+}, and interactions between aquaporin monomers, whereas the AQP substrate specificity is determined by its structure (Wang Y. et al., 2020).

AQP phylogeny and subfamilies

MIP family genes, including *AQPs*, are characterized by a larger number of isoforms in plants compared to animals, as plant cells are more compartmentalized. *AQPs* in plants vary by species, subcellular localization, solute water-transmitting capabilities, and function (Chaumont, Tyerman, 2014; Afzal et al., 2016).

Based on their structural features and functional diversity, AQPs can be roughly divided into four subfamilies. The first subfamily functions as a water-permeable channel, the so-called Classical AQP (C-AQP). The second subfamily is Aquaglyceroporin (AQGP); it is responsible for glycerol transport. The third subfamily includes proteins with highly degraded NPA motifs, the functions of which are yet to be identified. This subfamily has been termed superaquaporins or subcellular aquaporins (SAQPs). SAQPs may be involved in the transmembrane transport of ammonia and some other molecules. This subfamily is AQP-8 (Jia, Liu, 2020).

In model plants, the aquaporin family is represented by different numbers of transporters: *Avena sativa* has 45 AQP genes, *A. thaliana* – 38, *Solanum lycopersicum* – 47, *Physcomitrella patens* – 35, *Gossypium hirsutum* – 74, *Zea mays* – 41, *Oryza sativa* – 33, *Populus trichocarpa* – 54, and *Glycine max* – 66. Rapeseed (*Brassica napus*) contains the highest number of AQPs, 121; including, PIP – 43, TIP – 35, NIP – 32, and SIP – 11 (Hussain et al., 2020; Zhou X. et al., 2024).

M. truncatula, a well-known model plant, contains 46 identified putative loci encoding genes from five aquaporin subfamilies, including 10 PIPs, 12 TIPs, 18 NIPs, 4 SIPs, and 2 XIPs. The first four subfamilies were further divided into 2 (PIP1-PIP2), 5 (TIP1-TIP5), 7 (NIP1-NIP7), and 2 (SIP1-SIP2) corresponding subgroups; the XIP subfamily has one subgroup with only two representatives (Min et al., 2019). Based on homology, PIPs are divided into two subgroups, PIP1 and PIP2. The differences between these two

subgroups lie in the water-transmitting capabilities of these proteins; PIP1 has longer N-terminal but shorter C-terminal ends compared to PIP2 (C-terminal ends have an additional 4–10 amino acid region in the first extracytosolic loop). PIP1 and PIP2 have five and eight isoforms, respectively. These two subgroups interact through hetero-oligomerization, in which two PIP2 monomers form heterotetramers with two PIP1 monomers (Wang Y. et al., 2020).

TIPs have more isoforms than PIPs and are divided into five protein subgroups. For example, in *P. trichocarpa*, 17 TIPs are present among 55 MIP sequences (Kapilan et al., 2018). For *Cicer arietinum* L., all TIPs are phylogenetically divided into 14 subgroups. The phylogenetic tree shows that out of 21 branches, only four are interspecific, and the rest are intraspecific. Their functionality for plant species is expanding (Hussain et al., 2020).

NIPs in plants also have numerous isoforms and can be divided into five subgroups. NIP subgroups are found in all higher plants, although NIP3 is found mainly in monocotyledons (Lu et al., 2018). In particular, eleven NIPs have been discovered in *P. trichocarpa* (Gupta, Sankararamakrishnan, 2009). NIPs were found in *G. max*, *C. arietinum*, and *Phaseolus vulgaris* as a result of symbiosis with nitrogen-fixing bacteria. NIP sequences vary significantly both within and between species (Hussain et al., 2020). Most NIPs bear similarity to the nodulin 26 protein, which is expressed in the symbiosome membrane under conditions of inoculation with rhizobacteria (Kapilan et al., 2018).

SIPs are small, just like TIPs. The main reason for their compact size is a very short cytosolic N-terminal region, as compared to other plant AQPs. Based on the NPA motif, SIP1 is divided into SIP1;1 and SIP1;2. Different SIP isoforms have different water-transmitting capabilities for solutes (Kapilan et al., 2018).

The first identification of the relatively recently discovered XIP subfamily was carried out for *G. hirsutum* in 2010 (Park et al., 2010). Nineteen representatives of XIP are known to date, including five XIPs in *P. trichocarpa*. The remaining ten XIP representatives have been found in other dicotyledonous plants, three – in mosses, and one – in protozoa. XIP homologs have not been discovered in monocotyledonous plants. Expression analysis in *P. trichocarpa* demonstrates that XIPs in poplar do not manifest an abundance of tissue-specific transcripts (Gupta, Sankararamakrishnan, 2009; Kapilan et al., 2018).

The high number of AQP isoforms within subfamilies may contribute to enhanced transporter functions and plant adaptation to changing conditions. *M. truncatula* AQP molecular structure analysis revealed that among aquaporins, seven genes (15.2 %) exhibit tandem duplication, and ten genes (21.7 %) exhibit segmented duplication (Min et al., 2019). In other words, the presence of many isoforms and distinct subgroups of different aquaporin subfamilies underscores their substantial importance in living organisms.

Aquaporin localization and functions in plants

Maurel C. et al. (2015) showed that in addition to water, some members of the MIP superfamily, including aquaporins, can also transport glycerol, carbon dioxide, urea, ammonia, hydrogen peroxide, boron, silicon, arsenic, antimonite, lactic acid, and oxygen across membranes. Some AQPs are capable of transmitting univalent cations (Byrt et al, 2017). AQPs may be involved in the signaling of such hormones as auxin, gibberellins, ethylene, and abscisic acid (ABA) in plants (Wang C. et al., 2016). The aquaporin subfamilies localization and functions are listed in the Table.

Embryophytes are known to have evolved with a remarkably high degree of subcellular compartmentalization. In

Cellular localization and plant aquaporin functions*

Subfamily	Localization	Substances transported
Plasma membrane intrinsic proteins (PIPs)	Plasma membrane, chloroplast inner membrane, thylakoid membrane, endoplasmic reticulum (ER) membrane	Water, hydrogen peroxide, glycerin, carbon dioxide
Tonoplast intrinsic proteins (TIPs)	Tonoplast, plasma membrane, chloroplast inner membrane, thylakoid membrane, and mitochondria	Water, hydrogen peroxide, ammonium, urea, glycerin
Nodulin26-like intrinsic proteins (NIPs)	Membrane of endoplasmic reticulum, plasma membrane	Permeable to a large number of substrates, including metalloids, but only marginally permeable to water. NIPs are assumed to be involved in the exchange of metabolites between a host plant and a symbiont
Small basic intrinsic proteins (SIPs)	Membrane of endoplasmic reticulum	Water
Uncharacterized/ X intrinsic proteins (XIPs)	Plasma membrane	Water, hydrogen peroxide, metalloids, urea, boric acid

* According to (Ishikawa et al., 2005; Kruse et al., 2006; Ma et al., 2006; Maurel et al., 2015; Pommerrenig et al., 2015; Lopez et al., 2016; Noronha et al., 2016; Wang C. et al., 2016; Byrt et al., 2017; Kapilan et al., 2018; Zhou X. et al., 2024).

this respect, in contrast to animals, plant aquaporins show a wider range of subcellular localizations (plasma membrane, tonoplast, chloroplast, endoplasmic reticulum, Golgi apparatus, and mitochondria), with AQPs capable of simultaneous localization in different cells and on different membranes. Representatives of the same subfamily may have different subcellular localizations (Zhou X. et al., 2024).

PIPs are mainly localized on the plasma membrane, typically in tissues characterized by high water transport; for example, in conducting tissues (Yaneff et al., 2014). PIP transporters are also found in *A. sativa*, on the endoplasmic reticulum (Zhou X. et al., 2024). PIP1 subgroup transporters typically localize at the plasma membrane (Kaldenhoff, Fischer, 2006) and have low water-transmitting capabilities (Kapilan et al., 2018). Some PIP1 proteins are unable to act independently, and they must form heterotetramers with PIP2 monomers to be able to increase water-transmitting capabilities (Schuurmans et al., 2003). In tobacco (*Nicotiana tabacum*) plants, decreased expression of *NtAQP1*, a PIP1 family member, caused decreased water transport in roots and reduced plant resistance to water stress. In peas (*Lathyrus oleraceus*), PIP1 has been found to play an important role in seed water supply (Kaldenhoff, Fischer, 2006).

Some researchers have noted an additional function of the PIP1 subgroups in plant cells – the ability to transport CO_2 (Kapilan et al., 2018). Increased *OsPIP2;7* expression in rice improves its survival under low-temperature stress and affects the expression levels of other AQP genes (Zhou X. et al., 2024).

The PIP1 and PIP2 subgroups are believed to be localized in almost all plant parts, including roots and leaves (Maurel et al., 2015). The PIP2 subfamily aquaporins are more efficient as water channels than PIP1 group members; various PIP2 isoforms are considered to be major transporters of water across the cell membrane (Kaldenhoff, Fischer, 2006). The expression patterns of the *AQP* gene in *A. thaliana* differed significantly under drought stress (e. g., decreased *AtPIP2;1* and *AtPIP2;2* expressions under stress). The expression analysis for *OsPIP1* in rice under drought stress showed that the expression of these two genes was elevated, whereas the expression of all *OsPIP2* genes was reduced. These results indicate that the expression of *AQP* genes in plants under drought is regulated by a complex signaling network, and the mechanism of *AQP* gene regulation in plant drought tolerance requires further investigation (Zhou X. et al., 2024).

TIPs are mainly localized on vacuolar membranes (Johnson et al., 1990). The five TIP aquaporin subgroups (in *Arabidopsis*, maize, and rice) are mainly located in tonoplasts, but some TIP isoforms have also been found at the plasma cell membrane. In *Avena*, seven TIP representatives are of cytoplasmic origin, and five TIPs are localized in the tonoplast (Zhou X. et al., 2024). Because of the high aquaporin content in the tonoplast, the water transmission of the tonoplast is believed to be much higher than that of the plasma membrane. This contributes to the turgor pressure within the cell (Luo et al., 2022). In addition to functions of aquaporins in water transfer, TIPs are known to participate in the transport of urea, glycerol, and ammonia and are also involved in the plant response to abiotic stress (Loque et al., 2005). Several TIP isoforms show important roles in plant response to drought. In 2020, A. Lopez-Zaplana et al. observed the presence of a TIP on mitochondrial membranes (Lopez-Zaplana et al., 2020). *TIP1* and *TIP2* isoforms are believed to be expressed in plant vegetative tissues, *TIP3* isoforms are mainly expressed in seeds, and *TIP5* isoforms are associated with pollen grains (Hussain et al., 2020). TIP is a highly specific subfamily and can fulfill a variety of functions in different plant species.

Nodulin 26-like intrinsic proteins were initially identified in the peribacteroid membrane in *G. max* nodule. GmNOD26 was the first to be described (Fortin et al., 1987). NIPs are thought to be involved in metabolite exchange between a host plant and a microsymbiont (Kruse et al., 2006). Although the *GmNOD26* gene product localized exclusively in the peribacteroid membrane of nitrogen-fixing symbiotic legume nodule, NOD26-like proteins (NIPs), forming the third subfamily, can also occur in plant species other than legumes in which they localize in the plasma membrane (Ma et al., 2006) or the ER membrane (Mizutani et al., 2006).

Therefore, NIPs are widespread not only in Leguminosae (forming a rhizobia-legume symbiosis), indicating that they can function in the absence of symbiotic relationships (Wayne, Tazawa, 2010). For example, X. Zhou et al. showed that in *A. sativa* the expression of the NIP subfamily *6Ag0000836.1* gene was significantly upregulated under various abiotic stresses. This gene is suggested to be a marker of response to abiotic stress (Zhou X. et al., 2024). Although NOD26 and other NIPs have lower water-transmitting capabilities compared to other aquaporins, they also have a transport function to carry glycerol (Kaldenhoff, Fischer, 2006). This may indicate that the common ancestor of different AQP groups in plants did not have a glycerol transport function (Zhang et al., 2020).

The subcellular localization of most NIP transporters is not entirely clear. NIPs can localize on the membrane of the endoplasmic reticulum (Lopez et al., 2016). In *A. thaliana*, NIP5;1 is localized in the plasma membrane, while NIP2;1 is localized on the ER membrane (Zhou X. et al., 2024). In addition, NIPs can transport ammonia, lactic acid, boron, and silicon (Danielson, Johanson, 2008). NIPs are also metalloid transporters (Pommerrenig et al., 2015); NIPs not only facilitate metalloid diffusion from the soil but also play a key role in their transport in the plant. Therefore, we can hypothesize that the NIP family promotes the uptake and translocation of metalloids, thus regulating their amount.

The group of small SIPs is localized on the ER membrane (Ishikawa et al., 2005). The SIP subfamily in plants is inadequately studied in terms of their structure and functionality (Hussain et al., 2020). SIP structure differs from other AQP subfamilies since their cytosolic N-terminal region is relatively shorter (Kapilan et al., 2018). Due to their short-

ened NPA motif, SIPs can also transport different molecules, not just water. Currently, there is no consensus on the role of SIPs in transport in plants. In a study of *A. thaliana*, only the aquaporins SIP1.1 and SIP1.2 (out of three subfamily aquaporins) demonstrated minor water-transmitting capabilities, while SIP2.1 did not exhibit this function (Ishikawa et al., 2005). SIP1 proteins transport water across the ER membrane, while the SIP2 protein acts as an endoplasmic reticulum channel for small molecules or ions (Hussain et al., 2020). To date, *A. sativa* has been found to have two SIP genes (*4Dg0000047.3* and *5Ag0000631.1*) that show expression in tissues of the aboveground plant organs (Zhou X. et al., 2024).

Members of the XIP subfamily are found in protozoa, fungi, and plants, but their functions are to be studied further (Kapilan et al., 2018). In tobacco plants, the XIP family gene products, *NtXIP1;1α* and *NtXIP1;1β*, are localized at the plasma membrane. *NtXIP1;1*, however, showed expression in all plant tissues. This subfamily is found to be absent in plants such as *Arabidopsis*, maize, and rice. The sequences of the XIP subfamily are shorter compared to those of other MIP subfamilies. However, their structure remains highly conserved and similar to other subfamilies (Wang C. et al., 2016). XIPs exhibit contrasting transport functions in different plant species (Lopez-Garcia et al., 2018). For example, in grapevine, VvXIP1 is active in osmotic regulation in addition to H_2O_2 transport and metalloid concentration regulation (Noronha et al., 2016). Heterologous expression of solanaceous XIP family genes in *Xenopus laevis* oocytes and various yeast strains of *Saccharomyces cerevisiae* showed that these isoforms contribute to the transport of large molecules such as glycerol, urea, and boric acid. Water-transmitting capabilities, however, were not identified. This indicates that XIPs are involved in the transport of non-charged molecules across the plasma membrane of cells in certain plant tissues (Kapilan et al., 2018).

Aquaporins in plant-microbial systems

Microbial-plant relationships form the foundation of life on Earth. These interactions can be specific and evolutionarily reinforced or non-specific, temporary, and random. It is known that embryophytes can enter into symbiotic relationships with microorganisms. These relationships can be of different types: mutualism, commensalism, amensalism, parasitism, or neutralism (Yatsenko-Stepanova et al., 2014). Symbiotic relationships depend on the conditions in which they exist. In other words, the same host-symbiont combination can be mutually beneficial in one case but parasitic in another (Chiu, Paszkowski, 2019). For *A. thaliana*, the *Colletotrichum toefieldiae* fungus was found to be beneficial only under phosphorus-deficient conditions, while in other situations it acted as a parasite (Hiruma et al., 2016).

Plants face constant threats from pathogens such as viruses, bacteria, and fungi. These attacks result in various diseases affecting crucial crops, leading to significant food losses (Savary et al., 2019). Growing evidence suggests that AQPs play a role in plant defense against pathogens by modulating plant immunity and resistance to invasive diseases (Li et al., 2020). Bacterial pathogen-induced AtPIP1;4 transports water from the apoplast to the cytoplasm to activate systemic resistance and immune responses in *A. thaliana* (Tian et al., 2016). Plants can close their stomas to conserve moisture after perceiving molecular patterns associated with the pathogen to limit the invasion. The stress hormone ABA is found to be involved in the regulation of stoma closures. AtPIP1;2 has been shown to facilitate water transport across the plasma membrane, causing ABA- and pathogen-induced closure of stomas in *A. thaliana* (Exposito-Rodrigues et al., 2017).

During the initial stages of infection, the fungal pathogens regulate their development to send the special infectious hyphae into the host organism to obtain nutrients. In *Fusarium graminearum*, the FgAQP1 protein localized in the nuclear envelope in conidia is important for hyphal growth, development, and secondary metabolism. Deletion in *FgAQP1* affects gene expression, which reduces plant infection efficiency, suggesting *FgAQP1* may play a key role in the interaction between *F. graminearum* and the host (Ding et al., 2018).

At the same time, bacteria are known to stimulate plant growth under both favorable and stressful conditions (*Pseudomonas mandelii*, *Rhizobium leguminosarum* bv. *viciae*, etc). E. Martynenko et al. showed the relationship between AQP and the formation of apoplastic barriers in the plant-microbial system "*Pisum sativum* + *P. mandelii*". *P. mandelii* increases aquaporin activity, which compensates for a possible decrease in water-transmitting capabilities in pea roots (Martynenko et al., 2023).

Compared to pathogens, AMF in root cortex do not disrupt host plant cell integrity (Mosse et al., 1981; Spatafora et al., 2016). Nutrient exchange and water transport between symbionts occur in arbuscules (Zhang et al., 2019). In addition, AMF colonization of plants promotes the smooth closure of plant stomas during drought. AM symbiosis improves the transporting capabilities of stomas and leaf transpiration to adapt to arid environmental conditions (Ni et al., 2024). Acting as "extended plant roots", AMF improve photosynthetic efficiency and osmoregulation, and enhance plant antioxidant metabolism (Evelin et al., 2019). *AQPs* can be divergently expressed in response to mycorrhization (Asadollahi et al., 2023). When a plant is under stress, AMF could affect *AQP* expression at transcriptional, translational, and posttranscriptional (*AQP* phosphorylation, multimerization, cycling, and internalization) levels, which contributes to the active regulation of *AQP* expression and protein abundance, thereby improving the efficiency of transport of H_2O, CO_2, glycerol, NH_3, etc. (Kakouridis et al., 2022).

In the plant-microbial system (PMS) "*Triticum aestivum* + *R. irregularis* (formerly known as *Glomus intraradices* Schenck & Smith)", it has been shown that AMF colonization of plants activates genes involved in the phenylpropanoid biosynthesis pathway and transcription factors that

play a vital role in plant defense against biotic and abiotic stresses (Mashini et al., 2022).

It is suggested that AM colonization is associated with modifications of membrane transporters, especially aquaporin proteins. For example, differential expression of genes due to water deficit was analyzed in the "*T. aestivum* + *Funneliformis mosseae*" PMS: *TaPIP1-6* and *TaPIP1-8* from the PIP1 subfamily; PIP2 – *TaPIP2-2C1*, *TaPIP2-2C3*, *TaPIP2-3C1*, *TaPIP2-7*, *TaPIP2-22*; NIP3 – only *TaNIP3-1*; TIP4 – *TaTIP2-5*, *TaTIP4-1*, *TaTIP4-2*, *TaTIP4-6*. The *AQP* genes with downregulated expression belonged to the *PIP1*, *PIP2*, *TIP2*, *TIP4*, and *NIP* subfamilies. Gene products were localized at the plasma membrane or tonoplast. In contrast to other AQPs, in addition to its function as a water transporter, *TaNIP3-1* also showed activity in the transmembrane transport of arsenite and boronic acid salts. In the "*T. aestivum* + *F. mosseae*" PMS, water deficit did not affect SIP expression. In wheat, 25 out of the 96 known aquaporin genes changed their expression when inoculated with an AM fungus. At the same time, only four genes showed increased expression: *TaNIP1-10*, *TaNIP3-3*, *TaNIP3-4*, and *TaTIP1-5*. Half of the analyzed *AQPs* with reduced gene expression in wheat were localized at the plasma membrane, while the rest were localized at the tonoplast (TIP1, TIP2, TIP4, PIP1, PIP2, NIP2, SIP2, and NIP3). It is of interest that the expression of *TaPIP2-2C3*, *TaPIP2-2C1*, *TaTIP4-6*, *TaPIP1-6*, and *TaPIP2-3C1* was suppressed in mycorrhized plants under water deficit, while the expression of *TaNIP1-10*, *TaNIP3-3*, *TaNIP3-4*, *TaNIP1-5*, and *TaPIP2-7* was elevated under the same conditions (Asadollahi et al., 2023).

In maize, AM symbiosis is known to suppress several aquaporins, including *ZmPIP1-1*, *ZmPIP1-3*, *ZmPIP1-4*, *ZmPIP1-6*, *ZmPIP2-2*, *ZmPIP2-4*, *ZmTIP1-1*, and *ZmTIP2-3*, but enhances the expression of *TIP4*-1. At the same time, the drought-tolerant maize variety showed different results. Only three of the *AQP* genes under study (*ZmPIP1;6*, *ZmPIP2;2*, and *ZmTIP4;1*) changed expression upon symbiosis with AMF. The results of the experiment with a drought-tolerant maize strain are consistent with the hypothesis that reducing aquaporin gene expression under water deficit may be a way to minimize water loss (Quiroga et al., 2017).

Expression of *GiAQP1*, *RiAQPF1*, and *RiAQPF2* was also evaluated in "*Daucus carota* + *R. irregularis*" PMS under drought conditions. Only *RiAQPF2* was expressed differentially (Keller-Pearson et al., 2023).

Joint colonization of maize with *R. irregularis* and *Exophiala pisciphila* (dark septate endophyte, DSE) resulted in high water transmission through stomas and downregulation of the *ZmPIP1;1*, *ZmPIP1;2*, *ZmPIP2;1*, *ZmPIP2;5*, and *ZmPIP2;6* gene expression compared with reference plants without microbial infection and with individual fungus colonization. The *GintAQPF1* and *GintAQPF2* expression in *R. irregularis* has been shown to change significantly under drought stress conditions. The competitive relationship between AMF and DSE in mycorrhization during the experiment is also worth noting. On the other hand, AMF and DSE are also known to form synergistic relationships to regulate membrane electrolyte transport, oxidative stress, photosynthesis, and aquaporin expression; such a relationship has been studied in maize seedlings (Gong et al., 2023).

D. Wang et al. showed that the *ZmTIP2;3* gene expression in the "*Z. mays* + *R. irregularis* (previously attributed to *Glomus intraradices*)" PMS was significantly upregulated under drought conditions through AMF symbiosis. *ZmTIP2;3* is an aquaporin with six transmembrane domains and two highly conserved NPA motifs. Its promoter region contains many cis-acting elements associated with the induction of AM symbiosis. In the experiment, *ZMTIP2;3* gene mutation resulted in decreased biomass, colonization rate, photosynthesis, proline, and expression levels of several drought-related genes (*LEA3*, *P5CS4*, and *NECD1*) compared with the wild post-AMF-inoculation type under drought conditions. This suggestes that *ZmTIP2;3* enhances drought tolerance in maize through symbiosis with the AM fungus (Wang D. et al., 2024).

Conclusion

The discovery of aquaporins was a major event in biology and medicine. The main function of aquaporins is the regulation of transmembrane water transport both between and within cells (Maloy, Hughes, 2013). Aquaporin isoforms can vary greatly depending on the organism type and the living conditions. Depending on the environmental conditions, the aquaporin activity may change, as well as their function in regulating water transport. Such strategic patterns of changes in aquaporin gene expression and functional diversity are the basis for adaptation to environmental changes, including stress-induced ones (Jia, Liu, 2020). A significant number of host plant aquaporins are regulated in AMF symbiosis. The regulation of their genes may depend on the magnitude of drought stress. Some of these aquaporins can transport other molecules of crucial importance, in addition to water. The results of studies carried out under a wide variety of conditions confirm that mycorrhized plants grow and develop better than plants without mycorrhization. At the same time, mycorrhized plants are more efficient in the preservation and transfer of water between tissues, nitrogen compound mobilization efficiency is increased, as well as glycerol accumulation, synthesis of signaling molecules, and accumulation of metals that play a role in resistance to abiotic stresses.

The continued exploration of this topic will enhance our understanding of the specific roles played by aquaporin isoforms in response to arbuscular mycorrhizal symbiosis. This research will help us determine how this symbiosis influences plant adaptation mechanisms under stress conditions. By monitoring the transcriptional responses of aquaporin genes to various environmental factors, we can deepen our knowledge and contribute to the development of biotechnological programs aimed at improving crop resilience.

References

Abascal F., Irisarri I., Zardoya R. Diversity and evolution of membrane intrinsic proteins. *Biochim Biophys Acta*. 2014;1840(5):1468-1481. doi 10.1016/j.bbagen.2013.12.001

Agre P., Preston G.M., Smith B.L., Jung J.S., Raina S., Moon C., Guggino W.B., Nielsen S. Aquaporin CHIP: the archetypal molecular water channel. *Am J Physiol*. 1993;265(4):F463-F476. doi 10.1152/ajprenal.1993.265.4.F463

Afzal Z., Howton T., Sun Y., Mukhtar M. The roles of aquaporins in plant stress responses. *J Dev Biol*. 2016;4(1):9. doi 10.3390/jdb4010009

Ahanger M.A., Agarwal R.M. Salinity stress induced alterations in antioxidant metabolism and nitrogen assimilation in wheat (*Triticum aestivum* L.) as influenced by potassium supplementation. *Plant Physiol Biochem*. 2017;115:449-460. doi 10.1016/j.plaphy.2017.04.017

Aroca R., Bago A., Sutka M., Paz J.A., Cano C., Amodeo G., Ruiz-Lozano J.M. Expression analysis of the first arbuscular mycorrhizal fungi aquaporin described reveals concerted gene expression between salt-stressed and nonstressed mycelium. *Mol Plant Microbe Interact*. 2009;22(9):1169-1178. doi 10.1094/MPMI-22-9-1169

Asadollahi M., Iranbakhsh A., Ahmadvand R., Ebadi M., Mehregan I. Synergetic effect of water deficit and arbuscular mycorrhizal symbiosis on the expression of aquaporins in wheat (*Triticum aestivum* L.) roots: insights from NGS RNA-sequencing. *Physiol Mol Biol Plants*. 2023;29(2):195-208. doi 10.1007/s12298-023-01285-w

Byrt C.S., Zhao M., Kourghi M., Bose J., Henderson S.W., Qiu J., Gilliham M., Schultz C., Schwarz M., Ramesh S.A., Yool A., Tyerman S. Non-selective cation channel activity of aquaporin AtPIP2;1 regulated by Ca^{2+} and pH. *Plant Cell Environ*. 2017;40(6):802-815. doi 10.1111/pce.12832

Chaumont F., Tyerman S.D. Aquaporins: highly regulated channels controlling plant water relations. *Plant Physiol*. 2014;164(4):1600-1618. doi 10.1104/pp.113.233791

Chiu C.H., Paszkowski U. Mechanisms and impact of symbiotic phosphate acquisition. *Cold Spring Harb Perspect Biol*. 2019;11(6):a034603. doi 10.1101/cshperspect.a034603

Chrispeels M.J., Agre P. Aquaporins: water channel proteins of plant and animal cells. *Trends Biochem Sci*. 1994;19(10):421-425. doi 10.1016/0968-0004(94)90091-4

Danielson J.Å., Johanson U. Unexpected complexity of the aquaporin gene family in the moss *Physcomitrella patens*. *BMC Plant Biol*. 2008;8(1):45. doi 10.1186/1471-2229-8-45

Ding M., Li J., Fan X., He F., Yu X., Chen L., Zou S., Liang Y., Yu J. Aquaporin1 regulates development, secondary metabolism and stress responses in *Fusarium graminearum*. *Curr Genet*. 2018;64(5):1057-1069. doi 10.1007/s00294-018-0818-8

Evelin H., Devi T.S., Gupta S., Kapoor R. Mitigation of salinity stress in plants by arbuscular mycorrhizal symbiosis: current understanding and new challenges. *Front Plant Sci.* 2019;10:470. doi 10.3389/fpls.2019.00470

Exposito-Rodriguez M., Laissue P.P., Yvon-Durocher G., Smirnoff N., Mullineaux P.M. Photosynthesis-dependent H_2O_2 transfer from chloroplasts to nuclei provides a high-light signalling mechanism. *Nat Commun.* 2017;29;8(1):49. doi 10.1038/s41467-017-00074-w

Fortin M.G., Morrison N.A., Verma D.P.S. Nodulin-26, a peribacteroid membrane nodulin is expressed independently of the development of the peribacteroid compartment. *Nucleic Acids Res*. 1987;15(2):813-824. doi 10.1093/nar/15.2.813

Gong M., Bai N., Wang P., Su J., Chang Q., Zhang Q. Co-inoculation with arbuscular mycorrhizal fungi and dark septate endophytes under drought stress: synergistic or competitive effects on maize growth, photosynthesis, root hydraulic properties and aquaporins? *Plants*. 2023;12(14):2596. doi 10.3390/plants12142596

Gupta A., Sankararamakrishnan R. Genome-wide analysis of major intrinsic proteins in the tree plant *Populus trichocarpa*: characterization of XIP subfamily of aquaporins from evolutionary perspective. *BMC Plant Biol*. 2009;9(1):134. doi 10.1186/1471-2229-9-134

Hiruma K., Gerlach N., Sacristán S., Nakano R.T., Hacquard S., Kracher B., Neumann U., Ramírez D., Bucher M., O'Connell R.J., Schulze-Lefert P. Root endophyte *Colletotrichum tofieldiae* confers plant fitness benefits that are phosphate status dependent. *Cell*. 2016;165(2):464-474. doi 10.1016/j.cell.2016.02.028

Huey C.J., Gopinath S.C.B., Uda M.N.A., Zulhaimi H.I., Jaafar M.N., Kasim F.H., Yaakub A.R.W. Mycorrhiza: a natural resource assists plant growth under varied soil conditions. *3 Biotech*. 2020;10(5):204. doi 10.1007/s13205-020-02188-3

Hussain A., Tanveer R., Mustafa G., Farooq M., Amin I., Mansoor S. Comparative phylogenetic analysis of aquaporins provides insight into the gene family expansion and evolution in plants and their role in drought tolerant and susceptible chickpea cultivars. *Genomics*. 2020;112(1):263-275. doi 10.1016/j.ygeno.2019.02.005

Irisarri I., Lorente-Martínez H., Strassert J.F.H., Agorreta A., Zardoya R., San Mauro D., De Vries J. Early diversification of membrane intrinsic proteins (MIPs) in eukaryotes. *Genome Biol Evol*. 2024;16(8):evae164. doi 10.1093/gbe/evae164

Ishikawa F., Suga S., Uemura T., Sato M.H., Maeshima M. Novel type aquaporin SIPs are mainly localized to the ER membrane and show cell-specific expression in *Arabidopsis thaliana*. *FEBS Lett*. 2005;579(25):5814-5820. doi 10.1016/j.febslet.2005.09.076

Jia Y., Liu X. Polyploidization and pseudogenization in allotetraploid frog *Xenopus laevis* promote the evolution of aquaporin family in higher vertebrates. *BMC Genomics*. 2020;21(1):525. doi 10.1186/s12864-020-06942-y

Johnson K.D., Höfte H., Chrispeels M.J. An intrinsic tonoplast protein of protein storage vacuoles in seeds is structurally related to a bacterial solute transporter (GIpF). *Plant Cell*. 1990;2(6):525-532. doi 10.1105/tpc.2.6.525

Kakouridis A., Hagen J.A., Kan M.P., Mambelli S., Feldman L.J., Herman D.J., Weber P.K., Pett-Ridge J., Firestone M.K. Routes to roots: direct evidence of water transport by arbuscular mycorrhizal fungi to host plants. *New Phytol*. 2022;236(1):210-221. doi 10.1111/nph.18281

Kaldenhoff R., Fischer M. Functional aquaporin diversity in plants. *Biochim Biophys Acta*. 2006;1758(8):1134-1141. doi 10.1016/j.bbamem.2006.03.012

Kapilan R., Vaziri M., Zwiazek J.J. Regulation of aquaporins in plants under stress. *Biol Res*. 2018;51(1):4. doi 10.1186/s40659-018-0152-0

Keller-Pearson M., Bortolazzo A., Willems L., Smith B., Peterson A., Ané J.-M., Silva E.M. A dual transcriptomic approach reveals contrasting patterns of differential gene expression during drought in arbuscular mycorrhizal fungus and carrot. *Mol Plant Microbe Interact*. 2023;36(12):821-832. doi 10.1094/MPMI-04-23-0038-R

Knepper M.A., Nielsen S. Peter Agre, 2003 Nobel Prize winner in chemistry. *J Am Soc Nephrol*. 2004;15(4):1093-1095. doi 10.1097/01.ASN.0000118814.47663.7D

Krajinski F., Biela A., Schubert D., Gianinazzi-Pearson V., Kaldenhoff R., Franken P. Arbuscular mycorrhiza development regulates the mRNA abundance of *Mtaqp1* encoding a mercury-insensitive aquaporin of *Medicago truncatula*. *Planta*. 2000;211(1):85-90. doi 10.1007/s004250000263

Kruse E., Uehlein N., Kaldenhoff R. The aquaporins. *Genome Biol*. 2006;7(2):206. doi 10.1186/gb-2006-7-2-206

Kuila D., Ghosh S. Aspects, problems and utilization of Arbuscular Mycorrhizal (AM) application as bio-fertilizer in sustainable agriculture. *Curr Res Microb Sci*. 2022;3:100107. doi 10.1016/j.crmicr.2022.100107

Lagrée V., Froger A., Deschamps S., Hubert J.-F., Delamarche C., Bonnec G., Thomas D., Gouranton J., Pellerin I. Switch from an aquaporin to a glycerol channel by two amino acids substitution. *J Biol Chem*. 1999;274(11):6817-6819. doi 10.1074/jbc.274.11.6817

Li G., Chen T., Zhang Z., Li B., Tian S. Roles of aquaporins in plant-pathogen interaction. *Plants*. 2020;9(9):1134. doi 10.3390/plants9091134

Lopez D., Amira M.B., Brown D., Muries B., Brunel-Michac N., Bourgerie S., Porcheron B., Lemoine R., Chrestin H., Mollison E., Di Cola A., Frigerio L., Julien J.-L., Gousset-Dupont A., Fumanal B., Label P., Pujade-Renaud V., Auguin D., Venisse J.-S. The *Hevea brasiliensis* XIP aquaporin subfamily: genomic, structural and functional characterizations with relevance to intensive latex harvesting. *Plant Mol Biol*. 2016;91(4-5):375-396. doi 10.1007/s11103-016-0462-y

Lopez-Zaplana A., Nicolas-Espinosa J., Carvajal M., Bárzana G. Genome-wide analysis of the aquaporin genes in melon (*Cucumis melo* L.). *Sci Rep*. 2020;10(1):22240. doi 10.1038/s41598-020-79250-w

Loque D., Ludewig U., Yuan L., von Wirén N. Tonoplast intrinsic proteins AtTIP2;1 and AtTIP2;3 facilitate NH3 transport into the vacuole. *Plant Physiol*. 2005;137(2):671-680. doi 10.1104/pp.104.051268

Lu L., Dong C., Liu R., Zhou B., Wang C., Shou H. Roles of soybean plasma membrane intrinsic protein GmPIP2;9 in drought tolerance and seed development. *Front Plant Sci*. 2018;9:530. doi 10.3389/fpls.2018.00530

Luo Y., Ma L., Du W., Yan S., Wang Z., Pang Y. Identification and characterization of salt- and drought-responsive *AQP* family genes in *Medicago sativa* L. *Int J Mol Sci*. 2022;23(6):3342. doi 10.3390/ijms23063342

Ma J.F., Tamai K., Yamaji N., Mitani N., Konishi S., Katsuhara M., Ishiguro M., Murata Y., Yano M. A silicon transporter in rice. *Nature*. 2006;440(7084):688-691. doi 10.1038/nature04590

Maloy S., Hughes K. (Eds) Brenner's Encyclopedia of Genetics. London: Elsevier, 2013

Martynenko E., Arkhipova T., Akhiyarova G., Sharipova G., Galin I., Seldimirova O., Ivanov R., Nuzhnaya T., Finkina E., Ovchinnikova T., Kudoyarova G. Effects of a *Pseudomonas* strain on the lipid transfer proteins, appoplast barriers and activity of aquaporins associated with hydraulic conductance of pea plants. *Membranes*. 2023;13(2):208. doi 10.3390/membranes13020208

Mashini A.G., Oakley C.A., Grossman A.R., Weis V.M., Davy S.K. Immunolocalization of metabolite transporter proteins in a model cnidarian-dinoflagellate symbiosis. *Appl Environ Microbiol*. 2022;88(12):e00412-22. doi 10.1128/aem.00412-22

Maurel C., Reizer J., Schroeder J.I., Chrispeels M.J. The vacuolar membrane protein gamma-TIP creates water specific channels in Xenopus oocytes. *EMBO J*. 1993;12(6):2241-2247. doi 10.1002/j.1460-2075.1993.tb05877.x

Maurel C., Boursiac Y., Luu D.-T., Santoni V., Shahzad Z., Verdoucq L. Aquaporins in plants. *Physiol Rev*. 2015;95(4):1321-1358. doi 10.1152/physrev.00008.2015

Min X., Wu H., Zhang Z., Wei X., Jin X., Ndayambaza B., Wang Y., Liu W. Genome-wide identification and characterization of the aquaporin gene family in *Medicago truncatula*. *J Plant Biochem Biotechnol*. 2019;28(3):320-335. doi 10.1007/s13562-018-0484-4

Mizutani M., Watanabe S., Nakagawa T., Maeshima M. Aquaporin NIP2;1 is mainly localized to the ER membrane and shows root-specific accumulation in *Arabidopsis thaliana*. *Plant Cell Physiol*. 2006;47(10):1420-1426. doi 10.1093/pcp/pcl004

Mosse B., Stribley D.P., LeTacon F. Ecology of mycorrhizae and mycorrhizal fungi. In: Alexander M. (Ed.) Advances in Microbial Ecology, vol. 5. Boston, MA: Springer US, 1981;5137-5210. doi 10.1007/978-1-4615-8306-6_4

Ni Y., Bao H., Zou R., Wang Y., Xie K., Cheng B., Li X. Aquaporin ZmPIP2;4 promotes tolerance to drought during arbuscular mycorrhizal fungi symbiosis. *Plant Soil*. 2024. doi 10.1007/s11104-024-06778-5

Nielsen S., Frøkiær J., Marples D., Kwon T.-H., Agre P., Knepper M.A. Aquaporins in the kidney: from molecules to medicine. *Physiol Rev*. 2002;82(1):205-244. doi 10.1152/physrev.00024.2001

Noronha H., Araújo D., Conde C., Martins A.P., Soveral G., Chaumont F., Delrot S., Gerós H. The grapevine uncharacterized intrinsic protein 1 (VvXIP1) is regulated by drought stress and transports glycerol, hydrogen peroxide, heavy metals but not water. *PLoS One*. 2016;11(8):e0160976. doi 10.1371/journal.pone.0160976

Park W., Scheffler B.E., Bauer P.J., Campbell B.T. Identification of the family of aquaporin genes and their expression in upland cotton (*Gossypium hirsutum* L.). *BMC Plant Biol*.10;142(2010). doi 10.1186/1471-2229-10-142

Pommerrenig B., Diehn T.A., Bienert G.P. Metalloido-porins: essentiality of Nodulin 26-like intrinsic proteins in metalloid transport. *Plant Sci*. 2015;238:212-227. doi 10.1016/j.plantsci.2015.06.002

Preston G.M., Carroll T.P., Guggino W.B., Agre P. Appearance of water channels in *Xenopus* oocytes expressing red cell CHIP28 protein. *Science*. 1992;256(5055):385-387. doi 10.1126/science.256.5055.385

Quiroga G., Erice G., Aroca R., Chaumont F., Ruiz-Lozano J.M. Enhanced drought stress tolerance by the arbuscular mycorrhizal symbiosis in a drought-sensitive maize cultivar is related to a broader and differential regulation of host plant aquaporins than in a drought-tolerant cultivar. *Front Plant Sci*. 2017;8:1056. doi 10.3389/fpls.2017.01056

Savary S., Willocquet L., Pethybridge S.J., Esker P., McRoberts N., Nelson A. The global burden of pathogens and pests on major food crops. *Nat Ecol Evol*. 2019;3(3):430-439. doi 10.1038/s41559-018-0793-y

Schachtman D.P., Reid R.J., Ayling S.M. Phosphorus uptake by plants: from soil to cell. *Plant Physiol*. 1998;116(2):447-453. doi 10.1104/pp.116.2.447

Schuurmans J.A.M.J., Van Dongen J.T., Rutjens B.P.W., Boonman A., Pieterse C.M.J., Borstlap A.C. Members of the aquaporin family in the developing pea seed coat include representatives of the PIP, TIP, and NIP subfamilies. *Plant Mol Biol*. 2003;53(5):655-667. doi 10.1023/B:PLAN.0000019070.60954.77

Seka A.M., Zhang J., Prodhan F.A., Ayele G.T., Finsa M.M., Sharma T.P.P., Melesse A.M. Hydrological drought impacts on water storage variations: a focus on the role of vegetation changes in the East Africa region. A systematic review. *Environ Sci Pollut Res*. 2022;29(53):80237-80256. doi 10.1007/s11356-022-23313-0

Singh R.K., Deshmukh R., Muthamilarasan M., Rani R., Prasad M. Versatile roles of aquaporin in physiological processes and stress tolerance in plants. *Plant Physiol Biochem*. 2020;149:178-189. doi 10.1016/j.plaphy.2020.02.009

Spatafora J.W., Chang Y., Benny G.L., Lazarus K., Smith M.E., Berbee M.L., Bonito G., Corradi N., Grigoriev I., Gryganskyi A., James T.Y., O'Donnell K., Roberson R.W., Taylor T.N., Uehling J., Vilgalys R., White M.M., Stajich J.E. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia*. 2016;108(5):1028-1046. doi 10.3852/16-042

Tian S., Wang X., Li P., Wang H., Ji H., Xie J., Qiu Q., Shen D., Dong H. Plant aquaporin AtPIP1;4 links apoplastic H_2O_2 induction to disease immunity pathways. *Plant Physiol*. 2016;171(3):1635-1650. doi 10.1104/pp.15.01237

Wang C., Hu H., Qin X., Zeise B., Xu D., Rappel W.-J., Boron W.F., Schroeder J.I. Reconstitution of CO_2 regulation of SLAC1 anion channel and function of CO_2 – permeable PIP2;1 aquaporin as CARBONIC ANHYDRASE4 interactor. *Plant Cell*. 2016;28(2):568-582. doi 10.1105/tpc.15.00637

Wang D., Ni Y., Xie K., Li Y., Wu W., Shan H., Cheng B., Li X. Aquaporin ZmTIP2;3 promotes drought resistance of maize through symbiosis with arbuscular mycorrhizal fungi. *Int J Mol Sci*. 2024;25(8):4205. doi 10.3390/ijms25084205

Wang Y., Zhao Z., Liu F., Sun L., Hao F. Versatile roles of aquaporins in plant growth and development. *Int J Mol Sci*. 2020;21(24):9485. doi 10.3390/ijms21249485

Wayne R., Tazawa M. Nature of the water channels in the internodal cells of *Nitellopsis*. *J Membrain Biol*. 1990;116(1):31-39. doi 10.1007/BF01871669

Yaneff A., Sigaut L., Marquez M., Alleva K., Pietrasanta L.I., Amodeo G. Heteromerization of PIP aquaporins affects their intrinsic permeability. *Proc Natl Acad Sci USA*. 2014;111(1):231-236. doi 10.1073/pnas.1316537111

Yatsenko-Stepanova T.N., Nemtseva N.V., Ignatenko M.E. The diversity of simbioses and their part in the evolution of the organic world. *Vestnik Orenburgskogo Gosudarstvennogo Universiteta = Vestnik of the Orenburg State University*. 2014;13(174):142-147 (in Russian)

Zhang X., Han C., Gao H., Cao Y. Comparative transcriptome analysis of the garden asparagus (*Asparagus officinalis* L.) reveals the molecular mechanism for growth with arbuscular mycorrhizal fungi under salinity stress. *Plant Physiol Biochem*. 2019;141:20-29. doi 10.1016/j.plaphy.2019.05.013

Zhang X., Zhuang L., Liu Y., Yang Z., Huang B. Protein phosphorylation associated with drought priming-enhanced heat tolerance in a temperate grass species. *Hortic Res*. 2020;7(1):207. doi 10.1038/s41438-020-00440-8

Zhou X., Yi D., Ma L., Wang X. Genome-wide analysis and expression of the aquaporin gene family in *Avena sativa* L. *Front Plant Sci*. 2024;14:1305299. doi 10.3389/fpls.2023.1305299

Zhou Y., MacKinnon R. The occupancy of ions in the K+ selectivity filter: charge balance and coupling of ion binding to a protein conformational change underlie high conduction rates. *J Mol Biol*. 2003;333(5):965-975. doi 10.1016/j.jmb.2003.09.022

Conflict of interest. The authors declare no conflict of interest.

Received October 14, 2024. Revised December 9, 2024. Accepted December 10, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):248-258

doi 10.18699/vjgb-25-28

Transcriptomic analysis of the symbiotic responsivity trait in pea (*Pisum sativum* L.)

D.O. Kuzmina, E.A. Zorin, A.S. Sulima, D.A. Romanyuk, M.L. Gordon, A.I. Zhernakov, O.A. Kulaeva, G.A. Akhtemova, O.Y. Shtark, I.A. Tikhonovich, V.A. Zhukov

All-Russia Research Institute for Agricultural Microbiology, Pushkin 8, St. Petersburg, Russia
vzhukov@arriam.ru

Abstract. Pea (*Pisum sativum* L.) is an important crop culture and a model object for studying the molecular genetic bases of nitrogen-fixing symbiosis and arbuscular mycorrhiza (AM). Pea genotypes with high and low responsivity to inoculation with nodule bacteria (rhizobia) and AM fungi have been described: the 'responsive' genotypes demonstrate an increase in seed weight under inoculation, while 'non-responsive' ones do not show such a reaction. In order to get insight into the molecular genetic mechanisms underlying the symbiotic responsivity, a transcriptomic analysis of whole root systems of pea plants of the 'responsive' genotype k-8274 (cv. Vendevil, France) and 'non-responsive' genotype k-3358 (unnamed cultivar, Saratov region, Russia) grown in soil without inoculation (control) and inoculated either with rhizobia (single inoculation) or with rhizobia together with AM fungi (double inoculation) was performed. It was shown that the 'responsive' genotype, indeed, demonstrated a pronounced transcriptomic response to single and double inoculation, in contrast to the 'non-responsive' genotype. In k-8274, single inoculation led to specific up-regulation of genes related to catabolism of polyamines, lipid metabolism, and jasmonic acid and salicylic acid signaling. Under double inoculation, the specifically up-regulated genes in k-8274 were related to arbuscular mycorrhiza infection, and the down-regulated genes were related to nodulation. This fact matches the phenotype of the plants: the number of nodules was lower in k-8274 under double inoculation as compared to the control. Thus, strict control over the nodule number may be one of the mechanisms underlying the symbiotic responsivity of pea. Finally, a comparison of expression profiles in k-8274 and k-3358 roots under double inoculation also allowed us to identify the transcriptomic signatures characteristic of the symbiotically responsive genotype. Further work will be focused on validation of these transcriptomic markers of the symbiotic responsivity trait in pea.

Key words: pea; legume-rhizobial symbiosis; arbuscular mycorrhiza; symbiotic responsivity; transcriptomics

For citation: Kuzmina D.O., Zorin E.A., Sulima A.S., Romanyuk D.A., Gordon M.L., Zhernakov A.I., Kulaeva O.A., Akhtemova G.A., Shtark O.Y., Tikhonovich I.A., Zhukov V.A. Transcriptomic analysis of the symbiotic responsivity trait in pea (*Pisum sativum* L.). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):248-258. doi 10.18699/vjgb-25-28

Funding. The study was supported by the Russian Science Foundation (grant # 22-16-00109, https://rscf.ru/project/22-16-00109/).

Транскриптомный анализ признака симбиотической отзывчивости у гороха посевного (*Pisum sativum* L.)

Д.О. Кузьмина, Е.А. Зорин, А.С. Сулима, Д.А. Романюк, М.Л. Гордон, А.И. Жернаков, О.А. Кулаева, Г.А. Ахтемова, О.Ю. Штарк, И.А. Тихонович, В.А. Жуков

Всероссийский научно-исследовательский институт сельскохозяйственной микробиологии, Пушкин-8, Санкт-Петербург, Россия
vzhukov@arriam.ru

Аннотация. Горох посевной (*Pisum sativum* L.) является важной сельскохозяйственной культурой и модельным объектом для изучения молекулярно-генетических основ азотфиксирующего симбиоза и арбускулярной микоризы (АМ). Описаны генотипы гороха с высокой и низкой отзывчивостью на инокуляцию клубеньковыми бактериями (ризобиями) и АМ грибами: «отзывчивые» генотипы демонстрируют прибавку массы семян при инокуляции, а для «неотзывчивых» генотипов такая реакция не характерна. С целью описания молекулярно-генетических механизмов, лежащих в основе симбиотической отзывчивости, был проведен транскриптомный анализ целых корневых систем растений гороха «отзывчивого» генотипа к-8274 (сорт Vendevil, Франция) и «неотзывчивого» генотипа к-3358 (сорт из Саратовской области, Россия), выращенных в почве без инокуляции (контроль) и при инокуляции ризобиями (одиночная инокуляция) и ризобиями совместно с АМ грибами (двойная инокуляция). «Отзывчивый» генотип действительно продемонстрировал выраженный ответ на уровне транскриптома на одиночную и двойную инокуляцию, в отличие от «неотзывчивого» генотипа. Одиночная инокуляция привела у к-8274 к специфическому повы-

шению экспрессии генов, связанных с катаболизмом полиаминов, метаболизмом липидов и сигналингом на основе жасмоновой и салициловой кислот. При двойной инокуляции у к-8274 была повышена экспрессия генов, связанных с арбускулярно-микоризной инфекцией, и понижена экспрессия генов, связанных с клубенькообразованием. Данный факт соответствует фенотипу растений: число клубеньков у к-8274 было снижено при двойной инокуляции по сравнению с контролем. Таким образом, одним из механизмов, лежащих в основе симбиотической отзывчивости у гороха, может быть строгий контроль над числом клубеньков. Наконец, сравнение экспрессионных профилей корней растений к-8274 и к-3358 в условиях двойной инокуляции позволило идентифицировать «транскриптомные сигнатуры», характерные для симбиотически отзывчивого генотипа к-8274. Дальнейшая работа будет направлена на подтверждение возможности использования выявленных генов в качестве транскриптомных маркеров признака симбиотической отзывчивости у гороха посевного.
Ключевые слова: горох; бобово-ризобиальный симбиоз; арбускулярная микориза; симбиотическая отзывчивость; транскриптомика

Introduction

Legume plants (family Fabaceae) are an important component of modern agricultural practices due to their ability to fix atmospheric nitrogen in symbiosis with nodule bacteria (rhizobia) (Rubiales, Mikic, 2015). The symbiotic nitrogen fixation provides plants with combined nitrogen, and this feature allows legumes to grow on lower doses of mineral fertilizers, which is economically advantageous and beneficial for the environment (Goyal et al., 2021). The N_2 fixation occurs in specific organs called root nodules, which are formed on the roots of legume plants, where rhizobia are hosted within the plant cells (Yang et al., 2022). In addition to nitrogen fixation, legumes, like most terrestrial plants, form arbuscular mycorrhiza (AM), which helps plants cope with water deficiency and lack of mineral phosphorus (Smith, Read, 2008). This tendency to form mutualistic symbioses with beneficial microorganisms makes legumes profitable crops for use according to the concept of sustainable agriculture.

It is generally accepted that the root nodule (RN) symbiosis of legumes has evolved on the base of pre-existing AM (Parniske, 2008; Oldroyd, 2013). This idea is supported by the fact that some components of signaling systems are shared between the two symbioses. The first example of such shared signaling system is the so-called CSSP (common symbiosis signaling pathway), the signal transduction pathway that is activated during the early steps of the development of both AM and RN symbioses (Harrison, 2012; Wang D. et al., 2022). The second one is the autoregulation system, which exercises systemic control over the nodule number and the rate of mycorrhizal colonization of roots (autoregulation of nodulation (AON)/ autoregulation of mycorrhization (AOM)), depending on the amount of available nitrogen and phosphorus in the growth substrate, respectively (Reid et al., 2011; Ferguson et al., 2019; Müller, Harrison, 2019). Interestingly, nodulation systemically influences mycorrhization, and *vice versa*, as was shown in split-root experiments in alfalfa (*Medicago sativa* L.) (Catford et al., 2003).

Pea (*Pisum sativum* L.) is a profitable legume crop that has also served as a model to study the genetic system controlling the development of RN symbiosis and AM, similar to model legumes *Medicago truncatula* Gaertn., *Lotus japonicus* (Regel.) K. Larsen and *Glycine max* (L.) Merr. (Roy et al., 2020; Tsyganov, Tsyganova, 2020). Different genotypes have been described in pea with high and low responsivity to inoculation with rhizobia and AM fungi: the 'responsive' genotypes increase their seed productivity under inoculation, and the 'non-responsive' ones do not demonstrate such a reaction to inoculation (Shtark et al., 2006). This symbiotic responsivity trait was also called EIBSM, standing for Effectiveness of Interaction with Beneficial Soil Microorganisms, and was proposed for breeding of pea and other legumes (Shtark et al., 2012, 2015).

Recent development of pea genomics makes it possible to use post-genomic technologies such as transcriptomics and proteomics for studying agriculturally important traits (Parihar et al., 2022; Rubiales et al., 2023). Most such works, however, are devoted to studying resistance to pathogens rather than responsivity to symbionts (Castillejo et al., 2015; Cerna et al., 2017; Liu et al., 2023; Kälin et al., 2024), so the molecular mechanisms underlying the symbiotic responsivity in pea remain largely unexplored (Zhukov et al., 2021a). Therefore, the aim of this work was to reveal the molecular genetic bases of this trait by describing the transcriptomic changes in roots of two contrasting pea genotypes, the 'responsive' k-8274 and the 'non-responsive' k-3358, after inoculation with rhizobia (Rh) and rhizobia plus AM fungi (Rh+AM), as compared to a non-treated control.

Materials and methods

Vegetation experiments. The plant material (whole root systems) for RNA extraction was taken from the previously conducted vegetation experiment described in detail in (Zhukov et al., 2017). Briefly, the plants of the genotypes k-8274 and k-3358 (VIR collection of pea, St. Petersburg, Russia) were grown in non-sterile sod-podzolic light loamy soil (Leningrad Oblast, area of the Belogorka Science and Production Association, C_{humus} 1.27 % and N_{total} 0.11 %, pH_{salt} 4.92), three plants per 5-liter pot, in a greenhouse during the summer period at ARRIAM, St. Petersburg, Russia in the following variants: non-inoculated plants (control), plants inoculated with *Rhizobium leguminosarum* bv. *viciae* strain RCAM1026 (Rh), plants inoculated with *Rhizobium leguminosarum* bv. *viciae* strain RCAM1026 together with a mixture of arbuscular mycorrhizal fungi *Rhizophagus irregularis* BEG144, *R. irregularis* BEG53 and *Glomus* sp. ST3 (Rh+AM). After 4 weeks of growth, the plants were harvested; root systems were washed with water, immediately frozen in liquid nitrogen and stored at –80 °C until further processing. For a biological replicate, three plants from one pot were collected. In total, there were three biological replicates per treatment per plant genotype.

In the experiment on phenotypic characterization of k-8274 and k-3358, plants were grown in 2-liter pots with sterile sand, 4 plants per pot, without inoculation (control) and under inoculation with either *Rhizobium leguminosarum* bv. *viciae*

RCAM1026 (Rh) or *Rhizophagus irregularis* BEG144 (AM), or their combination (Rh+AM), in four replicates (pots) per treatment. The seeds were pre-sterilized with concentrated sulphuric acid, rinsed 5 times with autoclaved distilled water, and germinated in Petri dishes on filter paper at 28 °C in the dark. Inoculation with rhizobia was performed by pouring 3 ml of water suspension (10^6 CFU×l^{-1}) of *Rh. leguminosarum* bv. *viciae* RCAM1026 under each seedling; inoculation with AM fungi was performed by adding 60 g of dry roots of *Sorghum* sp. colonized by AM fungi (the details of the AM inoculation method are described in Shtark et al., 2019). At planting, the seedlings were supplied with mineral nutrition solution with a low content of phosphorus and nitrogen (Sulima et al., 2019), 150 ml per pot.

The plants were grown in a VB 1014 (Vötsch Industrietechnik, Germany) growth chamber under the following conditions: day/night 16/8 h, temperature 21 ± 1 °C, relative humidity 75 %, light irradiation 600 μmol of photons × m^{-2} × s^{-1}. The plants were watered by 200 ml of distilled water twice a week. At 4 weeks after planting/inoculation, the plants were harvested, and the root systems were washed with water. The number of nodules was counted during visual examination of the root systems; the shoots were air-dried and weighted.

Statistical analysis was carried out in the R environment using the core package 'stats' (version 4.3.0). A two-way ANOVA was used to assess the effects of inoculations and their action on shoot dry weight, and a one-way ANOVA was used to evaluate the impact of mycorrhiza on nodulation.

RNA extraction, library preparation and sequencing. The frozen whole root systems were grinded in liquid nitrogen using mortars and pistils; the resulting powder was used for total RNA extraction using Trizol reagent (Thermo Fisher Scientific, USA). The quality of RNA was assessed using a 2100 Bioanalyzer Instrument (Agilent Technologies, USA). RNA of sufficient quality (RNA integrity number (RIN) > 8) was obtained from only two replicates of the samples from the control conditions, single inoculation and combined (double) inoculation, which allowed us to study the effect of single and combined inoculation on plants of the k-8274 and k-3358 genotypes.

The extracted RNA was used for RNAseq library preparation using the MACE v1.0 kit (GenXPro GmbH, Frankfurt am Main, Germany). The libraries were sequenced in GenXPro GmbH (Frankfurt am Main, Germany) on Illumina HiSeq2000. Raw sequencing data were uploaded to the NCBI database (BioProject number PRJNA1154300).

Bioinformatic analysis. The quality of raw reads was assessed using FastQC (version 0.11.9) (Andrews, 2010) and multiqc (Ewels et al., 2016). Trimmomatic (version 0.39) with default parameters was employed to remove adapter sequences and low-quality sequences (Bolger et al., 2014). Clean reads were aligned to the reference genome of cv. Frisson (NCBI: JANEYU000000000; Zorin et al., 2022) with STAR (version 2.7.10b) (Dobin et al., 2013) and sorted using SAMtools (version 1.17) (Danecek et al., 2021). Using multi-mapping, featureCounts (version 2.0.3) was used to count the number of reads that were aligned to genes or exons (Liao et al., 2014).

The BLAST+ command line tool (version 2.9.0) was used to annotate genes to which reads were mapped, with an E-value threshold of 1e–5 against *Medicago truncatula* functional annotation (genomic assembly MedtrA17_4.0) (Camacho et al., 2009). PCA plots were generated using the R packages DESeq2 and ggplot2 (Love et al., 2014; Wickham, 2016). R (version 4.1.3) was used to perform differential gene expression analysis with the DESeq2 package. Genes were considered differentially expressed if the Wald test was passed with the False Discovery Rate (FDR) value of no more than 0.05 and a log2-fold change less than or more than 0.5. Additionally, a targeted analysis of differential gene expression was performed on gene lists (listed in the Supplementary Table S1)[1] that were acquired from earlier research projects. These gene lists included *Sym* genes and genes involved in the systemic process of autoregulation of nodulation.

GO enrichment analysis was carried out using the topGO packages (version 2.42.0) (Alexa, Rahnenfuhrer, 2024), with the use of the weight01 method and Fisher's exact test. Genes, the expression of which was considerably elevated or decreased, were used independently to search for biological processes. Biological pathways with statistically significant up/down-regulation were counted at p-value < 0.05 and depicted using ggplot2. The p-value indicates the probability of this value occurring by chance, and suggests that the biological process under investigation is enriched in the transcriptomic data.

The heatmap is based on a matrix containing the values of 1–R (R is the Pearson correlation coefficient). The correlation is calculated based on the values of normalized expression (the number of normalized reads per million was obtained using DESeq2, VST (Variance Stabilizing Transformation)), which were additionally logarithmically transformed (log2), and then transformed into a z-scale, which for each gene for each sample reflects the number of standard deviations from the average value for all samples for this gene. Further, the genes were clustered based on these values using the hierarchical clustering method. The obtained matrix was displayed as heat maps with pheatmap (version 1.0.12) (Kolde, 2015). The R packages VennDiagram (version 1.7.3) and EnhancedVolcano (version 1.18.0) were used to show the results of the differential gene expression study (Chen, Boutros, 2011).

Results

Transcriptomic response to single inoculation (Rh)

Inoculation with rhizobia changed the gene expression profiles in roots of both studied genotypes. The response to inoculation was more pronounced in k-8274: 440 unique genes were differentially expressed as compared to control; for k-3358, there were only 14 such genes. Additionally, 81 genes changed their expression similarly in both genotypes (Fig. 1*A*), and for all but one, the expression level increased. Enrichment analysis for the similarly up-regulated genes showed that these genes were related to the biological processes connected with nodulation and nitrogen fixation, such as biosynthesis of glutamate from proline, 1-aminocyclopropane-1-carboxylate biosynthesis, polyamine transmembrane transport, etc. (Table 1). The genes additionally activated in k-8274 were related to such biological processes as catabolism of polyamines, lipid metabolism, and jasmonic acid and salicylic acid signaling (Table 1).

[1] Supplementary Tables S1, S2 and Fig. S1 are available at: https://vavilovj-icg.ru/download/pict-2025-29/appx9.pdf

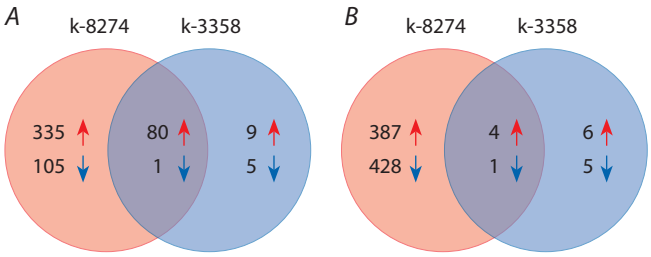


Fig. 1. The number of differentially expressed genes in roots of k-8274 and k-3358 under single (*A*) and double (*B*) inoculation as compared to uninoculated control.

Interestingly, the same genes in k-3358 were not classified as DEGs due to their low expression level both in control and under inoculation with rhizobia. Further, in k-8274, the down-regulated genes were related to ion transport, growth and response to fungi, which may indicate a decrease in the AM fungi spread in k-8274 roots. In k-3358, the down-regulated genes were related to the electron transport chain, as well to the response to abscisic acid and water deprivation.

Transcriptomic response to double inoculation (Rh+AM)

The transcriptomic response to double inoculation was different in the studied genotypes (Fig. 1*B*). Double inoculation of the responsive k-8274 genotype altered the expression of 815 unique genes in the roots, while only 11 unique genes in the non-responsive genotype k-3358 changed their expression level, and 5 genes showed identical expression changes in both genotypes (Fig. 1*B*). The absolute expression level of the same 815 genes in k-3358 in all tested conditions was comparable to that of k-8274 in control conditions and did not change significantly due to inoculation (the effect similar to that observed under mono-inoculation with rhizobia), reflecting the low responsivity of k-3358 to double inoculation.

Enrichment analysis showed that several up-regulated DEGs in k-8274 under double inoculation (Rh+AM) were related to the same biological processes as in k-8274 under single inoculation (Rh), namely, jasmonic acid signaling, defense response and response to wounding (Table 2). The common down-regulated biological processes included pectin catabolism, inorganic anion transport, and regulation of 1-deoxy-D-xylulose-5-phosphate synthase activity. Further, many genes specifically up-regulated in k-8274 under double inoculation (Rh+AM) are associated with arbuscular mycorrhiza infection (GO biological processes: oxylipin biosynthetic process, defense response, response to chitin). In turn, the specific down-regulated genes in k-8274 are related to nodulation and assimilation of combined nitrogen (GO biological processes: nodulation, nitrate transmembrane transport) (Table 2). The observed down-regulation of the nodulation-related genes matches the result of phenotypic analysis of the plants (Zhukov et al., 2017): k-8274 plants formed significantly less nodules under double inoculation than in control conditions and under single inoculation with rhizobia.

Table 1. Up-regulated biological processes under single inoculation with rhizobia (Rh) in the studied genotypes

GO category	Biological process	*p*-value
	Common **up-regulated** biological processes in **k-8274+Rh** and **k-3358+Rh** (as compared to control)	
GO:0098869	Cellular oxidant detoxification	0.012
GO:0010133	Proline catabolic process to glutamate	0.013
GO:0009873	Ethylene-activated signaling pathway	0.013
GO:0042218	1-Aminocyclopropane-1-carboxylate biosynthesis	0.017
GO:1902047	Polyamine transmembrane transport	0.017
GO:0055085	Transmembrane transport	0.032
GO:0006979	Response to oxidative stress	0.039
	Specific **up-regulated** biological processes in **k-8274+Rh** (as compared to control)	
GO:0016311	Dephosphorylation	0.0012
GO:0006598	Polyamine catabolic process	0.0025
GO:0009269	Response to desiccation	0.0025
GO:0006629	Lipid metabolic process	0.0030
GO:0009805	Coumarin biosynthetic process	0.0043
GO:0006538	Glutamate catabolic process	0.0079
GO:0009694	Jasmonic acid metabolic process	0.0114
GO:0009696	Salicylic acid metabolic process	0.0116
GO:0052746	Inositol phosphorylation	0.0116

Table 2. Up- and down-regulated biological processes under double inoculation with rhizobia (Rh) and AM fungi (AM) in the studied genotypes

GO category	Biological process	*p*-value
	Common **up-regulated** biological processes in **k-8274+Rh+AM** and **k-8274+Rh** (as compared to control)	
GO:0009611	Response to wounding	1.6e–07
GO:2000022	Regulation of jasmonic acid mediated signaling pathway	2.0e–07
GO:0031347	Regulation of defense response	6.7e–05
GO:0009695	Jasmonic acid biosynthetic process	0.00012
GO:0005992	Trehalose biosynthetic process	0.00064
GO:0052746	Inositol phosphorylation	0.00084
GO:0043086	Negative regulation of catalytic activity	0.00172
GO:0016311	Dephosphorylation	0.00240
GO:0030026	Cellular manganese ion homeostasis	0.00325
GO:0009269	Response to desiccation	0.00375
	Common **down-regulated** biological processes in **k-8274+Rh+AM** and **k-8274+Rh** (as compared to control)	
GO:0006873	Cellular ion homeostasis	0.00028
GO:0045490	Pectin catabolic process	0.00563
GO:0015698	Inorganic anion transport	0.00765
GO:0071577	Zinc ion transmembrane transport	0.00926
GO:0006355	Regulation of DNA-templated transcription	0.00984
GO:0046622	Positive regulation of organ growth	0.00985
GO:0071836	Nectar secretion	0.00985
GO:1902395	Regulation of 1-deoxy-D-xylulose-5-phosphate synthase activity	0.00985
GO:0061087	Positive regulation of histone H3-K27 methylation	0.00985
GO:0010322	Regulation of isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway	0.01959
	Specific **up-regulated** biological processes in **k-8274+Rh+AM** (as compared to control)	
GO:0015824	Proline transport	4.7e–05
GO:0080163	Regulation of protein serine/threonine phosphatase activity	0.0011
GO:0009269	Response to desiccation	0.0015
GO:0031408	Oxylipin biosynthetic process	0.0016
GO:0009695	Jasmonic acid biosynthetic process	0.0019
GO:0006952	Defense response	0.0024
GO:0010200	Response to chitin	0.0026
GO:0009805	Coumarin biosynthetic process	0.0034
GO:0043086	Negative regulation of catalytic activity	0.0034
GO:0010286	Heat acclimation	0.0051
	Specific **down-regulated** biological processes in **k-8274+Rh+AM** (as compared to control)	
GO:0009877	Nodulation	1.7e–05
GO:0006465	Signal peptide processing	5.0e–05
GO:0015671	Oxygen transport	0.00053
GO:0009228	Thiamine biosynthetic process	0.00083
GO:0010044	Response to aluminum ion	0.00375
GO:0015706	Nitrate transmembrane transport	0.00483
GO:1903401	L-lysine transmembrane transport	0.00614
GO:0006873	Cellular ion homeostasis	0.00615
GO:0034220	Ion transmembrane transport	0.00675
GO:0015713	Phosphoglycerate transmembrane transport	0.00905

Comparison of k-8274 and k-3358 under double inoculation

In order to get an insight into the molecular bases of the symbiotic responsivity in pea, we compared the transcriptomes of the 'responsive' k-8274 and 'non-responsive' k-3358 genotypes treated with Rh+AM. The biological processes up-regulated in k-8274 roots include lignin biosynthesis and cell wall biogenesis, as well as phosphatidylinositol biosynthesis (Table 3). In turn, responses to glucose and fructose, and biosynthesis and metabolism of glutathione and cysteine in k-8274 were down-regulated compared to those in k-3358 (Table 3). The corresponding genes with higher expression level in k-8274 or k-3358 roots are listed in Table S2.

In our previous study, root transcriptomes of three highly responsive (high-EIBSM) and three low-responsive (low-EIBSM) pea genotypes grown in experimental conditions similar to those of the present experiment (i. e., in pots with soil under combined inoculation with nodule bacteria and AM fungi) were compared, and 90 differentially expressed genes were identified (Afonin et al., 2021). The intersection of the lists of DEGs from the previous and the present experiment revealed 11 genes that were similarly up- or down-regulated in the following comparisons: 'three high-EIBSM *vs* three low-EIBSM genotypes' and 'k-8274 *vs* k-3358' (Table 4). They can be considered as transcriptomic markers of the symbiotic responsivity trait. Interestingly, 5 out of 9 up-regulated genes are related to abscisic acid, jasmonic acid and salicylic acid metabolism and signaling, which points towards the activation of defense reactions in the roots of the high-EIBSM pea plants.

Table 3. Biological processes that are differentially regulated in roots of k-8274 and k-3358 under double inoculation (Rh+AM)

GO category	Biological process	*p*-value
	Up-regulated biological processes in **k-8274+Rh+AM** as compared to **k-3358+Rh+AM**	
GO:1901430	Positive regulation of syringal lignin biosynthesis	0.0019
GO:2000652	Regulation of secondary cell wall biogenesis	0.0056
GO:0009664	Plant-type cell wall organization	0.0332
GO:0006661	Phosphatidylinositol biosynthetic processes	0.0495
	Down-regulated biological processes in **k-8274+Rh+AM** as compared to **k-3358+Rh+AM**	
GO:0016487	Farnesol metabolic process	0.0022
GO:0009269	Response to desiccation	0.0022
GO:0009788	Negative regulation of abscisic acid-activated signalling pathway	0.0130
GO:1990961	Xenobiotic detoxification by transmembrane export across the plasma membrane	0.0130
GO:0009749	Response to glucose	0.0173
GO:0009750	Response to fructose	0.0238
GO:0009853	Photorespiration	0.0259
GO:0006749	Glutathione metabolic process	0.0323
GO:0019344	Cysteine biosynthetic process	0.0365
GO:0006414	Translational elongation	0.0365

Table 4. Differentially expressed genes of k-8274 (Rh+AM) as compared to k-3358 (Rh+AM) that overlap in the present study and Afonin et al., 2021

Gene ID in the genome of *P. sativum* cv. Frisson	Functional annotation according to Mercator4 ver.6.0	Log2FoldChange	Adjusted *p*-value
evm.TU.contig_1396.103	Jasmonoyl-amino acid hydroxylase / (CYP94B) & Cytochrome P450 94B3	3.1922	1.28e–07
evm.TU.scaffold_1292.71	Jasmonic acid oxidase / (JOX/JAO) & Jasmonate-induced oxygenase 2	2.0252	1.01e–09
evm.TU.scaffold_443.110	EC_2.4 glycosyltransferase & Probable alpha,alpha-trehalose-phosphate synthase [UDP-forming] 11	1.9831	0.0456
evm.TU.contig_941.42	H-type thioredoxin / (Trx-H) & Thioredoxin H-type	1.8029	0.0053
evm.TU.scaffold_2090.181	Receptor component / (PYL/RCAR) of cytoplasm-localized abscisic acid receptor complex & Abscisic acid receptor PYL4	1.6877	0.0001
evm.TU.contig_1345.208	EC_2.3 acyltransferase & Benzyl alcohol O-benzoyltransferase	1.3998	0.0238
evm.TU.scaffold_1655.224	Proline dehydrogenase & Proline dehydrogenase 2	1.3108	0.0495
evm.TU.scaffold_1169.2	HD-ZIP I/II-type transcription factor & Homeobox-leucine zipper protein HAT22	0.9947	0.0006
evm.TU.scaffold_1535.577	Receptor component / (PYL/RCAR) of cytoplasm-localized abscisic acid receptor complex & Abscisic acid receptor PYL4	0.9145	0.0044
evm.TU.scaffold_1535.556	Iron storage protein / (FER) & Ferritin-3	–0.8532	0.0192
evm.TU.scaffold_783.526	Organic cation transporter / (OCT) & Organic cation/carnitine transporter 3	–1.0481	0.0256

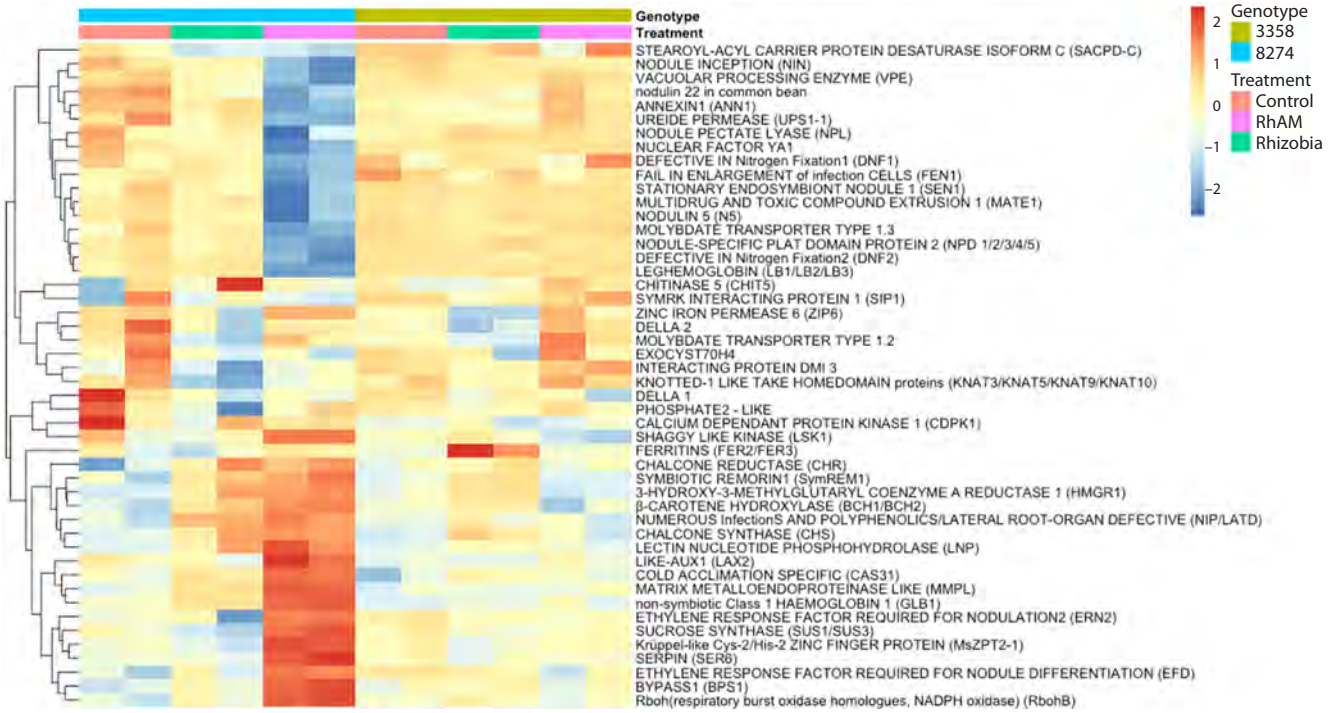


Fig. 2. Expression level of selected *Sym*-genes in roots of k-8274 and k-3358 in control conditions and under single and double inoculation.

The heatmap is based on a matrix containing the values of 1–R (R = Pearson correlation coefficient). The raw data were normalized to the size of the dataset, logarithmically transformed, and converted to a z-scale. The lowest gene expression value is represented in blue; the highest value is represented in red.

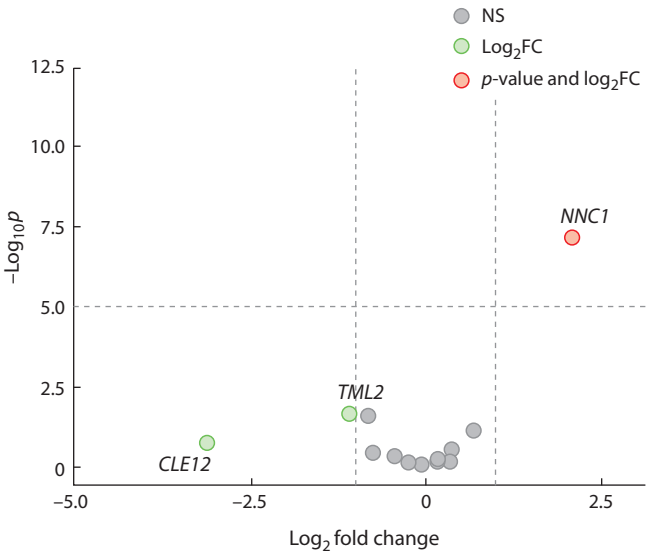


Fig. 3. Differential expression of genes related to autoregulation of nodulation in roots of k-8274 under double inoculation as compared to uninoculated control.

Log_2 fold change (Log_2FC) – the binary logarithm of the ratio of a transcript's expression values in two different conditions; $-Log_{10}p$ – negative logarithm to base 10 of the *p*-value. For ease of visualization, the *p*-values for each gene were log-transformed on the graph; NS – statistically non-significant difference in expression ($-Log_{10}p < 5.0$, or $p < 0.00001$). The threshold *p*-value ($p < 0.00001$) was chosen taking into account the correction for multiple comparisons.

The red color: the *NNC1* gene passed both the *p*-value ($-Log_{10}p > 5.0$) and expression level ($log_2FC > 1.5$) significance thresholds; the green color: the *CLE12* and *TML2* genes have significantly reduced expression levels ($log_2FC < -1.5$), but this reduction is not statistically significant ($-Log_{10}p < 5.0$).

Targeted analysis of marker genes

The expression level of the symbiotic (*Sym*) genes (i. e., genes, the role of which in AM and/or RN symbiosis was experimentally verified by mutation analysis or RNAi experiments in model legumes) was examined across all RNAseq samples (39 genes, see Table S1). Interestingly, the samples of k-8274 under double inoculation (Rh+AM) strikingly differed from all other samples, showing two clusters of *Sym* genes, the expression level of which was either higher or lower as compared to other samples (although several genes in the up-regulated cluster had similar expression level in the samples under mono-inoculation with rhizobia) (Fig. 2). The cluster of down-regulated genes contains 15 genes that are normally expressed during nodule development (which again reflects the decreased nodule number in k-8274 plants under double inoculation) and 2 genes (*VPE*, *ANN1*) involved in the infection process in both Rh and AM symbioses (Table S1). The cluster of 18 up-regulated genes includes the genes of the common symbiotic signaling pathway (CSSP) (*SymREM1*, *HMGR1*). Interestingly, the other genes belonging to CSSP do not demonstrate altered expression, which suggests that the early steps of both AM and nodule development are not disturbed in k-8274.

The expression level of the genes related to the AON system was also examined. Out of 13 genes tested, the expression level of only one gene, *NNC1*, was increased in roots of k-8274 under double inoculation (Fig. 3) compared to control. The NNC1 protein in soybean is a suppressor of the master transcriptional regulator NIN (Wang L. et al., 2019), so it seems that the observed down-regulation of *Nin* in k-8274 under

double inoculation (Fig. 2) is also controlled via up-regulation of *NNC1*, resulting in reduction of nodule formation.

Finally, the expression level of the *PsGLP2* gene was tested, which was recently identified as a transcriptional marker of high symbiotic responsivity of k-8274 and its descendent breeding line 'Triumph' (Zorin et al., 2023). We detected the trend of up-regulation of *PsGLP2* in k-8274 under both mono- and double inoculation, as compared to control, although this change in expression level was not statistically significant (Fig. S1). In k-3358, the expression of *PsGLP2* was not altered under either single or double inoculation.

The suppression of nodulation in k-8274 is condition-dependent

We reanalyzed the raw data of nodule number from Zhukov et al., 2017 by one-way ANOVA and confirmed that AM significantly influenced the nodulation in k-8274 but not k-3358 plants grown in a non-sterile soil (Table 5). However, this result was not reproduced in a new experiment in other experimental conditions (in sterile sand): both k-3358 and k-8274 genotypes demonstrated a decrease in nodulation under combined inoculation (rhizobia + AM fungi) as compared to single inoculation with rhizobia, but the effect was not statistically significant (Table 6).

In sand, AM treatment decreased the shoot dry weight of the non-responsive genotype k-3358 and did not alter the shoot dry weight of the responsive genotype k-8274, compared to the untreated control (Table 7). One can suggest that in some environments (sterile sand) mycorrhiza begins to harm an ineffective genotype, but not an effective one; thus, the trait of symbiotic responsiveness may manifest itself differently in different experimental conditions.

Discussion

The trait of symbiotic responsivity, or EIBSM (Efficiency of Interaction with Beneficial Soil Microorganisms), is a quantitative trait which is determined as an increase in seed biomass due to inoculation with rhizobia and arbuscular mycorrhizal fungi (Shtark et al., 2012). Apparently, the genetic control of this trait is complex; therefore, preliminary works aimed at molecular genetic characterization of plant genotypes with different symbiotic responsivity are required. In pea, two contrasting genotypes, the 'responsive' k-8274 and the 'non-responsive' k-3358, are used as models to study different aspects of manifestation of this trait. Previously, by proteomic analysis of pea seeds, it was shown that the high responsivity to combined inoculation with rhizobia and AM fungi is connected with prolongation of the seed filling period in the

Table 5. ANOVA results assessing the effect of arbuscular mycorrhiza on nodulation (non-sterile soil, data from Zhukov et al., 2017)

Genotype	Factor	Impact	Degrees of freedom	Mean square	F value	*p*-value
k-8274	Arbuscular mycorrhiza	–	1	5280.7	20.114	0.000185
	Residuals		22	262.5		
k-3358	Arbuscular mycorrhiza	+	1	600.25	2.7627	0.1159
	Residuals		16	217.27		

Note. Here and in Tables 6 and 7: the 'Impact' column indicates an increase (+) or decrease (–) in the mean value due to the estimated factor.

Table 6. ANOVA results assessing the effect of arbuscular mycorrhiza on nodulation (sterile sand)

Genotype	Factor	Impact	Degrees of freedom	Mean square	F value	*p*-value
k-8274	Arbuscular mycorrhiza	–	1	22.702	0.318	0.5778
	Residuals		25	71.379		
k-3358	Arbuscular mycorrhiza	–	1	896.53	3.0379	0.09232
	Residuals		28	295.12		

Table 7. Two-way ANOVA results assessing effects of rhizobia and arbuscular mycorrhiza on shoot weight (sterile sand)

Genotype	Factor	Impact	Degrees of freedom	Mean square	F value	*p*-value
k-8274	Rhizobia	+	1	0.004	2.377	0.1295
	Arbuscular mycorrhiza	+	1	0.000186	0.1103	0.7412
	Rhizobia × Arbuscular mycorrhiza		1	0.000751	0.4464	0.5071
	Residuals		50	0.001683		
k-3358	Rhizobia	+	1	0.01206	0.7278	0.397358
	Arbuscular mycorrhiza	–	1	0.1853	11.1826	**0.001506**
	Rhizobia × Arbuscular mycorrhiza		1	0.007641	0.4611	0.499996
	Residuals		54	0.01657		

'responsive' genotype (Mamontova et al., 2019). The results of the present work expand the description of the responsivity trait: it was demonstrated that the roots of the 'responsive' genotype showed a more pronounced reaction to inoculation at the transcriptomic level than the roots of the 'non-responsive' one, and that the reaction of the 'responsive' genotype to combined inoculation (rhizobia + AM fungi) involved down-regulation of the nodule-related genes, which is in line with the suppression of nodulation shown in the earlier experiments (Zhukov et al., 2017).

Although the bulk transcriptome analysis of the entire root system does not allow accurate assessment of gene expression (since the development of nodules and/or arbuscular mycorrhiza may be regulated differently in different zones of the roots), the effect of inoculation was clearly visible in the 'responsive' genotype in contrast to the 'non-responsive' one. The list of genes with an elevated expression in the roots of k-8274 (as compared to k-3358) under double inoculation includes the genes encoding Lipid Transfer Protein (LTP) family member, putative SOUL heme-binding protein, MYB-like transcription factor, expansin, and metallothionein protein (Table S2). Some of these genes are directly linked to nodulation and/or mycorrhization: one of the members of the LTP family in *M. truncatula* known as Nodulin 5 is required for the successful symbiosis with *Sinorhizobium meliloti* (Pii et al., 2009); expansins play a role in arbuscular mycorrhiza formation (Mohanty et al., 2018); a member of metallothionein family is involved in rhizobial infection and nodulation in *Phaseolus vulgaris* (Fonseca-García et al., 2022).

Interestingly, several DEGs obtained in this comparison were also found within the previously published list of transcriptomic signatures characteristic of roots of symbiotically responsive pea genotypes grown in a non-sterile soil (Afonin et al., 2021). It is important to note that the experiment of Afonin and colleagues did not include either k-8274 or k-3358, thus, the resulting intersection of the gene lists from Afonin et al., 2021 and the present study may be considered as a list of reliable transcriptomic markers of the EIBSM trait (Table 4). These markers include the up-regulated genes that encode the enzymes jasmonoyl-amino acid hydroxylase and jasmonic acid oxidase, two different genes encoding the abscisic acid receptor PYL4 (which is also involved in jasmonic acid signaling; Lackman et al., 2011), and a benzyl alcohol O-benzoyltransferase, which is involved in biosynthesis of salicylic acid (Kotera et al., 2023). Salicylic acid and jasmonates play a key role in plant defense and have a strong influence on plants metabolism (Jeyasri et al., 2023; Monte, 2023); thus, the manifestation of the EIBSM trait may be based on the fine-tuning of defense reactions accompanied by metabolic changes. Moreover, it can be hypothesized that interaction with beneficial soil microorganisms may have a positive effect on systemic resistance of k-8274 plants, and this aspect of the EIBSM trait manifests itself in non-sterile soil but not in sterile sand.

In k-8274, the double inoculation with nodule bacteria and AM fungi led to down-regulation of the genes normally expressed in nodules, which corresponds to the previously described phenotype of plants (suppression of nodulation; Zhukov et al., 2017) and perhaps reflects the optimization of the nitrogen nutrition of the plants. Interestingly, the early symbiotic genes were not suppressed under double inoculation, indicating that CSSP, which is common for both AM and RN symbioses, functioned normally in these conditions. This means that the down-regulation of RN symbiosis takes place after the common signaling pathway, apparently, in order not to block the development of both symbioses together. Probably, a similar block of the symbiosis development occurs when pea interacts with non-specific rhizobia (this is the case for Afghanistan peas, a group of varieties which can form nodules only with a low number of specific strains, as opposed to European peas, which are nodulated by a broad spectrum of strains (Lie, 1984; Firmin et al., 1993)). In this case, the phenotypic analysis suggests that the early steps of symbiosis (encoded by CSSP genes) proceed normally, but the penetration of rhizobia into the root hairs is blocked due to absence of the signal transduction mediated by the receptor kinase LykX (=Sym2) (Sulima et al., 2017, 2019). It would be interesting to assess the expression level of the *PsNNC1* gene, which was up-regulated in k-8274 under double inoculation, in Afghanistan peas interacting with non-specific rhizobia, in order to check whether it participates in signal transduction during specific and non-specific interactions with rhizobia.

The nodulation suppression detected in k-8274 under double inoculation is accompanied by a decrease in the expression level of the *PsCLE12* and *PsTML2* genes ($Log_2FC < -1.0$, although this decrease is not statistically significant, see Fig. 3). Orthologs of these genes in *M. truncatula* act as negative regulators of nodule development and are parts of the AON (autoregulation of nodulation) system (Gautrat et al., 2019). This observation is consistent with our previous suggestion regarding the possible connection between the AON system and symbiosis efficiency (Zhukov et al., 2021b).

The phenomenon of nodulation suppression was observed in non-sterile soil (Zhukov et al., 2017), where plenty of microorganisms occur, whereas in sterile sand, the decrease in nodule number under double inoculation (Rh+AM) was visible but non-significant for both studied genotypes. Also, in sand, the inoculation with AM fungi had a negative effect on the 'non-responsive' genotype k-3358 and was neutral to the 'responsive' genotype k-8274. One can conclude that the responsivity trait may be dependent on several environmental factors such as temperature, humidity, the presence of indigenous microorganisms in the growth substrate, etc. Therefore, large-scale experiments are required to estimate the percentage of genotype (G) effect on the manifestation of the symbiotic responsivity trait in comparison to environment (E) and genotype-environment (G × E) interaction.

Recently we showed that the plant's habitus plays a role in manifestation of the responsivity trait: pea genotypes bearing a natural mutation *le* (p.A229T) in the *Le* gene encoding gibberellin 3-beta-dioxygenase (Martin et al., 1997) have shortened internodes, lower biomass and are more responsive to double inoculation (Rh+AM) than wild-type genotypes (Zhukov et al., 2021a). One of the explanations for this phenomenon was that smaller plants could react more quickly to change in the nitrogen/phosphorous content in the roots and inhibit formation of new symbiotic structures, since this reaction is mediated by long-distance signaling involving CLE and CEP peptides (Okamoto et al., 2016). Indeed, we showed that pea genotypes with long stems had more AM in their roots than the

genotypes with short stems (Zhukov et al., 2021a), and in the present study, we found that the down-regulation of nodule-related genes in non-sterile soil is characteristic of k-8274, which has the le phenotype, as opposed to k-3358 with the Le phenotype. Thus, the pleiotropic effect of the *le* mutation may also include influence on the plants' symbiotic responsivity, probably due to quicker signaling, which leads to suppression of formation of excessive symbiotic structures; however, further experiments are required to prove this statement.

Conclusion

Due to the development of pea genomics, genome- and transcriptome-wide analyses became available, making it possible to uncover the molecular bases of the traits of interest, including the symbiotic responsivity trait. Here, we described the transcriptomic signatures characteristic of roots of the symbiotically responsive k-8274 genotype. The biological processes associated with the functions of the identified genes include lignin biosynthesis, cell wall biogenesis, and biosynthesis of phosphatidylinositol. Also, the 'responsive' genotype k-8274 demonstrated the pronounced change in the gene expression profiles in roots, as opposite to the 'non-responsive' genotype k-3358, which reflects the observed differences in the effect of inoculation with symbiotic microorganisms. Further work should be devoted to the search for specific genes that affect EIBSM, which will form the basis for marker-assisted selection of new pea cultivars with high effectiveness of interaction with nodule bacteria and arbuscular mycorrhizal fungi.

References

Afonin A.M., Gribchenko E.S., Zorin E.A., Sulima A.S., Romanyuk D.A., Zhernakov A.I., Shtark O.Y., Akhtemova G.A., Zhukov V.A. Unique transcriptome features of pea (*Pisum sativum* L.) lines with differing responses to beneficial soil microorganisms. *Ecol Genet.* 2021;19(2):131-141. doi 10.17816/ECOGEN54703

Alexa A., Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. R Package Version 2.58.0. 2024. doi 10.18129/B9.bioc.topGO. https://bioconductor.org/packages/topGO

Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bolger A.M., Lohse M., Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-2120. doi 10.1093/bioinformatics/btu170

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T. L. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421. doi 10.1186/1471-2105-10-421

Castillejo M.Á., Bani M., Rubiales D. Understanding pea resistance mechanisms in response to *Fusarium oxysporum* through proteomic analysis. *Phytochemistry.* 2015;115(1):44-58. doi 10.1016/j.phytochem.2015.01.009

Catford J., Staehelin C., Lerat S., Piché Y., Vierheilig H. Suppression of arbuscular mycorrhizal colonization and nodulation in split-root systems of alfalfa after pre-inoculation and treatment with Nod factors. *J Exp Bot.* 2003;54(386):1481-1487. doi 10.1093/jxb/erg156

Cerna H., Černý M., Habánová H., Šafářová D., Abushamsiya K., Navrátil M., Brzobohatý B. Proteomics offers insight to the mechanism behind *Pisum sativum* L. response to pea seed-borne mosaic virus (PSbMV). *J Proteomics.* 2017;153:78-88. doi 10.1016/j.jprot.2016.05.018

Chen H., Boutros P.C. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics.* 2011;12:35. doi 10.1186/1471-2105-12-35

Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., Whitwham A., Keane T., McCarthy S.A., Davies R.M. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10(2):giab008. doi 10.1093/gigascience/giab008

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21. doi 10.1093/bioinformatics/bts635

Ewels P., Magnusson M., Lundin S., Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* 2016;32(19):3047-3048. doi 10.1093/bioinformatics/btw354

Ferguson B.J., Mens C., Hastwell A.H., Zhang M., Su H., Jones C.H., Chu X., Gresshoff P.M. Legume nodulation: the host controls the party. *Plant Cell Environ*. 2019;42(1):41-51. doi 10.1111/pce.13348

Firmin J.L., Wilson K.E., Carlson R.W., Davies A.E., Downie J.A. Resistance to nodulation of cv. Afghanistan peas is overcome by *nodX*, which mediates an O-acetylation of the *Rhizobium leguminosarum* lipo-oligosaccharide nodulation factor. *Mol Microbiol*. 1993;10(2):351-360. doi 10.1111/j.1365-2958.1993.tb01961.x

Fonseca-García C., López-García C.M., Pacheco R., Armada E., Nava N., Pérez-Aguilar R., Solis-Miranda J., Quinto C. Metallothionein1A regulates rhizobial infection and nodulation in *Phaseolus vulgaris*. *Int J Mol Sci*. 2022;23(3):1491. doi 10.3390/ijms23031491

Gautrat P., Mortier V., Laffont C., De Keyser A., Fromentin J., Frugier F., Goormachtig S. Unraveling new molecular players involved in the autoregulation of nodulation in *Medicago truncatula*. *J Exp Bot*. 2019;70(4):1407-1417. doi 10.1093/jxb/ery465

Goyal R.K., Mattoo A.K., Schmidt M.A. Rhizobial–host interactions and symbiotic nitrogen fixation in legume crops toward agriculture sustainability. *Front Microbiol*. 2021;12:669404. doi 10.3389/fmicb.2021.669404

Harrison M.J. Cellular programs for arbuscular mycorrhizal symbiosis. *Curr Opin Plant Biol.* 2012;15(6):691-698. doi 10.1016/j.pbi.2012.08.010

Jeyasri R., Muthuramalingam P., Karthick K., Shin H., Choi S.H., Ramesh M. Methyl jasmonate and salicylic acid as powerful elicitors for enhancing the production of secondary metabolites in medicinal plants: an updated review. *Plant Cell Tissue Organ Cult.* 2023;153(3):447-458. doi 10.1007/s11240-023-02485-8

Kälin C., Piombo E., Bourras S., Brantestam A.K., Dubey M., Elfstrand M., Karlsson M. Transcriptomic analysis identifies candidate genes for Aphanomyces root rot disease resistance in pea. *BMC Plant Biol.* 2024;24(1):144. doi 10.1186/s12870-024-04817-y

Kolde R. Package 'pheatmap.' *R Package.* 2015;1(7):790

Kotera Y., Komori H., Tasaki K., Takagi K., Imano S., Katou S. The peroxisomal β-oxidative pathway and benzyl alcohol O-benzoyltransferase HSR201 cooperatively contribute to the biosynthesis of salicylic acid. *Plant Cell Physiol.* 2023;64(7):758-770. doi 10.1093/pcp/pcad034

Lackman P., González-Guzmán M., Tilleman S., Carqueijeiro I., Pérez A.C., Moses T., Seo M., Kanno Y., Häkkinen S.T., Van Montagu M.C.E. Jasmonate signaling involves the abscisic acid receptor PYL4 to regulate metabolic reprogramming in *Arabidopsis* and tobacco. *Proc Natl Acad Sci USA.* 2011;108(14):5891-5896. doi 10.1073/pnas.1103010108

Liao Y., Smyth G.K., Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923-930. doi 10.1093/bioinformatics/btt656

Lie T.A. Host genes in *Pisum sativum* L. conferring resistance to European *Rhizobium leguminosarum* strains. *Plant Soil.* 1984;82(3):415-425. doi 10.1007/BF02184279

Liu C., Han X., Steenwyk J.L., Shen X.-X. Temporal transcriptomics provides insights into host-pathogen interactions: a case study of *Didymella pinodella* and disease-resistant and disease-susceptible pea varieties. *Crop Health.* 2023;1(1):5. doi 10.1007/s44297-023-00005-w

Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014; 15(12):550. doi 10.1186/s13059-014-0550-8

Mamontova T., Afonin A.M., Ihling C., Soboleva A., Lukasheva E., Sulima A.S., Shtark O.Y., Akhtemova G.A., Povydysh M.N., Sinz A., Frolov A., Zhukov V.A., Tikhonovich I.A. Profiling of seed proteome in pea (*Pisum sativum* L.) lines characterized with high and low responsivity to combined inoculation with nodule bacteria and arbuscular mycorrhizal fungi. *Molecules.* 2019;24(8):1603. doi 10.3390/molecules24081603

Martin D.N., Proebsting W.M., Hedden P. Mendel's dwarfing gene: cDNAs from the *Le* alleles and function of the expressed proteins. *Proc Natl Acad Sci USA.* 1997;94(16):8907-8911. doi 10.1073/pnas.94.16.8907

Mohanty S.K., Arthikala M.-K., Nanjareddy K., Lara M. Plant-symbiont interactions: the functional role of expansins. *Symbiosis*. 2018; 74:1-10. doi 10.1007/s13199-017-0501-8

Monte I. Jasmonates and salicylic acid: evolution of defense hormones in land plants. *Curr Opin Plant Biol.* 2023;76:102470. doi 10.1016/j.pbi.2023.102470

Müller L.M., Harrison M.J. Phytohormones, miRNAs, and peptide signals integrate plant phosphorus status with arbuscular mycorrhizal symbiosis. *Curr Opin Plant Biol.* 2019;50:132-139. doi 10.1016/j.pbi.2019.05.004

Okamoto S., Tabata R., Matsubayashi Y. Long-distance peptide signaling essential for nutrient homeostasis in plants. *Curr Opin Plant Biol.* 2016;34:35-40. doi 10.1016/j.pbi.2016.07.009

Oldroyd G.E.D. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat Rev Microbiol.* 2013;11(4):252-263. doi 10.1038/nrmicro2990

Parihar A.K., Kumar J., Gupta D.S., Lamichaney A., Naik S.J.S., Singh A.K., Dixit G.P., Gupta S., Toklu F. Genomics enabled breeding strategies for major biotic stresses in pea (*Pisum sativum* L.). *Front Plant Sci.* 2022;13:861191. doi 10.3389/fpls.2022.861191

Parniske M. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nat Rev Microbiol.* 2008;6(10):763-775. doi 10.1038/nrmicro1987

Pii Y., Astegno A., Peroni E., Zaccardelli M., Pandolfini T., Crimi M. The *Medicago truncatula N5* gene encoding a root-specific lipid transfer protein is required for the symbiotic interaction with *Sinorhizobium meliloti*. *Mol Plant Microbe Interact*. 2009;22(12): 1577-1587. doi 10.1094/MPMI-22-12-1577

Reid D.E., Ferguson B.J., Hayashi S., Lin Y.H., Gresshoff P.M. Molecular mechanisms controlling legume autoregulation of nodulation. *Ann Botany*. 2011;108(5):789-795. doi 10.1093/aob/mcr205

Roy S., Liu W., Nandety R.S., Crook A., Mysore K.S., Pislariu C.I., Frugoli J., Dickstein R., Udvardi M.K. Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *Plant Cell*. 2020;32(1):15-41. doi 10.1105/tpc.19.00279

Rubiales D., Barilli E., Rispail N. Breeding for biotic stress resistance in pea. *Agriculture*. 2023;13(9):1825. doi 10.3390/agriculture13091825

Rubiales D., Mikic A. Introduction: legumes in sustainable agriculture. *Crit Rev Plant Sci*. 2015;34(1-3):2-3. doi 10.1080/07352689.2014.897896

Shtark O.Y., Danilova T.N., Naumkina T.S., Vasilchikov A.G., Chebotar V.K., Kazakov A.E., Zhernakov A.I., Nemankin T.A., Prilepskaya N.A., Borisov A.U. Analysis of pea (*Pisum sativum* L.) source material for breeding of cultivars with high symbiotic potential and choice of criteria for its evaluation. *Ecol Genet*. 2006;4(2):22-28. doi 10.17816/ecogen4222-28

Shtark O.Y., Borisov A.Y., Zhukov V.A., Tikhonovich I.A. Mutually beneficial legume symbioses with soil microbes and their potential for plant production. *Symbiosis*. 2012;58(1-3):51-62. doi 10.1007/s13199-013-0226-2

Shtark O.Y., Zhukov V.A., Sulima A.S., Singh R., Naumkina T.S., Borisov A.Y. Prospects for the use of multi-component symbiotic systems of the Legumes. *Ecol Genet*. 2015;13(1):33-46. doi 10.17816/ecogen13133-46

Shtark O.Y., Puzanskiy R.K., Avdeeva G.S., Yurkov A.P., Smolikova G.N., Yemelyanov V.V., Kliukova M.S., Shavarda A.L., Kirpichnikova A.A., Zhernakov A.I. Metabolic alterations in pea leaves during arbuscular mycorrhiza development. *PeerJ*. 2019;7:e7495. doi 10.7717/peerj.7495

Smith S.E., Read D. The symbionts forming arbuscular mycorrhizas. In: Smith S.E., Read D. Mycorrhizal Symbiosis. Academic Press, 2008;13-41. doi 10.1016/b978-012370526-6.50003-9

Sulima A.S., Zhukov V.A., Afonin A.A., Zhernakov A.I., Tikhonovich I.A., Lutova L.A. Selection signatures in the first exon of paralogous receptor kinase genes from the Sym2 region of the *Pisum sativum* L. genome. *Front Plant Sci*. 2017;8:1957. doi 10.3389/fpls.2017.01957

Sulima A.S., Zhukov V.A., Kulaeva O.A., Vasileva E.N., Borisov A.Y., Tikhonovich I.A. New sources of Sym2^{A} allele in the pea (*Pisum sativum* L.) carry the unique variant of candidate LysM-RLK gene *LykX*. *PeerJ*. 2019;7:e8070. doi 10.7717/peerj.8070

Tsyganov V.E., Tsyganova A.V. Symbiotic regulatory genes controlling nodule development in *Pisum sativum* L. *Plants*. 2020;9(12):1741. doi 10.3390/plants9121741

Wang D., Dong W., Murray J., Wang E. Innovation and appropriation in mycorrhizal and rhizobial Symbioses. *Plant Cell*. 2022;34(5): 1573-1599. doi 10.1093/plcell/koac039

Wang L., Sun Z., Su C., Wang Y., Yan Q., Chen J., Ott T., Li X. A GmNINa-miR172c-NNC1 regulatory network coordinates the nodulation and autoregulation of nodulation pathways in soybean. *Mol Plant*. 2019;12(9):1211-1226. doi 10.1016/j.molp.2019.06.002

Wickham H. Getting Started with ggplot2. In: ggplot2. Use R! Springer, 2016;11-31. doi 10.1007/978-3-319-24277-4_2

Yang J., Lan L., Jin Y., Yu N., Wang D., Wang E. Mechanisms underlying legume–rhizobium symbioses. *J Int Plant Biol*. 2022;64(2): 244-267. doi 10.1111/jipb.13207

Zhukov V.A., Akhtemova G.A., Zhernakov A.I., Sulima A.S., Shtark O.Y., Tikhonovich I.A. Evaluation of the symbiotic effectiveness of pea (*Pisum sativum* L.) genotypes in pot experiment. *Agric Biol*. 2017;52(3):607-614. doi 10.15389/agrobiology.2017.3.607eng

Zhukov V.A., Zhernakov A.I., Sulima A.S., Kulaeva O.A., Kliukova M.S., Afonin A.M., Shtark O.Y., Tikhonovich I.A. Association study of symbiotic genes in pea (*Pisum sativum* L.) cultivars grown in symbiotic conditions. *Agronomy*. 2021a;11(11):2368. doi 10.3390/agronomy11112368

Zhukov V., Zorin E., Zhernakov A., Afonin A., Akhtemova G., Bovin A., Dolgikh A., Gorshkov A., Gribchenko E., Ivanova K., Kirienko A., Kitaeva A., Kliukova M., Kulaeva O., Kusakin P., Leppyanen I., Pavlova O., Romanyuk D., Rudaya E., Serova T., Shtark O., Sulima A., Tsyganova A., Vasileva E., Dolgikh E., Tsyganov V., Tikhonovich I. Transcriptomic analysis of *sym28* and *sym29* supernodulating mutants of pea (*Pisum sativum* L.) under complex inoculation with beneficial microorganisms. *Biol Commun.* 2021b; 66(3):181-197. doi 10.21638/spbu03.2021.301

Zorin E.A., Kliukova M.S., Afonin A.M., Gribchenko E.S., Gordon M.L., Sulima A.S., Zhernakov A.I., Kulaeva O.A., Romanyuk D.A., Kusakin P.G., Tsyganova A.V., Tsyganov V.E., Tikhonovich I.A., Zhukov V.A. A variable gene family encoding nodule-specific cysteine-rich peptides in pea (*Pisum sativum* L.). *Front Plant Sci*. 2022;13: 884726. doi 10.3389/fpls.2022.884726

Zorin E.A., Sulima A.S., Zhernakov A.I., Kuzmina D.O., Rakova V.A., Kliukova M.S., Romanyuk D.A., Kulaeva O.A., Akhtemova G.A., Shtark O.Y., Tikhonovich I.A., Zhukov V.A. Genomic and transcriptomic analysis of pea (*Pisum sativum* L.) breeding line 'Triumph' with high symbiotic responsivity. *Plants*. 2023;13(1):78. doi 10.3390/plants13010078

Conflict of interest. The authors declare no conflict of interest.

Received August 30, 2024. Revised November 28, 2024. Accepted December 16, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):259-267

doi 10.18699/vjgb-25-29

Correction of GenBank's taxonomic entry error raises a new issue regarding intergeneric relationships among salangid fishes (Osmeriformes: Salangidae)

E.S. Balakirev

A.V. Zhirmunsky National Scientific Center of Marine Biology, Far Eastern Branch of the Russian Academy of Sciences, Vladivostok, Russia
esbalakirev@mail.ru

Abstract. The GenBank database of publicly available nucleotide sequences is the largest genetic repository providing vitally important resources for downstream applications in biology and medicine. The concern raised about reliability of GenBank data necessitates monitoring of possible taxonomic entry errors. A case of mitochondrial genome (or mitogenome) misidentification for a salangid fish belonging to the genus *Neosalanx* (Osmeriformes, Salangidae) is considered in this report. The GenBank database contains four complete mitogenome sequences of *N. taihuensis* with the accession numbers JX524196, KP170510, MH348204, and MW291630. The overall mean *p*-distance for these sequences is quite high (7.01 ± 0.14 %) but becomes 29-fold lower (0.24 ± 0.05 %) after excluding the MW291630 mitogenome. An analysis of all available nucleotide sequences of salangids has shown that the observed inconsistency in the level of divergence between *N. taihuensis* mitogenomes is due to species misidentification. It has turned out that the mitogenome MW291630 available in GenBank does not belong to *N. taihuensis*, but is, in fact, a mitogenome of *N. jordani* misidentified as *N. taihuensis*. The resolved taxonomic identity of the MW291630 mitogenome, as well as an extended sample of species with investigated single-marker sequences, has raised some new issues regarding intergeneric relationships in salangid fishes. In particular, the obtained data do not support synonymization of the genus *Neosalanx* with *Protosalanx*, as was suggested in the last revision of the salangid classification. As the comparative analysis of interspecific and intergeneric divergences shows, *Protosalanx* is not an all-inclusive clade that includes all *Neosalanx* species. Instead, it consists of (at least) two evolutionary distinct lineages with the level of genetic divergence between them matching well the mean value of divergence between the other salangid genera. Further analysis using nuclear genome-wide data is required to have new insights into the evolution of salangid fishes.

Key words: *Neosalanx*; *Protosalanx*; taxonomic misidentification; mitochondrial genomes; *CytB*; single-marker sequences; genetic divergence

For citation: Balakirev E.S. Correction of GenBank's taxonomic entry error raises a new issue regarding intergeneric relationships among salangid fishes (Osmeriformes: Salangidae). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):259-267. doi 10.18699/vjgb-25-29

Funding. This study received budget support within the framework of the Research Study entitled "World Ocean biodiversity: taxonomy, barcoding, phylogenetics, reproductive and evolutionary biology, biogeography" (State registry no. 121082600036-9, Ministry of Science and Higher Education of the Russian Federation) conducted by A.V. Zhirmunsky National Scientific Center of Marine Biology, Far Eastern Branch, Russian Academy of Sciences, Vladivostok, Russia.

Acknowledgements. The author greatly appreciates the encouragement and help from Elena I. Balakireva, Alla G. Oleinik, and Larisa G. Erofeeva. Evgeny P. Shvetsov (Zhirmunsky National Scientific Center of Marine Biology, FEB, RAS, Vladivostok, Russia) kindly proofread the manuscript.

Исправление таксономической ошибки в базе данных GenBank поднимает новый вопрос относительно межродовых отношений у саланксовых рыб (Osmeriformes: Salangidae)

Е.С. Балакирев

Национальный научный центр морской биологии им. А.В. Жирмунского Дальневосточного отделения Российской академии наук, Владивосток, Россия
esbalakirev@mail.ru

Аннотация. Генетическая база данных GenBank является крупнейшим генетическим хранилищем, предоставляющим жизненно важные ресурсы для последующего применения в биологии и медицине. Высказанная обеспокоенность надежностью GenBank обусловливает необходимость мониторинга возможных таксономических ошибок в записях этой базы данных. Здесь мы сообщаем о случае ошибочной идентификации митохондриального генома (или митогенома) у саланксовой рыбы, принадлежащей к роду *Neosalanx* (Osmeriformes, Salangidae). База данных GenBank содержит четыре полные последовательности митогенома *N. taihuensis* с номерами доступа JX524196, KP170510, MH348204 и MW291630. Средняя *p*-дистанция между этими последовательностями довольно велика (7.01 ± 0.14 %), но становится в 29 раз меньше (0.24 ± 0.05 %) после исключения митогенома MW291630. Анализ всех доступных нуклеотидных последовательностей салангид показал, что наблюдаемое несоответствие в уровне дивергенции между митогеномами *N. taihuensis* обусловлено ошибочной идентификацией видов. Оказалось, что митогеном MW291630 не принадлежит *N. taihuensis*, а в действительности представляет митогеном *N. jordani*, ошибочно идентифицированный как *N. taihuensis*. Установленная таксономическая идентичность митогенома MW291630, а также расширенная выборка видов с исследованными маркерными последовательностями выявили некоторые новые аспекты межродовых отношений у саланксовых рыб. В частности, полученные данные не подтверждают синонимизацию рода *Neosalanx* с *Protosalanx*, как это было предложено в последней ревизии классификации салангид. Как показывает настоящий анализ, *Protosalanx* не является кладой, включающей все виды *Neosalanx*. Напротив, эта клада состоит по крайней мере из двух эволюционно разных линий, уровень генетической дивергенции между которыми соответствует межродовым значениям дивергенции у салангид. Необходим дальнейший анализ с использованием полных ядерных геномов для выяснения эволюции саланксовых рыб.
Ключевые слова: *Neosalanx*; *Protosalanx*; таксономические ошибки идентификации; митохондриальные геномы; *CytB*; маркерные последовательности; генетическая дивергенция

Introduction

The value and reliability of the GenBank database (Sayers et al., 2023) depends on the accuracy of species identification of biological samples, which is quite often not provided when based solely on morphology with an insufficient number of diagnostic characters. Species identification errors have been increasingly referred to as a serious challenge limiting the utility and reliability of public databases. In fact, for organisms such as fungi, which are notoriously difficult to distinguish, up to 20 % (Nilsson et al., 2006) or even 30 % (Hofstetter et al., 2019) of DNA sequence records in GenBank may have erroneous lineage designations. Multiple taxonomic misidentifications were reported for nuclear genome-sequenced strains of medically important lower eukaryotes (e. g., Houbraken et al., 2021), for single-marker sequences of many fishes (e. g., Li et al., 2018), and for complete mitogenomes of many higher eukaryotes, including bivalve mollusks (Salvi et al., 2021; Cunha et al., 2022), ticks (Mohamed et al., 2022), insects (Ožana et al., 2022; Kim et al., 2023), parasitic nematodes (Nielsen et al., 2014), fishes (Cheng et al., 2012; Balakirev et al., 2017, 2024; Oleinik et al., 2019; Sangster, Luksenburg, 2021a; Teske, 2021), amphibians (Mulder et al., 2016), reptiles (Simonov et al., 2018), birds (Sangster, Luksenburg, 2021b), and placental mammals (Botero-Castro et al., 2016).

A taxonomic misidentification causes discordance between the species name and the nucleotide sequence, thus, compromising downstream inferences. Consequently, it is urgently important to disclose such problematic sequences and report them as fast as possible after their deposition in GenBank in order to prevent propagation of incorrect biological information among databases and subsequent publications (e. g., Balakirev et al., 2017, 2024; Sangster, Luksenburg, 2021b).

Here, we report a case of mitochondrial genome misidentification for a salangid fish belonging to the genus *Neosalanx* Wakiya, Takahashi, 1937 (Osmeriformes, Salangidae). Salangids are endemic to East Asia and inhabit a wide range of marine, brackish-water, and freshwater habitats in China, Vietnam, Korean Peninsula, Japan, and Russia (e. g., Roberts, 1984). These are small, neotenic fishes with early maturation, relatively high fecundity, and a life span of about one year. Species identification of salangid fishes remains a serious challenge.

The taxonomy of salangids, based on morphological, ecological, and genetic approaches, has been subject to various revisions with multiple known synonyms (Fu et al., 2005, 2012; Zhang et al., 2007; Guo et al., 2011). In particular, it was shown that *N. taihuensis* Chen, 1956, *N. tangkahkeii* (Wu, 1931), and *N. pseudotaihuensis* Zhang, 1987 are junior synonyms of *N. brevirostris* (Pellegrin, 1923) (Zhang et al., 2007; Guo et al., 2011). *Hemisalanx* Regan, 1908 was shown to be a junior synonym of *Salanx* Cuvier, 1816 (Guo et al., 2011). Somewhat close genetic relationships were also found (Zhang et al., 2007) between *Protosalanx chinensis* (Basilewsky, 1855), *N. anderssoni* (Rendahl, 1923), *N. argentea* (Lin, 1932), and *N. tangkahkeii*. Based on the morphological characters, ecological preferences, and genetic data (mitochondrial *CytB* gene), Zhang et al. (2007) identified a group of species within the genus *Neosalanx*, including *N. reganius* Wakiya, Takahashi, 1937, *N. jordani* Wakiya, Takahashi, 1937, *N. oligodontis* Chen, 1956, and *Neosalanx* sp., which they proposed to treat as a separate new undescribed genus "*Microsalanx*". Zhang et al. (2007) and Guo et al. (2011) assumed that *N. anderssoni* may also belong to the genus *Protosalanx* Regan, 1908. Using extensive morphological analysis and also genetic markers such as mitochondrial (*CytB*) and seven nuclear genes, Fu et al. (2012) suggested that the genus *Neosalanx* should be considered a junior synonym of *Protosalanx*. These authors also found a distant relationship between *Salangichthys ishikawae* Wakiya, Takahashi, 1913 and *S. microdon* Bleeker, 1860, which proved that the two species belong to different genera: *Salangichthys* Bleeker,

1859 and the newly established *Neosalangichthys* Fu, Li, Xia, Lei, 2012 including a single species, *N. ishikawae*. Fu et al. (2012) found that the genera *Leucosoma* Gray, 1831 and *Salanx* differ significantly in genetic and morphological diagnostic characters and are, therefore, not synonymous.

Yang et al. (2020) deposited a complete mitogenome of the salangid *N. taihuensis* to GenBank under the accession no. MW291630 (taxonomy ID NCBI:txid240825), providing the forth mitogenome for this species in addition to the already available ones: JX524196, KP170510, and MH348204. An analysis of the new *N. taihuensis* MW291630 mitogenome in comparison with all other available mitogenome sequences, as well as the use of single-marker sequences of salangid fishes, has shown that this mitogenome sequence does not belong to *N. taihuensis*. We found that the specimen investigated by Yang et al. (2020) was erroneously identified as *N. taihuensis* and actually represents *N. jordani*. Therefore, the aim of the present study was to document this GenBank entry error and use the correctly identified MW291630 mitogenome, as well as an extended sample of single-marker sequences, to clarify some challenging issues regarding intergeneric relationships among salangid fishes.

Materials and methods

Mitochondrial genomes and single-marker sequences. A total of 13 complete mitogenome sequences from fishes of the family Salangidae Bleeker, 1859 were accessed from the Genetic Sequence Data Bank (the National Center for Biotechnology Information; NCBI-GenBank Flat File Release 260.0, April 15, 2024). The outgroup species, including *Plecoglossus altivelis* (Temminck, Schlegel, 1846) (family Plecoglossidae Bleeker, 1859) and *Retropinna retropinna* (Richardson, 1848) (family Retropinnidae Gill, 1862), were selected based on the previous molecular evidence of their close relationship to the family Salangidae (Fu et al., 2005; Zhang et al., 2007; Guo et al., 2011) and on a screening of nucleotide sequences available in GenBank using the basic local alignment search tool (BLAST) procedure (Altschul et al., 1990). Additionally, we also analyzed 406 mitochondrial single-marker sequences, including *12S rRNA*, *16S rRNA*, *ND1*, *COI*, and *CytB* published in previous studies on salangids (see Supplementary Table S1 for accession numbers and references)[1].

DNA sequence analysis. Previously, we described the DNA sequence analysis in detail elsewhere (e. g., Balakirev et al., 2017, 2020; Balakirev, 2022). The main steps are summarized in brief below. The nucleotide sequences were aligned using the software MUSCLE (Edgar, 2004). The programs DnaSP v. 6 (Rozas et al., 2017) and MEGA v. 11 (Tamura et al., 2021) were used for intra- and interspecific analysis of polymorphism and divergence based on uncorrected *p*-distance (Kartavtsev, 2011; Collins et al., 2012). Phylogenetic reconstructions were inferred from an analysis of complete mitogenomes by the maximum likelihood methods available in IQ-TREE v. 2 (Nguyen et al., 2015). The trees were constructed using complete mitogenomes or mitochondrial single-marker sequences only (*12S rRNA*, *16S rRNA*, *ND1*, *COI*, and *CytB*). For all reconstructions, the best-fit model of nucleotide substitution was chosen with the Akaike Information Criterion and the Bayesian Information Criterion in MEGA and IQ-TREE. The ultrafast maximum likelihood bootstrap analysis (Hoang et al., 2018) consisted of 10,000 replicates.

[1] Supplementary Table S1 and Figs. S1–S3 are available at: https://doi.org/10.5281/zenodo.13455533

Results and discussion

Variability and divergence of salangid mitogenomes

Figure 1 displays a maximum likelihood tree of complete mitogenome sequences for the salangid species including representatives of the valid genera *Salanx*, *Leucosoma*, *Salangichthys*, *Protosalanx*, and *Neosalanx*. The tree shows the *N. taihuensis* (with synonyms) specimens present in two significantly diverged clusters (Lineage 1 and Lineage 2; Fig. 1) with the overall mean distance equal to 7.01 ± 0.14 %. The *N. taihuensis* mitogenome sequences from Lineage 1 (JX524196, KP170510, and MH348204) were very similar to each other (with an average *p*-distance of 0.24 ± 0.05 %), thus, demonstrating a typical level of intraspecific nucleotide diversity in fishes (e. g., Kartavtsev et al., 2016; Li et al., 2018). Lineage 1, except for *N. taihuensis*, also included *P. chinensis* and *N. anderssoni*. The overall mean distance between the species from Lineage 1 (using a single randomly picked sequence per species) was 7.70 ± 0.17 % with pairwise *p*-distances varying from 4.82 ± 0.17 % between *P. chinensis* and *N. anderssoni* to 9.21 ± 0.22 % between *P. chinensis* and *N. taihuensis*, which matched well the known interspecific nucleotide diversity in fishes (e. g., Kartavtsev et al., 2016; Li et al., 2018). Lineage 2 (Fig. 1) included the *N. taihuensis* MW291630 mitogenome only, which demonstrated a high level of divergence (14.08 ± 0.27 %) with the representative sequence of the *N. taihuensis* mitogenome from Lineage 1. With the use of all mitogenomes for the species from Lineage 1 for comparison (*P. chinensis*, *N. anderssoni*, and *N. taihuensis*; Fig. 1), the difference between Lineages 1 and 2 still remained markedly higher (13.78 ± 0.24 %) than the overall mean distance (7.70 ± 0.17 %) estimated for Lineage 1.

We found the diagnostic 15-bp deletion that occurs within the *ND5* gene (at coordinates 79–93, Supplementary Fig. S2) and the 1-bp and 2-bp diagnostic deletions that occur within the non-coding (control) region (at coordinates 534, 963, 1051–1052, and 1071; Supplementary Fig. S3). These are shared by the *P. chinensis*, *N. anderssoni*, and *N. taihuensis* mitogenomes (Lineage 1) and distinguish them clearly from the *N. taihuensis* MW291630 (Lineage 2) and the rest of the salangid mitogenomes. The 15-bp deletion within the *ND5* gene is the only sequence length variability detected for the protein-coding genes in the mitogenomes of salangid fishes. Taking into account the high phylogenetic informativeness of gaps (e. g., Giribet, Wheeler, 1999), these diagnostic deletions provide robust evidence for the close relationships of the species belonging to Lineage 1 and their distinct difference from Lineage 2.

To scale the value of full mitogenome divergence between Lineages 1 and 2, we estimated the average level of divergence based on the representative genera including *Protosalanx*, *Salanx*, *Leucosoma*, and *Salangichthys*. To be conservative, we excluded *N. anderssoni* and *N. taihuensis* (with synonyms) in order to prevent underestimation of *p*-distance values due

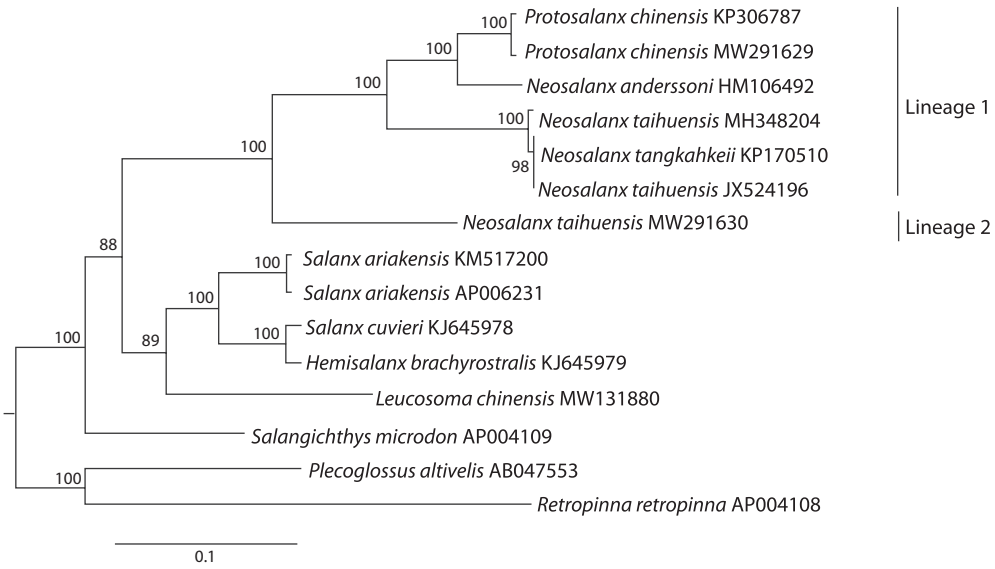


Fig. 1. Maximum likelihood tree inferred from an analysis of the complete mitochondrial genomes for fishes of the family Salangidae.

The TIM2+F+I+G4 model was used to construct the tree. The numerals at the nodes are bootstrap probability (percentage) values based on 10,000 replicates (values lower than 75 % are omitted). The tree includes all salangid mitogenomes available in GenBank except the three recombinant sequences of *Protosalanx chinensis* under the accession nos. HM106494, MH330683, and KJ499917 (Balakirev, 2022). The synonymous species names *N. taihuensis* or *N. tangkahkeii* were used for the originally published KP170510, MW291630, JX524196, and MH348204 mitogenomes. To avoid any confusion, we leave the names as they were originally assigned for the salangid species considered in this paper.

Table 1. Pairwise *p*-distances between salangid genera based on complete mitogenomes

Genus	*Salanx*	*Leucosoma*	*Salangichthys*	*Protosalanx*
Salanx		0.0026	0.0026	0.0028
Leucosoma	0.1251		0.0027	0.0029
Salangichthys	0.1342	0.1461		0.0029
Protosalanx	0.1507	0.1704	0.1658	

Note. The salangid genera *Salanx, Leucosoma*, *Salangichthys*, and *Protosalanx* are represented by the following species: *Salanx ariakensis* Kishinouye, 1902 (AP006231), *Leucosoma chinensis* (Osbeck, 1765) (MW131880), *Salangichthys microdon* (AP004109), and *Protosalanx chinensis* (KP306787). The *p*-distances are below the diagonal line. The standard errors, obtained with 10,000 bootstrap replications, are above the diagonal line.

to possible congeneric relationships of these species (Zhang et al., 2007; Guo et al., 2011). We also excluded the MW291630 mitogenome sequence with uncertain identity. The obtained overall mean *p*-distance for all available genera of salangid fishes was 14.87 ± 0.21 % with pairwise *p*-distances varying from 12.51 ± 0.26 % between *Leucosoma* and *Salanx* to 17.04 ± 0.29 % between *Leucosoma* and *Protosalanx* (Table 1), which was close to the value of divergence between Lineages 1 and 2 (13.78 ± 0.24 %).

It is worth noting that in pairwise comparisons the divergence between Lineages 1 and 2 (13.78 ± 0.24 %) was not markedly different from the divergence between *Leucosoma* and *Salangichthys*, or was even slightly higher than the divergence between *Leucosoma* and *Salanx*, as well as between *Salangichthys* and *Salanx* (Table 1). Thus, the pairwise comparisons showed that the mitogenome divergence between Lineage 1 and Lineage 2 matched well the intergeneric level of divergence in salangid fishes. The interlineage distance matched also the average value of divergence between different genera reported for the single-marker sequences or complete mitogenomes in other groups of fishes (e. g., Kartavtsev et al., 2016; Li et al., 2018; Balakirev et al., 2020).

Identification of the MW291630 mitogenome

According to a taxonomic hypothesis based on genetic data, *N. taihuensis*, *N. tangkahkeii*, *N. pseudotaihuensis*, and *N. brevirostris* are synonyms (Zhang et al., 2007; Guo et al., 2011). Consequently, the genus *Neosalanx* is represented in GenBank by only two species, *N. taihuensis* (with synonyms) and *N. anderssoni* (HM106492; Supplementary Table S1), which makes the identification of the problematic complete mitogenome MW291630 impossible. However, the GenBank database contains at least five more *Neosalanx* species, representing the full taxonomic diversity known for the genus *Neosalanx*, that were investigated using mitochondrial single-marker sequences: *N. argentea*, *N. jordani*, *N. oligodontis*,

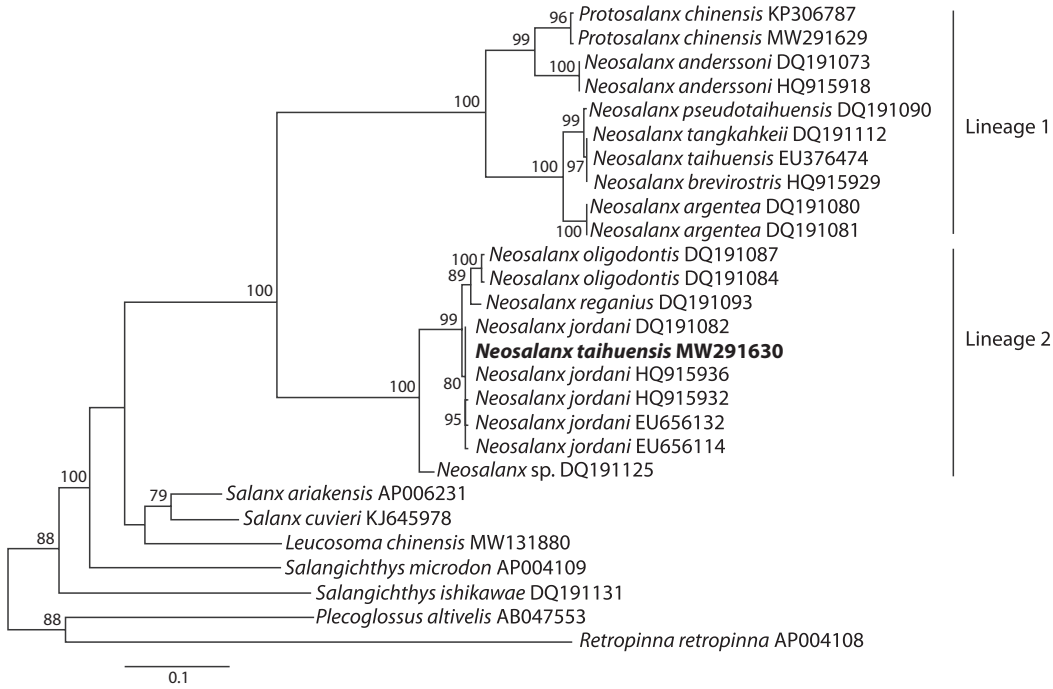


Fig. 2. Maximum likelihood tree for the members of the family Salangidae based on the *CytB* gene sequences.

The Tamura-Nei + gamma (TN93+G) model was used to infer the tree. The *N. jordani CytB* sequences are represented by three datasets investigated by Fu et al. (2012) (HQ915932 and HQ915936), Zhang et al. (2007) (DQ191082), and Zhao et al. (2010) (EU656114 and EU656132). The *N. taihuensis* MW291630 mitogenome is highlighted in bold. For tree reconstruction, we used only some representative samples from larger datasets (a full list of the *CytB* sequences is provided in Supplementary Table S1). For other comments, see Figure 1.

N. reganius, and *Neosalanx* sp. (the names of the species are listed as they were identified by the authors who submitted the respective nucleotide sequences to GenBank). The nucleotide sequences obtained for these species can be used to resolve the observed inconsistency detected for the *N. taihuensis* complete mitogenomes and to identify the taxonomically problematic MW291630 mitogenome.

We analyzed the GenBank mitochondrial single-marker sequences that are most frequently used in taxonomic and phylogenetic reconstructions of salangid fishes, including *12S rRNA*, *16S rRNA*, *ND1*, *COI*, and *CytB*. A preliminary analysis revealed that among the single-marker sequences, only *CytB* demonstrated noticeable divergence values. The other markers provided much lower resolution but were still not contradictory to the *CytB* data (see, e. g., the maximum likelihood tree based on the *COI* gene; Supplementary Fig. S1). Consequently, further analysis was based on the *CytB* gene only.

Figure 2 illustrates the maximum likelihood tree based on the *CytB* gene for *N. taihuensis* and other members of the family Salangidae representing almost the entire taxonomic diversity of the genus *Neosalanx*. There were two significantly different clusters that included the species name *N. taihuensis*. These clusters corresponded to Lineages 1 and 2 identified on the basis of mitogenome sequences (Fig. 1). The overall mean *p*-distance for Lineage 1 was 9.13 ± 0.66 % using a single randomly picked sequence per species (with pairwise *p*-distances varying from 3.33 ± 0.54 % between *N. taihuensis* and *N. argentea* to 12.09 ± 0.99 % between *N. taihuensis* and *N. anderssoni*). Lineage 1 included *P. chinensis* and part of the *Neosalanx* species (*N. anderssoni*, *N. taihuensis*, and *N. argentea*) that Fu et al. (2012) had combined with other *Neosalanx* species and synonymized with the genus *Protosalanx* (see Introduction). The second cluster contained *N. oligodontis*, *N. reganius*, *N. jordani*, *Neosalanx* sp., and the *CytB* portion of the MW291630 mitogenome with an overall mean *p*-distance of 4.69 ± 0.45 % and pairwise *p*-distances varying from 2.10 ± 0.43 % between *N. oligodontis* and *N. reganius* to 7.01 ± 0.76 % between *N. reganius* and *Neosalanx* sp. This cluster included a group of species that were placed in the genus "*Microsalanx*" erected by Zhang et al. (2007). The mean *p*-distance between the clusters (Lineage 1 and Lineage 2, Fig. 2) was 19.03 ± 1.06 % with a single randomly picked sequence per species or 18.97 ± 1.09 % with all 146 sequences available for Lineages 1 and 2 (Supplementary Table S1). This value fit well into the range of intergeneric divergences of fishes (e. g., Kartavtsev et al., 2016; Li et al., 2018; Balakirev et al., 2020).

An analysis of pairwise *p*-distances showed a surprisingly high level of sequence divergence (18.89 ± 1.16 %) (Table 2) between the GenBank *CytB* sequences of *N. taihuensis*, including 70 isolates obtained from different localities by various authors (Zhang et al., 2007; Zhao et al., 2008; see also Supplementary Table S1) with low intraspecific variability (0.44 ± 0.09 %), and the respective *CytB* portion of the MW291630 mitogenome. The result was consistent with the data obtained for the complete mitogenomes (see above), which showed a substantial difference between the MW291630

Table 2. Pairwise *p*-distances between the *CytB* sequences for *Neosalanx* species

	1	2	3	4	5	6	7	8
1		0.0003	0.0041	0.0044	0.0075	0.0116	0.0119	0.0120
2	0.0012		0.0040	0.0044	0.0075	0.0115	0.0119	0.0120
3	0.0189	0.0196		0.0042	0.0078	0.0114	0.0117	0.0119
4	0.0219	0.0229	0.0198		0.0077	0.0115	0.0117	0.0120
5	0.0657	0.0666	0.0686	0.0701		0.0112	0.0114	0.0115
6	0.1889	0.1890	0.1837	0.1853	0.1848		0.0052	0.0099
7	0.1906	0.1909	0.1855	0.1836	0.1862	0.0332		0.0097
8	0.2004	0.2003	0.1964	0.1986	0.1841	0.1215	0.1172	

Note. All available *CytB* sequences for each included species were used for this analysis (the number of sequences is in parentheses). 1: MW291630 (1); 2: *N. jordani* (25); 3: *N. oligodontis* (7); 4: *N. reganius* (1); 5: *Neosalanx* sp. (1); 6: *N. taihuensis* (70); 7: *N. argentea* (2); and 8: *N. anderssoni* (10). For other comments, see Table 1 and Figure 2.

mitogenome and the other *N. taihuensis* (with synonyms) mitogenomes, JX524196, KP170510, and MH348204, available in GenBank.

The *CytB* portion of the *N. taihuensis* MW291630 mitogenome demonstrated very close affinity to the *N. jordani* single-marker sequences obtained from different localities by various authors (25 isolates; Zhang et al., 2007; Zhao et al., 2010; Fu et al., 2012) with low intraspecific variability (0.24 ± 0.05 %; see also Fig. 2). The *p*-distance between the *CytB* gene of the MW291630 mitogenome and the GenBank *CytB* sequences of *N. jordani* was surprisingly low (0.12 ± 0.03 %; Table 2); it fit well into the range of intraspecific variability in fishes (e. g., Kartavtsev et al., 2016; Li et al., 2018). The maximum likelihood tree based on the *COI* gene showed a similar topology (Supplementary Fig. S1).

Three species, *N. oligodontis*, *N. reganius*, and *Neosalanx* sp., demonstrated more pronounced differences from the MW291630 mitogenome than *N. jordani* (Table 2, Fig. 2). Zhang et al. (2007) suggested that *N. reganius* and *N. oligodontis* could be considered as subspecies of *N. jordani*. Indeed, *N. jordani*, *N. oligodontis*, *N. reganius*, *Neosalanx* sp., and the *CytB* portion of the MW291630 mitogenome were all in the same cluster (Fig. 2) with an overall mean *p*-distance of 4.69 ± 0.45 %, which suggested close relationships for these salangid species.

Thus, the single-marker sequences clearly showed that the GenBank accession no. MW291630 represents the mitogenome of *N. jordani* mistaken for *N. taihuensis*. The observed inconsistency in the level of divergence between the *N. taihuensis* mitogenomes (see above) is due to incorrect species identification. The *CytB* analysis of within- and between lineage variability confirmed the data obtained with complete mitogenomes.

MtDNA data indicates a generic level of divergence between Lineage 1 and Lineage 2

The close relationships between *N. jordani*, *N. oligodontis*, *N. reganius*, and *Neosalanx* sp. had been reported previously, as well as the significant difference of this group from other *Neosalanx* and *Protosalanx* species including *N. taihuensis*, *N. argentea*, *N. anderssoni*, and *P. chinensis* (Fig. 2, Table 2) (Zhang et al., 2007; Guo et al., 2011). Based on integrative data, Zhang et al. (2007) erected the genus "*Microsalanx*" (see Introduction). The results of the present analysis do not contradict this hypothesis. Both the complete mitogenomes (Fig. 1) and the single-marker sequences (Fig. 2) clearly demonstrated two significantly diverged clusters (Lineage 1 and Lineage 2). The interlineage divergence based on the *CytB* gene was high (18.97 ± 1.09 %). It was significantly higher than the average divergences within each of the lineages: the overall mean distances for Lineage 1 and Lineage 2 separately were 9.13 ± 0.66 and 4.69 ± 0.45 %, respectively.

To evaluate the scale of divergence between Lineages 1 and 2, we estimated the average level of divergence based on the *CytB* gene for the salangid genera available in GenBank including *Protosalanx*, *Salanx*, *Leucosoma*, *Neosalangichthys*,

Table 3. Pairwise *p*-distances between salangid genera based on the *CytB* gene

Genus	*Protosalanx*	*Salanx*	*Leucosoma*	*Neosalangichthys*	*Salangichthys*
Protosalanx		0.0118	0.0123	0.0124	0.0125
Salanx	0.1876		0.0093	0.0106	0.0108
Leucosoma	0.2068	0.1174		0.0103	0.0113
Neosalangichthys	0.2077	0.1472	0.1411		0.0109
Salangichthys	0.2147	0.1490	0.1648	0.1613	

Note. The salangid genera *Protosalanx*, *Salanx*, *Leucosoma*, *Neosalangichthys*, and *Salangichthys* are represented by the following species: *Protosalanx chinensis* (KP306787), *Salanx ariakensis* (AP006231), *Leucosoma chinensis* (MW131880), *Neosalangichthys ishikawae* (Wakiya, Takahashi, 1913) (DQ191127), and *Salangichthys microdon* (AP004109). For other comments, see Table 1.

and *Salangichthys* (Table 3). The obtained overall mean *p*-distance was 16.98 ± 0.78 % (with pairwise *p*-distances varying from 11.74 ± 0.93 % between *Leucosoma* and *Salanx* to 21.47 ± 1.25 % between and *Protosalanx* and *Salangichthys*; Table 3), which was close to the value of divergence between Lineages 1 and 2 (18.97 ± 1.09 %) based on the multiple *CytB* gene sequences (see above). The divergence between Lineages 1 and 2 (18.97 ± 1.09 %) was not markedly different from that between *Protosalanx* and *Salanx* or it was even higher than the *p*-distances in pairwise comparisons of *Leucosoma* vs. *Salanx*, *Neosalangihthys*, and *Salangichthys*; *Salanx* vs. *Neosalangihthys* and *Salangichthys*; and *Neosalangihthys* vs. *Salangichthys* (Table 3).

Thus, an analysis of the multiple *CytB* sequences encompassing the full diversity of salangid fishes showed a high level of divergence between Lineage 1 and Lineage 2 (18.97 ± 1.09 %), which fit into the range of intergeneric distances reported for salangids (see above) and other fishes (see the references above). Lineage 1 included a group of species (*P. chinensis*, *N. anderssoni*, *N. taihuensis*, and *N. argentea*; Fig. 2) comprising a part of the reorganized genus *Protosalanx* (Fu et al., 2012). The group of species from Lineage 1 was previously divided in two sub-groups ("primitive lineages") (Zhang et al., 2007; Guo et al., 2011). Indeed, the pairwise *p*-distances for Lineage 1 varied within a relatively wide range from 3.32 ± 0.52 % between *N. taihuensis* and *N. argentea* to 12.15 ± 0.99 % between *N. taihuensis* and *N. anderssoni* (Table 2). However, the overall mean sub-group divergence (*P. chinensis* + *N. anderssoni* vs. *N. taihuensis* + *N. argentea*) within Lineage 1 was still markedly lower (11.42 ± 0.87 %) than the divergence between Lineages 1 and 2 (18.97 ± 1.09 %). Thus, unlike Zhang et al. (2007) and Guo et al. (2011), we did not find sufficient grounds to split Lineage 1 into two sub-groups and consider it a single evolutionary lineage representing the genus *Protosalanx*. This conclusion was supported by the diagnostic deletions detected within the *ND5* gene and the control region in the salangids' mitogenomes (see the "Variability and divergence of salangid mitogenomes" section above). Nevertheless, the relationships between the "primitive lineages" *P. chinensis* + *N. anderssoni* and *N. taihuensis* + *N. argentea* need to be further clarified using a more representative array of genetic markers (see below).

Lineage 2 contained a group of species (*N. oligodontis*, *N. reganius*, *N. jordani*, and *Neosalanx* sp.) placed in the genus "*Microsalanx*" by Zhang et al. (2007). This subdivision was reasonable (see Introduction) to distinguish this group of species from the rest of the *Neosalanx* species. However, the transfer of *N. taihuensis* (with synonyms), *N. anderssoni*, and *N. argentea* to the genus *Protosalanx*, as suggested earlier (Zhang et al., 2007; Guo et al., 2011; Fu et al., 2012) and supported by our data (Figs. 1 and 2, Tables 1 and 2), gives reason to abolish the genus name "*Microsalanx*" (at least until the generic heterogeneity is proven for Lineage 1; see above). Consequently, the original genus name *Neosalanx* is appropriate for the salangid species *N. oligodontis*, *N. reganius*, *N. jordani*, and *Neosalanx* sp. comprising Lineage 2 (Figs. 1 and 2).

Thus, in contrast to Fu et al. (2012), our analysis based on complete mitogenomes and mtDNA single-marker sequences, as well as the analysis of Zhang et al. (2007) based on morphological, ecological, and genetic data, did not support the synonymization of all *Neosalanx* species with *Protosalanx*. The data clearly show two substantially diverged evolutionary lineages (Figs. 1 and 2): (1) *P. chinensis*, *N. anderssoni*, *N. taihuensis* (with synonyms), and *N. argentea* representing the genus *Protosalanx* and (2) *N. oligodontis*, *N. reganius*, *N. jordani*, and *Neosalanx* sp. representing the genus *Neosalanx*.

For phylogenetic analysis of salangid fishes, Fu et al. (2012) used a concatenated multigene dataset including the mitochondrial *CytB* gene and seven nuclear sequences (*28S rRNA*, *RAG1*, *zic1*, *ENC1*, *RNF213*, *glyt*, and *SH3PX3*). As a result (among others), these authors (Fu et al., 2012, p. 853) discovered that "all species from the '*Neosalanx–Protosalanx*' complex belong to a same genus" and considered *Neosalanx* as a junior synonym of *Protosalanx*.

Compared to mtDNA markers and complete mitogenomes, the nuclear markers (*28S rRNA*, *RAG1*, *zic1*, *ENC1*, *RNF213*, *glyt*, and *SH3PX3*), mostly used by Fu et al. (2012), demonstrated a much lower divergence between the salangid genera. For the genera *Protosalanx* (except *Neosalanx*), *Salanx*, *Leucosoma*, *Neosalangichthys*, and *Salangichthys*, the values of the overall mean distance for the nuclear markers were low and varied in a narrow range (from 1.98 ± 0.36 % for *zic1* to 3.56 ± 0.54 % for *RAG1*). The low divergence of the nuclear markers can be explained by the fact that they mostly represent highly conserved sequences developed for analyzing deep phylogenetic relationships on a scale of dozens to hundreds of millions of years, e. g., to infer phylogenetic relationships of all bony fishes, which requires analysis of genomic regions with slow rates of evolution (e. g., Betancur-R et al., 2017). These markers might be not sensitive enough for salangid fishes that experienced most speciation events around 1.1–9.9 Ma (Zhang et al., 2007). As a consequence, we suggest that the phylogenetic signal of *CytB*, also used by Fu et al. (2012), was significantly "diluted" by the effect of strongly conserved nuclear sequences. Indeed, the overall mean *p*-distance between the genera *Protosalanx* (except *Neosalanx*), *Salanx*, *Leucosoma*, *Neosalangichthys*, and *Salangichthys* was equal to 16.98 ± 0.78 % based on the *CytB* gene only (see above). However, it decreased significantly, to 2.72 ± 0.17 %, when the nuclear multigene dataset of Fu et al. (2012) was used.

Although the suggested relationships in salangid fishes seem robust, we expect them to be modified, possibly, as new genetic data become available. In particular, the mitochondrial sequences have revealed a relatively high level of divergence between two sub-groups within the genus *Protosalanx* (*P. chinensis* + *N. anderssoni* and *N. taihuensis* + *N. argentea*; Fig. 2) (see also Zhang et al., 2007; Guo et al., 2011), which may indicate a supra-species taxonomical range. Consequently, more nuclear genome-wide data are necessary to further address this and other issues concerning the taxonomic composition and the evolutionary relationships among salangid fishes.

Conclusion

Misidentified nucleotide sequences, including complete mitogenomes, are becoming increasingly frequent in GenBank, which leads to an explosive spread of incorrect biological information in subsequent scientific publications over time.

The misidentified *N. taihuensis* MW291630 mitogenome has been revealed in our study. We argue that the GenBank accession no. MW291630 actually represents the mitogenome of *N. jordani* mistaken for *N. taihuensis*. Thus, GenBank users should be aware of the above-described entry error to avoid conflicting results in their downstream evolutionary and comparative genomic studies.

The data obtained have raised a new issue regarding intergeneric relationships among salangid fishes. In contrast to the study by Fu et al. (2012), our data from the comparative analyses of interspecific and intergeneric divergences do not support the synonymization of the genus *Neosalanx* with *Protosalanx* and oppose the suggestion to consider *Neosalanx* as a junior synonym of *Protosalanx*. Genome-wide studies are needed to further clarify the evolutionary relationships of salangid fishes.

The introduction and spread of misidentified nucleotide sequences in genetic databases, which compromises downstream applications, is unlikely to be completely curbed. However, some appropriate steps can be undertaken (see, e. g., Balakirev et al., 2017, 2024; Sangster, Luksenburg, 2021b) to minimize their massive accumulation and subsequent propagation in scientific publications, thereby increasing the reliability of findings reported in them.

References

Altschul S., Gish W., Miller W., Myers E., Lipman D. Basic local alignment search tool. *J Mol Biol.* 1990;215:403-410. doi 10.1016/S0022-2836(05)80360-2

Balakirev E.S. Recombinant mitochondrial genomes reveal recent interspecific hybridization between invasive salangid fishes. *Life.* 2022;5:661. doi 10.3390/life12050661

Balakirev E.S., Saveliev P.A., Ayala F.J. Complete mitochondrial genomes of the Cherskii's sculpin *Cottus czerskii* and Siberian taimen *Hucho taimen* reveal GenBank entry errors: incorrect species identification and recombinant mitochondrial genome. *Evol Bioinform Online*. 2017;13:1176934317726783. doi 10.1177/1176934317726783

Balakirev E.S., Kravchenko A.Y., Semenchenko A.A. Genetic evidence for a mixed composition of the genus *Myoxocephalus* (Cottoidei: Cottidae) necessitates generic realignment. *Genes*. 2020;11:1071. doi 10.3390/genes11091071

Balakirev E.S., Sharina S.N., Balanov A.A. Misidentified mitogenomes of two *Lycodes* species (Perciformes: Zoarcidae) in GenBank. *Russ J Genet*. 2024;60(10):1375-1382. doi 10.1134/S1022795424700911

Betancur-R R., Wiley E.O., Arratia G., Acero A., Bailly N., Miya M., Lecointre G., Ortí G. Phylogenetic classification of bony fishes. *BMC Evol Biol.* 2017;17:162. doi 10.1186/s12862-017-0958-3

Botero-Castro F., Delsuc F., Douzery E.J.P. Thrice better than once: quality control guidelines to validate new mitogenomes. *Mitochondrial DNA A DNA Mapp Seq Anal.* 2016;27(1):449-454. doi 10.3109/19401736.2014.900666

Cheng J., Ma G.-Q., Song N., Gao T.-X. Complete mitochondrial genome sequence of bighead croaker *Collichthys niveatus* (Perciformes, Sciaenidae): a mitogenomic perspective on the phylogenetic relationships of Pseudosciaeniae. *Gene*. 2012;491(2):210-223. doi 10.1016/j.gene.2011.09.020

Collins R.A., Boykin L.M., Cruickshank R.H., Armstrong K.F. Barcoding's next top model: an evaluation of nucleotide substitution models for specimen identification. *Methods Ecol Evol*. 2012;3:457-465. doi 10.1111/j.2041-210x.2011.00176.x

Cunha R.L., Nicastro K.R., Zardi G.I., Madeira C., McQuaid C.D., Cox C.J., Castilho R. Comparative mitogenomic analyses and gene rearrangements reject the alleged polyphyly of a bivalve genus. *PeerJ*. 2022;10:e13953. doi 10.7717/peerj.13953

Edgar R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792-1797. doi 10.1093/nar/gkh340

Fu C., Luo J., Wu J., López J.A., Zhong Y., Lei G., Chen J. Phylogenetic relationships of salangid fishes (Osmeridae, Salanginae) with comments on phylogenetic placement of the salangids based on mitochondrial DNA sequences. *Mol Phylogenet Evol*. 2005;35:76-84. doi 10.1016/j.ympev.2004.11.024

Fu C., Guo L., Xia R., Li J., Lei G. A multilocus phylogeny of Asian noodlefishes Salangidae (Teleostei: Osmeriformes) with a revised classification of the family. *Mol Phylogenet Evol*. 2012;62(3):848-855. doi 10.1016/j.ympev.2011.11.031

Giribet G., Wheeler W.C. On gaps. *Mol Phylogenet Evol.* 1999;13(1):132-143. doi 10.1006/mpev.1999.0643

Guo L., Li J., Wang Z., Fu C. Phylogenetic relationships of noodlefishes (Osmeriformes: Salangidae) based on four mitochondrial genes. *Acta Hydrobiol*. 2011;35:449-459. doi 10.3724/SP.J.1035.2011.00449

Hoang D.T., Chernomor O., von Haeseler A., Minh B.Q., Vinh L.S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018;35:518-522. doi 10.1093/molbev/msx281

Hofstetter V., Buyck B., Eyssartier G., Schnee S., Gindro K. The unbearable lightness of sequenced-based identification. *Fungal Divers*. 2019;96:243-284. doi 10.1007/s13225-019-00428-3

Houbraken J., Visagie C.M., Frisvad J.C. Recommendations to prevent taxonomic misidentification of genome-sequenced fungal strains. *Microbiol Resour Ann.* 2021;10:e01074-20. doi 10.1128/MRA.01074-20

Kartavtsev Y.P. Sequence divergence at mitochondrial genes in animals: applicability of DNA data in genetics of speciation and molecular phylogenetics. *Mar Genomics.* 2011;4(2):71-81. doi 10.1016/j.margen.2011.02.002

Kartavtsev Y.P., Rozhkovan K.V., Masalkova N.A. Phylogeny based on two mtDNA genes (*Co-1, Cyt-B*) among sculpins (Scorpaeniformes, Cottidae) and some other scorpionfish in the Russian Far East. *Mitochondrial DNA A DNA Mapp Seq Anal.* 2016;27(3):2225-2240. doi 10.3109/19401736.2014.984164

Kim D.E., Kim P., Lee H., Kim N.H., Kim D., Lee M.J., Ban Y.G., Jang B., Park J. Comprehensive analysis of the complete mitochondrial genome of *Melanoplus differentialis* (Acrididae: Melanoplinae) captured in Korea. *Entomol Res*. 2023;53:66-81. doi 10.1111/1748-5967.12633

Li X., Shen X., Chen X., Xiang D., Murphy R.W., Shen Y. Detection of potential problematic *Cytb* gene sequences of fishes in GenBank. *Front Genet*. 2018;9:30. doi 10.3389/fgene.2018.00030

Mohamed W.M.A., Moustafa M.A.M., Kelava S., Barker D., Matsuno K., Nonaka N., Shao R., Mans B.J., Barker S.C., Nakao R. Reconstruction of mitochondrial genomes from raw sequencing data provides insights on the phylogeny of *Ixodes* ticks and cautions for species misidentification. *Ticks Tick Borne Dis*. 2022;13(1):101832. doi 10.1016/j.ttbdis.2021.101832

Mulder K.P., Lourenço A., Carneiro M., Velo-Antón G. The complete mitochondrial genome of *Salamandra salamandra* (Amphibia: Urodela: Salamandridae). *Mitochondrial DNA Part B.* 2016;1:880-882. doi 10.1080/23802359.2016.1253042

Nguyen L.T., Schmidt H.A., von Haeseler A., Minh B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol Biol Evol*. 2015;32:268-274. doi 10.1093/molbev/msu300

Nielsen M.K., Wang J., Davis R., Bellaw J.L., Lyons E.T., Lear T.L., Goday C. *Parascaris univalens* – a victim of large-scale misidentification? *Parasitol Res*. 2014;113:4485-4490. doi 10.1007/s00436-014-4135-y

Nilsson R.H., Ryberg M., Kristiansson E., Abarenkov K., Larsson K.H., Koljalg U. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. *PLoS One*. 2006;1(1):e59. doi 10.1371/journal.pone.0000059

Oleinik A.G., Skurikhina L.A., Kukhlevsky A.D. Clarification of taxonomic assignment of smelt complete mitochondrial genome:

GenBank accession number KP281293.1 (NC_026566.1). *Mitochondrial DNA Part B*. 2019;4:1696-1697. doi 10.1080/23802359.2019.1607578

Ožana S., Dolný A., Pánek T. Nuclear copies of mitochondrial DNA as a potential problem for phylogenetic and population genetic studies of Odonata. *Syst Entomol*. 2022;47:591-602. doi 10.1111/syen.12550

Roberts T.R. Skeletal anatomy and classification of the neotenic Asian Salmoniform superfamily Salangoidea (icefishes or noodlefishes). *Proc Califor Acad Sci*. 1984;43:179-220

Rozas J., Ferrer-Mata A., Sánchez-DelBarrio J.C., Guirao-Rico S., Librado P., Ramos-Onsins S.E., Sánchez-Gracia A. DnaSP 6: DNA Sequence Polymorphism analysis of large datasets. *Mol Biol Evol*. 2017;34:3299-3302. doi 10.1093/molbev/msx248

Salvi D., Berrilli E., Garzia M., Mariottini P. Yet another mitochondrial genome of the Pacific cupped oyster: the published mitogenome of *Alectryonella plicatula* (Ostreinae) is based on a misidentified *Magallana gigas* (Crassostreinae). *Front Mar Sci*. 2021;8:741455. doi 10.3389/fmars.2021.741455

Sangster G., Luksenburg J.A. The published complete mitochondrial genome of the milk shark (*Rhizoprionodon acutus*) is a misidentified Pacific spadenose shark (*Scoliodon macrorhynchos*) (Chondrichthyes: Carcharhiniformes). *Mitochondrial DNA Part B*. 2021a;6:828-830. doi 10.1080/23802359.2021.1884019

Sangster G., Luksenburg J.A. Sharp increase of problematic mitogenomes of birds: causes, consequences, and remedies. *Genome Biol Evol*. 2021b;13:evab210. doi 10.1093/gbe/evab210

Sayers E.W., Cavanaugh M., Clark K., Pruitt K.D., Sherry S.T., Yankie L., Karsch-Mizrachi I. GenBank 2023 update. *Nucleic Acids Res*. 2023;51:D141-D144. doi 10.1093/nar/gkac1012

Simonov E., Lisachov A., Oreshkova N., Krutovsky K.V. The mitogenome of *Elaphe bimaculata* (Reptilia: Colubridae) has never been published: a case with the complete mitochondrial genome of *E. dione*. *Acta Herpetol*. 2018;13:185-189. doi 10.13128/Acta_Herpetol-23394

Tamura K., Stecher G., Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Mol Biol Evol*. 2021;38:3022-3027. doi 10.1093/molbev/msab120

Teske P.R. Mitochondrial genome announcements need to consider existing short sequences from closely related species to prevent taxonomic errors. *Conserv Genet Resour*. 2021;13:359-365. doi 10.1007/s12686-021-01214-7

The National Center for Biotechnology Information. Available online: https://www.ncbi.nlm.nih.gov/ (accessed on July 29, 2024)

Yang Y., Sui Z., Liu K., Liu Y. The complete mitochondrial DNA sequence of Linyi small icefish (*Neosalanx taihuensis*). GenBank submission: 24-NOV-2020. Genbank accession number: MW291630

Zhang J., Li M., Xu M., Takita T., Wei F. Molecular phylogeny of icefish Salangidae based on complete mtDNA *cytochrome b* sequences, with comments on estuarine fish evolution. *Biol J Linn Soc*. 2007;91:325-340. doi 10.1111/j.1095-8312.2007.00785.x

Zhao L., Zhang J., Liu Z., Funk S.M., Wei F., Xu M., Li M. Complex population genetic and demographic history of the Salangid, *Neosalanx taihuensis,* based on cytochrome *b* sequences. *BMC Evol Biol*. 2008;8:201. doi 10.1186/1471-2148-8-201

Zhao L., Zhang J., Liu Z., Xu M., Li M. Population genetic structure and demographic history of *Neosalanx jordani* based on cytochrome *b* sequences. *Biodiv Sci*. 2010;18(3):251-261. doi 10.3724/SP.J.1003.2010.251

Conflict of interest. The author declares no conflicts of interest.

Received June 9, 2024. Revised August 29, 2024. Accepted October 25, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):268-278

doi 10.18699/vjgb-25-30

Taxonomic diversity of microbial communities in the cold sulfur spring Bezymyanny (Pribaikalsky district, Republic of Buryatia)

T.G. Banzaraktsaeva [1], E.V. Lavrentyeva [1, 2], V.B. Dambaev[1], I.D. Ulzetueva[3], V.V. Khakhinov[2]

[1] Institute of General and Experimental Biology of the Siberian Branch of the Russian Academy of Sciences, Ulan-Ude, Russia
[2] D. Banzarov Buryat State University, Ulan-Ude, Russia
[3] Baikal Institute of Nature Management of the Siberian Branch of the Russian Academy of Sciences, Ulan-Ude, Russia
tuyana_banz@mail.ru

Abstract. The environmental conditions of cold sulfur springs favor the growth and development of abundant and diverse microbial communities with many unique sulfur cycle bacteria. In this work, the taxonomic diversity of microbial communities of three different biotopes (microbial mat, bottom sediment, and water) in the cold sulfur spring Bezymyanny located on the shore of Lake Baikal (Pribaikalsky district, Republic of Buryatia) was studied using high-throughput sequencing of the 16S rRNA gene. By sequencing the microbial mat, bottom sediment, and water samples, 76,972 sequences assigned to 1,714 ASVs (ASV, amplicon sequence variant) were obtained. Analysis of the ASV distribution by biotopes revealed a high percentage (66–93 %) of uniqueness in the three communities studied. An estimate of the alpha diversity index showed that bottom sediment community had higher indices, while microbial mat community was characterized by a lowest diversity. Bacteria of the phyla Pseudomonadota, Bacteroidota, Campylobacterota, Actinomycetota, Desulfobacterota dominated in different proportions in the studied communities. The features of the community structure of the studied biotopes were established. The microbial mat community was represented mainly by *Thiothrix* (43.2 %). The bottom sediment community was based on *Sulfurovum* (11.2 %) and co-dominated by unclassified taxa (3.2–1 %). Sequences assigned to the genera *Novosphingobium*, *Nocardioides*, *Legionella*, *Brevundimonas*, *Sphingomonas*, *Bacillus*, *Mycobacterium*, *Sphingopyxis*, *Bradyrhizobium* and *Thiomicrorhabdus* were found only in the water microbial community. Sulfur-oxidizing bacteria (SOB) and sulfate-reducing bacteria (SRB) were identified in all the communities studied, which indicates the ongoing processes of the sulfur cycle in the Bezymyanny spring ecosystem. It should be noted that sequences of unclassified and uncultivated sulfur cycle bacteria were present in all communities and a significant proportion of sequences (20.3–53.9 %) were not classified.

Key words: cold sulfur spring; microbial community diversity; microbial mat; bottom sediment; water; sulfur-oxidizing bacteria; sulfate-reducing bacteria

For citation: Banzaraktsaeva T.G., Lavrentyeva E.V., Dambaev V.B., Ulzetueva I.D., Khakhinov V.V. Taxonomic diversity of microbial communities in the cold sulfur spring Bezymyanny (Pribaikalsky district, Republic of Buryatia). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):268-278. doi 10.18699/vjgb-25-30

Funding. This work was financially supported by the Russian Science Foundation, No. 23-27-00131.

Таксономическое разнообразие микробных сообществ холодного сероводородного источника Безымянный (Прибайкальский район, Республика Бурятия)

Т.Г. Банзаракцаева [1], Е.В. Лаврентьева [1, 2], В.Б. Дамбаев[1], И.Д. Ульзетуева[3], В.В. Хахинов[2]

[1] Институт общей и экспериментальной биологии Сибирского отделения Российской академии наук, Улан-Удэ, Россия
[2] Бурятский государственный университет им. Доржи Банзарова, Улан-Удэ, Россия
[3] Байкальский институт природопользования Сибирского отделения Российской академии наук, Улан-Удэ, Россия
tuyana_banz@mail.ru

Аннотация. Экологические условия холодных серосодержащих источников благоприятствуют росту и развитию богатых микробных сообществ со множеством уникальных бактерий цикла серы. В настоящей работе с использованием высокопроизводительного секвенирования гена 16S рРНК было изучено таксономическое разнообразие микробных сообществ трех различных биотопов (микробный мат, донный осадок и вода) в холодном сероводородном источнике Безымянный, расположенном на побережье озера Байкал (Прибайкальский район, Республика Бурятия). В результате секвенирования проб микробного мата, донного осадка и воды получено 76 972 последовательности, отнесенных к 1714 ASV (amplicon sequence variant – варианты последовательностей ампликонов).

Анализ распределения ASV по биотопам выявил высокий показатель (66–93 %) уникальности трех исследуемых сообществ. Оценка индекса альфа-разнообразия показала, что сообщество донного осадка имело более высокие индексы, сообщество микробного мата отличалось наименьшим разнообразием. В исследуемых сообществах в разных пропорциях доминировали бактерии филумов Pseudomonadota, Bacteroidota, Campylobacterota, Actinomycetota, Desulfobacterota. Установлены особенности структуры сообществ исследуемых биотопов. Сообщество микробного мата было представлено преимущественно бактериями рода *Thiothrix* (43.2 %). В сообществе донного осадка основу составляли бактерии рода *Sulfurovum* (11.2 %), содоминировали неклассифицируемые таксоны (3.2–1 %). Микробное сообщество воды характеризовалось присутствием последовательностей, обнаруженных только в воде. Данные последовательности были отнесены к родам *Novosphingobium*, *Nocardioides*, *Legionella*, *Brevundimonas*, *Sphingomonas*, *Bacillus*, *Mycobacterium*, *Sphingopyxis*, *Bradyrhizobium* и *Thiomicrorhabdus*. Во всех изучаемых сообществах были идентифицированы сероокисляющие бактерии (SOB) и серовосстанавливающие бактерии (SRB), что свидетельствует о протекающих процессах цикла серы в экосистеме источника Безымянный. Необходимо отметить, что во всех сообществах присутствовали последовательности неклассифицированных и некультивируемых бактерий цикла серы, и в целом значительную долю последовательностей (20.3–53.9 %) не удалось классифицировать.

Ключевые слова: холодный сероводородный источник; разнообразие микробных сообществ; микробный мат; донный осадок; вода; сероокисляющие бактерии; сульфатвосстанавливающие бактерии

Introduction

The Baikal region has a large area of natural water bodies, among which a significant part is represented by numerous vents of mineral springs. On the territory of Buryatia there are almost all known types of mineral waters, which are formed in the hypergenesis zone of rock (Borisenko, Zamana, 1978; Namsaraev et al., 2005). Sulfur springs are enriched with hydrogen sulfide as a result of biochemical sulfate reduction coming with mineral water or mineral suspension containing sulfates in exchange and bound state (Borisenko, Zamana, 1978; Kononov, 1983).

The environmental conditions of cold sulfur springs favor the growth and development of abundant and diverse microbial communities with many unique sulfur cycle bacteria (Douglas S., Douglas D., 2001; Rudolph et al., 2004; Chaudhary et al., 2009; Headd, Engel, 2014; Hahn et al., 2022). Cold springs are characterized by a slow change of parameters such as pH, temperature, dissolved gases and other factors and are more stable for bacterial life compared to other environments (Nosalova et al., 2023c).

Previous studies in cold sulfur springs, using culturable and non-culturable approaches, have mainly focused on the composition and community structure of microbial mats (Douglas S., Douglas D., 2001; Chaudhary et al., 2009; Klatt et al., 2016; Sapers et al., 2017; Nosalova et al., 2023b). Microscopy study in microbial mats from the cold sulfur spring Ancaster (Ontario, Canada) showed the development of all major groups of sulfide-oxidizing bacteria, purple, green, cyanobacteria and colorless sulfur-oxidizing bacteria (Douglas D., Douglas S., 2001). Using 16S rRNA gene sequencing, a clear dominance of phyla Pseudomonadota and Campylobacterota was shown in microbial mats of Slovak cold sulfur springs, and genera *Thiothrix* and *Sulfurovum* were identified as the core microbiota of microbial mats (Nosalova et al., 2023b).

In the present work, the taxonomic diversity of microbial communities of microbial mat, bottom sediment, and water in the cold sulfur spring Bezymyanny, located on the coast of Lake Baikal, was studied for the first time using high-throughput sequencing of the 16S rRNA gene. The aim of the study was to determine the bacterial composition of microbial communities in different biotopes of the Bezymyanny spring (Pribaikalsky district, Republic of Buryatia).

Materials and methods

The Bezymyanny sulfur spring is located in the forest area close to the coastline of Lake Baikal and is situated at an altitude of 638 m above sea level (53°02′48.95″ N, 108°19′57.68″ E) (Fig. 1). The water of the mineral spring seeps through a 25–30 cm thick layer of loose sediments to form a stream. In the stream, microbial mats were found on the surface of the bottom sediments.

Microbial mat, bottom sediment and water samples were collected in October, 2023. Water samples for chemical analysis were collected from the mineral spring vent and the outflow stream, into clean polyethylene and sterile glass bottles. Microbial mat, bottom sediment and water samples for microbiological studies were collected into sterile 50 ml Falcon tubes.

At the sampling sites, pH and temperature were measured with a portable pH-meter with a sensor thermometer pH-200 HM Digital (South Korea). Total salinity was measured with a COM-100 conductometer. A portable redox potential meter ORP (Portugal) was used to determine redox potential. The concentration of carbonates and hydrocarbonates in the analyzed waters was determined in the field environment at the moment of sampling by titration with 0.1 N HCl in the presence of phenolphthalein and methyl orange indicators. Total hardness, calcium and magnesium ions, chloride ions were determined by titrimetric method. Concentrations of ammonium ions, nitrite, nitrate, phosphate, sulfate and fluoride ions, silicic acid were determined by spectrophotometric method. A single beam spectrophotometer CECIL 1000 (UK) was used for analysis; ion concentrations were calculated by regression equations. The

Fig. 1. Map-scheme of Bezymyanny spring location (*a*), Bezymyanny spring photo (*b*), spring microbial mat photo (*c*).

content of hydrogen sulfide and sulfides was determined by calorimetric method with the addition of iron-ammonium alum at 670 nm. Sulfites and thiosulfates were determined by reverse iodometric titration (Fomin, 2000).

In accordance with the manufacturer's instructions, a reagent kit (MACHEREY-NAGEL NucleoSpin Soil) from MACHEREY-NAGEL (Duren, Germany) was used to isolate DNA from microbial mat, sediment and water samples.

Purified DNA preparations were used to create libraries of 16S rRNA gene fragments by PCR using universal primers for V4 variable region: F515/R806 (GTGCCAGCMGCCG GCGGTAA/GGACTACVSGGGTATCTAAT) (Bates et al., 2011), with attached adapters and unique Illumina barcodes. PCR was carried out in 15 μL of reaction mixture containing 0.5–1 unit of activity of Q5 High-Fidelity DNA Polymerase (NEB, Ipswich, MA, USA), 5 pM each of forward and reverse primers, 10 ng of DNA matrix and 2 nM of each dNTP (LifeTechnologies, Carlsbad, CA, USA). The mixture was denatured at 94 °C for 1 min, followed by 35 cycles of 94 °C for 30 s, 50 °C for 30 s and 72 °C for 30 s. The final elongation was carried out at 72 °C for 3 min. PCR products were purified according to the method recommended by Illumina using AMPureXP magnetic particles (Beckman-Coulter, Brea, CA, USA).

Library preparation and sequencing were carried out in accordance with the manufacturer's recommendations for operation of the Illumina MiSeq instrument (Illumina, San Diego, CA, USA) using MiSeq® ReagentKit v3 (600 cycles). Initial data processing, namely, sample demultiplexing and removal of adapters, was carried out using Illumina software (Illumina, USA). For subsequent denoising, sequence merging, deletion of chimeric reads, restoration of the original phylotypes (ASV, amplicon sequence variant), and further taxonomic classification of the obtained ASVs, the software packages DADA2 (Callahan et al., 2016), PHYLOSEQ (McMurdie, Holmes, 2013) and SILVA (Quast et al., 2013) were used; the work was carried out in the R software environment. Raw sequences were pre-filtered for quality with the filterAndTrim function with a sequence trim length of 250 and 200 n for forward and reverse reads respectively; the

acceptable level of expected error was 2 and 3 for forward and reverse reads respectively. The learnErrors function with MAX_CONSIST parameter equal to 20 was used to build the error model. The dereplication process was carried out using the derepFastq function with parameter *n* equal to 10e7. Denoising was carried out with the dada function with the pool parameter set to 'pseudo'. The recovered sequences were combined using the mergePairs function. The table of numbers of the obtained phylotypes was constructed using the makeSequenceTable function. A check for the presence of chimeras was conducted and they were filtered using the removeBimeraDenovo function by the 'consensus' method with the parameter minFoldParentOverAbundance equal to 2 and allowOneOff set to 'TRUE'. The OTU table was constructed using the otu_table function with the taxa_are_rows parameter set to 'FALSE'. The file with representations for each phylotype was generated using the getSequences, DNAStringSet and writeXStringSet functions.

Classification of the resulting phylotypes was performed with the assignTaxonomy function using the SILVA release 128 database and with the minBoot parameter set to 70. Fragments related to plastid and mitochondrial DNA were removed from the list of phylotypes. BIOM table construction was performed using the python 3 programming language, with the biom, numpy and pandas packages. The tools of the QIIME 1 software package (Caporaso et al., 2010) were used to present the data of the taxonomic analysis. The research was carried out at the Core Centrum 'Genomic Technologies, Proteomics and Cell Biology' in All-Russia Research Institute for Agricultural Microbiology (St. Petersburg, Russia).

The NCBI server (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to search for closest homologues. Alpha diversity indices were calculated using the software package Past 4.16 (Hammer et al., 2001). The iNEXT Online resource (https://chao.shinyapps.io/iNEXTOnline/) was used to plot rarefaction curves (Chao et al., 2014, 2016). The Venn diagram and heat map were created with the SRplot Online resource using the pheatmap R package, where a standard scaling approach (https://www.bioinformatics.com.cn/plot_basic_cluster_heatmap_plot_024_en) was applied to normalize the data (Tang et al., 2023).

Results

Physical and chemical characteristics of the Bezymyanny spring

At the time of sampling, the water temperature was 5.7 °C, the pH value, 8.4, mineralization, 0.12 g/dm^3 and redox potential, –113 mV. Hydrochemical analysis of water composition showed hydrogen carbonate and chloride ions content of 30.5 and 3.5 mg/dm^3 respectively. Calcium cations were 36.6 mg/dm^3 and magnesium cations were 15.6 mg/dm^3. Nitrate and nitrite ions in the amount of 4.7 and 0.01 mg/dm^3 and phosphate and fluoride ions in the amount of 0.35 and 0.32 mg/dm^3 were found. Carbonate and ammonium ions were not detected and iron ions were present in the amount of 0.14 mg/dm^3. Of the sulfur-containing compounds, 42.0 mg/dm^3 of sulfate ions, 11.0 mg/dm^3 of sulfide ions, and the presence of small amounts of sulfite and thiosulfate ions were detected.

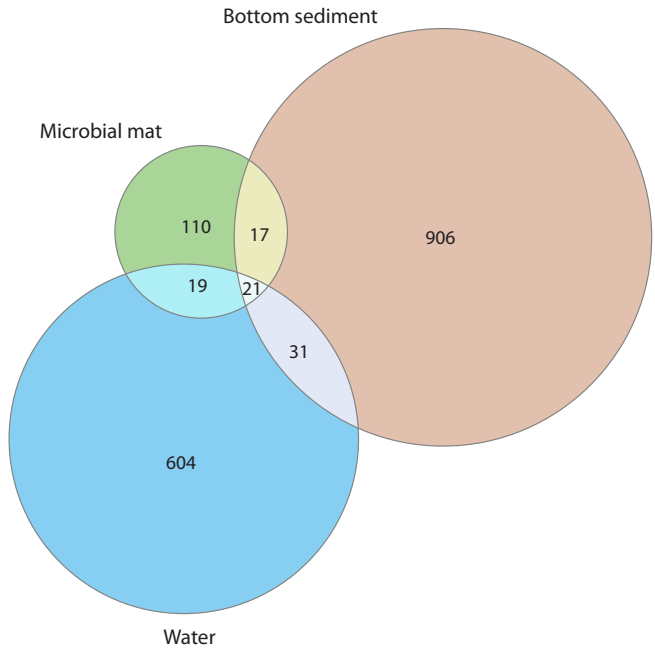


Fig. 2. Venn diagram of ASV bacterial communities of microbial mat, bottom sediment and water of the Bezymyanny spring.

Analysis of microbial communities diversity in different biotopes

Pyrosequencing of the 16S rRNA gene fragment from microbial mat, bottom sediment and water samples produced a total of 143,192 sequence reads. After their filtration, alignment, pre-clustering and removal of chimeric sequences and singletons, 76,972 reads were included in the analysis. The sequences were assigned to 1,714 ASVs and their distribution in three biotopes of the Bezymyanny spring is shown in the Venn diagram (Fig. 2). Only 21 ASVs were shared for all three biotopes. Most of the microbial communities in the biotopes are represented by unique sequences.

Alpha diversity indices were evaluated to assess diversity in each sample (Table 1). The bottom sediment microbial community was characterized by the highest number of ASVs, the highest diversity, and its unique ASVs constituted 93 %.

The rarefaction curves showed similar results as in Table 1 (Fig. 3).

Taxonomic composition of the microbial community

Microbial communities of microbial mat, bottom sediment and water of the Bezymyanny sulfide spring are represented by the Bacteria domain. Single sequences were assigned to the Archaea domain. Representatives of the phylum Pseudomonadota dominated in all samples (12.9–53.8 % of the total number of reads) (Fig. 4).

Table 1. Number of sequences and ASVs in the Bezymyanny spring samples and alpha diversity indices of the samples

Samples	Number of raw sequences	Number of sequences in the analysise	Number of ASVs	Unique ASV, %	Simpson	Shannon	Chao 1
Microbial mat	48,694	20,324	167	66	0.8175	3.159	167.2
Bottom sediment	45,061	27,011	975	93	0.9857	5.897	975
Water	49,437	29,637	675	89	0.9892	5.354	675.1

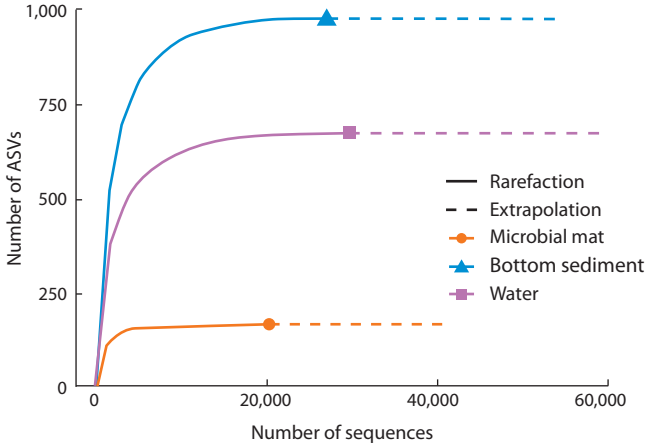


Fig. 3. ASV rarefaction curves in the samples of microbial mat, bottom sediment and water.

Pseudomonadota were represented by Gammaproteobacteria, which formed the core of the microbial communities (52 % in microbial mat, 29.4 % in water and 11.9 % in bottom sediment). The number of Alphaproteobacteria ranged from 0.9 % in bottom sediment to 23.4 % in water. Bacteroidota (2.3–9.8 %) was found in all samples. Sequences of the phylum Campylobacterota were widely distributed in the microbial mat and bottom sediment, 8.5 and 11.8 %, respectively. The water community was characterized by the high abundance of Actinomycetota (10.1 %) and Bacillota (2.2 %). The mat community was characterized by a high proportion of Verrucomicrobiota (3.8 %) and Cyanobacterota (2.6 %). The bottom sediment community showed a high percentage of Desulfobacterota (6.3 %) and Chloroflexota (2.9 %). It should also be noted that in all communities, 0.9–6 % of sequences were identified to the domain level, and a significant proportion of sequences (20.3–53.9 %) were not classified.

The analysis of taxonomic composition at a deeper level showed that representatives of the genera *Sulfuricurvum*, *Sulfurovum*, *Thiothrix*, *Flavobacterium* and unclassified sequences of unclassified_Comamonadaceae, unclassified_Burkholderiales, unclassified_Gammaproteobacteria were present in all investigated microbial communities. However, representatives of these taxa were unevenly distributed over communities and varied from clear dominance to representation of 0.1 % or less. It was found that each microbial community had its own characteristics at the level considered (Fig. 5).

The microbial mat community was heavily dominated by bacteria of the genus *Thiothrix* (43.2 %). The mat com-

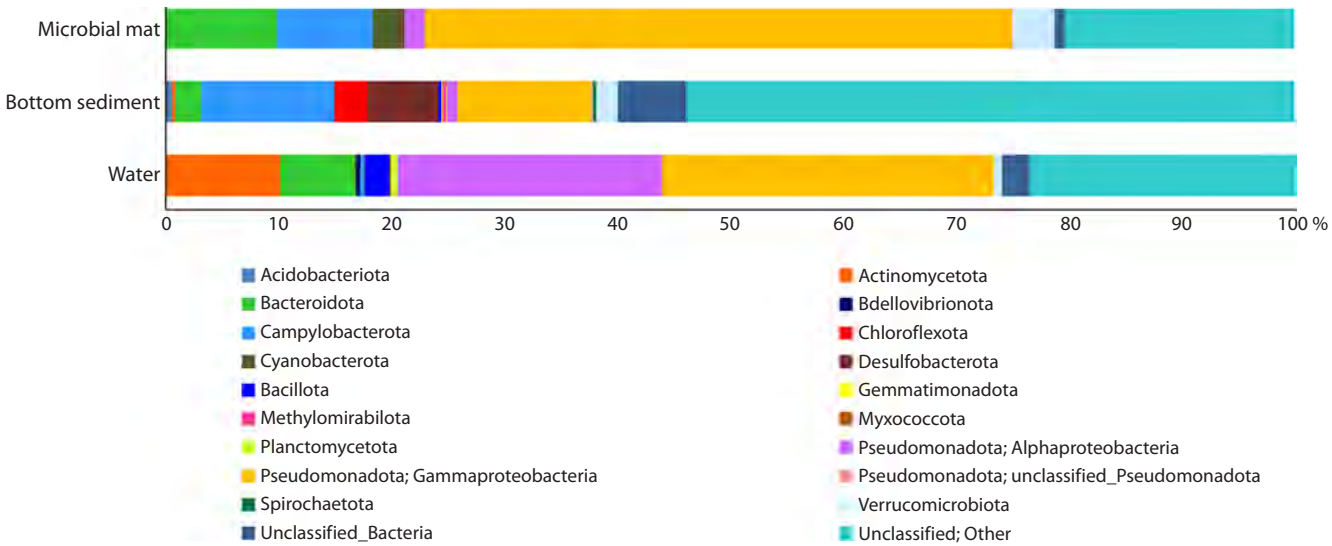


Fig. 4. Taxonomic diversity of prokaryotes at the phylum level (classes for Pseudomonadota) in microbial mat, bottom sediment and water of Bezymyanny spring.

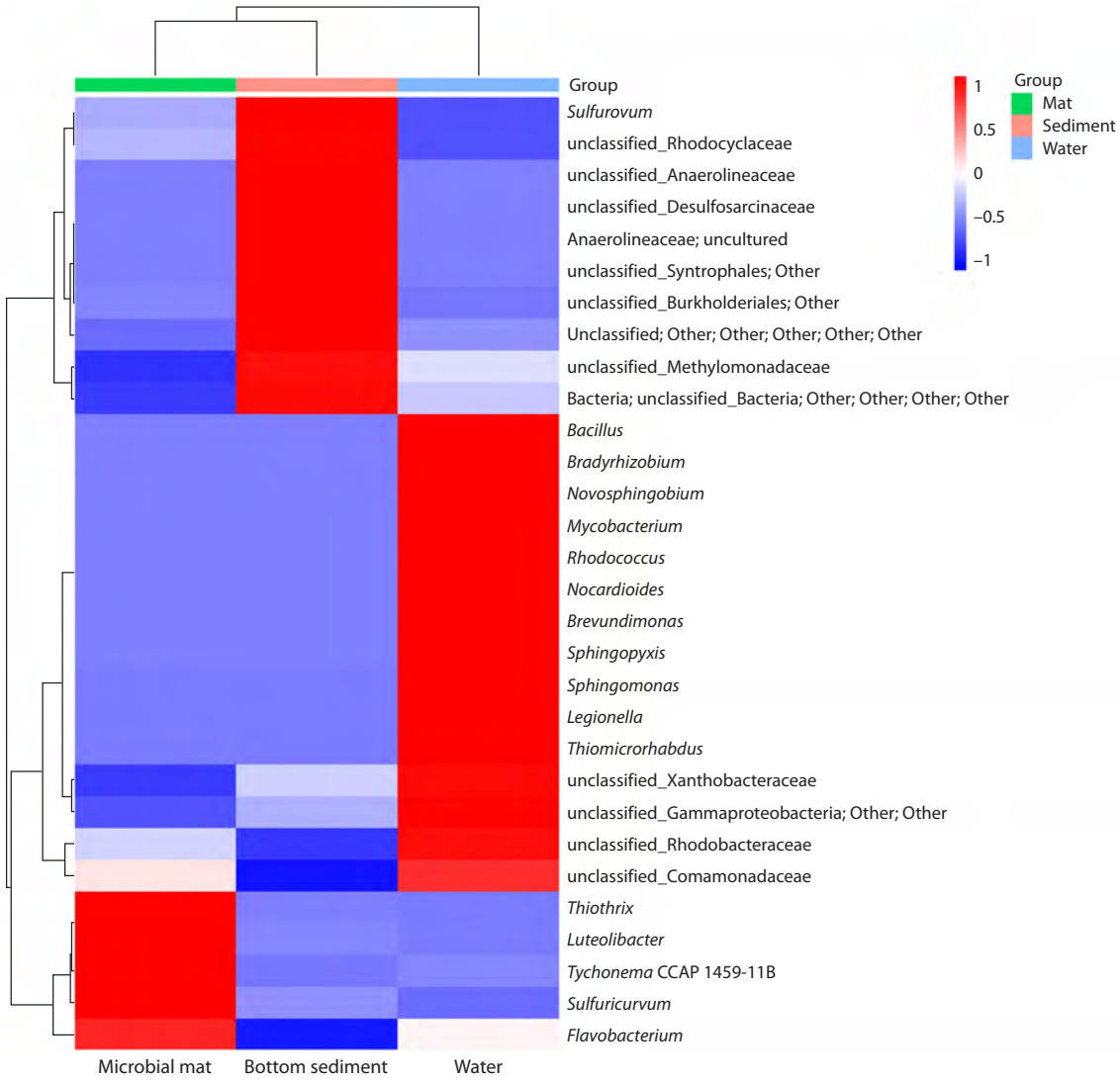


Fig. 5. Heat map of taxa (number of reads ≥ 1 %) in microbial mat, bottom sediment and water of the Bezymyanny spring. The color scale reflects the distance of each value from the mean in the standard deviation units.

munity also included bacteria of genus *Flavobacterium*, *Sulfuricurvum*, *Luteolibacter* and cyanobacteria *Tychonema* CCAP 1459-11B. The core bottom sediment community was formed by *Sulfurovum* (11.2 %) and representatives of unclassified_Burkholderiales, unclassified_Anaerolineaceae, unclassified_Desulfosarcinaceae, unclassified_Rhodocyclaceae, unclassified_Methylomonadaceae, unclassified_Syntrophales co-dominated (3.2–1 %). In the water community, among the dominants (>1 % of all sequences obtained), sequences typical only for this community and assigned to the genera *Novosphingobium*, *Nocardioides*, *Legionella*, *Brevundimonas*, *Sphingomonas*, *Bacillus*, *Mycobacterium*, *Sphingopyxis*, *Bradyrhizobium* and *Thiomicrorhabdus* were present.

Sulfur cycle bacteria

Microorganisms involved in the sulfur cycle were found in the taxonomic composition of the Bezymyanny spring. *Thiothrix* and *Sulfurovum* comprised the majority of the sequences in all of the investigated spring biotopes. Representatives of the genus *Thiothrix* of the family Thiotrichaceae, class Gammaproteobacteria, were the key taxon in the microbial mat (43.2 %). Genus *Sulfurovum* of the family Sulfurovaceae, class Campylobacteria, constituted a significant part (11.2 %) of the taxonomic composition of the bottom sediment community (Table 2).

The identified sulfate-reducing bacteria belong mainly to uncultivated unclassified_Desulfosarcinaceae (family Desulfosarcinaceae) and reach 1.7 % in bottom sediments. Other representatives of sulfate-reducing bacteria account for <1 % in the bottom sediment and microbial mat and are assigned to the genera *Desulfobulbus*, *Desulfomonile*, *Desulfocapsa*, *Desulfatiglans* (family Desulfatiglandaceae) and unclassified_Syntrophobacteraceae, [*Desulfobacterium*] catecholicum group (family Desulfobacteraceae).

Table 2. Comparative characteristics of sulfur cycle bacteria representation in microbial communities of Bezymyanny spring biotopes (number of sequences, %)

Phylotype	Microbial mat	Bottom sediment	Water
Sulfate-reducing bacteria			
Desulfobulbus (family Desulfobulbaceae)	0.0	0.1	0.0
Desulfomonile (family Syntrophaceae)	0.0	0.4	0.0
unclassified_Syntrophobacteraceae (family Syntrophobacteraceae)	0.0	0.1	0.0
[*Desulfobacterium*] catecholicum group (family Desulfobacteraceae)	0.0	0.2	0.0
Desulfocapsa (family Desulfobulbaceae)	0.2	0.1	0.0
Desulfatiglans (family Desulfatiglandaceae)	0.0	0.4	0.0
SEEP-SRB1 (family Desulfosarcinaceae)	0.0	0.1	0.0
unclassified_Desulfosarcinaceae (family Desulfosarcinaceae)	0.0	1.7	0.0
uncultured (family Desulfosarcinaceae)	0.0	0.2	0.0
unclassified_Desulfobacterales; Other	0.0	0.4	0.0
unclassified_Desulfobacteria;Other; Other	0.0	0.4	0.0
unclassified_Desulfobulbaceae (family Desulfobulbaceae)	0.1	0.0	0.0
unclassified_Desulfocapsaceae (family Desulfocapsaceae)	0.0	0.7	0.0
unclassified_Desulfobacterales	0.0	0.4	0.0
Sulfur-oxidizing bacteria			
Thiothrix (family Thiotrichaceae, Gammaproteobacteria)	43.2	1.1	0.1
Bradyrhizobium (family Xanthobacteraceae; Alphaproteobacteria)	0.0	0.0	1.3
Sulfurovum (family Sulfurovaceae, Campylobacterota)	2.4	11.2	0.1

Discussion

The territory of the Republic of Buryatia is extremely rich in mineral waters with different physical properties, chemical and gas composition (Mikhailov, Tolstikhin, 1946; Tkachuk et al., 1957; Namsaraev et al., 2005). Cold mineral springs are formed both as a result of interaction of water with the host rocks, and owing to input of some constituents from deep zones of the earth. The Bezymyanny sulfur spring is characterized by regular low temperature with reduced conditions. Water is alkalescent and weakly mineralized with predominance of hydrocarbonate, sulfate and sulfide ions.

For the first time, studies of microbial mat, bottom sediment and water by 16S rRNA gene sequencing in the cold sulfur spring Bezymyanny revealed 15 bacterial phyla, five of which were most abundant: Pseudomonadota, Bacteroidota, Campylobacterota, Actinomycetota, Desulfobacterota. Pseudomonadota, represented by class Gammaproteobacteria, and Bacteroidota were dominant in all samples, which coincides with previous reports on the communities of different types of sulfur habitats (Elshahed et al., 2003; Gulecal-Pektas, Temel, 2016; Nosalova et al., 2023a). The chemolithotrophic Gammaproteobacteria and Bacteroidota play a major role in the formation of primary production by iron and sulfide oxidation in inactive sulfide ecosystems (Dong et al., 2021). Campylobacterota were present in all samples and dominated in the bottom sediment (11.8 %) and microbial mat (8.5 %) communities. Similar results were obtained when communities in hydrothermal vents and aphotic (cave) sulfide springs were studied, where it was shown that in these ecosystems, sulfur-based chemolithoautotrophy is mainly carried out by Campylobacterota (formerly Epsilonproteobacteria) (Karl et al., 1980; Sarbu et al., 1996; Engel et al., 2003, 2004).

A distinctive feature of the Bezymyanny spring water community was the significant presence of Actinomycetota representatives. Some new actinobacteria from geothermal environments are known to be able to grow autotrophically with sulfur as an energy source (Norris et al., 2011). Using a culturable approach, sulfur-oxidizing bacteria phylogenetically related to Actinomycetota have been isolated from cold, high sulfide and sulfate springs at Gypsum Hill (Arctic, Canada) (Perreault et al., 2008). In the bottom sediment of the Bezymyanny spring, Desulfobacterota (6.3 %) made a significant contribution to the community. The abundance of sulfate-reducing bacteria belonging to Desulfobacterota has been described in a number of publications on the mi-

crobiota of cold saline springs in the Canadian Arctic and has also been found in the high-sulfide wetland Solodovka (Samara region, Russia) (Perreault et al., 2008; Sapers et al., 2017; Colangelo-Lillis et al., 2019; Gorbunov et al., 2022).

It has now been noted that cold sulfur springs harbor unique, not yet explored bacterial communities (Hamilton et al., 2015; Nosalova et al., 2023a). A high number of sequences in our studies remained unclassified, suggesting the presence of many undiscovered and unstudied communities and indicating potentially novel microorganisms in the cold sulfur spring ecosystem.

Analyses of the taxonomic composition of microbial communities at the genus level showed characteristic features of each community in all three biotopes studied. The microbial mat community was represented mainly by bacteria of the genus *Thiothrix* (43.2 %). NCBI database sequence analysis revealed similarity (100 % homology) with *Thiothrix fructosivorans*, which is able to deposit intracellular elemental sulfur in the presence of reduced inorganic sulfur compound (Howarth et al., 1999). A recent study on the Baikal region described an unculturable *Thiothrix* sp. from the mixing zone of the waters of Lake Baikal and the Zmeiny geothermal spring (northern basin of Lake Baikal, Russia) (Chernitsyna et al., 2024). Comparative analysis of *Thiothrix* sequences from our study and *Thiothrix* from the geothermal spring Zmeiny revealed 99 % similarity. The mat community of the Bezymyanny spring included sequences assigned to the genera *Flavobacterium, Sulfuricurvum, Luteolibacter* and the cyanobacterium *Tychonema* CCAP 1459-11B, the closest homologues of which have been isolated mainly from low-temperature habitats (Kodama, Watanabe, 2004; Jiang et al., 2012; Yang et al., 2019; Conklin et al., 2020).

The bottom sediment community was based on bacteria of the genus *Sulfurovum* (11.2 %), the closest homologue (98.81 %) of which was the mesophilic, facultatively anaerobic sulfur- and thiosulfate-oxidizing *Sulfurovum lithotrophicum* (Inagaki et al., 2004). Co-dominants with the proportion ranging from 1 to 3.2 % of all sequences were classified only up to the order and family levels. A search for closely related species in the NCBI database for unclassified_Burkholderiales (3.2 % presence) showed 98 % similarity to *Georgfuchsia toluolica* (Pseudomonadota; Betaproteobacteria; Nitrosomonadales; Sterolibacteriaceae), which is able to use Fe(III), Mn(IV) and nitrate as terminal electron acceptors for growth on aromatic compounds (Weelink et al., 2009). Unclassified_Anaerolineaceae typical for bottom sediments had 89 % homology with the marine thermophilic, anaerobic and heterotrophic bacterium *Thermomarinilinea lacunofontalis* (Nunoura et al., 2013). The closest homologue (95 % similarity) for unclassified_Desulfosarcinaceae was *Desulfosarcina widdelii*, hydrocarbon-degrading sulfate-reducing bacteria (Watanabe et al., 2017).

In our study, bacteria of the genera *Novosphingobium, Nocardioides, Legionella, Brevundimonas, Sphingomonas, Bacillus, Mycobacterium, Sphingopyxis, Bradyrhizobium* and *Thiomicrorhabdus* dominated among the sequences found only in the water community. Representatives of these genera are found in various natural environments and belong to heterotrophic prokaryotes that utilize various carbon, nitrogen and sulfur compounds as energy sources (Fliermans, 1996; Kumar R. et al., 2017; Tóth et al., 2017; Song et al., 2022; Kuang et al., 2023). Bacteria of the genera *Novosphingobium, Nocardioides, Sphingomonas* and *Sphingopyxis* are known to be able to grow under low nutrient conditions and are important agents in the biodegradation of various persistent and toxic organic substances, including aromatic compounds, hydrocarbons, halogenoalkanes, nitrogen heterocycles and polymeric polyesters (Song et al., 2022; Ma et al., 2023). In the water community, the closest homologues of the dominant sequences were bacteria involved in the sulfur cycle. For example, sequences assigned to the genus *Bacillus* showed 100 % similarity to the chemolithoautotrophic thiosulfate-oxidizing bacterium *Bacillus thioparus* (Pérez-Ibarra et al., 2007). Sequences identified as *Thiomicrorhabdus* found close affinity to *Thiomicrorhabdus aquaedulcis*, freshwater obligate sulfur-oxidizing chemolithotroph (Kojima, Fukui, 2019).

Microbial oxidation and reduction of sulfur are the most active and ancient metabolic processes in the sulfur cycle that occur in various ecosystems. These processes are carried out by sulfur-oxidizing (SOB) and sulfate-reducing bacteria (SRB) in all ecosystems and are considered as a key phenomenon in the biogeochemical sulfur cycle (Kumar U. et al., 2018). At the genus level, the microbial mat of the Bezymyanny sulfur spring was found to be predominantly composed of the colorless sulfur-oxidizing bacteria *Thiothrix*. *Thiothrix* species are considered to be typical of sulfur-oxidizing microbial communities in sulfur-rich habitats. Using a non-cultivation approach, the genus *Thiothrix* was identified in cold sulfur springs in Slovakia (Nosalova et al., 2023b).

The bottom sediment was dominated by members of the genus *Sulfurovum*. Previous studies have found this genus to be part of the microbial community in a variety of sulfur environments including springs, caves, vents and oxygen-free sediments (Nosalova et al., 2023b). These facultative anaerobic chemolithotrophs succeed under extreme conditions and are primary producers in sulfur-rich habitats. In the work of Wright et al. (2013), it was noted that all sulfur redox genes are present in the genomes of sequenced representatives of *Sulfurovum* and their genetic ability allows them to use various sulfur compounds.

In addition, obligate chemolithoautotrophic sulfur-oxidizing bacterial species related to *Thiomicrorhabdus* have been found only in cold spring water. *Thiomicrorhabdus* has previously been found in cold saline environments, including Arctic marine sediments and Antarctic subglacial brines (Knittel et al., 2005). *Thiomicrorhabdus* has also been found in abundance in cold saline spring sediments on Axel Heiberg Island, Canada (Magnuson et al., 2023). The authors mark that *Thiomicrorhabdus* aerobically and chemolithoautotrophically oxidizes sulfide to elemental sulfur.

It is known that in oxygen-free, sulfate-saturated layers beneath the sediment surface, sulfate-reducing microorganisms are among the most important participants that mediate a significant fraction of organic matter degradation (Yin et al., 2024). In the Bezymyanny cold spring, the highest distribution of sulfate-reducing bacteria in the bottom sediment, mainly belonging to phylum Thermodesulfobacteriota, was found. It should be noted that the large number of unclassified and uncultured representatives of sulfate-reducing bacteria suggests the presence of new species of sulfate-reducing bacteria in the Bezymyanny cold spring.

Conclusion

For the first time, the taxonomic diversity of microorganisms was studied, and characteristic features of microbial community structure of different biotopes (microbial mat, bottom sediment and water) in the cold sulfur spring Bezymyanny (Pribaikalsky district, Republic of Buryatia) were revealed by using molecular biological methods. According to the results of studies, sulfur-oxidizing (SOB) and sulfate-reducing bacteria (SRB) were identified in the community, which indicates that sulfur cycling processes are taking place in the ecosystem of the Bezymyanny spring. On the whole, the analysis of taxonomic composition showed a high percentage of unclassified sequences in the communities studied. The obtained data indicate that the microbiota of cold sulfur springs is still a hidden resource of new taxa, including sulfur cycle bacteria. Studies of cold sulfur springs will further expand our knowledge of bacteria involved in the biogeochemical cycle of sulfur, their metabolism and evolution, and may indicate the ecological features of the development of sulfur microbial communities and their relationship to the environment of their habitat.

Data availability: all raw sequences obtained from 16S rRNA gene sequencing are available in the NCBI SRA archive under BioProject number PRJNA1202704.

References

Bates S.T., Berg-Lyons D., Caporaso J.G., Walters W.A., Knight R., Fierer N. Examining the global distribution of dominant archaeal populations in soil. *ISME J*. 2011;5(5):908-917. doi 10.1038/ismej.2010.171

Borisenko I.M., Zamana L.V. Mineral Waters of Buryat ASSR. Ulan-Ude: Buryat Publ., 1978 (in Russian)

Callahan B.J., McMurdie P.J., Rosen M.J., Han A.W., Johnson A.J., Holmes S.P. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13(7):581-583. doi 10.1038/nmeth.3869

Caporaso J.G., Kuczynski J., Stombaugh J., Bittinger K., Bushman F.D., Costello E.K., Fierer N., … Walters W.A., Widmann J., Yatsunenko T., Zaneveld J., Knight R. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7(5):335-336. doi 10.1038/nmeth.f.303

Chao A., Gotelli N.J., Hsieh T.C., Sander E.L., Ma K.H., Colwell R.K., Ellison A.M. Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecol Monogr*. 2014;84(1):45-67. doi 10.1890/13-0133.1

Chao A., Ma K.H., Hsieh T.C. iNEXT (iNterpolation and EXTrapolation) Online: Software for Interpolation and Extrapolation of Species Diversity. 2016. Program and User's Guide published at http://chao.stat.nthu.edu.tw/wordpress/software_download/

Chaudhary A., Haack S.K., Duris J.W., Marsh T.L. Bacterial and archaeal phylogenetic diversity of a cold sulfur-rich spring on the shoreline of Lake Erie, Michigan. *Appl Environ Microbiol*. 2009;75(15):5025-5036. doi 10.1128/AEM.00112-09

Chernitsyna S.M., Elovskaya I.S., Bukin S.V., Bukin Y.S., Pogodaeva T.V., Kwon D.A., Zemskaya T.I. Genomic and morphological characterization of a new *Thiothrix* species from a sulfide hot spring of the Zmeinaya bay (Northern Baikal, Russia). *Antonie van Leeuwenhoek*. 2024;117(1):23. doi 10.1007/s10482-023-01918-w

Colangelo-Lillis J., Pelikan C., Herbold C.W., Altshuler I., Loy A., Whyte L.G., Wing B.A. Diversity decoupled from sulfur isotope fractionation in a sulfate-reducing microbial community. *Geobiology*. 2019;17(6):660-675. doi 10.1111/gbi.12356

Conklin K.Y., Stancheva R., Otten T.G., Fadness R., Boyer G.L., Read B., Zhang X., Sheath R.G. Molecular and morphological characterization of a novel dihydroanatoxin-a producing *Microcoleus* species (cyanobacteria) from the Russian River, California, USA. *Harmful Algae*. 2020;93:101767. doi 10.1016/j.hal.2020.101767

Dong X., Zhang C., Li W., Weng S., Song W., Li J., Wang Y. Functional diversity of microbial communities in inactive seafloor sulfide deposits. *FEMS Microbiol Ecol*. 2021;97(8):fiab108. doi 10.1093/femsec/fiab108

Douglas S., Douglas D.D. Structural and geomicrobiological characteristics of a microbial community from a cold sulfide spring. *Geomicrobiol J*. 2001;18(4):401-422. doi 10.1080/01490450175321056 7

Elshahed M.S., Senko J.M., Najar F.Z., Kenton S.M., Roe B.A., Dewers T.A., Spear J.R., Krumholz L.R. Bacterial diversity and sulfur cycling in a mesophilic sulfide-rich spring. *Appl Environ Microbiol*. 2003;69(9):5609-5621. doi 10.1128/AEM.69.9.5609-5621.2003

Engel A.S., Lee N., Porter M.L., Stern L.A., Bennett P.C., Wagner M. Filamentous "*Epsilonproteobacteria*" dominate microbial mats from sulfidic cave springs. *Appl Environ Microbiol*. 2003;69(9):5503-5511. doi 10.1128/AEM.69.9.5503-5511.2003

Engel A.S., Porter M.L., Stern L.A., Quinlan S., Bennett P.C. Bacterial diversity and ecosystem function of filamentous microbial mats from aphotic (cave) sulfidic springs dominated by chemolithoautotrophic "*Epsilonproteobacteria*". *FEMS Microbiol Ecol*. 2004;51(1):31-53. doi 10.1016/j.femsec.2004.07.004

Fliermans C.B. Ecology of *Legionella*: from data to knowledge with a little wisdom. *Microb Ecol*. 1996;32(2):203-228. doi 10.1007/BF00185888

Fomin G.S. Water. Control of Chemical, Bacterial and Radiation Safety According to International Standards. Encyclopedic reference book. Moscow: Protector Publ., 2000 (in Russian)

Gorbunov M.Y., Khlopko Y.A., Kataev V.Y., Umanskaya M.V. Bacterial diversity in attached communities of a cold high-sulfide water body in European Russia. *Microbiology*. 2022;91:77-90. doi 10.1134/S0026261722010040

Gulecal-Pektas Y., Temel M. A window to the subsurface: microbial diversity in hot springs of a sulfidic cave (Kaklik, Turkey). *Geomicrobiol J*. 2016;34(4):374-384. doi 10.1080/01490451.2016.1204374

Hahn C.R., Farag I.F., Murphy C.L., Podar M., Elshahed M.S., Youssef N.H. Microbial diversity and sulfur cycling in an early earth analogue: from ancient novelty to modern commonality. *mBio*. 2022;13(2):e0001622. doi 10.1128/mbio.00016-22

Hamilton T.L., Jones D.S., Schaperdoth I., Macalady J.L. Metagenomic insights into S(0) precipitation in a terrestrial subsurface lithoautotrophic ecosystem. *Front Microbiol*. 2015;5:756. doi 10.3389/fmicb.2014.00756

Hammer Ø., Harper D.A.T., Ryan P.D. PAST: PAleontological STatistics software package for education and data analysis. *Palaeontol Electronica*. 2001;4(1):1-9

Headd B., Engel A.S. Biogeographic congruency among bacterial communities from terrestrial sulfidic springs. *Front Microbiol*. 2014;5:473. doi 10.3389/fmicb.2014.00473

Howarth R., Unz R.F., Seviour E.M., Seviour R.J., Blackall L.L., Pickup R.W., Jones J.G., Yaguchi J., Head I.M. Phylogenetic relationships of filamentous sulfur bacteria (*Thiothrix* spp. and Eikelboom type 021N bacteria) isolated from wastewater-treatment plants and description of *Thiothrix eikelboomii* sp. nov., *Thiothrix unzii* sp. nov., *Thiothrix fructosivorans* sp. nov. and *Thiothrix defluvii* sp. nov. *Int J Syst Bacteriol*. 1999;49(Pt.4):1817-1827. doi 10.1099/00207713-49-4-1817

Inagaki F., Takai K., Nealson K.H., Horikoshi K. *Sulfurovum lithotrophicum* gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the ε-*Proteobacteria* isolated from Okinawa Trough hydrothermal sediments. *Int J Syst Evol Microbiol*. 2004;54(Pt.5): 1477-1482. doi 10.1099/ijs.0.03042-0

Jiang F., Li W., Xiao M., Dai J., Kan W., Chen L., Li W., Fang C., Peng F. *Luteolibacter luojiensis* sp. nov., isolated from Arctic tundra soil, and emended description of the genus *Luteolibacter*. *Int J Syst Evol Microbiol*. 2012;62(Pt.9):2259-2263. doi 10.1099/ijs.0.037309-0

Karl D.M., Wirsen C.O., Jannasch H.W. Deep-sea primary production at the Galápagos hydrothermal vents. *Science*. 1980;207(4437):1345-1347. doi 10.1126/science.207.4437.1345

Klatt J.M., Meyer S., Häusler S., Macalady J.L., de Beer D., Polerecky L. Structure and function of natural sulphide-oxidizing microbial mats under dynamic input of light and chemical energy. *ISME J*. 2016;10(4):921-933. doi 10.1038/ismej.2015.167

Knittel K., Kuever J., Meyerdierks A., Meinke R., Amann R., Brinkhoff T. *Thiomicrospira arctica* sp. nov. and *Thiomicrospira psychrophila* sp. nov., psychrophilic, obligately chemolithoautotrophic, sulfur-oxidizing bacteria isolated from marine Arctic sediments. *Int J Syst Evol Microbiol*. 2005;55(Pt.2):781-786. doi 10.1099/ijs. 0.63362-0

Kodama Y., Watanabe K. *Sulfuricurvum kujiense* gen. nov., sp. nov., a facultatively anaerobic, chemolithoautotrophic, sulfur-oxidizing bacterium isolated from an underground crude-oil storage cavity. *Int J Syst Evol Microbiol*. 2004;54(Pt.6):2297-2300. doi 10.1099/ijs.0.63243-0

Kojima H., Fukui M. *Thiomicrorhabdus aquaedulcis* sp. nov., a sulfur-oxidizing bacterium isolated from lake water. *Int J Syst Evol Microbiol*. 2019;69(9):2849-2853. doi 10.1099/ijsem.0.003567

Kononov V.I. Geochemistry of Thermal Areas of Modern Volcanism (rift zones and island arcs). Moscow: Nauka Publ., 1983 (in Russian)

Kuang B., Xiao R., Hu Y., Wang Y., Zhang L., Wei Z., Bai J., Zhang K., Acuña J.J., Jorquera M.A., Pan W. Metagenomics reveals biogeochemical processes carried out by sediment microbial communities in a shallow eutrophic freshwater lake. *Front Microbiol*. 2023;13: 1112669. doi 10.3389/fmicb.2022.1112669

Kumar R., Verma H., Haider S., Bajaj A., Sood U., Ponnusamy K., Nagar S., Shakarad M.N., Negi R.K., Singh Y., Khurana J.P., Gilbert J.A., Lal R. Comparative genomic analysis reveals habitat-specific genes and regulatory hubs within the genus *Novosphingobium*. *mSystems*. 2017;2(3):e00020-17. doi 10.1128/mSystems.00020-17

Kumar U., Panneerselvam P., Gupta V.V.S.R., Manjunath M., Priyadarshinee P., Sahoo A., Dash S.R., Kaviraj M., Annapurna K. Diversity of sulfur-oxidizing and sulfur-reducing microbes in diverse ecosystems. In: Advances in Soil Microbiology: Recent Trends and Future Prospects. Microorganisms for Sustainability. Vol. 3. Singapore: Springer, 2018;3:65-89. doi 10.1007/978-981-10-6178-3_4

Ma Y., Wang J., Liu Y., Wang X., Zhang B., Zhang W., Chen T., Liu G., Xue L., Cui X. *Nocardioides*: "specialists" for hard-to-degrade pollutants in the environment. *Molecules*. 2023;28(21):7433. doi 10.3390/molecules28217433

Magnuson E., Altshuler I., Freyria N.J., Leveille R.J., Whyte L.G. Sulfur-cycling chemolithoautotrophic microbial community dominates a cold, anoxic, hypersaline Arctic spring. *Microbiome*. 2023;11(1): 203. doi 10.1186/s40168-023-01628-5

McMurdie P.J., Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8(4):e61217. doi 10.1371/journal.pone.0061217

Mikhailov M.P., Tolstikhin N.I. Mineral Springs and Mud Lakes of Eastern Siberia, Their Hydrology, Balneochemistry and Balneological Significance. Irkutsk, 1946 (in Russian)

Namsaraev B.B., Danilova E.V., Barkhutova D.D., Khakhinov V.V. Mineral Springs and Therapeutic Lakes of South Buryatia. Ulan-Ude: Buryat State University Publ., 2005 (in Russian)

Norris P.R., Davis-Belmar C.S., Brown C.F., Calvo-Bado L.A. Autotrophic, sulfur-oxidizing actinobacteria in acidic environments. *Extremophiles*. 2011;15(2):155-163. doi 10.1007/s00792-011-0358-3

Nosalova L., Fecskeova L.K., Piknova M., Bonova K., Pristas P. Unique populations of sulfur-oxidizing bacteria in natural cold sulfur springs in Slovakia. *Geomicrobiol J*. 2023a;40(4):315-324. doi 10.1080/01490451.2023.2167021

Nosalova L., Mekadim C., Mrazek J., Pristas P. *Thiothrix* and *Sulfurovum* genera dominate bacterial mats in Slovak cold sulfur springs. *Environ Microbiome*. 2023b;18(1):72. doi 10.1186/s40793-023-00527-4

Nosalova L., Piknova M., Kolesarova M., Pristas P. Cold sulfur springs-neglected niche for autotrophic sulfur-oxidizing bacteria. *Microorganisms*. 2023c;11(6):1436. doi 10.3390/microorganisms 11061436

Nunoura T., Hirai M., Miyazaki M., Kazama H., Makita H., Hirayama H., Furushima Y., Yamamoto H., Imach H., Takai K. Isolation and characterization of a thermophilic, obligately anaerobic and heterotrophic marine *Chloroflexi* bacterium from a *Chloroflexi*-dominated microbial community associated with a Japanese shallow hydrothermal system, and proposal for *Thermomarinilinea lacunofontalis* gen. nov., sp. nov. *Microbes Environ*. 2013;28(2):228-235. doi 10.1264/jsme2.me12193

Pérez-Ibarra B.M., Flores M.E., García-Varela M. Isolation and characterization of *Bacillus thioparus* sp. nov., chemolithoautotrophic, thiosulfate-oxidizing bacterium. *FEMS Microbiol Lett*. 2007; 271(2):289-296. doi 10.1111/j.1574-6968.2007.00729.x

Perreault N.N., Greer C.W., Andersen D.T., Tille S., Lacrampe-Couloume G., Lollar B.S., Whyte L.G. Heterotrophic and autotrophic microbial populations in cold perennial springs of the high arctic. *Appl Environ Microbiol*. 2008;74(22):6898-6907. doi 10.1128/AEM.00359-08

Quast C., Pruesse E., Yilmaz P., Gerken J., Schweer T., Yarza P., Peplies J., Glöckner F.O. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41:D590-D596. doi 10.1093/nar/gks1219

Rudolph C., Moissl C., Henneberger R., Huber R. Ecology and microbial structures of archaeal/bacterial strings-of-pearls communities and archaeal relatives thriving in cold sulfidic springs. *FEMS Microbiol Ecol*. 2004;50:1-11. doi 10.1016/j.femsec.2004.05.006

Sapers H.M., Ronholm J., Raymond-Bouchard I., Comrey R., Osinski G.R., Whyte L.G. Biological characterization of microenvironments in a hypersaline cold spring Mars analog. *Front Microbiol*. 2017;8:2527. doi 10.3389/fmicb.2017.02527

Sarbu S.M., Kane T.C., Kinkle B.K. A chemoautotrophically based cave ecosystem. *Science*. 1996;272(5270):1953-1955. doi 10.1126/science.272.5270.1953

Song D., Chen X., Xu M. Characteristics and functional analysis of the secondary chromosome and plasmids in sphingomonad. *Int Biodeterior Biodegrad*. 2022;171:105402. doi 10.1016/j.ibiod.2022.105402

Tang D., Chen M., Huang X., Zhang G., Zeng L., Zhang G., Wu S., Wang Y. SRplot: a free online platform for data visualization and graphing. *PLoS One*. 2023;18(11):e0294236. doi 10.1371/journal.pone.0294236

Tkachuk V.G., Yasnitskaya N.V., Ankudinova G.A. Mineral Waters of the Buryat-Mongolian ASSR. Irkutsk, 1957 (in Russian)

Tóth E., Szuróczki S., Kéki Z., Kosztik J., Makk J., Bóka K., Spröer C., Márialigeti K., Schumann P. *Brevundimonas balnearis* sp. nov., isolated from the well water of a thermal bath. *Int J Syst Evol Microbiol*. 2017;67(4):1033-1038. doi 10.1099/ijsem.0.001746

Watanabe M., Higashioka Y., Kojima H., Fukui M. *Desulfosarcina widdelii* sp. nov. and *Desulfosarcina alkanivorans* sp. nov., hydrocar-

bon-degrading sulfate-reducing bacteria isolated from marine sediment and emended description of the genus *Desulfosarcina*. *Int J Syst Evol Microbiol*. 2017;67(8):2994-2997. doi 10.1099/ijsem.0.002062

Weelink S.A., van Doesburg W., Saia F.T., Rijpstra W.I., Röling W.F., Smidt H., Stams A.J. A strictly anaerobic betaproteobacterium *Georgfuchsia toluolica* gen. nov., sp. nov. degrades aromatic compounds with Fe(III), Mn(IV) or nitrate as an electron acceptor. *FEMS Microbiol Ecol*. 2009;70(3):575-585. doi 10.1111/j.1574-6941.2009.00778.x

Wright K.E., Williamson C., Grasby S.E., Spear J.R., Templeton A.S. Metagenomic evidence for sulfur lithotrophy by Epsilonproteobacteria as the major energy source for primary productivity in a sub-aerial arctic glacial deposit, Borup Fiord Pass. *Front Microbiol*. 2013;4:63. doi 10.3389/fmicb.2013.00063

Yang L.L., Liu Q., Liu H.C., Zhou Y.G., Xin Y.H. *Flavobacterium laiguense* sp. nov., a psychrophilic bacterium isolated from Laigu glacier on the Tibetan Plateau. *Int J Syst Evol Microbiol*. 2019;69(6):1821-1825. doi 10.1099/ijsem.0.003400

Yin X., Zhou G., Wang H., Han D., Maeke M., Richter-Heitmann T., Wunder L.C., Aromokeye D.A., Zhu Q.Z., Nimzyk R., Elvert M., Friedrich M.W. Unexpected carbon utilization activity of sulfate-reducing microorganisms in temperate and permanently cold marine sediments. *ISME J*. 2024;18(1):wrad014. doi 10.1093/ismejo/wrad014

Conflict of interest. The authors declare no conflict of interest.

Received October 24, 2024. Revised December 26, 2024. Accepted December 26, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):279-289

doi 10.18699/vjgb-25-31

Cystic fibrosis therapy: from symptoms to the cause of the disease

T.N. Kireeva, D.I. Zhigalina, N.A. Skryabin

Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia
tatyana.kireeva@medgenetics.ru

Abstract. Cystic fibrosis (CF) is a disease with a broad clinical and genetic spectrum of manifestations, significantly impacting the quality and duration of life of patients. At present, a diagnosis of CF enables the disease to be identified at the earliest stages of its development. The accelerated advancement of scientific knowledge and contemporary research techniques has transformed the methodology employed in the treatment of CF, encompassing a spectrum of approaches from symptomatic management to pathogenetic therapies. Pathogenetic therapy represents an approach to treatment that aims to identify methods of restoring the function of the *CFTR* gene. The objective of this review was to analyse and summarize the available scientific data on the pathogenetic therapy of CF. This paper considers various approaches to the pathogenetic therapy of CF that are based on the use of targeted drugs known as CFTR modulators. The article presents studies employing gene therapy techniques for CF, which are based on the targeted delivery of a normal copy of the *CFTR* gene cDNA to the respiratory tract via viral or non-viral vectors. Some studies have demonstrated the efficacy of RNA therapeutic interventions in restoring splicing, promoting the production of mature RNA, and increasing the functional expression of the CFTR protein. The review also analyzes literature data that consider methods of etiotropic therapy for CF, which consists of targeted correction of the *CFTR* gene using artificial restriction enzymes, the CRISPR/Cas9 system and a complex of peptide-nucleic acids. In a prospective plan, the use of cell therapy methods in the treatment of lung damage in CF is considered.

Key words: cystic fibrosis (CF); *CFTR*; *CFTR* mutations; CFTR modulators; gene therapy; genome editing; CRISPR/Cas9

For citation: Kireeva T.N., Zhigalina D.I., Skryabin N.A. Cystic fibrosis therapy: from symptoms to the cause of the disease. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):279-289. doi 10.18699/vjgb-25-31

Acknowledgements. The authors express their deep gratitude to A.A. Agafonova, Head of the Department of Hereditary Diseases, and V.V. Petrova, Pediatrician at the Department of Hereditary Diseases of the Medical Genetic Center (Genetic Clinic) Research Institute of Medical Genetics of the Tomsk National Research Medical Center, R.R. Savchenko, Researcher at the Laboratory of Genomics of Orphan Diseases, Research Institute of Medical Genetics, Tomsk National Research Medical Center, for consultative assistance and review of the manuscript.

Терапия муковисцидоза: от симптомов к причине заболевания

Т.Н. Киреева, Д.И. Жигалина, Н.А. Скрябин

Томский национальный исследовательский медицинский центр Российской академии наук, Томск, Россия
tatyana.kireeva@medgenetics.ru

Аннотация. Муковисцидоз (МВ) – заболевание с широким клиническим и генетическим спектром проявлений, оказывающее значительное влияние на качество и продолжительность жизни пациента. В настоящее время диагностика МВ позволяет выявлять заболевание на самых ранних стадиях. Стремительное развитие науки и современные методы исследования изменили подходы в лечении МВ, начиная от симптоматического лечения до методов патогенетической терапии. Подходы патогенетической терапии направлены на поиск способов восстановления функции гена *CFTR*. Целью обзора стали анализ и обобщение имеющихся научных сведений о патогенетической терапии МВ. Рассмотрены подходы патогенетической терапии МВ на основе приема пациентами таргетных препаратов – CFTR-модуляторов. Приведены исследования с использованием методов генной терапии МВ, в основе которых лежит целенаправленная доставка нормальной копии кДНК гена *CFTR* в дыхательные пути с помощью вирусных или невирусных агентов. В некоторых исследованиях показано применение методов РНК-терапии для восстановления сплайсинга, продукции зрелой РНК и функционального белка CFTR. Также в обзоре проведен анализ литературных данных, в которых рассмотрены методы этиотропной терапии МВ, заключающейся в направленной коррекции гена *CFTR* с использованием искусственных ферментов рестрикции, системы CRISPR/Cas9 и комплекса пептидно-нуклеиновых кислот. В перспективном плане обсуждаются методы клеточной терапии в лечении поражения легких при муковисцидозе.

Ключевые слова: муковисцидоз (МВ); ген *CFTR*; мутации *CFTR*; модуляторы CFTR; генная терапия; геномное редактирование; CRISPR/Cas9

Introduction

Cystic fibrosis (CF) (OMIM 219700) is a monogenic orphan disease with autosomal recessive type of inheritance, systemic organ damage with a severe course of the disease and prognosis (https://www.omim.org/). The incidence of CF is on average one case per 2,500–3,000 newborns (Kashirskaya, Kapranov, 2014). CF is most frequently registered among Caucasians, for example, in the USA and Europe, there are about 70,000 patients with CF; in Russia, there are about 4,000 patients with this disease (Simonova et al., 2020; Lomunova, Gershovich, 2023).

Cystic fibrosis is caused by pathogenic variants in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. The *CFTR* gene was identified and cloned in 1989 (Gembitskaya et al., 2012; Elborn, 2016; Spielberg, Clancy, 2016). The *CFTR* gene contains 27 exons and is located in region 31.1 of the long arm of chromosome 7 (7q31.1). The protein encoded by this gene, the CFTR transmembrane conductance regulator, is a member of the ABC transporter (ATP-binding cassette) superfamily of proteins. The structural organization of the CFTR protein includes two transmembrane domains (TMD1 and TMD2), two nucleotide-binding domains (NBD1 and NBD2), and a central intracellular regulatory domain (R-domain). Localized in the apical *membrane* of epithelial cells, the CFTR protein creates a chloride channel regulated by cyclic adenosine monophosphate (cAMP). The CFTR protein regulates not only chloride ions (Cl^-), but also bicarbonate (HCO_3^-) secretion, which regulates the pH of the fluid on the surface of airway cells. CFTR also plays an important role in the hydration of secretions and mucins through inhibition of the epithelial sodium channel (ENaC) (Ginter, 2000; Moran, 2014; Kondratieva et al., 2018; Bell et al., 2020; Hanssens et al., 2021).

Mutations in the *CFTR* gene lead to disruption of ion channels, causing a decrease in conductivity for Cl^- ions and an increase in conductivity for Na^+ ions. These disorders cause changes in hydration processes in the apical membrane of epithelial cells and changes in the viscoelastic properties of substances produced by exocrine glands. These changes have a greater impact on the functioning of the respiratory system, pancreas, liver, bile ducts, gastrointestinal tract, sweat glands and organs of the male reproductive system (Dechecchi et al., 2018; Smirnikhina, Lavrov, 2018; Lomunova, Gershovich, 2023).

CFTR gene variants

The identification and characterization of *CFTR* gene variants is carried out by the international Cystic Fibrosis Genetic Analysis Consortium (CFGAC), which unites laboratories, the activities of which are aimed at genetic diagnostics and research of CF around the world. For general access, the obtained data are placed in the database of *CFTR* gene variants "CFTR1" (http://www.genet.sickkids.on.ca/cftr/) and the subsequently created database "CFTR2" (http://www.cftr2.org/). The CFTR2 database includes current information on recently discovered *CFTR* gene variants (Rommens et al., 2006; Dechecchi et al., 2018; Kondratieva et al., 2018). Currently, the CFGAC database contains more than 2,000 variants in the *CFTR* gene, which are divided into seven classes, depending on the mechanism of their effect on the function of the CFTR protein (Fanen et al., 2014; Elborn, 2016; Kondratieva et al., 2018; Bell et al., 2020; Lee et al., 2021; Krasnova et al., 2023).

In **class I** of genetic variants (R553X, W1282X, 2143delT, G542X, 1677delTA), there is no functional CFTR protein, resulting in impaired transcription and translation. Approximately 22 % of patients with CF have at least one mutant allele of this class (Lee et al., 2021). As a result of genetic variants of **class II** (F508del, I507del, N1303K, S549N), the maturation of the CFTR protein is blocked due to an incorrect configuration of its molecule. Misfolded protein molecules do not reach the surface of epithelial cells because they undergo endoplasmic reticulum-associated protein degradation (ERAD). Approximately 88 % of CF patients have at least one mutant allele and the main variant is F508del, caused by a deletion of phenylalanine at position 508. With the genetic variant F508del, there is a disruption of post-translational modification of the CFTR protein, which leads to the protein molecule becoming functionally defective and unstable, or being completely destroyed (Van Goor et al., 2006; Smirnikhina, Lavrov, 2018).

Genetic variants of **class III** (G1224E, S1255P, G551D) which are localized in the regulatory domain of the CFTR protein and its nucleotide binding domains, lead to a disruption in the regulation of the chloride channel. The defect in the chloride channel in this case is due to the fact that the CFTR protein is synthesized and transported to the cell membrane, but does not respond to cAMP stimulation. Missense mutations, related to genetic variants of **class IV** in the *CFTR* gene, (R117H, R347P, R334W) lead to a decrease in ion flow as a result of changes in the conductivity of the chloride channel. These variants are located in transmembrane domains and affect the reduction of ion channel opening time. About 6 % of patients with CF have this type of genetic variant. *CFTR* gene variants of **class V** reduce the levels of functional protein and its transport to the apical membrane surface, which is characteristic of 5 % of patients with CF. **Class VI** includes *CFTR* gene variants that alter protein stability, resulting in a decrease in the time the protein remains on the membrane surface. It has been noted that 5 % of patients with CF have at least one allele of this variant (Kondratieva et al., 2018; Dechecchi et al., 2018). **Class VII** is also distinguished: its genetic variants affect the expression of CFTR protein mRNA. The absence of mRNA is caused by a genetic variant characterized by a large deletion – CFTRdele2,3 (21 kb) (Lee et al., 2021).

It is important to start therapy in a timely manner to prevent the development of severe complications in CF and generally improve the prognosis of the disease. The results of fundamental research have allowed us to expand our understanding of the main pathogenetic and pathophysio-

logical mechanisms of CF, which contributed to the rapid development and emergence of new approaches in the treatment of this disease. Currently, the basis of treatment for patients with CF is complex therapy, combining methods of both symptomatic and pathogenetic treatment. Methods based on the use of tools for *CFTR* gene correction are also being considered in the future (Gembitskaya et al., 2012; Bell et al., 2020).

Symptomatic treatment of CF

Symptomatic treatment is aimed at combating infection, improving mucus clearance from the bronchi and preventing nutritional deficiencies, including macro- and micronutrient deficiencies. Patients with CF are prescribed antibiotics, mucolytic and bronchodilator drugs in combination with enzymes, vitamins and a course of kinesiotherapy (Kashirskaya, Kapranov, 2014; Simonova et al., 2020). To treat respiratory lesions in patients with CF, anti-inflammatory and massive antibacterial therapy is used, while it is noted that the inhalation route of drug administration (mucolytics, bronchodilators, antibiotics and glucocorticoids) is highly effective (Gembitskaya et al., 2012; Olveira et al., 2017; Kondratieva et al., 2018; Simonova et al., 2020).

Methods of optimized antibiotic therapy have a significant impact on the course of CF, where the choice of antibiotic depends on the microbiological status of the patient. Antibiotic resistance is overcome by aerosol delivery of antibiotics into the bronchial lumen, which also reduces side effects during long-term treatment and the use of high doses, since the concentration of drugs in the blood serum is low in this case (Gorinova et al., 2015; Kondratieva et al., 2018; Simonova et al., 2020).

In the treatment of CF, mucolytic drugs are prescribed to normalize the viscous-elastic properties of sputum and improve its transport. In this group of drugs, a great advantage is possessed by the genetically engineered drug – the mucolytic dornase alpha, which has a complex effect on the infection, inflammation and obstruction observed in CF. The use of this drug is of great importance in the complex treatment of the bronchopulmonary process in patients with CF, especially immediately after diagnosis (Sherman et al., 2011). Together with mucolytic drugs, patients with CF are prescribed special active breathing exercises (kinesiotherapy) to remove phlegm from the respiratory tract (Simonova et al., 2020).

No less important in the therapy of CF is the correction of exocrine pancreatic insufficiency and treatment of hepatobiliary disorders, as well as maintaining the nutritional status of patients with the help of diet therapy. In patients with pancreatic insufficiency, in addition to diet therapy, enzyme replacement therapy and fat-soluble vitamins are also prescribed (Kashirskaya, Kapranov, 2011, 2014; Kondratieva et al., 2018).

All the developed methods and applied drugs for symptomatic treatment affect not only the life expectancy of patients with CF, but also their quality of life, significantly improving it. However, symptomatic treatment is aimed only at controlling symptoms and limiting complications in CF, while not affecting the functioning of the defective CFTR protein in any way (Smirnikhina, Lavrov, 2018; Simonova et al., 2020).

Pathogenetic therapy of CF

The development and testing of new methods and drugs aimed at finding ways to restore the function of the *CFTR* gene is becoming relevant. In this direction, pathogenetic therapy methods are considered promising (Gembitskaya et al., 2012; Rafeeq, Murad, 2017; Bell et al., 2020). Given the diversity of genetic variants in the *CFTR* gene and their various clinical manifestations, studies have been conducted to find drugs that suppress premature termination of protein translation for patients with nonsense mutations of class I, drugs for carriers of the common F508del variant and other genetic variants of class II, as well as drugs that work with all classes of genetic variants (see the Table) (Kondratieva et al., 2018; Dechecchi et al., 2018).

In the search for drugs that facilitate the “reading” of CFTR-mRNA stop codons and prevent premature termination of protein molecule synthesis, the drug ataluren (PTC Therapeutics, USA) was proposed, prescribed for the treatment of Duchenne muscular dystrophy caused by nonsense mutations. However, in the group of patients with nonsense mutations in the *CFTR* gene, this drug was ineffective. Currently, drugs for correcting *CFTR* gene variants of class I have not yet been developed (Kerem et al., 2014; Zainal Abidin et al., 2017; Smirnikhina, Lavrov, 2018).

The most promising therapeutic agents for the treatment of CF turned out to be a group of modulators, which are small molecule drugs that were identified as a result of high-throughput screening to correct impaired CFTR protein transport to the plasma membrane or to increase the conductance of the chloride channel (Dechecchi et al., 2018; Sui et al., 2022; Krasnova et al., 2023). In CF therapy, the choice of modulator drug depends on the class of *CFTR* gene variant and the direction of their compensatory actions. In this regard, modulators are divided into potentiators, correctors, amplifiers and stabilizers (Lee et al., 2021).

Potentiators

The action of potentiators is aimed at enhancing the opening of the ion channel formed by the mutant CFTR protein on the cell surface. The effect on the ion channel is achieved through activation of the adenylate cyclase pathway (*CFTR* gene variants of classes III–IV). Ivacaftor (Vertex Pharmaceuticals, Germany) is one of the drugs in this group. Phase I trials of ivacaftor were conducted in healthy volunteers and showed the safety of the drug (Van Goor et al., 2009). Then, in 2011, based on the results of tests on 112 CF patients (USA), data were presented that in individuals with the G551D, G178R, G551S, G1244E, G1349D mutations, a reliable increase in the transport of chloride ions was established (Flume et al., 2012). In 2012, Food and Drug Administration

Cystic fibrosis therapy

Type of therapy	Method	Drug/complex/combinations	Effectiveness shown in studies
		Pathogenetic therapy	
Pharmacotherapy: use of drugs (medicaments) – CFTR modulators	Use of CFTR premature translation suppressors	Ataluren	Not effective for patients with nonsense mutations
	Using CFTR ion channel potentiators	Ivacaftor	Effective for patients with the genetic variant G551D and patients with the G461E/N1303K genotype. Insignificant effect for patients with the F508del/F508del genotype
	Using CFTR folding and processing correctors	Lumacaftor	Partially restores the function of the mutant CFTR protein
	Using a "potentiator+corrector" combination	Ivacaftor/Lumacaftor	Effective for patients with the F508del/F508del genotype
		Ivacaftor/Tezacaftor	
	Using the combination "corrector+corrector+potentiator"	Elexacaftor/Tezacaftor/ Ivacaftor	Effective for patients with both one and two F508del alleles
	Use of CFTR stabilizers at the membrane	Kavosonstat	Low efficiency compared to potentiators
	Use of CFTR molecule synthesis enhancers	Nesolicaftor	Effective in inflammatory processes in the respiratory tract of patients with several genetic variants of the *CFTR* gene
Gene therapy: targeted delivery of a normal copy of the *CFTR* gene cDNA to the respiratory tract	Viral vector-based	Recombinant adenoviral vector (rAd)	Transient *CFTR* expression, significant immune response
		Helper-dependent adenoviral vector (Hd-Ad)	Loses effectiveness *in vivo*, does not induce an immune response
		Adeno-associated vector (AAV)	Not effective in transducing human cells
		Adeno-associated capsid AAV204	Effectively restores the functioning of chlorine channels
		Spiro-2101: Adeno-associated capsid carrying a functional copy of the *CFTR* gene	Assigned the status of "Orphan drug" for the treatment of CF
		Retroviral vectors	Low transducing efficiency
		Lentiviral vectors	High transducing efficiency
	Non-viral vector-based	pGM169/GL67A: cDNA/cationic lipid complex	Low effectiveness in restoring lung function
RNA therapy: restoration of splicing and production of mature RNA and functional CFTR protein	Use of mRNA as therapeutic agents	*CFTR* mRNA/lipid nanoparticles (LNPs) complex	Restoration of chloride channels and a significant increase in the amount of CFTR protein on the surface of the cell membrane
	Use of short RNA molecules as therapeutic agents	Antisense oligonucleotides (ASOs)	Low efficiency for correcting CFTR mRNA
		Eluforsen, QR-010: single-stranded antisense RNA	Effective in improving the functioning of chloride channels in patients with the F508del/ F508del genotype
	Use of spliceosome-mediated trans-splicing technique	SMaRT	Temporary restoration of CFTR function

Table (end)

Type of therapy	Method	Drug/complex/combinations	Effectiveness shown in studies
		Etiotropic therapy	
Genome editing: targeted correction of the *CFTR* gene	Use of artificial restriction enzymes	Zinc finger nucleases (ZFNs)	Restoration of CFTR function at therapeutically significant levels in basal cells of the respiratory epithelium
		Transcription activator-like effector nucleases (TALENs)	The editing efficiency in iPSCs from CF patients was 10 %
	Use of programmable nucleases CRISPR/Cas9	Cas9/guide RNAs (sgRNAs) and single-stranded oligodeoxyribonucleotides (ssODN)	The efficiency of correction of the F508del variant in HEK293T cells ranged from 0.08 to 0.7 % of alleles
	Use of peptide nucleic acids (PNA)	PNA/donor DNA/biodegradable polymer nanoparticles	*In vivo* correction rate of the F508del genetic variant in mouse epithelial cells ranges from ~0.1 to ~2 %
		Cell therapy	
Cell therapy: recovery of lung lesions	Autologous transplantation of CFTR-expressing cells into diseased areas of the respiratory tract	iPSCs differentiated into CFTR-expressing respiratory epithelial cells	No data

(FDA) approved ivacaftor for use in patients with at least one out of 38 point mutations, including five splicing mutations (Van Goor et al., 2006, 2009; Smirnikhina, Lavrov, 2018). Ivacaftor is recommended worldwide, including in Russia, for patients with the G551D mutation (Sui et al., 2022). In 2017, a case of successful treatment of a patient with CF with the G461E/N1303K genotype was described; after six months of using ivacaftor, the patient's clinical course of the disease changed significantly (Amelina et al., 2017). Since 2018, data have been published evaluating the efficacy of ivacaftor in a group of pediatric patients (2–3 years old), where more preserved lung function and a lower level of complications observed in CF were noted (Bessonova et al., 2018).

Correctors

Correctors are pharmacological substances that bind to the mutant CFTR protein, promoting its "maturation" by adapting protein homeostasis and reducing the degradation of the mutant protein in the intracellular quality control system (*CFTR* gene variants of class II) (Smirnikhina, Lavrov, 2018). Among the group of corrective drugs, the use of 4-phenylbutyrate/genistin, curcumin, tezacaftor, and lumacaftor is known. The largest number of studies is devoted to the evaluation and analysis of the effectiveness of lumacaftor in stabilizing the mutant CFTR protein and its movement from the endoplasmic reticulum (ER) to the surface of the cell membrane. Moreover, it was shown that lumacaftor is able to partially restore the function of the mutant CFTR protein by stabilizing its N-terminal domain (Ren et al., 2013; Lee et al., 2021).

It was later shown that the use of lumacaftor or ivacaftor alone only slightly reduced sweat chloride levels for patients homozygous for the F508del mutation. This suggests that monotherapy with either modulator is ineffective in improving lung function (Flume et al., 2012; Hanssens et al., 2021). Subsequently, the effectiveness of various combinations of modulators was assessed to restore CFTR protein function in patients with the F508del/F508del genotype. A number of larger studies have shown that long-term combination therapy with lumacaftor and ivacaftor was effective in patients over 12 years of age who were homozygous for the F508del mutation (Boyle et al., 2014; Wainwright et al., 2015). For Russian patients, a combination of lumacaftor and ivacaftor is used in the treatment of CF together with basic therapy, which results in such improvements as a decrease in the level of chlorides in sweat fluid, an increase in the forced expiratory volume in 1 second (FEV1), an improvement in the general condition and weight gain (Amelina et al., 2019).

As a result of the fact that the combination of a potentiator and a corrector has a positive effect on the clinical effect in patients with the F508del mutation, in 2015, the FDA approved the combination drug lumacaftor/ivacaftor for use in the treatment of CF. This drug is approved for use in children over 6 years of age and adults with the F508del/F508del genotype (Dechecchi et al., 2018; Smirnikhina, Lavrov, 2018; Simonova et al., 2020). Despite the efficacy demonstrated in clinical trials, the use of this drug entails a number of side effects, and, in addition, a positive effect is observed only in the case of one genetic variant, F508del, which is in a homozygous state (Lee et al., 2021).

The combination of ivacaftor with another drug, tezacaftor, has shown a positive therapeutic effect in improving lung function in the treatment of patients who are homozygous for F508del. The combination drug tezacaftor+ivacaftor/ivacaftor is used to treat CF in children 12 years of age and older and adults with the homozygous F508del mutation (Taylor-Cousar et al., 2017).

According to Vertex, a combination of three new-generation modulators, elexacaftor/tezacaftor/ivacaftor (ETI), has shown the greatest effectiveness in treating patients with the F508del/F508del genotype (Smirnikhina, Lavrov, 2018). This combination drug increases the activity of the CFTR protein and reduces mortality and morbidity in patients with CF, and is applicable both to CF patients homozygous for F508del (in 90 % of cases) and to the group of patients heterozygous for F508del and the variant with residual function (Keating et al., 2018). In clinical studies, the use of the ETI combination has been shown to improve mutant CFTR protein function to levels of 40–50 % of normal CFTR protein activity in airway and intestinal epithelial cells. This combination has also been shown to be highly effective in improving lung function, reducing sweat chloride concentration, and reducing pulmonary exacerbation frequency (Piehler et al., 2023).

Stabilizers and amplifiers

In the treatment of patients with CF, it is necessary to use compounds that stabilize and enhance the CFTR protein. By fixing the CFTR protein to the plasma membrane, stabilizers prevent its detachment and degradation in lysosomes. Nivalis Therapeutics has developed a compound that stabilizes the protein, cavosonstat, which was clinically tested in 138 patients homozygous for F508del. Patients received cavosonstat in combination with ivacaftor. However, in phase II, this study was completed due to the lack of advantages of the stabilizer compared to potentiators (Krasnovidova et al., 2023).

Enhancers are used to increase the amount of CFTR protein molecules synthesized in cells, available for subsequent modulation by protein-active small molecules. This group includes the drug nesolicaftor, which was developed by Proteostasis Therapeutics. Nesolicaftor enhances CFTR synthesis and, in combination with other existing CF treatments, has shown a positive effect on protein activity *in vitro*, nearly doubling its activity in bronchial epithelial cells of patients with multiple genetic variants of the *CFTR* gene. When using nesolicaftor in combination with ETIs in primary human bronchial epithelial F508del cells, it was shown to reverse cytokine transforming growth factor beta 1 (TGF-β1)-mediated inhibition of corrected CFTR function, likely through mRNA stabilization. Nesolicaftor also indirectly increases the level of secreted cytokines through its effect on apical ion channel function. The use of enhancers has been shown to be effective in treating inflammatory processes in the airways of patients with CF (Bengtson et al., 2022).

Thus, the considered pharmacological agents for pathogenetic therapy of CF significantly increased the life expectancy of patients with this diagnosis. However, CFTR modulators do not eliminate the cause of the disease, but only correct the functioning of the defective protein. CFTR modulator therapy requires lifelong drug administration, and their long-term potential side effects remain unclear (Sui et al., 2022). It is also noted that approximately 10 % of patients are resistant to modulators due to the absence or low levels of the CFTR protein. Also, according to clinical studies, about 10–20 % of patients with CF have individual intolerance to modulator drugs (Smirnikhina, Lavrov, 2018; Lee et al., 2021; Lomunova, Gershovich, 2023). In this regard, new methods of treating CF are being developed, aimed at eliminating the pathological changes underlying the development of this disease. First of all, these are gene therapy methods (Maule et al., 2020).

Gene therapy for CF

The monogenic and recessive type of inheritance in CF has led to the emergence of treatment methods for this disease using gene therapy methods (see the Table) (Sui et al., 2022). Gene therapy for CF involves the targeted delivery of a normal copy of complementary DNA (cDNA) of the *CFTR* gene to the most affected areas of the respiratory tract of patients using viral particles carrying the target transgene and non-viral agents, such as liposomes, nanoparticles, etc. (Ginter, 2000; Smirnikhina, Lavrov, 2018; Lomunova, Gershovich, 2023).

In 1993, a study was initiated to deliver a normal copy of *CFTR* cDNA to the nasal epithelium of CF patients using a recombinant adenoviral vector (rAd). This study demonstrated the potential of using recombinant adenoviral vectors to temporally correct Cl^- ion transport in CF. However, it was subsequently shown that rAd-mediated *CFTR* expression in postmitotic airway epithelial cells is transient and promotes robust cellular and humoral immune responses (Van Goor et al., 2009). Subsequently, a helper-dependent adenoviral vector (Hd-Ad) was developed to eliminate the problem of immune response. Hd-Ad delivers DNA (up to 37 kb) to airway cells, excluding host T cell responses to the expression of foreign viral protein, i. e. without causing inflammation (Lee et al., 2021). A study in CF mouse and pig airway basal cells showed restoration of CFTR function to levels seen in normal wild-type cells after correction of *CFTR* with Hd-Ad. In lung cells from *CFTR* gene knockout mice, the effectiveness of Hd-Ad vectors for *CFTR* gene correction was also demonstrated. However, due to airway cell turnover, the use of Hd-Ad vectors *in vivo* for *CFTR* gene correction loses its therapeutic efficacy (Koehler et al., 2003; Cao et al., 2020).

From 1998 to 2007, clinics led by Targeted Genetics Corporation evaluated the potential of using adeno-associated vectors (AAV) in the treatment of CF lung disease, of which rAAV2 was the only available vector of this serotype. Preclinical studies have demonstrated the ability of rAAV2 to

productively transduce lung cells from rhesus macaques and rabbits. However, more recent studies of rAAV2 transduction biology in a polarized human airway epithelium (HAE) cell culture model at the air-liquid interface (ALI) have found that rAAV2 poorly transduces human airway epithelial cells. Another limitation of the use of rAAV vectors in *CFTR* gene transfer is their relatively small packaging capacity (~4.9 kb) (Sui et al., 2022).

In recent years, several pharmaceutical companies have been developing AAV-based gene therapy agents. For example, Abeona Therapeutics has developed a next-generation capsid, AAV204, which carries a functional copy of the human mini-*CFTR* gene. The use of this agent in therapy allows for the effective restoration of the functioning of chloride channels in cells, both *in vitro* and *in vivo*. In 2020, Spirovant Sciences introduced another adeno-associated capsid with improved tropism for airway epithelial cells for delivering a functional copy of the *CFTR* gene. FDA granted Spirovant Sciences Orphan Drug Designations for Spiro-2101 for Treatment of CF (Lee et al., 2021; Lomunova, Gershovich, 2023).

Retroviral and lentiviral vectors have also been shown to be useful in CF gene therapy. In studies on rabbits, the use of retroviruses carrying the *CFTR* gene demonstrated persistent expression of this gene in their respiratory tract for up to three weeks, but low transduction efficiency was observed (Lee et al., 2021). The advantage of lentiviral vectors derived from immunodeficiency viruses is their ability to transduce both dividing and non-dividing cells, and transgene expression from the integrated viral genome is likely to be maintained throughout the life cycle of recipient cells. In this case, lentiviral vectors used for transduction into respiratory epithelial cells must be pseudotyped with appropriate protein coats. Studies have shown higher transduction efficiency into airway cells using a lentiviral strategy compared to a non-viral one (Alton et al., 2015; Sui et al., 2022). However, non-viral methods of delivering the normal *CFTR* gene are safer and better tolerated due to the absence of insertional mutagenesis and secondary effects of altered transgene expression levels (Lee et al., 2021).

Another advantage of using non-viral vectors is the ability to use larger fragments of donor DNA for gene repair. For efficient non-viral delivery of *CFTR*, a cDNA/cationic lipid complex is used. According to a study published by the UK CF Gene Therapy Consortium, CFTR function increased by up to 3.7 % in lung cells from CF patients after treatment with the nebulized cationic lipid pGM169/GL67A, which delivers donor DNA from the normal *CFTR* gene. However, this improvement was still not sufficient to restore lung function in CF (Alton et al., 2015; Spielberg, Clancy, 2016).

Thus, for almost three decades now, the search for suitable gene therapy methods for the treatment of CF has been ongoing. There have been approximately 36 clinical trials of gene therapy involving a significant number of patients with CF; however, due to the low clinical effect, these studies have not been further developed. Nevertheless, these studies have shown the promise of the concept of gene therapy for CF and have created a great foundation in this field (Sui et al., 2022).

CFTR gene editing

New approaches to targeted gene correction have come into use thanks to the emergence and improvement of experimental cellular and animal models. One of these effective methods is genome editing methods (see the Table).

To correct genes, tools are used based on targeted DNA cleavage using artificial restriction enzymes: zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and with the help of a programmable nuclease (most often Cas9), the specificity of which is achieved using guide RNA (sgRNA). The mechanisms functioning in the cell – non-homologous end joining (NHEJ) and homologous-directed repair (HDR), a common form of which is homologous recombination (HR), – ensure DNA repair (Smirnikhina et al., 2020; Lee et al., 2021).

Using ZFNs, the feasibility of editing the *CFTR* locus in airway basal cells derived from CF patients was assessed using two approaches. The first approach, based on sequence replacement to correct F508del, demonstrated restoration of mature CFTR protein and its function in ALI border cultures derived from massively edited basal cells. The second approach aimed to integrate partial cDNA into an intron of the endogenous *CFTR* gene to correct all genetic variants of the *CFTR* gene. As a result, highly efficient site-specific targeted integration into basal cells harboring different genetic variants of the *CFTR* gene was observed and restoration of CFTR function at therapeutically relevant levels was demonstrated (Suzuki et al., 2020). The use of TALEN in experiments shows better affinity than ZFNs. In one study, Hd-Ad vectors were used to deliver TALENs with donor DNA into cells, resulting in approximately 5 % targeted gene integration. TALEN-mediated editing of F508del has also been demonstrated in induced pluripotent cells (iPSCs) derived from CF patients, with editing efficiency in this case being no greater than 10 %. It was noted that manipulations with the iPSC genome did not affect their properties and ability to differentiate (Holkers et al., 2013; Xia et al., 2019).

Genome editing using CRISPR/Cas9 (Clustered Regulatory Interspaced Short Palindromic Repeats/CRISPR associated protein 9) allows editing a pathogenic variant in a gene with high efficiency and allows fixing the “corrected” allele in the genome. CRISPR/Cas9 is a very promising technology for creating valuable experimental tools for testing treatments for a wide range of pathogenetic variants that cause CF (Smirnikhina, Lavrov, 2018). The first use of the CRISPR/Cas9 editing system to correct the *CFTR* gene locus was applied in cultured intestinal stem cells from patients homozygous for F508del. The genetically modified stem cells formed organoids that responded functionally to forskolin through changes in volume. In another study, iPSCs were generated from fibroblast cells from CF patients

(F508del), which were also subsequently modified to contain the *CFTR* gene using the CRISPR/Cas9. Corrected iPSCs were able to differentiate into mature airway epithelial cells and demonstrated restoration of chloride transport (Wang, 2023).

To develop methods for editing genetic variants in the *CFTR* gene, a variety of cell lines have been generated using Cas9 nucleases, representing alternative models. These models are cell cultures into which plasmids carrying synthetic vectors with a fragment of the *CFTR* gene containing a target mutation, including a rare one, have been introduced. Based on this approach, the following cell lines were created: human lung cancer (Calu-3 CF), human leukemia (HL-60 F508del-CF), human carcinoma (T84 F508del-CF), human bronchial epithelial cells (16HBE14o-CF with F508del), as well as isogenic cell models with the G542X, W1282X mutations (Wang, 2023). A model of CF was created in HEK293T cell culture by introducing the synthetic plasmid pGEM-CFTR, carrying the *CFTR* locus with the F508del mutation. The efficiency of correction of this genetic variant was then assessed using six different combinations of Cas9/guide RNA (sgRNA) and single-stranded oligodeoxyribonucleotides (ssODN). The efficiency of correction of the F508del mutation ranged from 0.08 to 0.7 % of alleles, depending on the combination of CRISPR/Cas9 components used (Smirnikhina et al., 2020).

In addition to the considered editing systems, the possibility of correcting the *CFTR* gene was demonstrated using peptide nucleic acids (PNA) not based on CRISPR. In studies on F508del airway epithelial cells, triplex-forming peptide nucleic acids and donor DNA packaged in biodegradable polymer nanoparticles were used. The results show that intranasal delivery of nanoparticles to CF mice induces changes in the nasal epithelial potential difference assay as a consequence of corrected CFTR function. Another study demonstrated *in vivo* correction of the F508del mutation in multiple epithelial cells, including nasal epithelium, trachea, lung, ileum, colon and rectum in CF mice with systemic delivery of PNA. The correction level ranged from ~0.1 to ~2 % (Wang, 2023).

The approaches considered in targeted correction of the *CFTR* gene are aimed at the causes underlying the disease, i. e. they have the potential to provide a permanent cure for patients with CF. Despite this significant advantage, these approaches are currently not used in clinical practice due to bioethical restrictions.

RNA therapy for CF

In the therapy of CF, the use of methods based on the use of RNA is considered: messenger RNA (mRNA), transfer RNA (tRNA) and smaller RNA molecules – oligonucleotides, as therapeutic agents (see the Table). Clinical trials are currently underway investigating the potential of mRNA in CF therapy. The RESTORE-CF study (NCT03375047) tested specialized lipid nanoparticles (LNPs) as mRNA carriers. The results of these tests are measured by changes in lung function, i. e. changes in FEV1. After introducing chemically modified *CFTR* mRNA into cells using relevant liposomal nanoparticles, restoration of the functioning of chloride channels and a significant increase in the amount of CFTR protein on the surface of the cell membrane of the respiratory epithelium of patients with CF were noted (Lomunova, Gershovich, 2023).

In order to restore splicing and production of mature RNA and functional CFTR protein, the use of antisense oligonucleotides (ASOs) is being considered (Egan, 2021). More than 40 clinical trials have been conducted to study the therapeutic potential of ASOs in the treatment of CF. In cell models with the F508del genetic variant, ASO was used to insert missing bases at position 508 of *CFTR* at the RNA transcript level, but this method of mRNA correction was not stable (Maule et al., 2020). ProQR Therapeutics conducted studies on intranasal administration of single-stranded antisense RNA (eluforsen, QR-010) to mice. This drug was designed to restore CFTR function in respiratory epithelium through specific binding to the F508del region of mRNA. Studies have shown that QR-010 successfully diffuses into cells and causes positive changes in chloride transport. Thus, after three intranasal administrations of QR-010 over four weeks, patients with F508del/F508del showed a clinically significant improvement in the functioning of the chloride channel due to the restoration of CFTR function (Lomunova, Gershovich, 2023).

Spliceosome-mediated RNA trans-splicing (SMaRT) was also used to restore nascent mRNA by replacing part of the altered transcript with the correct exogenous mRNA. This technique was used in cell models with F508del to restore correct transcripts. However, this method only temporarily restored CFTR function (Maule et al., 2020).

The RNA therapies discussed above are considered possible treatments for patients with CF; however, these treatments require lifelong administration of therapeutic agents, as does CFTR modulator therapy.

Cell therapy for CF

In the future, the use of cell therapy methods in the treatment of lung damage in CF is being considered (see the Table). However, the method of delivering donor cells to human lungs poses significant difficulties.

Experiments on mice have shown the possibility of delivering cells to their lungs, e. g. embryonic stem cells (ESCs) were introduced into the lungs of mice by intravenous administration, and bone marrow (BM) cells were introduced by intratracheal administration. However, in these cases, the efficacy was low (Lee et al., 2021). Several studies have been conducted on the introduction of multipotent mesenchymal stromal cells (MMSCs) into the affected lungs of mice, where it was shown that the introduction of intact MMSCs into the body activates anti-inflammatory immunity in animals with various forms of lung inflammation (Smirnikhina,

Lavrov, 2018). A study conducted at Stanford University involved editing the mutant *CFTR* gene in primary airway basal cells using the CRISPR/Cas9 system delivered to these cells using AAV vectors. The corrected basal cells were then placed into rat sinus cavities, where the ability of these cells to proliferate into *CFTR*-normal cells was further assessed (Egan, 2021).

To date, protocols for differentiation of iPSCs into *CFTR*-expressing respiratory epithelial cells have already been developed, which allows iPSCs to be considered a promising material for autogenous transplantation in lung lesions. However, at present, clinical trials using iPSCs as part of cell therapy for patients with CF are not being conducted (Lomunova, Gershovich, 2023).

Conclusion

The ultimate goal of research into the discovery and development of treatments for CF is to provide all patients with therapy early enough in life to delay or even prevent many of the disease's manifestations, and to personalize the overall therapy itself based on patients' needs.

The advent of a number of targeted drugs in 2012 gave rise to a personalized approach to the treatment of patients with CF. Some drugs have already passed clinical trials and are used in therapy; these drugs include first-generation CFTR modulators: ivacaftor, lucamaftor/ivacaftor, tezacaftor+ivacaftor/ivacaftor, elexacaftor/tezacaftor/ivacaftor+ivacaftor. The use of modulators in CF therapy has made it possible to restore the functions of the mutant CFTR protein and improve the functioning of chloride channels on the surface of cells. However, this modulator therapy is not curative and does not cover all mutations in the *CFTR* gene. For the 10 % of CF patients with missense mutations, where cells produce little to no CFTR protein, therapy with CFTR modulators is not an option, making research into CF gene therapy, including genome editing, of great importance.

The advantage of gene therapy is that it is suitable for all CF patients, regardless of their genotype. There have been large research programs in the area of gene therapy for CF, developing potential agents for this type of therapy, and numerous clinical trials have been conducted to deliver the normal *CFTR* gene into respiratory epithelial cells. However, the long road to using gene therapy as a treatment for CF has not resulted in significant consistent clinical efficacy, even though there may have been some level of correction. Approaches using methods of genomic editing of the *CFTR* gene in CF are considered, using such tools as CRISPR/Cas9, ZFNs, TALEN and peptide nucleic acids. Research on genome editing in CF is in the preclinical phase.

Thus, patients with CF have been given the opportunity to significantly increase their life expectancy, along with improving its quality, thanks to the huge amount of research into the pathogenesis of CF and developments using innovative gene-directed personalized treatment methods.

References

Alton E.W.F.W., Armstrong D.K., Ashby D., Bayfield K.J., Bilton D., Bloomfield E.V., Boyd A.C., … Waller M.D., Wasowicz M.Y., Wilson J.M., Wolstenholme-Hogg P., UK Cystic Fibrosis Gene Therapy Consortium. Repeated nebulisation of non-viral *CFTR* gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial. *Lancet Respir Med*. 2015;3(9):684-691. doi 10.1016/S2213-2600(15)00245-3

Amelina E.L., Krasovskiy S.A., Usacheva M.V., Krylova N.A. Pathogenic treatment of cystic fibrosis: the first clinical case in Russia. *Pulmonologiya = Russian Pulmonology*. 2017;27(2):298-301. doi 10.18093/0869-0189-2017-27-2-298-301 (in Russian)

Amelina E.L., Krasovskiy S.A., Shumkova G.L., Krylova N.A. Targeted therapy for CF patients with F508del/F508del genotype. *Pulmonologiya = Russian Pulmonology*. 2019;29(2):235-238. doi 10.18093/0869-0189-2019-29-2-235-238 (in Russian)

Bell S.C., Mall M.A., Gutierrez H., Macek M., Madge S., Davies J.C., Burgel P.R., … Southern K.W., Sivam S., Stephenson A.L., Zampoli M., Ratjen F. The future of cystic fibrosis care: a global perspective. *Lancet Respir Med*. 2020;8(1):65-124. doi 10.1016/S2213-2600(19)30337-6

Bengtson C., Silswal N., Baumlin N., Yoshida M., Dennis J., Yerrathota S., Kim M., Salathe M. The CFTR amplifier nesolicaftor rescues TGF-β1 inhibition of modulator-corrected F508del CFTR function. *Int J Mol Sci*. 2022;23(18):10956. doi 10.3390/ijms231810956

Bessonova L., Volkova N., Higgins M., Bengtsson L., Tian S., Simard C., Konstan M.W., Sawicki G.S., Sewall A., Nyangoma S., Elbert A., Marshall B.C., Bilton D. Data from the US and UK cystic fibrosis registries support disease modification by CFTR modulation with ivacaftor. *Thorax*. 2018;73(8):731-740. doi 10.1136/thoraxjnl-2017-210394

Boyle M.P., Bell S.C., Konstan M.W., McColley S.A., Rowe S.M., Rietschel E., Huang X., Waltz D., Patel N.R., Rodman D.; VX09-809-102 study group. A CFTR corrector (lumacaftor) and a CFTR potentiator (ivacaftor) for treatment of patients with cystic fibrosis who have a phe508del *CFTR* mutation: a phase 2 randomised controlled trial. *Lancet Respir Med*. 2014;2(7):527-538. doi 10.1016/S2213-2600(14)70132-8

Cao H., Ouyang H., Laselva O., Bartlett C., Zhou Z.P., Duan C., Gunawardena T., Avolio J., Bear C.E., Gonska T., Hu J., Moraes T.J. A helper-dependent adenoviral vector rescues CFTR to wild-type functional levels in cystic fibrosis epithelial cells harbouring class I mutations. *Eur Respir J*. 2020;56(5):2000205. doi 10.1183/13993003.00205-2020

Dechecchi M.C., Tamanini A., Cabrini G. Molecular basis of cystic fibrosis: from bench to bedside. *Ann Transl Med*. 2018;6(17):334. doi 10.21037/atm.2018.06.48

Egan M.E. Emerging technologies for cystic fibrosis transmembrane conductance regulator restoration in all people with CF. *Pediatr Pulmonol*. 2021;56(1):32-39. doi 10.1002/ppul.2496

Elborn J.S. Cystic fibrosis. *Lancet*. 2016;388(10059):2519-2531. doi 10.1016/S0140-6736(16)00576-6

Fanen P., Wohlhuter-Haddad A., Hinzpeter A. Genetics of cystic fibrosis: CFTR mutation classifications toward genotype-based CF therapies. *Int J Biochem Cell Biol*. 2014;52:94-102. doi 10.1016/j.biocel.2014.02.023

Flume P.A., Liou T.G., Borowitz D.S., Li H., Yen K., Ordoñez C.L., Geller D.E.; VX 08-770-104 Study Group. Ivacaftor in subjects with cystic fibrosis who are homozygous for the F508del-CFTR mutation. *Chest*. 2012;142(3):718-724. doi 10.1378/chest.11-2672

Gembitskaya T.E., Chermensky A.G., Boytsova E.P. Cystic fibrosis today: progress and problems, promises of etiopathogenetic therapy. *Vrach = The Doctor*. 2012;2:5-8 (in Russian)

Ginter E.K. Gene therapy of hereditary diseases. *Voprosy Meditsinskoi Khimii*. 2000;46(3):264-278 (in Russian)

Gorinova Yu.V., Simonova O.I., Lazareva A.V., Chernevich V.P., Smirnov I.E. Experience of the sustainable use of inhalations of tobramycin solution in chronic *Pseudomonas aeruginosa* infection in children with cystic fibrosis. *Rossijskij Pediatricheskij Zhurnal = Russ Pediatr J*. 2015;18(3):50-53 (in Russian)

Hanssens L.S., Duchateau J., Casimir G.J. CFTR protein: not just a chloride channel? *Cells*. 2021;10(11):2844. doi 10.3390/cells10112844

Holkers M., Maggio I., Liu J., Janssen J.M., Miselli F., Mussolino C., Recchia A., Cathomen T., Gonçalves M.A. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res*. 2013;41(5):e63. doi 10.1093/nar/gks1446

Kashirskaya N.Yu., Kapranov N.I. Experience in the treatment of exocrine pancreatic insufficiency in cystic fibrosis in Russia. *Russ Med J*. 2011;19(7):476-484 (in Russian)

Kashirskaya N.Yu., Kapranov N.I. Modern pharmacotherapeutic approaches to the treatment of cystic fibrosis. *Farmateka*. 2014; 3(276):38-43 (in Russian)

Keating D., Marigowda G., Burr L., Daines C., Mall M.A., McKone E.F., Ramsey B.W., Rowe S.M., Sass L.A., Tullis E., McKee C.M., Moskowitz S.M., Robertson S., Savage J., Simard C., Van Goor F., Waltz D., Xuan F., Young T., Taylor-Cousar J.L.; VX16-445-001 Study Group. VX-445 – Tezacaftor-ivacaftor in patients with cystic fibrosis and one or two Phe508del alleles. *N Engl J Med*. 2018;379(17):1612-1620. doi 10.1056/NEJMoa1807120

Kerem E., Konstan M.W., De Boeck K., Accurso F.J., Sermet-Gaudelus I., Wilschanski M., Elborn J.S., Melotti P., Bronsveld I., Fajac I., Malfroot A., Rosenbluth D.B., Walker P.A., McColley S.A., Knoop C., Quattrucci S., Rietschel E., Zeitlin P.L., Barth J., Elfring G.L., Welch E.M., Branstrom A., Spiegel R.J., Peltz S.W., Ajayi T., Rowe S.M.; Cystic Fibrosis Ataluren Study Group. Ataluren for the treatment of nonsense-mutation cystic fibrosis: a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet Respir Med*. 2014;2(7):539-547. doi 10.1016/S2213-2600(14)70100-6

Koehler D.R., Sajjan U., Chow Y.H., Martin B., Kent G., Tanswell A.K., McKerlie C., Forstner J.F., Hu J. Protection of *Cftr* knockout mice from acute lung infection by a helper-dependent adenoviral vector expressing *Cftr* in airway epithelia. *Proc Natl Acad Sci USA*. 2003; 100(26):15364-15369. doi 10.1073/pnas.2436478100

Kondratieva E.I., Kashirskaya N.Yu., Kapranov N.I. (Eds.) Cystic Fibrosis: Definition, Diagnostic Criteria, Therapy. Moscow: BORGES Company Publ., 2018 (in Russian)

Krasnova M.G., Melianovskaya Y.L., Krasovskiy S.A., Bulatenko N.V., Efremova A.S., Bukharova T.B., Goldshtein D.V. Description of the clinical picture and assessment of functional activity of the CFTR channel in a patient with a complex allele [S466X; R1070Q]. *Pulmonologiya = Russ Pulmonology J*. 2023;33(2):233-242. doi 10.18093/0869-0189-2023-33-2-233-242 (in Russian)

Krasnovidova A.E., Simonova O.I., Chernevich V.P., Pakhomov A.V., Reykh A.P., Pushkov A.A. Genotype-phenotype correlation in siblings with cystic fibrosis. *Rossijskij Pediatricheskij Zhurnal = Russ Pediatr J*. 2023;26(3):159-167. doi 10.46563/1560-9561-2023-26-3-159-167 (in Russian)

Lee J.A., Cho A., Huang E.N., Xu Y., Quach H., Hu J., Wong A.P. Gene therapy for cystic fibrosis: new tools for precision medicine. *J Transl Med*. 2021;19(1):452. doi 10.1186/s12967-021-03099-4

Lomunova M.A., Gershovich P.M. Gene therapy for fibrosis: recent advances and future prospects. *Acta Naturae*. 2023;15(2):20-31. doi 10.32607/actanaturae.11708

Maule G., Arosio D., Cereseto A. Gene therapy for cystic fibrosis: progress and challenges of genome editing. *Int J Mol Sci*. 2020;21(11): 3903. doi 10.3390/ijms21113903

Moran O. On the structural organization of the intracellular domains of CFTR. *Int J Biochem Cell Biol*. 2014;52:7-14. doi 10.1016/j.biocel.2014.01.024

Olveira C., Padilla A., Dorado A., Contreras V., Garcia-Fuentes E., Rubio-Martin E., Porras N., Doña E., Carmona A., Olveira G. Inflammation and oxidation biomarkers in patients with cystic fibrosis: the influence of azithromycin. *Eurasian J Med*. 2017;49(2):118-123. doi 10.5152/eurasianjmed.2017.17010

OMIM.org [Internet]. Online Mendelian Inheritance in Man®. [cited 2023 Aug 28]. Available from: https://www.omim.org/

Piehler L., Thalemann R., Lehmann C., Thee S., Röhmel J., Syunyaeva Z., Stahl M., Mall M.A., Graeber S.Y. Effects of elexacaftor/tezacaftor/ivacaftor therapy on mental health of patients with cystic fibrosis. *Front Pharmacol*. 2023;14:1179208. doi 10.3389/fphar.2023.1179208

Rafeeq M.M., Murad H.A.S. Cystic fibrosis: current therapeutic targets and future approaches. *J Transl Med*. 2017;15(1):84-92. doi 10.1186/s12967-017-1193-9

Ren H.Y., Grove D.E., De La Rosa O., Houck S.A., Sopha P., Van Goor F., Hoffman B.J., Cyr D.M. VX-809 corrects folding defects in cystic fibrosis transmembrane conductance regulator protein through action on membrane-spanning domain 1. *Mol Biol Cell*. 2013;24(19):3016-3024. doi 10.1091/mbc.E13-05-0240

Rommens J.M., Iannuzzi M.C., Kerem B., Drumm M.L., Melmer G., Dean M., Rozmahel R., Cole J.L., Kennedy D., Hidaka N., Zsiga M., Buchwald M., Tsui L.-C., Riordan J.R., Collins F.S. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science*. 2006;245(4922):1059-1065. doi 10.1126/science.2772657

Sherman V.D., Kapranov N.I., Kashirskaya N.Yu. Dornase alpha (Pulmozyme) for the complex treatment of bronchopulmonary process in cystic fibrosis patients. *Farmateka*. 2011;11(224):42-45 (in Russian)

Simonova O.I., Gorinova Yu.V., Chernevich V.P. Cystic fibrosis: a breakthrough in 21st-century therapy. *Rossijskij Pediatricheskij Zhurnal = Russ Pediatr J*. 2020;23(1):35-41. doi 10.18821/1560-9561-2020-23-1-35-41 (in Russian)

Smirnikhina S.A., Lavrov V.A. Modern pathogenesis-based methods and development of new gene and cell-based methods for cystic fibrosis treatment. *Genes Cells*. 2018;13(3):23-31. doi 10.23868/201811029 (in Russian)

Smirnikhina S.A., Kondratyeva E.V., Anuchina A.A., Zaynitdinova M.I., Lavrov A.V. Modeling of cystic fibrosis in HEK293T cell culture and development of a method for the correction of F508del mutation. *Medicinskij Vestnik Severnogo Kavkaza = Medical News of North Caucasus*. 2020;15(2):158-162. doi 10.14300/mnnc.2020.15038 (in Russian)

Spielberg D.R., Clancy J.P. Cystic fibrosis and its management through established and emerging therapies. *Annu Rev Genomics Hum Genet*. 2016;17:155-175. doi 10.1146/annurev-genom-090314-050024

Sui H., Xu X., Su Y., Gong Z., Yao M., Liu X., Zhang T., Jiang Z., Bai T., Wang J., Zhang J., Xu C., Luo M. Gene therapy for cystic fibrosis: challenges and prospects. *Front Pharmacol*. 2022;13:1015926. doi 10.3389/fphar.2022.1015926

Suzuki S., Crane A.M., Anirudhan V., Barillà C., Matthias N., Randell S.H., Rab A., Sorscher E.J., Kerschner J.L., Yin S., Harris A., Mendel M., Kim K., Zhang L., Conway A., Davis B.R. Highly efficient gene editing of cystic fibrosis patient-derived airway basal cells results in functional CFTR correction. *Mol Ther*. 2020;28(7): 1684-1695. doi 10.1016/j.ymthe.2020.04.021

Taylor-Cousar J.L., Munck A., McKone E.F., van der Ent C.K., Moeller A., Simard C., Wang L.T., Ingenito E.P., McKee C., Lu Y., Lekstrom-Himes J., Elborn J.S. Tezacaftor-ivacaftor in patients with cystic fibrosis homozygous for Phe508del. *N Engl J Med*. 2017; 377(21):2013-2023. doi 10.1056/NEJMoa1709846

Van Goor F., Straley K.S., Cao D., González J., Hadida S., Hazlewood A., Joubran J., Knapp T., Makings L.R., Miller M., Neuberger T., Olson E., Panchenko V., Rader J., Singh A., Stack J.H., Tung R., Grootenhuis P.D., Negulescu P. Rescue of ΔF508-CFTR trafficking and gating in human cystic fibrosis airway primary cultures by small molecules. *Am J Physiol Lung Cell Mol Physiol*. 2006;290(6):L1117-L1130. doi 10.1152/ajplung.00169.2005

Van Goor F., Hadida S., Grootenhuis P.D., Burton B., Cao D., Neuberger T., Turnbull A., Singh A., Joubran J., Hazlewood A., Zhou J., McCartney J., Arumugam V., Decker C., Yang J., Young C., Olson E.R., Wine J.J., Frizzell R.A., Ashlock M., Negulescu P. Rescue of CF airway epithelial cell function in vitro by a CFTR potentiator, VX-770. *Proc Natl Acad Sci USA*. 2009;106(44):18825-18830. doi 10.1073/pnas.0904709106

Wainwright C.E., Elborn J.S., Ramsey B.W., Marigowda G., Huang X., Cipolli M., Colombo C., Davies J.C., De Boeck K., Flume P.A., Konstan M.W., McColley S.A., McCoy K., McKone E.F., Munck A., Ratjen F., Rowe S.M., Waltz D., Boyle M.P.; TRAFFIC Study Group; TRANSPORT Study Group. Lumacaftor-ivacaftor in patients with cystic fibrosis homozygous for Phe508del *CFTR*. *N Engl J Med*. 2015;373(18):1783-1784. doi 10.1056/NEJMc1510466

Wang G. Genome editing for cystic fibrosis. *Cells*. 2023;12(12):1555. doi 10.3390/cells12121555

Xia E., Zhang Y., Cao H., Li J., Duan R., Hu J. TALEN-mediated gene targeting for cystic fibrosis-gene therapy. *Genes* (*Basel*). 2019;10(1):39. doi 10.3390/genes10010039

Zainal Abidin N., Haq I.J., Gardner A.I., Brodlie M. Ataluren in cystic fibrosis: development, clinical studies and where are we now? *Expert Opin Pharmacother*. 2017;18(13):1363-1371. doi 10.1080/14656566.2017.1359255

Conflict of interest. The authors declare no conflict of interest.
Received August 16, 2024. Revised October 16, 2024. Accepted October 17, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):290-300

doi 10.18699/vjgb-25-32

The role of retroelements in Parkinson's disease development

R.N. Mustafin

Bashkir State Medical University, Ufa, Russia
ruji79@mail.ru

Abstract. Parkinson's disease is the second most common neurodegenerative disease characterized by accumulation of alpha-synuclein and Lewy bodies in the brain's substantia nigra. Genetic studies indicate an association of various SNPs, many of which are located in intergenic and intronic regions, where retrotransposons and non-coding RNA genes derived from them reside, with this disease. Therefore, we hypothesize the influence of SNPs in retroelement genes on Parkinson's disease development. A susceptibility factor is retrotransposons activation with age, since the disease is associated with aging. We hypothesized that alpha-synuclein accumulates in the brain due to its interaction with transcripts of activated retroelements. As a result of a defective antiviral response and a large number of RNA targets for this protein, its aggregates form Lewy bodies in neurons with inflammation and neurodegeneration development in the substantia nigra. As evidence, data are presented on the role of alpha-synuclein in the antiviral response with binding to RNA viruses, which are characterized by the ability to activate retroelements that have evolved from exogenous viruses integrated into the human genome. Activation of LINE1s in the brain, endogenous retroviruses, and LINE1s in the blood serum of Parkinson's disease patients was detected. An additional mechanism contributing to the progression of the disease is mitochondrial dysfunction due to insertions of Alu elements into their genomes using LINE1 enzymes. Mechanisms of activated retrotransposons' influence on microRNAs that evolved from them are described. Analysis of the scientific literature allowed us to identify 35 such microRNAs (miR-1246, -1249, -1271, -1273, -1303, -151, -211, -28, -31, -320b, -320d, -330, -335, - 342, -374a, -374b, -421, -4293, -4317, -450b, -466, -487b, -493, -495, -5095, -520d, -576, -585, -6088, -619, -625, -626, -769, -885, -95) associated with Parkinson's disease, which may become promising targets for its treatment and diagnosis.

Key words: Parkinson's disease; viruses; microRNA; retroelements

For citation: Mustafin R.N. The role of retroelements in Parkinson's disease development. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):290-300. doi 10.18699/vjgb-25-32

Роль ретроэлементов в развитии болезни Паркинсона

Р.Н. Мустафин

Башкирский государственный медицинский университет, Уфа, Россия
ruji79@mail.ru

Аннотация. Болезнь Паркинсона – второе по распространенности нейродегенеративное заболевание, характеризующееся накоплением альфа-синуклеина и телец Леви в черной субстанции головного мозга. Генетические исследования свидетельствуют об ассоциации с болезнью различных SNP, многие из которых расположены в межгенных и интронных областях, где локализованы также ретротранспозоны и произошедшие от них гены некодирующих РНК. В связи с этим сделано предположение о влиянии SNP в генах ретроэлементов на развитие болезни Паркинсона. Фактором предрасположенности является активация ретротранспозонов с возрастом, поскольку заболевание ассоциировано со старением. Предложена гипотеза о том, что альфа-синуклеин накапливается в головном мозге вследствие его взаимодействия с транскриптами активированных ретроэлементов. В результате дефектного противовирусного ответа и большого количества РНК-мишеней для данного белка его агрегаты образуют тельца Леви в нейронах с последующим воспалением черной субстанции и активацией нейродегенеративных процессов. В качестве доказательства приведены данные о роли альфа-синуклеина в противовирусном ответе со связыванием с РНК вирусов, которые характеризуются способностью активировать ретроэлементы, произошедшие в эволюции от встроенных в геном человека экзогенных вирусов. Обнаружены также активированные LINE1-ретроэлементы в головном мозге, эндогенные ретровирусы и LINE1 в сыворотке крови пациентов с болезнью Паркинсона. Дополнительный механизм, способствующий прогрессированию болезни, представляет собой дисфункция митохондрий вследствие инсерций в их геномы Alu-элементов с помощью ферментов LINE1. Описаны механизмы влияния активированных ретротранспозонов на произошедшие от них в эволюции микроРНК. Анализ научной литературы позволил выявить 35 таких микроРНК (miR-1246, -1249, -1271, -1273, -1303, -151, -211, -28, -31, -320b, -320d, -330, -335, -342, -374a, -374b, -421, -4293, -4317, -450b, -466, -487b, -493, -495, -5095, -520d, -576, -585, -6088, -619, -625, -626, -769, -885, -95), ассоциированных с болезнью Паркинсона, которые могут стать перспективными мишенями для ее лечения и диагностики.

Ключевые слова: болезнь Паркинсона; вирусы; микроРНК; ретроэлементы

Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disease after Alzheimer's disease, affecting 2 % of the world's population over 65 years of age (Morais et al., 2016). PD is characterized by the degeneration of dopaminergic neurons in the substantia nigra of the brain due to the accumulation of alpha-synuclein (AS) and Lewy bodies in them (Leblanc, Vorberg, 2022). This disease is characterized by prion-like spread of AS (Park et al., 2021). As a result, symptoms such as rigidity, tremors, gait disturbances, and slowness of movement progress clinically slowly. Subsequently, speech, gait, and the performance of daily activities are impaired, and dementia develops (Hossain et al., 2022). The overall heritability of PD risk ranges from 0.27 (Blauwendraat et al., 2019) to 0.36 (Nalls et al., 2019). In most cases, PD is a multifactorial disease associated with polymorphic variants of various genes (Blauwendraat et al., 2019). However, 10 % of patients with PD have monogenic forms of the disease, the most common cause of which are mutations in the *LRRK2* gene, which encodes leucine-rich repeat kinase (Oliveira et al., 2021).

A GWAS conducted in 2019 on DNA samples from 28,568 patients with PD identified more than 40 loci reliably associated with PD, including SNPs located in the *GBA, INPP5F/SCARB2, LRRK2, MCC1, SNCA, VPS13C* genes (Blauwendraat et al., 2019). In another GWAS of the same year, 78 PD-associated polymorphic loci were identified in 37,688 PD patients (Nalls et al., 2019). Most of these SNPs are located in intergenic, promoter and intronic regions (Ohnmacht et al., 2020), where the bulk of retroelement (REs) and non-coding RNA (ncRNA) genes are located (Nurk et al., 2022). Therefore, it can be assumed that the influence of many PD-associated polymorphisms is due to changes in the functioning of REs and ncRNAs, which play a role in regulating the expression of brain neuronal genes (Mustafin, Khusnutdinova, 2020). This is supported by both indirect and direct evidence of the role of REs in the pathogenesis of PD. In particular, the characteristic strong association of PD with aging (only 4 % of PD patients worldwide are under 50 years of age (Hossain et al., 2022)) may be due to the activation of REs during aging (Gorbunova et al., 2021) due to DNA methylation and heterochromatin destruction changes (Ravel-Godreuil et al., 2021).

REs are transposable elements (TEs), which are specific regions of the genome that move to new loci by a "copy and paste" mechanism. TEs also include another class, DNA transposons, which use a "cut and paste" mechanism (Gorbunova et al., 2021). In total, transposons occupy about 1.4 billion bp in the human genome, which is 46.7 % of all DNA sequences. The largest share is made up of autonomous LINEs (0.63 billion bp) that do not contain long terminal repeats (LTR) and non-autonomous SINEs (0.39 billion bp) containing LTR REs (human endogenous retroviruses (HERVs)), which make up 0.27 billion bp (Nurk et al., 2022). About 0.13 % of the human genome is occupied by non-autonomous SVA (SINE-VNTR-Alu) REs in the amount of about 3,000 elements (Fröhlich et al., 2024). DNA transposons occupy 0.108 billion bp (Nurk et al., 2022). REs are important sources of evolutionary emergence of ncRNAs such as microRNAs (Mustafin, Khusnutdinova, 2023). This may explain the results of the analysis of the human genome using specific oligonucleotides complementary to transposons, which showed that RE sequences (not only the REs themselves, but also the regulatory elements derived from them, introns, ncRNA genes and tandem repeats) occupy at least 2/3 of the entire human genome (de Koning et al., 2011).

List of abbreviations

AS – alpha-synuclein
GWAS – Genome Wide Association Study
HERV – Human Endogenous RetroVirus
HIV – Human Immunodeficiency Virus
HLA – Human Leukocyte Antigen
LINE – Long Interspersed Nuclear Element
LTR – Long Terminal Repeat, ncRNA – non-coding RNA
ncRNA – non-coding RNA
NHEJ – non-homologous end joining
ORF – Open Reading Frame
PD – Parkinson's disease
RC-LINE1 – retrotransposition-competent LINE1
RdDM – RNA-dependent DNA methylation
REs – retroelements
SINE – Short Interspersed Nuclear Element
siRNA – small interfering RNA
SNP – Single Nucleotide Polymorphism
SVA – SINE-VNTR-Alu
SV-SVA – structurally variable SVA
TEs – transposable elements
TLR3 – Toll-like receptor 3
WEEV – Western equine encephalitis virus
WNV – West Nile virus

The close relationship between the functioning of REs and the ncRNAs they generate in regulating gene expression suggests the role of transposons as drivers of epigenetic regulation. Therefore, the failure of evolutionarily programmed species-specific control due to individual RE sequence polymorphisms detected by GWAS (Nalls et al., 2019; Ohnmacht et al., 2020; Bantle et al., 2021) under the influence of aging (Gorbunova et al., 2021) and environmental factors (such as past viral infections (Jang et al., 2009; Batman et al., 2015; Marreiros et al., 2020; Park et al., 2021; Leblanc, Vorberg, 2022)) can cause epigenetic dysregulation in the brain, characterized by the most pronounced TEs activity (Mustafin, Khusnutdinova, 2020). As a result, a neurodegenerative process develops, in which the accumulation of AS and Lewy bodies may reflect a failure in the protective mechanisms of cells against hyperactivated REs, which is due to the role of AS in antiviral processes.

The role of alpha-synuclein in antiviral defense

REs evolved from exogenous viruses (Mustafin, 2018), which explains one of the modern concepts of aging being caused by hyperactivation of REs (Gorbunova et al., 2021), which stimulate the antiviral interferon response with the development of systemic aseptic inflammation, progressive degeneration of organs and tissues (De Cecco et al., 2019). Therefore, the role

of REs in the development of PD may be evidenced by both the influence of viruses on PD and the protective function of AS against viruses. Indeed, according to meta-analyses and systematic reviews of the scientific literature, PD is caused by influenza viruses, Coxsackie, HIV, Japanese encephalitis B, West Nile virus (WNV), St. Louis (Jang et al., 2009), influenza A viruses, herpes viruses and flaviviruses. An increased risk of developing PD after hepatitis B and C infections has been identified (Wang et al., 2020; Leblanc, Vorberg, 2022). Influenza A H1N1 virus has been found to promote proteostasis disruption and AS aggregation (Marreiros et al., 2020). Coxsackie virus B3 induces formation of AC-associated inclusion bodies in neurons acting as PD triggers (Park et al., 2021). Neuroinvasive WNV activates AS expression in neurons (Beatman et al., 2015).

A model was presented in which WNV-induced AS localized to endoplasmic reticulum membranes, modulating virus-induced stress signaling and inhibiting viral replication (Beatman et al., 2015). Experiments with infection of mice with the WEEV (western equine encephalitis virus) revealed protein aggregation in many areas of the brain, including the substantia nigra, with loss of dopaminergic neurons, persistent activation of microglia and astrocytes (Bantle et al., 2021). HIV promotes accumulation of AS in neurons, which explains the development of cognitive and motor disorders in HIV-infected patients, among whom the frequency of SNCA/alpha-synuclein staining is higher than in healthy people of the same age (Santerre et al., 2021).

AS has many biophysical characteristics of antiviral peptides, binding to virus-carrying vesicles. AS promotes neuronal resistance to viral infections by signaling the immune system and recruiting neutrophils, macrophages, and activating dendritic cells. It has been noted that chronic gastrointestinal infections can lead to the accumulation of AS forming neurotoxic aggregates, as from there AS enters the brain, providing immunity before infection (Barbut et al., 2019).

The mechanism of AS-induced immune responses to RNA viral infections was investigated and it was determined that AS is required for neuronal expression of interferon-stimulated genes. Human AS knockout neurons failed to induce a broad range of interferon-stimulated genes. In the nuclei of interferon-treated human neurons, AS accumulates, with interferon-mediated phosphorylation of STAT2 depending on its expression and localized together with AS after such stimulation. Increased levels of phosphoserine129 alpha-synuclein are expressed in brain tissue from patients with viral (WNV and VEEV) encephalitis (Monogue et al., 2022). A systematic review of the scientific literature in 2024 showed that SARS-CoV-2 induces AS aggregation, promoting the development of PD by stably binding alpha-synuclein to the S1 protein and activating AS as part of the immune response to infection (Iravanpour et al., 2024).

Direct role of transposable elements in the development of Parkinson's disease

AS plays a critical physiological role in immune responses and inflammation. Similar to amyloid-beta in Alzheimer's disease, AS fibrillation represents the brain's innate immunity against viruses (Vojtechova et al., 2022). Since REs have an evolutionary relationship with viruses (Mustafin, 2018), it can be assumed that mRNA of pathologically activated REs also contributes to the fibrillization of AS. This is evidenced by the results of a study of the abdominal cavity, in which AS is involved in the normal functioning of the immune system, being a mediator of immune responses and inflammation (Alam et al., 2022). Similar to exogenous viruses, degradation and processing products of Res transcripts are stimulators of the interferon response, which contributes to the development of inflammation (Gazquez-Gutierrez et al., 2021). This can be induced not only by LINE1, but also by non-autonomous Alu, which use the enzymes of activated LINE1 for their own transpositions (Elbarbary, Maquat, 2017). As a result, aseptic inflammation characteristic of aging develops (De Cecco et al., 2019), which has been detected in the brain of mice modeled for PD (Ghosh et al., 2016).

In the brain of patients with PD, activation of the immune cytokine network and increased levels of toll-like receptor 3 in response to double-stranded RNA are detected. A C3 complement antisense oligonucleotide, which switches splicing and promotes splicing of unproductive C3 mRNA, has been shown to prevent AS changes (Thomas et al., 2023). The accumulation of pathological AS aggregates (Lewy bodies) in PD may be due to the ineffectiveness of AS action on pathologically activated REs. In the normal brain, REs are also activated, but the interaction of proteins with them may play a role in specific functions of neurons and glial cells. However, in pathological interactions caused by the activation of REs that are not specific to certain structures of the brain (which is due to the spatiotemporal features of REs activation during neuronal differentiation (Mustafin, Khusnutdinova, 2020)), protein conglomerates are formed, especially under the influence of aging (Gorbunova et al., 2021), viruses (Jang et al., 2009; Beatman et al., 2015; Marreiros et al., 2020; Park et al., 2021; Leblanc, Vorberg, 2022) and in the presence of a genetic predisposition caused by polymorphisms in the loci of the location of TEs (Blauwendraat et al., 2019; Nalls et al., 2019; Ohnmacht et al., 2020) (Fig. 1).

Despite the enormous number of REs in the human genome, only a small fraction of them have retained the ability to transpose. This is due to the accumulation of many inactivating mutations during evolution, and the conservation of sequences is due to the use of retroelements by the "hosts" as sources of regulatory elements and ncRNA genes (Mustafin, Khusnutdinova, 2017). For example, LINE1s are distributed in the human genome as over 1 million copies, of which less than 100 have been confirmed to be capable of retrotransposition. Such REs are called "RC-LINE1" (retrotransposition competent LINE1). In addition to these RC-LINE1s, which are contained in the reference genome, there are a small number of non-reference LINE1 insertions (Pfaff et al., 2020).

However, the persistence of activity of even hundreds of REs causes significant insertional polymorphism between individuals, meaning the presence or absence of REs in certain regions of the human genome. Statistical analysis has shown that new Alu insertions occur in every 40th newborn, new LINE1 insertions, in every 63rd, and those of SVA, in every 63rd (Feusier et al., 2019). Whole-genome sequencing showed association of 16 highly active RC-LINE1s with PD compared to healthy controls (Pfaff et al., 2020). 81 reference SVAs were also identified that were polymorphic in presence

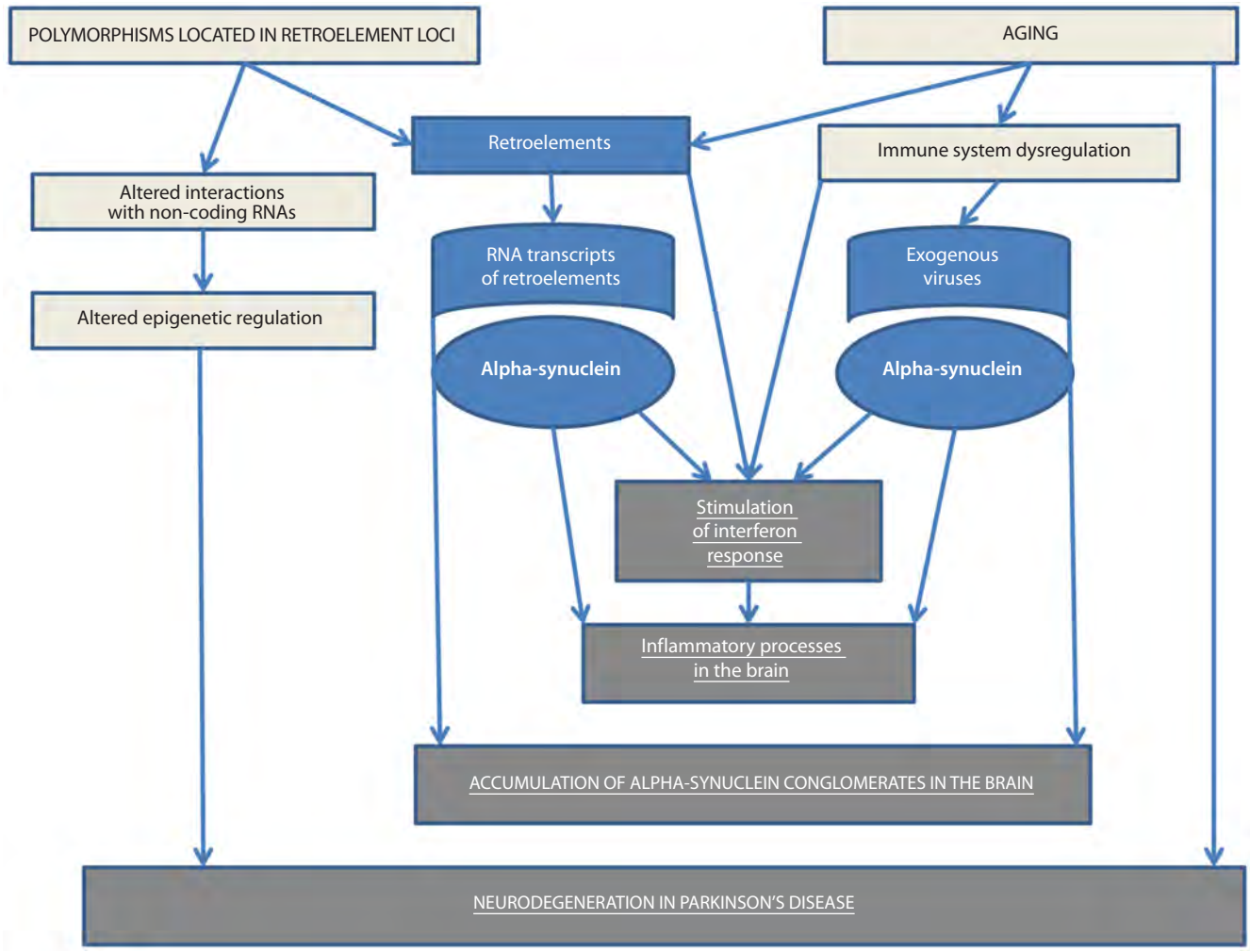


Fig. 1. Scheme of retroelements' involvement in Parkinson's disease pathogenesis.

or absence in PD patients, of which seven were associated with disease progression and PD-specific gene expression changes (Pfaff et al., 2021).

The presence or absence of human-specific SVA_67 correlates with PD progression. SVA_67 exerts a regulatory effect throughout the human genome, being polymorphic in its variable-number tandem repeat (VNTR) domain (Fröhlich et al., 2024). The analysis of polymorphic 2886 Alu, 360 L1, 128 SVA, which are not included in the reference human genome, by their presence or absence in PD compared with healthy controls allowed us to detect REs that have a significant effect on longitudinal changes in clinically significant outcomes of PD (Koks et al., 2022).

LINE1 insertional polymorphisms influence PD progression, as most novel LINE1 insertions are able to regulate gene expression *in trans*. An association with longitudinal changes in PD progression has been identified for 70 LINE1 markers of degeneration and disease severity (Fröhlich et al., 2023). Using bioinformatics studies and whole-genome sequencing data from 1,000 genomes from different populations, 46 polymorphic HERV-K insertions have been identified. Further analysis of experimental factor ontology enrichment has shown that polymorphic HERV-K insertions (rs12185268, rs17577094, rs17649553, rs183211, rs199515, rs199533, rs415430, rs8070723, rs2395163, rs9275326) are associated with PD features (Wallace et al., 2018).

Non-allelic recombination between homologous repeat elements Alu and LINE1 is widespread in the human genome with tissue-specific features that may act as recombination hotspots. An association between recombination of these REs and genomic instability in PD has been identified (Pascarella et al., 2022). REs are also the cause of most large deletions due to non-homologous end joining (NHEJ) in monogenic forms of PD caused by mutations in the *PARK2* gene (Morais et al., 2016). Structurally variable SVAs (SV-SVA) associated with PD and differential gene expression in this disease were identified, which are associated with SNP and differential expression of the *BCKDK* gene associated with the risk of developing PD. The *BCKDK* gene encodes branched-chain keto acid dehydrogenase kinase.

The minor risk allele rs14235, located in the *BCKDK* exon, is associated with a 1.36-fold increase in the mean number of Lewy bodies in PD (Van Bree et al., 2022). Experiments in En+/– mice, a model of PD, revealed loss of heterochromatin

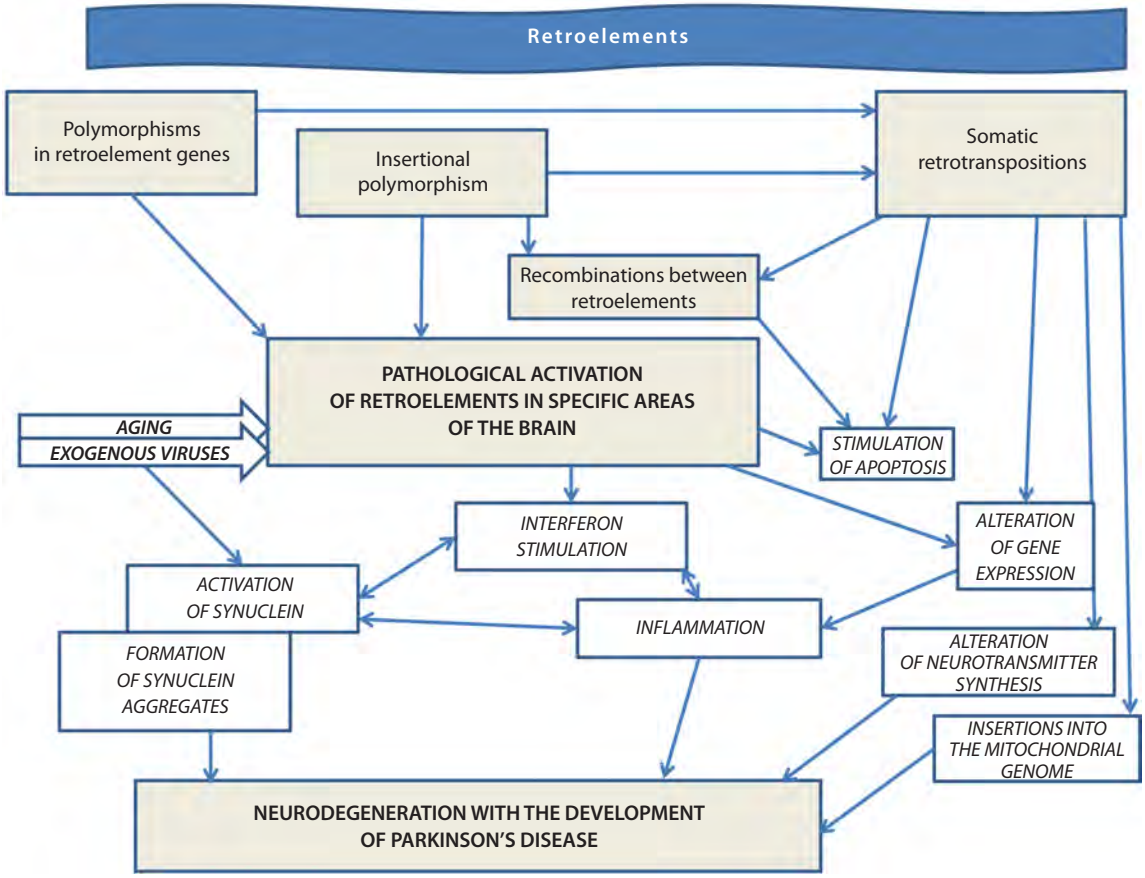


Fig. 2. Mechanisms of retroelements' influence on the development of Parkinson's disease.

and increased LINE1 expression in dopamine neurons. Degeneration of these cells was blocked by direct transcriptional repression using the nucleoside analogue reverse transcriptase inhibitor stavudine, LINE1-targeted small interfering RNAs and expression of viral Piwi1, as well as the specific protein Engrailed, which directly suppresses LINE1 in dopaminergic neurons. LINE1 activation promoted DNA double-strand breaks (Blaudin de Thé et al., 2018). In another study, overexpression of multifactorial protein Gadd45b, involved in DNA demethylation, was induced in the midbrain. In these mechanisms of neurodegeneration, DNA damage was preceded by activated LINE1s with changes characteristic of PD. It has been suggested that aging-related changes in the brain contribute to dopaminergic neurons degeneration with potential implications for PD (Ravel-Godreuil et al., 2021). REs are also sources of DNA damage during aging, which leads to neurodegeneration in PD (Peze-Heidsieck et al., 2022).

The development of PD is also influenced by somatic transpositions in the brain, which affect the biosynthesis of dopamine, serotonin, 3-methoxytyramine, homovanillate, phenethylamine and taurine (Abrusán, 2012). In PD patients, Alu integration into mitochondrial genomes disrupts populations of these organelles in neurons, contributing to the progression of neuronal dysfunction (Larsen et al., 2017). Inhibition of mitochondrial chain complex I in a PD model results in a significant increase in LINE1 element ORF1 protein expression in human dopaminergic LUHMES cells. Activation of these REs was accompanied by loss of DNA cytosine methylation. These mechanisms were blocked by the mitochondrial antioxidant phenothiazine. Such activation of LINE1 is a consequence of mitochondrial distress, which is characteristic of PD (Baeken et al., 2020).

A study of the SVA influence in the composition of the genes of the major histocompatibility complex HLA in patients with PD showed that the expressed alleles of the SVA and HLA genes in circulating leukocytes are differently coordinated in the regulation of immune responses, as well as in the progression of PD (Kulski et al., 2024). Thus, the development of PD can be influenced by structural polymorphisms in the REs genes, the characteristics of the distribution of REs in the genome, reflected in their recombinations and somatic transpositions (Fig. 2).

Role of retroelement-derived microRNAs in Parkinson's disease development

An analysis of the scientific literature on changes in the expression of microRNAs originating from REs (according to a published systematic review (Mustafin, Khusnutdinova, 2023)) in PD revealed 35 such microRNAs (see the Table).

Pathological activation of REs in PD may influence the expression of their derived microRNAs in several ways (Fig. 3). First, activated REs act as "sponges" for microRNAs by complementarily binding to nucleotide sequences due to their evolutionary relationship, thus blocking the effects of

Retroelement-derived microRNAs associated with Parkinson's disease

Source of microRNA	microRNA/ change in level in the disease/references	Function of microRNA/references
ERVL-MaRL	miR-1246/ increased/(Hossain et al., 2022)	Inhibits the expression of the *CKS2* (regulatory subunit of cyclin-dependent kinase 2), *TAPBP* (TAP-binding protein) genes/(Hossain et al., 2022)
LINE2	miR-1249/ increased/(Soreq et al., 2013)	Regulates the *VEGFA* and *HMGA2* genes/(Chen et al., 2019)
LINE2	miR-1271/ decreased/(Ma, Zhao, 2023)	Suppresses *PAX4*, *Grb2*, *NADPH* genes expression, inhibits the Wnt/beta-catenin pathways/(Ma, Zhao, 2023)
SINE/Alu	miR-1273/ decreased/(Kamenova et al., 2021)	Regulates *PDP2* gene expression/(Kamenova et al., 2021)
SINE/Alu	miR-1303/ decreased/(Boros et al., 2021)	Interacts with lncRNA NEAT1/(Boros et al., 2021)
LINE2	miR-151/ decreased/(Martins et al., 2011)	Regulates *CRK*, *FAM5C*, *RBM5*, *TWIST1* genes expression/ (Martins et al., 2011)
LINE2	miR-211/ increased/(Motawi et al., 2022)	Regulates *CHOP* gene expression/(Motawi et al., 2022)
LINE2	miR-28/ increased/(He S. et al., 2021)	Suppresses *FOXO* gene expression/(He S. et al., 2021)
LINE2	miR-31/ increased/(Li L. et al., 2021)	Regulates apoptosis by potentiating PI3K/AKT signaling/(Li L. et al., 2021)
LINE2	miR-320b/ decreased/(Soreq et al., 2013)	Inhibits the *FOXM1* gene (encodes a transcriptional activator that regulates cell proliferation)/(Jingyang et al., 2021)
LINE1	miR-320d/ decreased/(Chatterjee, Roy, 2017)	Suppresses the expression of TUSC3 (tumor suppressor)/ (Yufeng et al., 2021)
SINE/MIR	miR-330/ increased/(Ravandis et al., 2020)	Targets mRNAs of proteins involved in activity-dependent synaptic plasticity in the hippocampus/(Ravandis et al., 2020)
SINE/MIR	miR-335/ decreased/(Oliveira et al., 2021)	Suppresses *LRRK2* gene expression/(Oliveira et al., 2021)
SINE/tRNA-RTE	miR-342/ increased/(Wu et al., 2019)	Suppresses the expression of *PAK1*, *GLT1*, *GLAST*, *TH* genes, Wnt signaling pathways and anti-apoptotic genes/(Wu et al., 2019)
LINE2	miR-374a/ increased/(He S. et al., 2021)	Inhibits translation of the *Wnt5a* gene mRNA/(Sun et al., 2018)
LINE2	miR-374b/ increased/(He S. et al., 2021)	Inhibits translation of the *Wnt5a* gene mRNA/(Sun et al., 2018)
LINE2	miR-421/ increased/(Dong et al., 2021)	Inhibits translation of the mRNA of the *MEF2D* gene (encodes myocyte-specific enhancer factor 2)/(Dong et al., 2021)
SINE/tRNA	miR-4293/ decreased/(Soreq et al., 2013)	Inhibits *WFDC21P* gene expression/(Zhang Q. et al., 2021)
SINE/MIR	miR-4317/ increased/(Soreq et al., 2013)	Inhibits *FGF9* and *CCND2* genes expression/(He X. et al., 2018)
LINE1	miR-450b/ increased/(Khoo et al., 2012)	Inhibits the *KIF26B* gene (encodes an intracellular protein that transports organelles along microtubules)/(Li H. et al., 2019)
LINE1	miR-466/ increased/(Kamenova et al., 2021)	Inhibits *PPARGC1A* and *GSK3B* genes expression/(Kamenova et al., 2021)
SINE/MIR	miR-487b/ decreased/(Kern et al., 2021)	Suppresses inflammation and neuronal apoptosis by targeting the mRNA of the *Ifitm3* gene/(Tong et al., 2022)
LINE2	miR-493/ decreased/(Kern et al., 2021)	Directly affects mRNA of the *Wnt5A* gene, inhibits p-PI3K/p-AKT and c-JUN with an increase in p21/(Bian et al., 2021)
ERV-L	miR-495/ increased/(Ravandis et al., 2020)	Inhibits the expression of the *CDK1* gene encoding the serine/threonine protein kinase factor G2/M transition in the cell cycle/(Tang et al., 2021)

Table (end)

Source of microRNA	microRNA/ change in level in the disease/references	Function of microRNA/references
SINE/Alu	miR-5095/ increased/(Kamenova et al., 2021)	Inhibits the expression of the *LRP10*, *PRKN*, *RBBP5*, *SLC14A1* genes/ (Kamenova et al., 2021)
SINE/Alu	miR-520d/ increased/(Jin et al., 2018)	Inhibits ceruloplasmin expression/(Jin et al., 2018)
LINE1	miR-576/ increased/(Liu et al., 2023)	Inhibits the expression of the *SGK1* gene, which encodes serine/threonine protein kinase, responsible for stress responses and neuronal excitability/ (Greenawalt et al., 2019)
ERV-L/MaLR	miR-585/ increased/(Zhang Y. et al., 2020)	Regulates PIK3R3 (phosphatidylinositol 3-kinase), influencing apoptosis/ (Zhang Y. et al., 2020)
SINE/Alu	miR-6088/ increased/(Marsh et al., 2016)	Regulates DNA polymerase eta (POLH)/(Sonobe et al., 2024)
LINE1	miR-619/ increased/(Cai et al., 2021)	Inhibits the expression of the *LRP10*, *PRKN*, *RBBP5*, *SLC14A1* genes/ (Kamenova et al., 2021)
LINE1	miR-625/ decreased/(Zhong et al., 2023)	Inhibits the expression of the *HMGA1* gene/(Zhong et al., 2023)
LINE1	miR-626/ decreased/(Qin et al., 2021)	Inhibits the expression of the *LRRK2* gene/(Qin et al., 2021)
LINE/CR1	miR-769/ decreased/(Soreq et al., 2013)	Regulates *HEY1* gene expression (encodes a protein of the helix-loop-helix family of basic transcriptional repressors)/(Han et al., 2018)
SINE/MIR	miR-885/ increased/(Behbahanipour et al., 2019)	Inhibits *IGF-1* expression by affecting the PI3K/Atk/GSK-3β, CTNNB1 (key regulatory protein of Wnt signaling) signaling pathways/ (Behbahanipour et al., 2019)
LINE2	miR-95/ increased/(Nair, Ge, 2016)	Regulates the expression of genes of glutamate ionotropic receptors *GR1D1* and *GR1A2*, metabotropic receptors GRM4/(Nair, Ge, 2016)

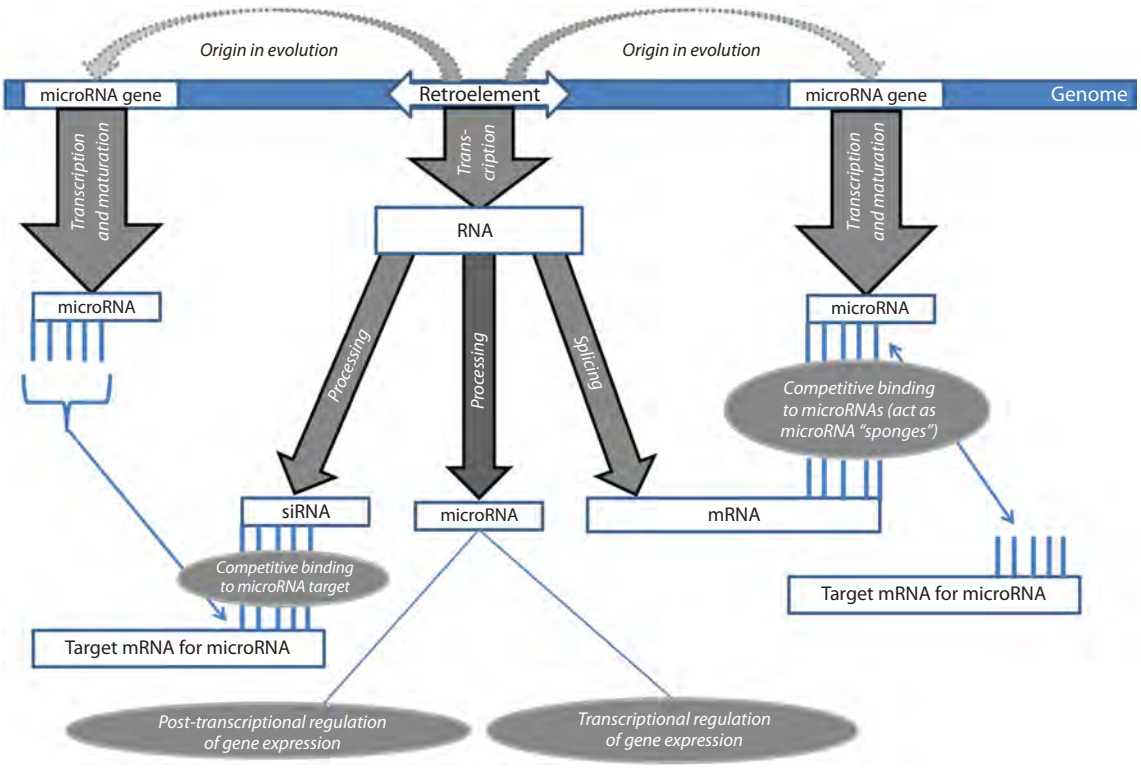


Fig. 3. Scheme of the pathways of influence of retroelements on microRNAs derived from them.

RNA interference on the mRNAs of the target genes of these microRNAs (Cornec, Poirier, 2023). This regulatory principle has been identified not only in animals but also in plants. For example, the transcript of the LTR-containing retroelement *MIKKI* (translated from Korean as "bait"), expressed in rice roots, is a mimic for miR-171, which destabilizes the mRNA of root transcription factors like SCARECROW. Processed *MIKKI* transcripts act as decoys for miR-171, triggering their degradation and promoting the accumulation of root-specific mRNA transcription factors (Cho, Paszkowski, 2017).

Second, LTR-containing REs transcripts (Lu et al., 2014) and LINE1s function as long ncRNA molecules, interacting with specific chromatin regions and regulating the expression of genes controlled by microRNA molecules (Honson, Macfarlan, 2018).

Third, some miRNAs are formed directly from REs genes, which are the basis for pre-miRNA hairpin structures. Such miRNAs lead to spatiotemporal dynamic expression networks, for the analysis of which the Brain miRTExplorer web application was created (Playfoot et al., 2022). Therefore, pathological activation of REs leads to the formation of various microRNAs from their transcripts, which affect the regulatory networks of other microRNAs in the body.

Fourth, REs exert regulatory effects on miRNAs by generating small interfering RNAs (siRNAs) from REs transcripts. In these mechanisms, siRNAs are competitive molecules for binding to mRNA targets of microRNAs, neutralizing their effect on gene expression. This effect is associated with the host cells' defense systems against activated REs in their genomes, triggering the degradation of REs transcripts by ribonucleases to miRNAs. The latter exert post-transcriptional inhibition of gene mRNAs due to partial complementarity (McCue et al., 2013).

Fifth, one of the ways in which microRNAs interact with REs in regulating gene activity is also the suppression of their expression when microRNAs bind to specific DNA structures formed by REs embedded in these regions.

In the human genome, the Z-form of DNA is formed by endogenous retroviruses, which provide functional genes with alternative promoters (Lee et al., 2022). In addition, the phenomenon of RNA-directed DNA methylation (RdDM) has been described in humans, due to which microRNAs (Playfoot et al., 2022) and miRNAs (McCue et al., 2013) formed from REs transcripts can affect the expression of REs through complementary interactions of sequences in the genome structure (Chalertpet et al., 2019).

Conclusion

The data presented in the review suggest that the development of PD is caused by the activation of REs as a result of individual characteristics of their distribution and the presence of polymorphisms associated with PD in them. This is evidenced by the following:

1) The results of scientific studies on the association of specific RC-LINE1 sets with PD were obtained.
2) The influence of LINE1 insertional polymorphism on the development of PD was revealed.
3) The significance of 360 LINE1s, 128 SVAs and 2886 Alu in the progression of PD was determined.
4) PD is associated with aging, which is characterized by the activation of REs and the associated inflammation and neurodegeneration.
5) 35 RE-derived microRNAs, the expression of which was significantly altered in PD, were identified.
6) The role of Alu distribution in the genome as a source of mutations in PD was discovered.
7) The influence of Alu insertions into mitochondrial genomes on the progression of PD was determined.
8) The role of synuclein in antiviral protection, with the influence of viruses on the formation of aggregates of this protein, was described.

Similarly, transcripts of pathologically activated REs, evolutionarily related to and interacting with exogenous and viral REs, can stimulate synuclein expression and fibrillization. The probable cause of damage to the substantia nigra is the spatiotemporal features of activation of specific REs in neurons of the brain, which is reflected in the results of their pathological activation in certain most vulnerable areas.

References

Abrusán G. Somatic transposition in the brain has the potential to influence the biosynthesis of metabolites involved in Parkinson's disease and schizophrenia. *Biol Direct*. 2012;7:41. doi 10.1186/1745-6150-7-41

Alam M.M., Yang D., Li X.-Q., Liu J., Back T.C., Trivett A., Karim B., Barbut D., Zasloff M., Oppenheim J.J. Alpha synuclein, the culprit in Parkinson disease, is required for normal immune function. *Cell Rep*. 2022;38(2):110090. doi 10.1016/j.celrep.2021.110090

Baeken M.W., Moosmann B., Hajieva P. Retrotransposon activation by distressed mitochondria in neurons. *Biochem Biophys Res Commun*. 2020;525(3):570-575. doi 10.1016/j.bbrc.2020.02.106

Bantle C.M., Rocha S.M., French C.T., Phillips A.T., Tran K., Olson K.E., Bass T.A., Aboellail T., Smeyne R.J., Tjalkens R.B. Astrocyte inflammatory signaling mediates α-synuclein aggregation and dopaminergic neuronal loss following viral encephalitis. *Exp Neurol*. 2021;346:113845. doi 10.1016/j.expneurol.2021.113845

Barbut D., Stolzenberg E., Zasloff M. Gastrointestinal immunity and alpha-synuclein. *J Parkinsons Dis*. 2019;9(s2):S313-S322. doi 10.3233/JPD-191702

Beatman E.L., Massey A., Shives K.D., Burrack K.S., Chamanian M., Morrison T.E., Beckham J.D. Alpha-synuclein expression restricts RNA viral infections in the brain. *J Virol*. 2015;90(6):2767-82. doi 10.1128/JVI.02949-15

Behbahanipour M., Peymani M., Salari M., Hashemi M.S., Nasr-Esfahani M.H., Ghaedi K. Expression profiling of blood microRNAs 885, 361, and 17 in the patients with the Parkinson's disease: integrating interaction data to uncover the possible triggering age-related mechanisms. *Sci Rep*. 2019;9(1):13759. doi 10.1038/s41598-019-50256-3

Bian W., Li Y., Zhu H., Gao S., Niu R., Wang C., Zhang H., Qin X., Li S. miR-493 by regulating of c-Jun targets Wnt5a/PD-L1-inducing esophageal cancer cell development. *Thorac Cancer*. 2021;12(10):1579-1588. doi 10.1111/1759-7714.13950

Blaudin de Thé F.X., Rekaik H., Peze-Heidsieck E., Massiani-Beaudoin O., Joshi R.L., Fuchs J., Prochiantz A. Engrailed homeoprotein blocks degeneration in adult dopaminergic neurons through LINE-1 repression. *EMBO J*. 2018;37(15):e97374. doi 10.15252/embj.201797374

Blauwendraat C., Heilbron K., Vallerga C.L., Bandres-Ciga S., von Coelln R., Pihlstrom L., Simón-Sánchez J., Schulte C., Sharma M., Krohn L., Siitonen A., Iwaki H., Leonard H., Noyce A.J., Tan M., Gibbs J.R., Nalls M.A., Singleton A.B.; International Parkinson's Disease Genomics Consortium (IPDGC). Parkinson's disease age at onset genome-wide association study: defining heritability, genetic

loci, and α-synuclein mechanisms. *Mov Disord*. 2019;34(6):866-875. doi 10.1002/mds.27659

Boros F.A., Vecsei L., Klivenyi P. NEAT1 on the field of Parkinson's disease: offence, defense, or a player on the bench? *J. Parkinson's Dis*. 2021;11(1):123-138. doi 10.3233/JPD-202374

Cai M., Chai S., Xiong T., Wei J., Mao W., Zhu Y., Li X., Wei W., Dai X., Yang B., Liu W., Shu B., Wang M., Lu T., Cai Y., Zheng Z., Mei Z., Zhou Y., Yang J., Zhao J., Shen L., Ho J.W.K., Chen J., Xiong N. Aberrant expression of circulating microRNA leads to the dysregulation of alpha-synuclein and other pathogenic genes in Parkinson's disease. *Front Cell Dev Biol*. 2021;9:695007. doi 10.3389/fcell.2021.695007

Chalertpet K., Pin-On P., Aporntewan C., Patchsung M., Ingrungruanglert P., Israsena N., Mutirangura A. Argonaute 4 as an effector protein in RNA-directed DNA methylation in human cells. *Front Genet*. 2019;10:645. doi 10.3389/fgene.2019.00645

Chatterjee P., Roy D. Comparative analysis of RNA-Seq data from brain and blood samples of Parkinson's disease. *Biochem Biophys Res Commun*. 2017;484(3):557-564. doi 10.1016/j.bbrc.2017.01.121

Chen X., Zeng K., Xu M., Liu X., Hu X., Xu T., He B., Pan Y., Sun H., Wang S. P53-induced miR-1249 inhibits tumor growth, metastasis, and angiogenesis by targeting VEGFA and HMGA2. *Cell Death Dis*. 2019;10(2):131. doi 10.1038/s41419-018-1188-3

Cho J., Paszkowski J. Regulation of rice root development by a retrotransposon acting as a microRNA sponge. *eLife*. 2017;6:e30038. doi 10.7554/eLife.30038

Cornec A., Poirier E.Z. Interplay between RNA interference and transposable elements in mammals. *Front Immunol*. 2023;14:1212086. doi 10.3389/fimmu.2023.1212086

De Cecco M., Ito T., Petrashen A.P., Elias A.E., Skvir N.J., Sedivy J.M. L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature*. 2019;566(7742):73-78. doi 10.1038/s41586-018-0784-9

de Koning A.P., Gu W., Castoe T.A., Batzer M.A., Pollock D.D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet*. 2011;7:e1002384. doi 10.1371/journal.pgen.1002384

Dong Y., Xiong J., Ji L., Xue X. MiR-421 aggravates neurotoxicity and promotes cell death in Parkinson's disease models by directly targeting MEF2D. *Neurochem Res*. 2021;46(2):299-308. doi 10.1007/s11064-020-03166-0

Elbarbary R.A., Maquat L.E. Distinct mechanisms obviate the potentially toxic effects of inverted-repeat Alu elements on cellular RNA metabolism. *Nat Struct Mol Biol*. 2017;24(6):496-498. doi 10.1038/nsmb.3416

Feusier J., Watkins W.S., Thomas J., Farrell A., Witherspoon D.J., Baird L., Ha H., Xing J., Jorde L.B. Pedigree-based estimation of human mobile element retrotransposition rates. *Genome Res*. 2019;29(10):1567-1577. doi 10.1101/gr.247965.118

Fröhlich A., Pfaff A.L., Bubb V.J., Quinn J.P., Koks S. Reference LINE-1 insertion polymorphisms correlate with Parkinson's disease progression and differential transcript expression in the PPMI cohort. *Sci Rep*. 2023;13(1):13857. doi 10.1038/s41598-023-41052-1

Fröhlich A., Pfaff A.L., Middlehurst B., Hughes L.S., Bubb V.J., Quinn J.P., Koks S. Deciphering the role of a SINE-VNTR-Alu retrotransposon polymorphism as a biomarker of Parkinson's disease progression. *Sci Rep*. 2024;14(1):10932. doi 10.1038/s41598-024-61753-5

Gazquez-Gutierrez A., Witteveldt J., Heras S., Macias S.R. Sensing of transposable elements by the antiviral innate immune system. *RNA*. 2021;27(7):735-752. doi 10.1261/rna.078721.121

Ghosh A., Tyson T., George S., Hildebrandt E.N., Steiner J.A., Madaj Z., Schulz E., Machiela E., McDonald W.G., Escobar Galvis M.L., Kordower J.H., Van Raamsdonk J.M., Colca J.R., Brundin P. Mitochondrial pyruvate carrier regulates autophagy, inflammation, and neurodegeneration in experimental models of Parkinson's disease. *Sci Transl Med*. 2016;8(368):368ra174. doi 10.1126/scitranslmed.aag2210

Gorbunova V., Seluanov A., Mita P., McKerrow W., Fenyö D., Boeke J.D., Linker S.B., Gage F.H., Kreiling J.A., Petrashen A.P., Woodham T.A., Taylor J.R., Helfand S.L., Sedivy J.M. The role of retrotransposable elements in ageing and age-associated diseases. *Nature*. 2021;596(7870):43-53. doi 10.1038/s41586-021-03542-y

Greenawalt E.J., Edmonds M.D., Jain N., Adams C.M., Mitra R., Eischen C.M. Targeting of SGK1 by miR-576-3p inhibits lung adenocarcinoma migration and invasion. *Mol Cancer Res*. 2019;17(1):289-298. doi 10.1158/1541-7786.MCR-18-0364

Han C., Song Y., Lian C. MiR-769 inhibits colorectal cancer cell proliferation and invasion by targeting HEY1. *Med Sci Monit*. 2018;24:9232-9239. doi 10.12659/MSM.911663

He S., Huang L., Shao C., Nie T., Xia L., Cui B., Lu F., Zhu L., Chen B., Yang Q. Several miRNAs derived from serum extracellular vesicles are potential biomarkers for early diagnosis and progression of Parkinson's disease. *Transl Neurodegener*. 2021;10(1):25. doi 10.1186/s40035-021-00249-y

He X., Chen S.Y., Yang Z., Zhang J., Wang W., Liu M.Y., Niu Y., Wei X.M., Li H.M., Hu W.N., Sun G.G. miR-4317 suppresses non-small cell lung cancer (NSCLC) by targeting fibroblast growth factor 9 (FGF9) and cyclin D2 (CCND2). *J Exp Clin Cancer Res*. 2018;37(1):230. doi 10.1186/s13046-018-0882-4

Honson D.D., Macfarlan T.S. A lncRNA-like role for LINE1s in development. *Dev Cell*. 2018;46:132-134. doi 10.1016/j.devcel.2018.06.022

Hossain M.B., Islam M.K., Adhikary A., Rahaman A., Islam M.Z. Bioinformatics approach to identify significant biomarkers, drug targets shared between Parkinson's disease and bipolar disorder: a pilot study. *Bioinform Biol Insights*. 2022;16:11779322221079232. doi 10.1177/11779322221079232

Iravanpour F., Farrokhi M.R., Jafarinia M., Oliaee R.T. The effect of SARS-CoV-2 on the development of Parkinson's disease: the role of α-synuclein. *Hum Cell*. 2024;37(1):1-8. doi 10.1007/s13577-023-00988-2

Jang H., Boltz D.A., Webster R.G., Smeyne R.J. Viral parkinsonism. *Biochim Biophys Acta*. 2009;1792(7):714-721. doi 10.1016/j.bbadis.2008.08.001

Jin L., Wan W., Wang L., Wang C., Xiao J., Zhang F., Zhao J., Wang J., Zhan C., Zhong C. Elevated microRNA-520d-5p in the serum of patients with Parkinson's disease, possibly through regulation of ceruloplasmin expression. *Neurosci Lett*. 2018;687:88-93. doi 10.1016/j.neulet.2018.09.034

Jingyang Z., Jinhui C., Lu X., Weizhong Y., Yunjiu L., Haihong W., Wuyuan Z. Mir-320b inhibits pancreatic cancer cell proliferation by targeting FOXM1. *Curr Pharm Biotechnol*. 2021;22(8):1106-1113. doi 10.2174/1389201021999200917144704

Kamenova S., Aralbayeva A., Kondybayeva A., Akimniyazova A., Pyrkova A., Ivashchenko A. Evolutionary changes in the interactions of miRNA with mRNA of candidate genes for Parkinson's disease. *Front Genet*. 2021;12:647288. doi 10.3389/fgene.2021.647288

Kern F., Fehlmann T., Violich I., Alsop E., Hutchins E., Kahraman M., Grammes N.L., Guimarães P., Backes C., Poston K.L., Casey B., Balling R., Geffers L., Krüger R., Galasko D., Mollenhauer B., Meese E., Wyss-Coray T., Craig D.W., Van Keuren-Jensen K., Keller A. Deep sequencing of sncRNAs reveals hallmarks and regulatory modules of the transcriptome during Parkinson's disease progression. *Nat Aging*. 2021;1(3):309-322. doi 10.1038/s43587-021-00042-6

Khoo S.K., Petillo D., Kang U.J., Resau J.H., Berryhill B., Linder J., Forsgren L., Neuman L.A., Tan A.C. Plasma-based circulating MicroRNA biomarkers for Parkinson's disease. *J Parkinsons Dis*. 2012;2(4):321-331. doi 10.3233/JPD-012144

Koks S., Pfaff A.L., Singleton L.M., Bubb V.J., Quinn J.P. Non-reference genome transposable elements (TEs) have a significant impact on the progression of the Parkinson's disease. *Exp Biol Med* (*Maywood*). 2022;247(18):1680-1690. doi 10.1177/15353702221117147

Kulski J.K., Suzuki S., Shiina T., Pfaff A.L., Kõks S. Regulatory SVA retrotransposons and classical HLA genotyped-transcripts asso-

ciated with Parkinson's disease. *Front Immunol*. 2024;15:1349030. doi 10.3389/fimmu.2024.1349030

Larsen P.A., Lutz M.W., Hunnicutt K.E., Mihovilovic M., Saunders A.M., Yoder A.D., Roses A.D. The *Alu* neurodegeneration hypothesis: a primate-specific mechanism for neuronal transcription noise, mitochondrial dysfunction, and manifestation of neurodegenerative disease. *Alzheimers Dement.* 2017;13(7):828-838. doi 10.1016/j.jalz.2017.01.017

Leblanc P., Vorberg I.M. Viruses in neurodegenerative diseases: more than just suspects in crimes. *PLoS Pathog*. 2022;18(8):e1010670. doi 10.1371/journal.ppat.1010670

Lee D.H., Bae W.H., Ha H., Park E.G., Lee Y.J., Kim W.R., Kim H.S. Z-DNA-containing long terminal repeats of human endogenous retrovirus families provide alternative promoters for human functional genes. *Mol Cells*. 2022;45(8):522-530. doi 10.14348/molcells.2022.0060

Li H., Shen S., Chen X., Ren Z., Li Z., Yu Z. miR-450b-5p loss mediated KIF26B activation promoted hepatocellular carcinoma progression by activating PI3K/AKT pathway. *Cancer Cell Int.* 2019;19:205. doi 10.1186/s12935-019-0923-x

Li L., Ren J., Pan C., Li Y., Xu J., Dong H., Chen Y., Liu W. Serum miR-214 serves as a biomarker for prodromal Parkinson's disease. *Front Aging Neurosci.* 2021;13:700959. doi 10.3389/fnagi.2021.700959

Liu T., Yang Z., Liu S., Wei J. Parkinson's disease as a risk factor for prostate adenocarcinoma: a molecular point of view. *Gerontology*. 2023;69(8):986-1001. doi 10.1159/000530088

Lu X., Sachs F., Ramsay L., Jacques P.E., Goke J., Bourque G., Ng H.H. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol*. 2014;21(4):423-425. doi 10.1038/nsmb.2799

Ma Y.M., Zhao L. Mechanism and therapeutic prospect of miRNAs in neurodegenerative diseases. *Behav Neurol.* 2023;2023:8537296. doi 10.1155/2023/8537296

Marreiros R., Muller-Schiffmann A., Trossbach S.V., Prikulis I., Hansch S., Weidtkamp-Peters S., Moreira A.R., Sahu S., Soloviev I., Selvarajah S., Lingappa V.R., Korth C. Disruption of cellular proteostasis by H1N1 influenza A virus causes alpha-synuclein aggregation. *Proc Natl Acad Sci USA.* 2020;117(12):6741-6751. doi 10.1073/pnas.1906466117

Marsh A.G., Cottrell M.T., Goldman M.F. Epigenetic DNA methylation profiling with MSRE: a quantitative NGS approach using a Parkinson's Disease test case. *Front Genet*. 2016;7:191. doi 10.3389/fgene.2016.00191

Martins M., Rosa A., Guedes L.C., Fonseca B.V., Oliveira S.A. Convergence miRNA expression profiling, α-synuclein interaction and GWAS in Parkinson's disease. *PLoS One*. 2011;6(10):e25443. doi 10.1371/journal.pone.0025443

McCue A.D., Nuthikattu S., Slotkin R.K. Genome-wide identification of genes regulated in trans by transposable element small interfering RNAs. *RNA Biol*. 2013;10:1379-1395. doi 10.4161/rna.25555

Monogue B., Chen Y., Sparks H., Behbehani R., Chai A., Rajic A.J., Massey A., Kleinschmidt-Demasters B.K., Vermeren M., Kunath T., Beckham J.D. Alpha-synuclein supports type 1 interferon signalling in neurons and brain tissue. *Brain*. 2022;145(10):3622-3636. doi 10.1093/brain/awac192

Morais S., Bastos-Ferreira R., Sequeiros J., Alonso I. Genomic mechanisms underlying *PARK2* large deletions identified in a cohort of patients with PD. *Neurol Genet*. 2016;2:e73. doi 10.1212/NXG.0000000000000073

Motawi T.K., Al-Kady R.H., Abdelraouf S.M., Senousy M.A. Empagliflozin alleviates endoplasmic reticulum stress and augments autophagy in rotenone-induced Parkinson's disease in rats: Targeting the GRP78/PERK/eIF2α/CHOP pathway and miR-211-5p. *Chem Biol Interact.* 2022;362:110002. doi 10.1016/j.cbi.2022.110002

Mustafin R.N. The hypothesis of the origin of viruses from transposons. *Mol. Genet. Microbiol. Virol.* 2018;33(4):223-232. doi 10.3103/S0891416818040067

Mustafin R.N., Khusnutdinova E.K. Involvement of transposable elements in neurogenesis. *Vavilov J Genet Breed*. 2020;24(2):209-218. doi 10.18699/VJ20.613

Mustafin R.N., Khusnutdinova E.K. Non-coding parts of genomes as the basis of epigenetic heredity. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2017;2017;21(6):742-749. doi 10.18699/VJ17.30-o (in Russian)

Mustafin R.N., Khusnutdinova E. Perspective for studying the relationship of miRNAs with transposable elements. *Curr Issues Mol Biol*. 2023;45(4):3122-3145. doi 10.3390/cimb45040204

Nair V.D., Ge Y. Alterations of miRNAs reveal a dysregulated molecular regulatory network in Parkinson's disease striatum. *Neurosci Lett.* 2016;629:99-104. doi 10.1016/j.neulet.2016.06.061

Nalls M.A., Blauwendraat C., Vallerga C.L., Heilbron K., Bandres-Ciga S., Chang D., Tan M., Iwaki H.; 23andMe Research Team; System Genomics of Parkinson's Disease Consortium; International Parkinson's Disease Genomics Consortium. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2019;18(12):1091-1102. doi 10.1016/S1474-4422(19)30320-5

Nurk S., Koren S., Rhie A., Rautiainen M., Bzikadze A.V., Mikheenko A., Vollger M.R., Altemose N., Uralsky L., Gershman A., Miga K.H., Philippy A.M. The complete sequence of a human genome. *Science*. 2022;376(6588):44-53. doi 10.1126/science.abj6987

Ohnmacht J., May P., Sinkkonen L., Krüger R. Missing heritability in Parkinson's disease: the emerging role of non-coding genetic variation. *J Neural Transm* (*Vienna*). 2020;127(5):729-748. doi 10.1007/s00702-020-02184-0

Oliveira S.R., Dionísio P.A., Gaspar M.M., Correia Guedes L., Coelho M., Rosa M.M., Ferreira J.J., Amaral J.D., Rodrigues C.M.P. miR-335 targets LRRK2 and mitigates inflammation in Parkinson's disease. *Front Cell Dev Biol.* 2021;9:661461. doi 10.3389/fcell.2021.661461

Park S.J., Jin U., Park S.M. Interaction between coxsackievirus B3 infection and alpha-synuclein in models of Parkinson's disease. *PLoS Pathog*. 2021;17(10):e1010018. doi 10.1371/journal.ppat.1010018

Pascarella G., Hon C.C., Hashimoto K., Buscho A., Luginbuhl J., Parr C., Yip W.H., Abe K., Kratz A., Bonetti A., Agostini F., Severin J., Murayama S., Suzuki Y., Gustincich S., Frith M., Carninci P. Recombination of repeat elements generates somatic complexity in human genomes. *Cell*. 2022;185(16):3025-3040.e6. doi 10.1016/j.cell.2022.06.032

Peze-Heidsieck E., Bonnifet T., Znaidi R., Ravel-Godreuil C., Massiani-Beaudoin O., Joshi R.L., Fuchs J. Retrotransposons as a source of DNA damage in neurodegeneration. *Front Aging Neurosci*. 2022;13:786897. doi 10.3389/fnagi.2021.786897

Pfaff A.L., Bubb V.J., Quinn J.P., Koks S. An increased burden of highly active retrotransposition competent l1s is associated with Parkinson's disease risk and progression in the PPMI cohort. *Int J Mol Sci.* 2020;21(18):6562. doi 10.3390/ijms21186562

Pfaff A.L., Bubb V.J., Quinn J.P., Koks S. Reference SVA insertion polymorphisms are associated with Parkinson's disease progression and differential gene expression. *NPJ Parkinsons Dis*. 2021;7(1):44. doi 10.1038/s41531-021-00189-4

Playfoot C.J., Sheppard S., Planet E., Trono D. Transposable elements contribute to the spatiotemporal microRNA landscape in human brain development. *RNA*. 2022;28:1157-1171. doi 10.1261/rna.079100.122

Qin L.X., Tan J.Q., Zhang H.N., Tang J.G., Jiang B., Shen X.M., Guo J.F., Tan L.M., Tang B., Wang C.Y. Preliminary study of hsa-mir-626 change in the cerebrospinal fluid in Parkinson's disease. *Neurol India*. 2021;69(1):115-118. doi 10.4103/0028-3886.310102

Ravanidis S., Bougea A., Papagiannakis N., Koros C., Simitsi A.M., Pachi I., Breza M., Stefanis L., Doxakis E. Validation of differentially expressed brain-enriched microRNAs in the plasma of PD patients. *Ann Clin Transl Neurol*. 2020;7(9):1594-1607. doi 10.1002/acn3.51146

Ravel-Godreuil C., Massiani-Beaudoin O., Mailly P., Prochiantz A., Joshi R.L., Fuchs J. Perturbed DNA methylation by Gadd45b induces chromatin disorganization, DNA strand breaks and dopaminergic neuron death. *iScience*. 2021;24(7):102756. doi 10.1016/j.isci.2021.102756

Santerre M., Arjona S.P., Allen C.N., Callen S., Buch S., Sawaya B.E. HIV-1 Vpr protein impairs lysosome clearance causing SNCA/alpha-synuclein accumulation in neurons. *Autophagy*. 2021;17(7):1768-1782. doi 10.1080/15548627.2021.1915641

Sonobe R., Yang P., Suzuki M.M., Shinjo K., Iijima K., Nishiyama N., Miyata K., Kataoka K., Kajiyama H., Kondo Y. Long noncoding RNA *TUG1* promotes cisplatin resistance in ovarian cancer via upregulation of DNA polymerase eta. *Cancer Sci*. 2024;115(6):1910-1923. doi 10.1111/cas.16150

Soreq L., Salomonis N., Bronstein M., Greenberg D.S., Israel Z., Bergman H., Soreq H. Small RNA sequencing-microarray analyses in Parkinson leukocytes reveal deep brain stimulation-induced splicing changes that classify brain region transcriptomes. *Front Mol Neurosci*. 2013;6:10. doi 10.3389/fnmol.2013.00010

Sun Z., Chen J., Zhang J., Ji R., Xu W., Zhang X., Qian H. The role and mechanism of miR-374 regulating the malignant transformation of mesenchymal stem cells. *Am J Transl Res*. 2018;10(10):3224-3232

Tang J., Pan H., Wang W., Qi C., Gu C., Shang A., Zhu J. MiR-495-3p and miR-143-3p co-target CDK1 to inhibit the development of cervical cancer. *Clin Transl Oncol*. 2021;23(11):2323-2334. doi 10.1007/s12094-021-02687-6

Thomas R., Connolly K.J., Brekk O.R., Hinrich A.J., Hastings M.L., Isacson O., Hallett P.J. Viral-like TLR3 induction of cytokine networks and α-synuclein are reduced by complement C3 blockade in mouse brain. *Sci Rep*. 2023;13(1):15164. doi 10.1038/s41598-023-41240-z

Tong D., Zhao Y., Tang Y., Ma J., Wang M., Li B., Wang Z., Li C. MiR-487b suppressed inflammation and neuronal apoptosis in spinal cord injury by targeted Ifitm3. *Metab Brain Dis*. 2022;37(7):2405-2415. doi 10.1007/s11011-022-01015-3

Van Bree E.J., Guimarães R.L.F.P., Lundberg M., Blujdea E.R., Rosenkrantz J.L., White F.T.G., Poppinga J., Ferrer-Raventós P., Schneider A.E., Clayton I., Haussler D., Reinders M.J.T., Holstege H., Ewing A.D., Moses C., Jacobs F.M.J. A hidden layer of structural variation in transposable elements reveals potential genetic modifiers in human disease-risk loci. *Genome Res*. 2022;32(4):656-670. doi 10.1101/gr.275515.121

Vojtechova I., Machacek T., Kristofikova Z., Stuchlik A., Petrasek T. Infectious origin of Alzheimer's disease: amyloid beta as a component of brain antimicrobial immunity. *PLoS Pathog*. 2022;18(11):e1010929. doi 10.1371/journal.ppat.1010929

Wallace A.D., Wendt G.A., Barcellos L.F., de Smith A.J., Walsh K.M., Metayer C., Costello J.F., Wiemels J.L., Francis S.S. To ERV is human: a phenotype-wide scan linking polymorphic human endogenous Retrovirus-K insertions to complex phenotypes. *Front Genet*. 2018;9:298. doi 10.3389/fgene.2018.00298

Wang H., Liu X., Tan C., Zhou W., Jiang J., Peng W., Zhou X., Mo L., Chen L. Bacterial, viral, and fungal infection-related risk of Parkinson's disease: meta-analysis of cohort and case-control studies. *Brain Behav*. 2020;10(3):e01549. doi 10.1002/brb3.1549

Wu D.M., Wang S., Wen X., Han X.R., Wang Y.J., Shen M., Fan S.H., Zhuang J., Zhang Z.F., Shan Q., Li M.Q., Hu B., Sun C.H., Lu J., Chen G.Q., Zheng Y.L. Suppression of microRNA-342-3p increases glutamate transporters and prevents dopaminergic neuron loss through activating the Wnt signaling pathway via p21-activated kinase 1 in mice with Parkinson's disease. *J Cell Physiol*. 2019;234(6):9033-9044. doi 10.1002/jcp.27577

Yufeng Z., Ming Q., Dandan W. MiR-320d inhibits progression of EGFR-positive colorectal cancer by targeting TUSC3. *Front Genet*. 2021;12:738559. doi 10.3389/fgene.2021.738559

Zhang Q., Yan Y.F., Lv Q., Li Y.J., Wang R.R., Sun G.B., Pan L., Hu J.X., Xie N., Zhang C., Tian B.C., Jiao F., Xu S., Wang P.Y., Xie S.Y. miR-4293 upregulates lncRNA WFDC21P by suppressing mRNA-decapping enzyme 2 to promote lung carcinoma proliferation. *Cell Death Dis*. 2021;12(8):735. doi 10.1038/s41419-021-04021-y

Zhang Y., Xia Q., Lin J. LncRNA H19 attenuates apoptosis in MPTP-induced Parkinson's disease through regulating miR-585-3p/PIK3R3. *Neurochem Res*. 2020;45:1700-1710. doi 10.1007/s11064-020-03035-w

Zhong C., Zhang Q., Bao H., Li Y., Nie C. Hsa_circ_0054220 upregulates HMGA1 by the competitive RNA pattern to promote neural impairment in MPTP model of Parkinson's disease. *Appl Biochem Biotechnol*. 2023;47:40-42. doi 10.1007/s12010-023-04740-2

Conflict of interest. The author declares no conflict of interest.
Received May 28, 2024. Revised August 19, 2024. Accepted October 24, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):301-309

doi 10.18699/vjgb-25-33

Genetic variants of the *DLK1*, *KISS1R*, *MKRN3* genes in girls with precocious puberty

E.A. Sazhenova [1] ✉, O.Yu. Vasilyeva [1], E.A. Fonova [1], M.B. Kankanam Pathiranage [2], A.Yu. Sambyalova [3], E.E. Khramova [3], L.V. Rychkova [3], S.A. Vasilyev [1], I.N. Lebedev [1]

[1] Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia
[2] Tomsk State University, Tomsk, Russia
[3] Scientific Center for Family Health and Human Reproduction Problems, Irkutsk, Russia
✉ elena.sazhenova@medgenetics.ru

Abstract. Precocious puberty (PP, E30.1, E22.8, E30.9 according to ICD 10, MIM 176400, 615346) in children is a disorder in which secondary sexual characteristics appear earlier than the age norm. The timing of puberty is regulated by a complex interaction of genetic and epigenetic factors, as well as environmental and nutritional factors. This study aimed to search for pathogenic, likely pathogenic variants or variants of uncertain significance (VUS) in the *KISS1*, *GPR54*, *DLK1*, and *MKRN3* genes in patients with the clinical picture of PP and normal karyotype by massive parallel sequencing. All identified genetic variants were confirmed by Sanger sequencing. The pathogenicity of identified genetic variants and the functional significance of the protein synthesized by them were analyzed according to recommendations for interpretation of NGS analysis results using online algorithms for pathogenicity prediction (Variant Effect Predictor, Franklin, Varsome, and PolyPhen2). Clinically significant genetic variants were detected in the heterozygous state in the *KISS1R*, *DLK1*, and *MKRN3* genes in 5 of 52 probands (9.6 %) with PP, including 3 of 33 (9.1 %) in the group with central PP and 2 of 19 (10.5 %) in the group with gonadotropin-independent PP. Two children with gonadotropin-independent PP had VUS in the *KISS1R* gene (c.191T>C, p.Ile64Thr and c.233A>G, p.Asn78Ser), one of which was inherited from the father and the second, from the mother. The remaining patients with central PP had likely pathogenic genetic variants: *DLK1*:c.373delC(p.Gln125fs) *de novo* and *DLK1*:c.480delT(p.Gly161Alafs*49) of paternal origin. The third proband had a VUS variant in the *MKRN3* gene (c.1487A>G, p.His496Arg), inherited from the father. All identified genetic variants were described for the first time in PP. Thus, in the present study, genetic variants in the *KISS1R*, *DLK1*, and *MKRN3* genes in girls with PP were characterized.

Key words: precocious puberty; hypothalamic-pituitary-gonadal axis; *DLK1*, *KISS1*, *KISS1R*, *MKRN3* genes

For citation: Sazhenova E.A., Vasilyeva O.Yu., Fonova E.A., Kankanam Pathiranage M.B., Sambyalova A.Yu., Khramova E.E., Rychkova L.V., Vasilyev S.A., Lebedev I.N. Genetic variants of the *DLK1*, *KISS1R*, *MKRN3* genes in girls with precocious puberty. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):301-309. doi 10.18699/vjgb-25-33

Funding. The study was carried out with the financial support of the Russian Science Foundation grant No. 23-25-00359.

Генетические варианты генов *DLK1*, *KISS1R*, *MKRN3* у девочек с преждевременным половым созреванием

Е.А. Саженова [1] ✉, О.Ю. Васильева [1], Е.А. Фонова [1], М.Б. Канканам Патиранаге [2], А.Ю. Самбялова [3], Е.Е.Храмова [3], Л.В. Рычкова [3], С.А. Васильев [1], И.Н. Лебедев [1]

[1] Научно-исследовательский институт медицинской генетики, Томский национальный исследовательский медицинский центр Российской академии наук, Томск, Россия
[2] Национальный исследовательский Томский государственный университет, Томск, Россия
[3] Научный центр проблем здоровья семьи и репродукции человека, Иркутск, Россия
✉ elena.sazhenova@medgenetics.ru

Аннотация. Преждевременное половое созревание (ППС, E30.1, E22.8, E30.9 по МКБ 10, MIM 176400, 615346) у детей – заболевание, при котором вторичные половые признаки появляются раньше возрастной нормы. Сроки полового созревания регулируются сложным взаимодействием генетических и эпигенетических факторов, а также факторов окружающей среды и питания. Цель настоящего исследования – поиск генетических причин формирования у девочек клинической картины ППС. Поиск клинически значимых генетических вариантов (патогенных, вероятно патогенных вариантов или вариантов с неопределенным клиническим значением (variant of uncertain significance, VUS)) проведен в генах *KISS1*, *KISS1R* (*GPR54*), *DLK1* и *MKRN3* у девочек с клинической картиной ППС и нормальным кариотипом методом таргетного массового параллельного секвенирования. Все найденные генетические варианты были подтверждены методом секвенирования ДНК по Сэнгеру. Патогенность

идентифицированных генетических вариантов и функциональная значимость кодируемого ими белка проанализированы с использованием онлайн-алгоритмов прогнозирования патогенности Variant Effect Predictor, Franklin и Varsome, а также PolyPhen2 (согласно рекомендациям по интерпретации результатов анализа NGS). Клинически значимые генетические варианты были обнаружены в гетерозиготном состоянии в генах *KISS1R*, *DLK1* и *MKRN3* у 5 из 52 пробандов (9.6 %) с ППС, из них 3 из 33 (9.1 %) – в группе с центральным ППС и 2 из 19 (10.5 %) – в группе с гонадотропин-независимой формой ППС. Два ребенка с гонадотропин-независимой формой ППС имели VUS в гене *KISS1R* (c.191T>C, p.Ile64Thr и c.233A>G, p.Asn78Ser), один из которых был унаследован от отца, второй – от матери. У остальных пациентов с центральным ППС были вероятно патогенные генетические варианты *DLK1*:c.373delC(p.Gln125fs) *de novo* и *DLK1*:c.480delT(p.Gly161Alafs*49) отцовского происхождения. Еще один пробанд имел вариант VUS в гене *MKRN3* (c.1487A>G, p.His496Arg), унаследованный от отца. Все выявленные генетические варианты описаны впервые при ППС. Таким образом, в настоящем исследовании найдены новые генетические варианты в генах *KISS1R*, *DLK1* и *MKRN3* у девочек с преждевременным половым созреванием.
Ключевые слова: преждевременное половое созревание; гипоталамо-гипофизарно-гонадная ось; гены *DLK1*, *KISS1*, *KISS1R*, *MKRN3*

Introduction

Precocious puberty (PP, E30.1, E22.8, E30.9 according to ICD 10, MIM 176400, 615346) is a disorder in which secondary sexual characteristics appear before the age of 8 in girls and before the age of 9 in boys, and, as a rule, there is an advancement of bone age by more than 2 years (Maione et al., 2021). The incidence of PP is 10–20 times higher in girls and varies widely across geographic regions, ranging from 0.217 to 26.28 per 10,000 girls and from 0.02 to 0.9 per 10,000 boys. The prevalence of familial cases of PP is 27.5 % (Brito et al., 2023).

PP can be gonadotropin-dependent (true, central), complete and incomplete, caused by premature reactivation of the hypothalamic-pituitary-gonadal (HPG) axis, and gonadotropin-independent (peripheral), developing as a result of excessive secretion of sex hormones by the gonads or adrenal glands, ovarian cysts, or human chorionic gonadotropin. The second form of PP is much less common, accounting for only 20 % of all PP (Shim et al., 2022). The causes of PP cannot be identified in most girls, therefore it is called idiopathic. If untreated, early puberty can lead to several serious complications, including short stature (caused by premature closure of the growth zones of tubular bones) and the formation of a dysplastic constitution (short limbs, elongated trunk, wide pelvis), psychological discomfort for girls and their parents. In girls, menstrual cycle disorders are observed, manifested in abnormal uterine bleeding, the development of polycystic ovary syndrome, premature ovarian failure, and, accordingly, early menopause. Earlier menarche in girls is also associated with an increased risk of breast cancer, endometrial cancer, obesity, type II diabetes, and cardiovascular diseases. PP can also be associated with organic brain lesions such as hypothalamic hamartoma, suprasellar arachnoid cysts, and hydrocephalus (Lagno et al., 2018; Peterkova et al., 2021).

Clinical features of PP include advanced growth spurt, progressive breast development in girls, and increased testicular volume in boys, and reflect high gonadotropin-releasing hormone (GnRH) levels and gonadotropin-stimulated sex steroid action (gonadarche). Accelerated growth velocity (>6 cm/year) and advanced bone age relative to biological age (>1 year or 2 SDS (standard deviation) points of chronological age) are common features of advanced PP. Hormonal findings supporting the diagnosis of PP include pubertal basal levels of luteinizing hormone (LH) or GnRH (Brito et al., 2023).

The timing of puberty depends on genetic, epigenetic, and environmental factors. In recent years, genetic variants in the *DLK1* (14q32), *MKRN3* (15q11.2), *KISS1* (1q32.1), and *KISS1R* (*GPR54*, 19p13.3) genes have been identified as hereditary causes of PP (Shim et al., 2022). However, in sporadic forms of PP, genetic variants in these genes are detected in only 10 % of cases (Canton et al., 2021, 2024). These genes primarily affect premature reactivation of the HPG axis and are directly involved in the formation of central PP. However, based on gene function, the clinical picture of peripheral PP may subsequently lead to central PP. For example, the presence of thelarche in girls after 2 years of age with a gonadotropin-independent form of PP increases the risk of subsequently developing central PP (Peterkova et al., 2021).

The *KISS1* gene (MIM 603286) and its receptor *KISS1R* (MIM 604161) are responsible for the secretion of GnRH, participating in the regulation of endocrine function and the onset of puberty. The *KISS1* gene encodes the kisspeptin protein, which stimulates the secretion of GnRH, and *KISS1R* is a regulator of this process and a key factor in the initiation of puberty, acting as a potent stimulator of the secretion of GnRH-dependent luteinizing hormone. It is expressed in various endocrine and gonadal tissues (Teles et al., 2008).

DLK1 (MIM 176290) is an imprinted gene expressed only on the paternal homologue and encodes an EGF-like growth factor. It is a membrane-binding protein that is involved in the Notch signaling pathway and promotes cell proliferation signals during neurogenesis. The product of this gene is also involved in osteogenesis, adipogenesis, hematopoiesis, and hepatocyte proliferation (Gomes et al., 2019; Macedo, Kaiser, 2019). In mice, *Dlk1* is expressed prenatally in neuroendocrine tissues, including the pituitary gland, and postnatally in the hypothalamus, including the mediobasal hypothalamus, the control center for GnRH secretion (Shim et al., 2022). The product of this gene is also important for adipose tissue homeostasis. Genome-wide association studies have shown that paternally inherited single nucleotide variants in the *DLK1* gene are associated with an earlier onset of menarche (Perry et al., 2014).

The imprinted and also paternally expressed *MKRN3* gene (MIM 603856) encodes the macorine protein RING-finger 3,

which belongs to the macorine family and is involved in controlling the onset of puberty by blocking the release of GnRH from the hypothalamus, thereby delaying the onset of puberty (Abreu et al., 2020). MKRN3 is responsible for protein ubiquitination, in which a ubiquitin moiety is attached to an intracellular protein to transfer it to the proteasome. Ubiquitination can also be an indicator of signal transmission for regulation of the cell cycle, differentiation, and morphogenesis (Abreu et al., 2020). Pathogenic and likely pathogenic variants in the *MKRN3* gene are the most commonly known factors in genetic etiology of central PP, accounting for 19–33 % in familial and 2–3.9 % in sporadic cases (Valadares et al., 2019; Roberts, Kaiser, 2020).

This study aimed to identify clinically significant genetic variants in the *KISS1*, *KISS1R* (*GPR54*), *DLK1*, and *MKRN3* genes in girls with a clinical picture of PP.

Material and methods

In the course of this study, a sample of 52 families (202 people in total) was formed based on the Scientific Center for Family Health and Human Reproduction Problems, Irkutsk. Each family consisted of a female proband with a clinical picture of PP, her parents and, in some cases, sisters and grandmothers. The study was conducted according to the provisions of the Helsinki Declaration of the World Medical Association. The study was approved by the Bioethics Committee of the Scientific Center for Family Health and Human Reproduction Problems (Protocol No. 1.1 dated 12.01.2023). Informed consent for participation in the study and DNA diagnostics was obtained from the patients' parents. The clinical picture in the probands included PP with isosexual gonadotropin-dependent (ICD-10: E22.8, n = 33, age 7.4 ± 1.6 years) and gonadotropin-independent (ICD-10: E30.9, n = 19, age 6.9 ± 0.8 years) forms. Girls with organic lesions of the central nervous system were not included in the study.

Description of patient subgroups:

- girls with the isosexual gonadotropin-dependent form of PP, under 8 years old, exhibiting accelerated physical development (height SDS +1 or more), with their sexual development corresponding to Tanner stages 2–4, levels of pituitary gonadotropic hormones corresponding to pubertal values, and a positive buserelin test. Additionally, they have enlarged mammary glands and uterus confirmed by ultrasound, and their biological age does not match their chronological (passport) age;
- girls with the gonadotropin-independent form of PP, under 8 years old, with either accelerated or normal physical development (height SDS +1 or more), advanced sexual development corresponding to Tanner stage 2, levels of pituitary gonadotropic hormones corresponding to prepubertal values, and a negative buserelin test. Additionally, they exhibit enlarged mammary glands and uterus, which is confirmed by ultrasound.

All probands underwent standard cytogenetic analysis, which showed a normal karyotype in all cases. Karyotyping was performed using a research-grade microscope AxioImager (Carl Zeiss, Germany).

Genomic DNA was isolated from venous blood by phenol-chloroform extraction. The concentration of the original samples was estimated using a Nanodrop 1000 spectrophotometer (ThermoFisher Scientific, USA). Genotyping of all exons in the *KISS1*, *GPR54* (*KISS1R*), *DLK1*, and *MKRN3* genes was performed using targeted massive parallel sequencing (NGS) of these genes using a MiSeq sequencer and a MicroKit (2x150) (Illumina, USA). For this purpose, amplification of long DNA fragments (Long-range PCR) was used. To obtain the nucleotide sequence, the UCSC In-Silico PCR genome browser was used, which contains information on genome sequences (hg38 assembly). The obtained nucleotide sequence was then used to select primers using the Primer-BLAST bioinformatics program provided by the National Center for Biotechnological Information (NCBI) (Table 1).

Amplification of target fragments was performed using the BioMaster HS-Taq PCR (2x) kit (Biolabmix, Russia) according to the manufacturer's protocol with the following PCR conditions: 95 °C for 5 min; 36 cycles: 95 °C for 40 s, 60 °C for 50 s, 68 °C for 1 min. The concentration of target fragments was determined using a Qubit 4.0 fluorimeter (ThermoFisher Scientific, USA). The reaction products were purified from impurities using a Sephadex G50 solution (Sigma, USA). The quality of reads was assessed using FastQC v0.11.8, after which trimming of the remaining adapter sequences and low-quality reads was performed using Trim-Galore.

All detected genetic variants were confirmed using Sanger sequencing. The primer sequences are presented in Table 2. The pathogenicity of the identified genetic variants was analyzed using online pathogenicity prediction algorithms: Variant Effect Predictor (http://www.ensembl.org/Tools/VEP), Provean (http://provean.jcvi.org/genomesubmit_2.php?species=human), Franclin (https://franklin.genoox.com/clinical-db/variant/snp/chr15-23621174-GC-G-hg38), VarSome (https://varsome.com/variant/hg19) and PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/) according to the recommendations for interpreting the results of NGS analysis (Eijkelenboom et al., 2019; Ryzhkova et al., 2019). The following databases were used to determine the frequency of identified mutations in population samples in order to exclude polymorphic variants in patients: Exome Aggregation Consortium (http://exac.broadinstitute.org/), Exome Variant Server (http://evs.gs.washington.edu/EVS), 1000 Genomes Project (http://browser.1000genomes.org/index.html), which are recommended for interpreting data obtained using NGS (Eijkelenboom et al., 2019; Ryzhkova et al., 2019).

The study was conducted using equipment from the Core Medical Genomics Facility of the Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk.

Results

Clinically significant genetic variants (likely pathogenic and variants of uncertain significance (VUS)) were identified in the *KISS1R*, *DLK1* and *MKRN3* genes in five of 52 probands (9.6 %) with PP, including three of 33 (9.1 %) in the group with central PP and two of 19 (10.5 %) in the group with gonadotropin-independent PP.

Table 1. Sequences of the oligonucleotide primers used to generate libraries for targeted massive parallel sequencing of the *KISS1*, *KISS1R*, *DLK1*, and *MKRN3* genes

Primer name	Primer sequence, 5′–3′	Length of product, bp	Position in genome (UCSC In-Silico PCR, hg38)
		DLK1	
DF1	TATGGCTAAGATGGGAAATCTGTGC	6,196	chr14:100725325-100731520
DR1	CCGTCAGGAATCAAGAAACCTGTTA		
DF2	GCTCAATAGTTCTAATTTCCCTGGC	4,040	chr14:100731378-100735417
DR2	CCGCTAAATCTCAAATCAATCGGAA		
DF3	GCTATCTCTTGTGTCAAATCTGGTG	4,689	chr14:100734980-100739668
DR3	CCTTCAGTGTGGTCATGTTATTTCC		
		MKRN3	
MF1	GGCAGACAGATACGAAAATACAACG	3,642	chr15:23565341-23568982
MR1	ATTTGCAGTTGATGCAGATCATACG		
MF2	GCTGCTCATCTGTTTGTTTACAGTT	5,062	chr15:23567720-23572781
MR2	AACTGTGATTCCCTCATCGTTTGTA		
MF3	TTCTAAACTGACTGTGACTAGGTGC	4,998	chr15:23626497-23631494
MR3	ATACCGAAATCTCATCCCATCTTGG		
MF4	AATGTCTCACCTTCCCTCTACAAAC	5,629	chr15:23620915-23626543
MR4	GTGGGAGATGATAGCAGAATAAGCA		
MF5	TAGTTCTATTATCAGCCATTGCCCC	6,048	chr15:23614951-23620998
MR5	TGGCAGAACTCTACAGAAAATCGAA		
		KISS1	
KF5	GCAAGGCTCATTAAGTTCACTG	6,203	chr1:204191506-204197708
KR5	CAGCCCTAATGGGTGTGATAAT		
KF7	CTGGAAGATGGTTAGAGGAACC	4,106	chr1:204188621-204192726
KR7	GCAAAATGAGCTTTCCCGTATT		
		KISS1R	
KRF2	CAAGTTCGTCAACTACATCCAG	3,436	chr19:918644-922079
KRR2	AAAAGTAAAGTGCCTAAGACCG		
KRF3	CAAATGGAAGCACCTTTTTCTTC	5,316	chr19:915605-920921
KRR3	ATCAATAGCAAACTTCACAACGA		

Table 2. Sequences of oligonucleotide primers used for Sanger sequencing of the *KISS1R*, *DLK1* and *MKRN3* genes

Primer name	Primer sequence, 5′–3′	Length of product, bp	Position in genome (UCSC In-Silico PCR, hg38)
		DLK1	
c373del_F	TAAACCCTCTTACTCCAGACCC	294	chr14:100731920-100732213
c373del_R	CATTAGATCACACACAGGAAGGA		
c480delF	GTGTTTTAAGCACCTGCCCCTTA	329	chr14:100734030-100734358
c480delR	CAGGTCTTGTCGATGAAGCCG		
		MKRN3	
c1487A>GF	GGAGAGGGCAACATGCTCTATAA	254	chr15:23567136-23567389
c1487A>GR	CAGTAAGAGTGTCAACACAGGGA		
		KISS1R	
c233A>GF	GTGCCGCTCTTCTTCGC	265	chr19:917635-917899
c233A>GR	CCACAGGGAAAAGATTCGAGG		
c191T>CF	GGGCTATAAACGCTCGGC	448	chr19:917300-917747
c191T>CR	CCGATGTAGAAGTTGGTCACG		

Table 3. The main clinical characteristics of the phenotype of patients with the identified genetic variants

Patient No.	PP, clinical form	Age of onset of PP, years	Clinical features	Presence of obesity	Age, years	
					biological	bone
19	Gonadotropin-dependent, incomplete form	6	Telarche and pubarche at 6 years old	–	6.5	9
45	Gonadotropin-dependent, complete form	7	Thelarche, pubarche and menarche from 7 years old	–	7	11–11.5
47	Gonadotropin-dependent, complete form	6	Pubarche from 6 years of age, thelarche from 8 years, menarche from 9 years old	Stage 2	9	11–11.5
10	Gonadotropin-dependent, incomplete form	5	Thelarche, adrenarche	Stage 1	5	7
14	Gonadotropin-dependent, incomplete form	6	Pubarche from 5 years, thelarche from 7 years old	Stage 1	7	8.5–9

Table 4. Position and characteristics of the identified genetic variants in patients with precocious puberty

Patient No.	Gene	Genetic variants	Amino acid variants	Presence of polymorphic variants (rs, dbSNP No.)	Frequency in population (GnomAD)	Genetic Variant Identifier (Varsome, Franklin)	ACMG Pathogenicity Score (Varsome, Franklin)	Inheritance
19	*DLK1*	c.373delC	p.Gln125fs*8	–	–	Frameshift variant	LP	*De novo*
45	*DLK1*	c.480delT	p.Gly161Alafs*49	–	–		LP	Paternal
47	*MKRN3*	c.1487A>G	p.His496Arg	rs749506944	0.000016	Missense variant	VUS	Paternal
14	*KISS1R*	c.191T>C	p.Ile64Thr	–	0.0004		VUS	Paternal
10	*KISS1R*	c.233A>G	p.Asn78Ser	rs540538484	0.000013		VUS	Maternal

Note. LP – likely pathogenic variant, VUS – variant of uncertain significance.

The main clinical characteristics of the phenotype of patients with the identified genetic variants are presented in Table 3. It is noteworthy that in three cases, the patients had obesity, which could contribute to the development of PP (Song et al., 2023).

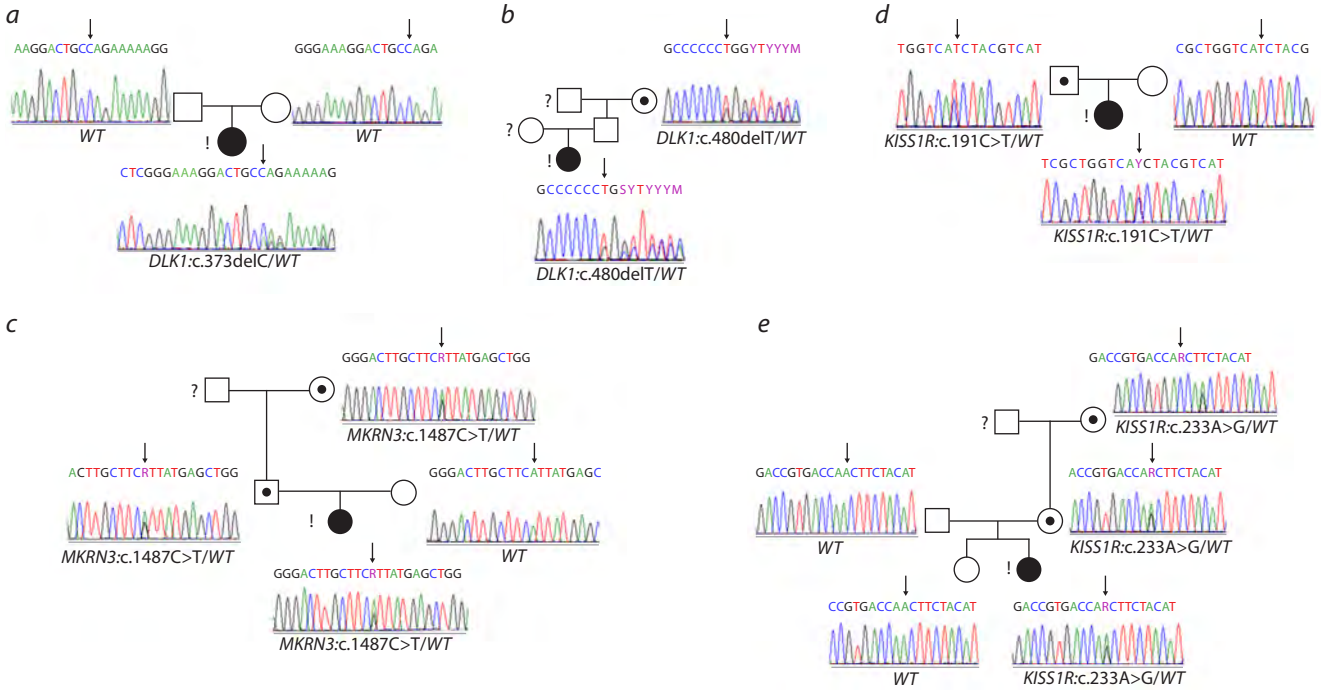
The clinically significant variants identified resulted in missense amino acid substitutions in three cases. Two variants were represented by single nucleotide deletions resulting in a reading frame shift. Table 4 and the Figure describe the spectrum of genetic variants identified in patients, which were registered in the heterozygous state in all cases, and also present their pedigrees.

A total of five genetic variants located in the coding region of the studied genes were identified. Two likely pathogenic variants were identified in the *DLK1* gene (c.373delC, p.Gln125fs and c.480delT p.Gly161Alafs*49) (Table 4, the Figure *a*, *b*). The first was located in the exon 4, and the second, in the exon 5. In both cases, these variants led to a reading frame shift and the formation of a stop codon and, as a consequence, to a shortening of the synthesized protein.

The *DLK1* gene has five exons. The DLK1 protein structure consists of a transmembrane domain with six epidermal growth factor (EGF)-like repeats and a protease-sensitive sequence – target of tumor necrosis factor α-converting enzyme (TACE), a transmembrane domain, and a short cytoplasmic domain (Sánchez-Solana et al., 2011).

In the present study, both the first and second variants are located in the region containing EGF repeats, which are crucial for inhibiting the activity of Notch transmembrane proteins. These proteins act as transcriptional activators in complex with CSL family transcription factors (Baladrón et al., 2005; Gomes et al., 2019). The first variant is located in the third repeat, and the second, in the fourth repeat. Both genetic variants are described for the first time in PP. Previously, the c.479delC(p.Pro160fs*50) variant was described for this pathology (Gomes et al., 2019; Yuan et al., 2022); it is located near the *DLK1*:c.480delT(p.Gly161Alafs*49) variant described by us.

One missense variant of uncertain clinical significance (c.1487A>G, p.His496Arg, rs749506944) was found in the exon 4 of the *MKRN3* gene. The frequency of this variant in the GnomAD database is extremely low (0.000016), and it is found only in the European population. The MKRN3 protein has four zinc finger domains: three RNA-binding C3H1 motifs and one protein-binding domain C3HC4, which is responsible for the activity of ubiquitin ligase. The MKRN-



Pedigrees of patients with mutations in the *KISS1R*, *DLK1*, and *MKRN3* genes available for familial segregation analysis.

Pedigrees of patients: *a* – No. 19; *b* – No. 45; *c* – No. 47; *d* – No. 14; *e* – No. 10; squares represent male family members; circles represent female family members; black symbols represent clinically affected family members; white symbols represent clinically unaffected carriers; black dot represents clinically unaffected carriers with a detected genetic variant; question mark represents unknown phenotype; exclamation mark represents the proband in each family; *WT* represents the wild-type genotype status.

specific Cys-His domain, which is part of the protein, has an unknown function. The genetic variant identified in this study is located in the region of RNA-binding C3H1motifs. Predictive programs prediction and the low frequency of this variant in the population indicate that this missense variant can be associated with the development of PP.

Genetic variants in the *DLK1* and *MKRN3* genes were inherited from fathers and paternal grandmothers in two cases, and *de novo* inheritance was observed in one family (No. 19) (Table 4, the Figure *a–c*). These genes are imprinted and are expressed only on the paternal chromosome. Fathers inherit this variant from their mothers, so this genetic variant is not active in fathers and there are no clinical manifestations of this disease. Indeed, fathers in families No. 45 and 47 did not have PP. At the same time, paternal grandmothers should have clinical signs of PP, since they have an active homologue. However, according to the survey, paternal grandmothers also did not have such disorders, which indicates incomplete penetrance of the identified genetic variants.

The other two genetic variants were found in the *KISS1R* gene, c.191T>C, p.Ile64Thr and c.233A>G, p.Asn78Ser; these were missense variants located in the first exon. In the first case, the proband inherited the variant from her father, who had no clinical manifestations of PP (the Figure *d*, *e*). In the second family, in addition to the proband, the mother and maternal grandmother had this genetic variant, and also had no cases of PP. The proband's sister was not diagnosed with this disease. With regard to PP, all this indicates incomplete penetrance of the clinical phenotype.

The *KISS1R* gene has five exons. The protein coded by this gene, GPR54, is located in the cell membrane and has an extracellular N-terminal domain, followed by seven transmembrane helices with three intracellular and three extracellular loops, and ends with a C-terminal cytoplasmic domain. The variants identified in this gene are located in the first transmembrane helix.

In two families, unique genetic variants were found that were not repeated in unrelated patients (Table 4). In the remaining families, the genetic variants found in this study were observed in population samples with frequencies ranging from 0.000013 to 0.0004 (according to GnomAD). As can be seen, these variants are extremely rare in populations, which may indicate their pathogenic nature.

Thus, in this study, it was shown that in the group of girls with PP in the *KISS1R*, *DLK1* and *MKRN3* genes, in 9.6 % (9.1 % in the group with central PP and 10.5 % in the group with gonadotropin-independent PP) of cases, likely pathogenic variants and variants with uncertain clinical significance (VUS) are present, which may be a potential cause of PP formation. All genetic variants found in this study are described in this disease for the first time. Identification of new genetic variants will allow a better understanding of the contribution of genetic causes to the development of PP.

Discussion

Female reproductive process is a well-organized and tightly controlled system governed by the HPG axis. The main element of this axis is the pulsatile secretion of GnRH, which

regulates the production of gonadotropins (follicle-stimulating hormone (FSH) and luteinizing hormone (LH)) by the anterior pituitary gland during puberty and maintains normal cycles in adults. GnRH and gonadotropin production are also regulated by negative feedback from estrogens secreted by developing ovarian follicles.

In most cases, PP is associated with variants in the *DLK1*, *MKRN3*, *KISS1*, and *KISS1R* genes. Indeed, in our study, we found likely pathogenic variants and variants with uncertain clinical significance in the *DLK1*, *MKRN3*, and *KISS1R* genes in five of 52 probands (9.6 %) in a sample of girls with a clinical picture of PP. This finding aligns with literature reports on the frequency of detection of genetic variants in PP (Canton et al., 2021, 2024). No genetic variants were identified in the *KISS1* gene in the sample of probands with a clinical picture of PP. This may be due to the limited number of patients in this study, as well as the low frequency of genetic disorders in this gene associated with PP. Indeed, only a few genetic variants in the *KISS1* gene have been described in this pathology (Silveira et al., 2010; Rhie et al., 2014). All identified genetic variants were in the heterozygous state, which is consistent with the literature data on the autosomal dominant nature of inheritance of genetic variants of these genes in PP.

In the present study, we found genetic variants in the *KISS1R* gene in girls with gonadotropin-independent PP, in all cases accompanied by thelarche. Thelarche refers to the development of the mammary glands and is a response to estrogen synthesis. It has been determined that the *KISS1* and *KISS1R* genes are expressed in various tissues, including the gonads, and are able to influence the level of these hormones either through temporary activation of the HPG axis or directly through stimulation of the gonads (Hu K. et al., 2018; Yarmolinskaya et al., 2016).

The low frequency (approximately 10 %) of detection of genetic variants in the *DLK1*, *MKRN3*, *KISS1*, and *KISS1R* genes, primarily in sporadic cases of PP, suggests that some other mechanisms or genes may also be involved in the formation of PP. Indeed, epimutations (changes in the methylation status of CpG dinucleotides) in the *DLK1/MEG3*:*IG-DMR* and *MKRN3*:*TSS-DMR* imprinting centers, which control the expression of the imprinted *DLK1* and *MKRN3* genes, may also be the cause of formation of the clinical picture of PP. In support of this, A.P.M. Canton et al. (2021) identified various genetic and epigenetic disruptions in 36 (18 %) of 197 unrelated patients with PP, among which: in 24 cases (67 %), genetic disruptions were found in the *KISS1R*, *KISS1*, *MKRN3* and *DLK1* genes; in 7 cases (19 %), CNVs were detected (3 patients had a *de novo* deletion of 7q11.23 (Williams–Beuren syndrome), 3 probands had an inherited deletion of Xp22.33 and one patient had a *de novo* duplication of 1p31.3); epigenetic abnormalities of imprinted centers of the *DLK1* and *MKRN3* genes accounted for 3 cases (9 %); identification of the genetic variants of genes using whole exome sequencing revealed rare *de novo* variants of loss of gene function in a dominant state in two probands (5 %) such as pathogenic deletion with a reading frameshift in the *TNRC6B* gene (p.Gly665Leufs*35) and a likely pathogenic variant of a reading frameshift in the *AREL1* gene (p.Ser229Phefs*3).

The *TNRC6B* gene (trinucleotide repeat containing adaptor 6B, region 22q13.1, OMIM 610740) encodes a protein with RNA-binding activity, which is involved in the regulation of gene expression. This gene plays a role in RNA-mediated gene silencing by both micro-RNAs (miRNAs) and short interfering RNAs (siRNAs). The *AREL1* gene (apoptosis-resistant E3 ubiquitin protein ligase 1, region 14q24.3, OMIM 615380) encodes a protein that activates ubiquitin protein transferases, is involved in the negative regulation of apoptosis, protein ubiquitination, and is located in the cytosol.

Meta-analysis of association studies has also expanded the range of genes that could potentially cause the development of PP. These include genes such as *LIN28B* and *PROKR2*, although their roles in this process are not so obvious (Perry et al., 2009).

The *LIN28B* gene (6q16.3, OMIM 611044) encodes a highly conserved RNA-binding protein that blocks LET7 family microRNAs and helps maintain the pluripotent state of embryonic stem cells by preventing differentiation, and is involved in metabolism and oncogenesis. It may also play a role in pubertal development. Several studies have shown that *LIN28B* is involved in forming the clinical presentation of PP, in particular, earlier development of thelarche, menarche, and pubarche (Ong et al., 2009; Perry et al., 2009; Hu Z. et al., 2016). However, another study assessed the association between *LIN28B* variants in 178 Brazilian children with PP, but did not find a causal relationship (Silveira-Neto et al., 2012). Moreover, genetic variants in *LIN28B* such as rs314276 have been reported to be associated with obesity, which is closely linked with PP (Ong et al., 2011). Thus, the role of the *LIN28B* gene in forming the clinical presentation of PP remains to be determined.

The *PROKR2* gene (prokineticin receptor 2, 20p12.3, OMIM 607123) is a G protein-coupled receptor that is involved in the development of GnRH neurons, but neither developing nor mature GnRH neurons express prokineticin receptors. M. Fukami et al. (2017) reported that a *PROKR2* variant is associated with the formation of central PP. In this case, a girl presented with thelarche at the age of 3 years and 5 months with blood gonadotropin and estradiol (E2) levels consistent with puberty. Molecular analysis revealed a heterozygous deletion c.724_727delTGCT in this gene, resulting in transcription terminations. This variant was also detected in the patient's mother, who did not have PP. It has been shown that in the heterozygous state, this variant forms a heterodimer with the wild type, which acts as a gain-of-function variant leading to PP. Moreover, S. Sposini et al. (2015) demonstrated that in the absence of the 6th and 7th transmembrane domains in the *PROKR2* gene, ligand-dependent signal transduction is enhanced. Thus, only certain variants in the *PROKR2* gene in the heterozygous state can lead to the development of PP.

Conclusion

The onset of puberty is controlled by the interaction between genetic, epigenetic and non-hereditary factors. PP is a result of premature activation of these interactions. In the present

study, it was shown that in the group of girls with PP, 9.6 % (9.1 % in the group with central PP and 10.5 % in the group with gonadotropin-independent PP) of cases had likely pathogenic variants and variants with uncertain clinical significance in the *KISS1R*, *DLK1* and *MKRN3* genes, which can be a potential cause of PP. All genetic variants detected in this study are described in PP for the first time. Analysis of familial segregation showed that in all cases, the probands had genetically significant variants in the heterozygous state, which confirms the autosomal dominant nature of inheritance. In all cases where family material was available, only the probands exhibited the clinical picture of PP, indicating incomplete penetrance of the disease.

Identification of genetic variants is necessary not only for molecular genetic confirmation of the diagnosis, but also for choosing the right tactics for patient management and medical genetic counseling of the family. A comprehensive and step-by-step study of genetic, epigenetic and non-hereditary factors can improve our understanding of the exact mechanism of PP.

References

Abreu A.P., Toro C.A., Song Y.B., Navarro V.M., Bosch M.A., Eren A., Liang J.N., Carroll R.S., Latronico A.C., Ronnekleiv O.K. *MKRN3* inhibits the reproductive axis through actions in kisspeptin-expressing neurons. *J Clin Invest*. 2020;130(8):4486-4500. doi 10.1172/JCI136564

Baladrón V., Ruiz-Hidalgo M.J., Nueda M.L., Díaz-Guerra M.J.M., García-Ramírez J.J., Bonvini E., Gubina E., Laborda J. Dlk acts as a negative regulator of Notch1 activation through interactions with specific EGF-like repeats. *Exp Cell Res*. 2005;303(2):343-359. doi 10.1016/j.yexcr.2004.10.001

Brito V.N., Canton A.P.M., Seraphim C.E., Abreu A.P., Macedo D.B., Mendonca B.B., Kaiser U.B., Argente J., Latronico A.C. The congenital and acquired mechanisms implicated in the etiology of central precocious puberty. *Endocr Rev*. 2023;44(2):193-221. doi 10.1210/endrev/bnac020

Canton A.P.M., Krepischi A.C.V., Montenegro L.R., Costa S., Rosenberg C., Steunou V., Sobrier M.L., … Jorge A.A.L., Mendonca B.B., Netchine I., Brito V.N., Latronico A.C. Insights from the genetic characterization of central precocious puberty associated with multiple anomalies. *Hum Reprod*. 2021;36(2):506-518. doi 10.1093/humrep/deaa306

Canton A.P.M., Seraphim C.E., Montenegro L.R., Krepischi A.C.V., Mendonca B.B., Latronico A., Brito V.N. The genetic etiology is a relevant cause of central precocious puberty. *Eur J Endocrinol*. 2024;190(6):479-488. doi 10.1093/ejendo/lvae063

Eijkelenboom A., Tops B.B.J., van den Berg A., van den Brule A.J.C., Dinjens W.N.M., Dubbink H.J., Ter Elst A., … Vogel M.J., van Wezel T., Nederlof P.M., Schuuring E., Ligtenberg M.J.L. Recommendations for the clinical interpretation and reporting of copy number gains using gene panel NGS analysis in routine diagnostics. *Virchows Arch*. 2019;474(6):673-680. doi 10.1007/s00428-019-02555-3

Fukami M., Suzuki E., Izumi Y., Torii T., Narumi S., Igarashi M., Miyado M., … Hata K., Umezawa A., Matsubara Y., Yamauchi J., Ogata T. Paradoxical gain-of-function mutant of the G-protein-coupled receptor *PROKR2* promotes early puberty. *J Cell Mol Med*. 2017;21:2623-2626. doi 10.1111/jcmm.13146

Gomes L.G., Cunha-Silva M., Crespo R.P., Ramos C.O., Montenegro L.R., Canton A., Lees M., … Baracat E.C., Jorge A.A.L., Mendonca B.B., Brito V.N., Latronico A.C. *DLK1* is a novel link between reproduction and metabolism. *J Clin Endocrinol Metab*. 2019;104(6):2112-2120. doi 10.1210/jc.2018-02010

Hu K., Zhao H., Chang H.M., Yu Y., Qiao J. Kisspeptin/kisspeptin receptor system in the ovary. *Front Endocrinol*. 2018;4(8):365. doi 10.3389/fendo.2017.00365

Hu Z., Chen R., Cai C. Association of genetic polymorphisms around the *LIN28B* gene and idiopathic central precocious puberty risks among Chinese girls. *Pediatr Res*. 2016;80:521-525. doi 10.1038/pr.2016.107

Lagno O.V., Plotnikova E.V., Shabalov N.P. To the question about premature adrenarche at girls (the literature review). *Pediatr = Pediatrician*. 2018;9(5):66-74. doi 10.17816/PED9566-74 (in Russian)

Macedo D.B., Kaiser U.B. *DLK1*, Notch signaling and the timing of puberty. *Semin Reprod Med*. 2019;37(4):174-181. doi 10.1055/s-0039-3400963

Maione L., Bouvattier C., Kaiser U.B. Central precocious puberty: recent advances in understanding the etiology and in the clinical approach. *Clin Endocrinol*. 2021;95(4):542-555. doi 10.1111/cen.14475

Ong K.K., Elks C.E., Li S., Zhao J.H., Luan J., Andersen L.B., Bingham S.A., ... Deloukas P., Barroso I., Mooser V., Loos R.J., Wareham N.J. Genetic variation in *LIN28B* is associated with the timing of puberty. *Nat Genet*. 2009;41:729-733. doi 10.1038/ng.382

Ong K.K., Elks C.E., Wills A.K., Wong A., Wareham N.J., Loos R.J., Kuh D., Hardy R. Associations between the pubertal timing-related variant in *LIN28B* and *BMI* vary across the life course. *J Clin Endocrinol Metab*. 2011;96:E125-E129. doi 10.1210/jc.2010-0941

Perry J.R., Stolk L., Franceschini N., Lunetta K.L., Zhai G., McArdle P.F., Smith A.V., Aspelund T., Bandinelli S., Boerwinkle E., Cherkas L., Eiriksdottir G., Estrada K. Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. *Nat Genet*. 2009;41:648-650. doi 10.1038/ng.386

Perry J.R., Day F., Elks C.E., Sulem P., Thompson D.J., Ferreira T., He C., … Feenstra B., Franceschini N., Ganna A., Johnson A.D., Kjellqvist S. Parent-of-origin specific allelic associations among 106 genomic loci for age at menarche. *Nature*. 2014;514(7520):92-97. doi 10.1038/nature13545

Peterkova V.A., Alimova I.L., Bashnina E.B., Bezlepkina O.B., Bolotova N.V., Zubkova N.A., Kalinchenko N.Yu., Kareva M.A., Kiyaev A.V., Kolodkina A.A., Kostrova I.B., Makazan N.V., Malievskiy O.A., Orlova E.M., Petryaykina E.E., Samsonova L.N., Taranushenko T.E. Clinical guidelines «Precocious puberty». *Problemy Endocrinologii = Problems of Endocrinology*. 2021;67(5):84-103. doi 10.14341/probl12821 (in Russian)

Rhie Y.J., Lee K.H., Ko J.M., Lee W.J., Kim J.H., Kim H.S. *KISS1* gene polymorphisms in Korean girls with central precocious puberty. *J Korean Med Sci*. 2014;29(8):1120-1125. doi 10.3346/jkms.2014.29.8.1120

Roberts S.A., Kaiser U.B. Genetic etiologies of central precocious puberty and the role of imprinted genes. *Eur J Endocrinol*. 2020;183(4):R107-R117. doi 10.1530/EJE-20-0103

Ryzhkova O.P., Kardymon O.L., Prokhorchuk E.B., Konovalov F.A., Maslennikov A.B., Stepanov V.A., Afanasyev A.A., Zaklyazminskaya E.V., Rebrikov D.V., Savostyanov K.V., Glotov A.S., Kostareva A.A., Pavlov A.E., Golubenko M.V., Polyakov A.V., Kutsev S.I. Guidelines for the interpretation of massive parallel sequencing variants (update 2018, v2). *Meditsinskaya Genetika = Medical Genetics*. 2019;18(2):3-24. doi 10.25557/2073-7998.2019.02.3-24 (in Russian)

Sánchez-Solana B., Nueda M.L., Ruvira M.D., Ruiz-Hidalgo M.J., Monsalve E.M., Rivero S., García-Ramírez J.J., Díaz-Guerra M.J.M., Baladrón V., Labordaet J. The EGF-like proteins *DLK1* and *DLK2* function as inhibitory non-canonical ligands of NOTCH1 receptor that modulate each other's activities. *Biochim Biophys Acta*. 2011;1813(6):1153-1164. doi 10.1016/j.bbamcr.2011.03.004

Shim Y.S., Lee H.S., Hwang J.S. Genetic factors in precocious puberty. *Clin Exp Pediatr*. 2022;65(4):172-181. doi 10.3345/cep.2021.00521

Silveira L.G., Noel S.D., Silveira-Neto A.P., Abreu A.P., Brito V.N., Santos M.G., Bianco S.D.C., … Escobar M.E., Arnhold I.J.P., Mendonca B.B., Kaiser U.B., Latronico A.C. Mutations of the *KISS1* gene in disorders of puberty. *J Clin Endocrinol Metab*. 2010;95(5): 2276-2280. doi 10.1210/jc.2009-2421

Silveira-Neto A.P., Leal L.F., Emerman A.B., Henderson K.D., Piskounova E., Henderson B.E. Absence of functional *LIN28B* mutations in a large cohort of patients with idiopathic central precocious puberty. *Horm Res Paediatr.* 2012;78:144-150. doi 10.1159/000342212

Song Y., Kong Y., Xie X., Wang Y., Wang N. Association between precocious puberty and obesity risk in children: a systematic review and meta-analysis. *Front Pediatr*. 2023;11:1226933. doi 10.3389/fped.2023.1226933

Sposini S., Caltabiano G., Hanyaloglu A.C., Miele R. Identification of transmembrane domains that regulate spatial arrangements and activity of prokineticin receptor 2 dimers. *Mol Cell Endocrinol*. 2015; 399:362-372. doi 10.1016/j.mce.2014.10.024

Teles M.G., Bianco S.D., Brito V.N., Trarbach E.B., Kuohung W., Xu S., Seminara S.B., Mendonca B.B., Kaiser U.B., Latronico A.C. *AGPR54*-activating mutation in a patient with central precocious puberty. *N Engl J Med*. 2008;358(7):709-715. doi 10.1056/NEJMoa073443

Valadares L.P., Meireles C.G., De Toledo I.P., de Oliveira R.S., de Castro L.C.G., Abreu A.P., Carroll R.S., Latronico A.C., Kaiser U.B., Guerra E.N.S., Lofrano-Porto A. *MKRN3* mutations in central precocious puberty: a systematic review and meta-analysis. *J Endocr Soc*. 2019;3:979-995. doi 10.1210/js.2019-00041

Yarmolinskaya M.I., Ganbarli N.F., Aylamazyan E.K. Role of kisspeptine in regulation of reproductive function. *Zhurnal Akusherstva i Zhenskikh Boleznej = J Obstetrics Women's Diseases*. 2016; LXV(6):4-18. doi 10.17816/JOWD6564-18 (in Russian)

Yuan G., Zhang X., Liu S., Chen T. Chinese familial central precocious puberty with hyperuricemia due to recurrent *DLK1* mutation: case report and review of the literature. *Mol Genet Genomic Med*. 2022;10(12):e2087. doi 10.1002/mgg3.2087

Conflict of interest. The authors declare no conflict of interest.
Received September 26, 2024. Revised November 18, 2024. Accepted November 26, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):310-319

doi 10.18699/vjgb-25-34

Identification of fungal diseases in strawberry by analysis of hyperspectral images using machine learning methods

A.F. Cheshkova

Siberian Federal Scientific Centre of Agro-BioTechnologies of the Russian Academy of Sciences, Krasnoobsk, Novosibirsk region, Russia
cheshanna@yandex.ru

Abstract. Leaf spot, leaf scorch and phomopsis leaf blight are the most common fungal diseases of strawberry in Western Siberia, which significantly reduce its yield and quality. Accurate, fast and non-invasive diagnosis of these diseases is important for strawberry production. This article explores the ability of hyperspectral imaging to detect and differentiate symptoms caused to strawberry leaves by pathogenic fungi *Ramularia tulasnei* Sacc., *Marssonina potentillae* Desm. and *Dendrophoma obscurans* Anders. The reflection spectrum of leaves was acquired with a Photonfocus MV1-D2048x1088-HS05-96-G2-10 hyperspectral camera under laboratory conditions using the line scanning method. Five machine learning methods were considered to differentiate between healthy and diseased leaf areas: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Partial Least Squares Discriminant Analysis (PLS-DA), and Random Forest (RF). In order to reduce the high dimensionality of the extracted spectral data and to increase the speed of their processing, several subsets of optimal wavelengths were selected. The following dimensionality reduction methods were explored: ROC curve analysis method, derivative analysis method, PLS-DA method, and ReliefF method. In addition, 16 vegetation indices were used as features. The support vector machine method demonstrated the highest classification accuracy of 89.9 % on the full range spectral data. When using vegetation indices and optimal wavelengths, the overall classification accuracy of all methods decreased slightly compared to the classification on the full range spectral data. The results of the study confirm the potential of using hyperspectral imaging methods in combination with machine learning for differentiating fungal diseases of strawberries.

Key words: hyperspectral imaging; fungal diseases of strawberries; machine learning methods; dimensionality reduction

For citation: Cheshkova A.F. Identification of fungal diseases in strawberry by analysis of hyperspectral images using machine learning methods. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):310-319. doi 10.18699/vjgb-25-34

Идентификация грибных болезней земляники садовой на основе анализа гиперспектральных изображений методами машинного обучения

А.Ф. Чешкова

Сибирский федеральный научный центр агробиотехнологий Российской академии наук, р. п. Краснообск, Новосибирская область, Россия
cheshanna@yandex.ru

Аннотация. Белая, бурая и угловатая пятнистости являются наиболее распространенными грибными болезнями земляники садовой в Западной Сибири, значительно влияющими на ее урожайность и качество. Точная, быстрая и неинвазивная диагностика этих заболеваний имеет важное значение в промышленном производстве земляники. В настоящей статье исследуются возможности применения методов машинного обучения и гиперспектральной визуализации для обнаружения и дифференциации на листьях земляники симптомов, вызванных патогенными грибами *Ramularia tulasnei* Sacc., *Marssonina potentillae* Desm. и *Dendrophoma obscurans* Anders. Спектр отражения листьев регистрировали гиперспектральной камерой Photonfocus MV1-D2048x1088-HS05-96-G2-10 в лабораторных условиях методом линейного сканирования. Для дифференциации здоровых и пораженных областей листьев изучено пять методов машинного обучения: метод опорных векторов (SVM), метод K-ближайших соседей (KNN), линейный дискриминантный анализ (LDA), дискриминантный анализ частичных наименьших квадратов (PLS-DA) и случайный лес (RF). С целью уменьшения высокой размерности извлеченных спектральных данных и увеличения скорости их обработки было отобрано несколько подмножеств оптимальных длин волн, несущих наиболее важную спектральную информацию. Рассмотрены следующие методы сокращения размерности: метод анализа ROC-кривых, метод анализа производных, метод PLS-DA, метод ReliefF. Кроме того, 16 вегетационных индексов задействовано в качестве информативных признаков. Наиболь-

шую точность классификации, 89.9 %, показал метод опорных векторов на полном спектре значений. При использовании вегетационных индексов и наборов оптимальных длин волн общая точность классификации всех методов снизилась незначительно по сравнению с классификацией на полном спектре значений. Результаты исследования подтверждают перспективность применения методов гиперспектральной визуализации в сочетании с методами машинного обучения для дифференциации грибных болезней земляники садовой.
Ключевые слова: гиперспектральные изображения; грибные болезни земляники; методы машинного обучения; сокращение размерности

Introduction

Strawberry is one of the most popular fruits among consumers by flavor, nutritional value and health benefits (Zheng et al., 2021). Strawberry has high productivity and profitability and is capable of rapid vegetative reproduction. One of the limiting factors for increasing the production of strawberries is the significant damage to cultivated varieties by fungal diseases, which leads to a decrease in yield and economic losses. The most common fungal diseases of strawberries in Western Siberia are leaf spot, leaf scorch and phomopsis leaf blight (Govorova, Govorov, 2015). Early detection of these diseases is crucial for targeted application of appropriate plant protection measures.

Traditional disease diagnostic methods such as visual assessment and microbiological laboratory analysis are time-consuming, error-prone and labor-intensive, which limits their application in precision agriculture. Recently, hyperspectral image analysis (Mishra et al., 2017; Mahlein et al., 2018; Cheshkova, 2022) has demonstrated great potential as an effective and non-invasive method for monitoring plant biotic and abiotic stress. The influence of pathogens causes changes of the physiological and biochemical parameters in the process of disease occurrence, creating a reflectance spectrum that is different from the spectrum of healthy plants. Modern optical sensors register up to several hundred bands of the electromagnetic spectrum over a wide range of wavelengths and form a spectral profile for each pixel combining spectral and spatial information (Mishra et al., 2017). Hyperspectral imaging combines the advantages of computer vision and optical spectroscopy, allowing simultaneous assessment of both physiological and morphological parameters. Currently, scientific publications provide examples of the successful use of hyperspectral imaging for the recognition of various strawberry diseases, such as powdery mildew (Mahmud et al., 2020), anthracnose (Lu et al., 2017; Jiang et al., 2021), verticillium wilt (Cockerton et al., 2019), gray mold (Wu et al., 2023), and spotting (Cheshkova, 2023).

Machine learning is one of the most effective ways to process and analyze the vast amounts of data obtained by remote sensing techniques (Nagaraju et al., 2020; Benos et al., 2021). Numerous studies show that using vegetation indices as features for building machine learning models allows achieving good results in identifying and recognizing diseases of agricultural crops (Mahlein, 2013; Lu et al., 2017).

Hyperspectral data has high collinearity. A large number of wavelengths complicates models and reduces performance. Dimension reduction is specific and significant for hyperspectral-based plant disease analysis, the purpose of which is to remove spectral redundancy while preserving important information. Optimal waveband selection has always been a primary concern in hyperspectral data analysis (Liu et al., 2014; Sun, Du, 2019). Уменьшение размерности может быть достигнуто за счет выбора определенных длин волн либо выделения информативных признаков.

The objective of this study was to determine the efficiency of hyperspectral imaging techniques for differentiating symptoms on strawberry leaves caused by pathogenic fungi *Ramularia Tulasnei* Sacc., *Marssonina potentillae* Desm. and *Dendrophoma obscurans* Anders.; to assess the accuracy of different machine learning methods for identifying fungal diseases of strawberry; to explore the possibility of using dimensionality reduction methods and vegetation indices to optimize the machine learning models.

Materials and methods

Plant material and fungal diseases. In our study, three types of fungal diseases of strawberry, most common in Western Siberia, were considered: leaf spot, leaf scorch and phomopsis leaf blight.

Strawberry leafspot is caused by *Mycosphaerella fragariae* (Tul.) Lindau (conidial stage: *Ramularia Tulasnei* Sacc.). The disease is first noticed as small, purplish circular spots on the surface of young leaflets. As the lesion enlarges, the center of the spot becomes gray to white and is surrounded by distinct reddish-brown borders.

Strawberry leaf scorch is caused by *Diplocarpon earliana* (Ell. et Ev.) Wolf (conidial stage: *Marssonina potentillae* (Desm.) P. Magn., M. fragariae (Lib.) Ohl.). The marks of the disease consist of many small irregular purple spots that appear on the outward leaf's surface. The lesions may enlarge to 5 mm across and appear irregular.

Phomopsis leaf blight is caused by *Dendrophoma obscurans* (Ell. et Ev.) H.W. Anderson (synonym: *Phomopsis obscurans* (Ell. et Ev.) Sutton). Lesions begin as circular to elliptical, purple spots that can appear identical to those of common leaf spot or leaf scorch. The purple spots develop dark brown centers as they enlarge. Some infected leaves display large V-shaped lesions, with the widest part at the leaf edge.

The strawberry plants grown at the experimental field of Siberian Federal Scientific Centre of Agro-BioTechnologies of the Russian Academy of Sciences (Krasnoobsk, Novosibirsk region, Russia) in 2021–2023 were used in the overall experiment. During the growing seasons, 120 plants were selected, including 30 healthy plants and 30 plants for each disease: with visible symptoms of leaf spot, leaf scorch or phomopsis leaf blight. One leaf from each plant was detached for further research in the laboratory. Identification of the disease was carried out through visual expertise by symptoms of the disease (Garrido et al., 2011; Govorova, Govorov, 2015).

Image acquisition and calibration. Imaging was performed by a Photonfocus MV1-D2048x1088-HS05-96-G2-10 hyperspectral camera, with an IMEC CMV2K-LS150-VNIR

sensor (Photonfocus AG, Switzerland, wavelength range 470–900 nm, spectral resolution 3 nm, spatial resolution 2048 × 1088 pixels) by the linear scanning method using a moving platform. The software and hardware equipment as well as related parameters can be referred to article (Maximov et al., 2023). The strawberry leaves were placed on a white platform for imaging. The illumination source was two halogen lamps. The scanning step, exposure time and camera mounting height were determined experimentally. Additionally, reference images were created for radiometric correction. The dark image was obtained by covering the camera lens with a non-reflective opaque black cap, and the white reference image was obtained by surveying the spectral image of the Teflon white board with 99.9 % reflectance. The calibrated image was calculated using the following formula:

$$R = \frac{I_S - I_D}{I_W - I_D},$$

where I_S – intensity value of the sample image, I_D – intensity value of the dark reference image, I_W – intensity value of the white reference image, R – the corrected hyperspectral reflectance image.

Three-dimensional data sets (hypercubes) containing two dimensions of spatial information and additionally one dimension of spectral information (2048 × 1088 × 131) were formed from the scanning results.

Spectral features extraction and processing. The resulting image files were divided into two groups: 96 leaf images to form a training set (24 in each of the four classes) and 24 leaf images to form a validation set (six in each of the four classes). Spectral data extraction was performed using ENVI 5.2 (NV5 Geospatial Solutions, Inc., USA). In each strawberry leaf image, regions of interest (ROIs) corresponding to healthy leaf tissue and to color spots of the diseased tissue were manually selected (Fig. 1). From each region, 250 pixels were ran-

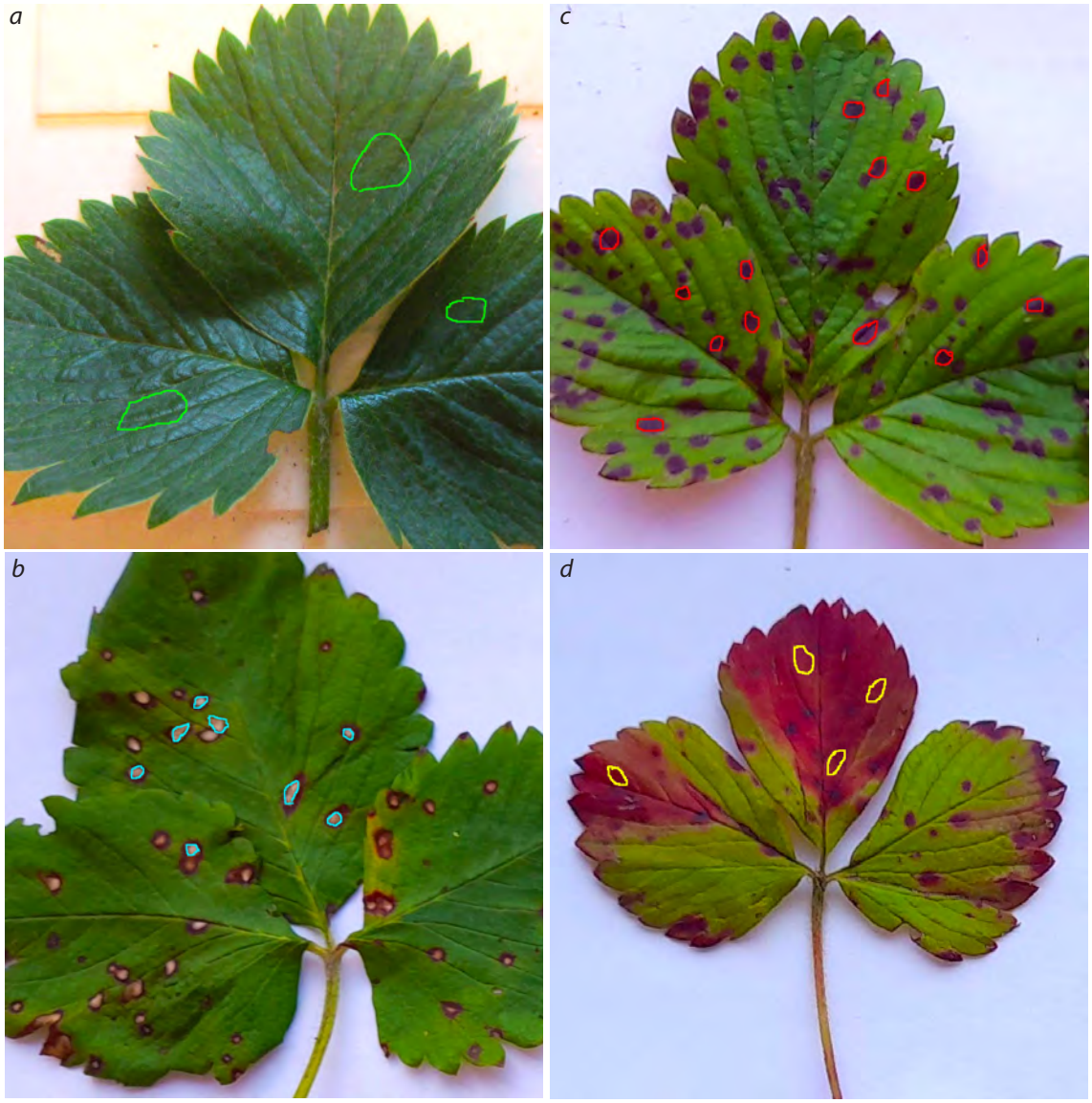


Fig. 1. BROI selection: *a* – healthy regions, *b* – regions with symptoms of leaf spot, *c* – regions with symptoms of leaf scorch, *d* – regions with symptoms of phomopsis leaf blight.

Table 1. Vegetation indices used as features

No.	Index	Equation
1	NDVI (normalized difference vegetation index)	$(R_{800} - R_{670})/(R_{800} + R_{670})$
2	RENDVI (red edge normalized difference vegetation index)	$(R_{750} - R_{705})/(R_{750} + R_{705})$
3	PhRI (physiological reflectance index)	$(R_{550} - R_{531})/(R_{550} + R_{531})$
4	RVSI (red-edge vegetation stress index)	$[(R_{712} + R_{752})/2] - R_{732}$
5	MCARI (modified chlorophyll absorption ratio index)	$[(R_{700} - R_{670}) - 0.2\cdot(R_{700} - R_{550})](R_{700}/R_{670})$
6	TVI (triangular vegetation index)	$0.5\cdot[120\cdot(R_{750} - R_{550}) - 200\cdot(R_{670} - R_{550})]$
7	VOG1 (Vogelman index 1)	R_{740}/R_{720}
8	VOG2 (Vogelman index 2)	$(R_{734} - R_{747})/(R_{715} + R_{726})$
9	VOG3 (Vogelman index 3)	R_{715}/R_{705}
10	PSRI (plant senescence reflectance index)	$(R_{680} - R_{500})/R_{750}$
11	NRI (nitrogen reflectance index)	$(R_{570} - R_{670})/(R_{570} + R_{670})$
12	PSSRa	R_{800}/R_{680}
13	PSSRb (pigments specific simple ratio)	R_{800}/R_{635}
14	CRI (carotenoid reflectance index)	$1/R_{510} - 1/R_{550}$
15	ARI (anthocyanin reflectance index)	$1/R_{550} - 1/R_{700}$
16	PRI (photochemical/physiological reflectance index)	$(R_{531} - R_{570})/(R_{531} + R_{570})$

domly selected. As a result, a training dataset of 24,000 spectrum values (6,000 px for each class) and a validation dataset of 6,000 spectrum values (1,500 px for each class) were formed.

To smooth the spectrum and correct for scattering, the Savitzky–Golay filter (Savitzky, Golay, 1964) and standard normal variate normalization (Vidal, Amigo, 2012) were applied to the spectral data.

Optimal wavelengths selection. In order to decrease the dimension of the raw spectral information and to find the optimal wavelengths for classification we examined the following dimensionality reduction methods:

– receiver operating characteristic (ROC) analysis (Luo et al., 2012); in this method, the area under curve (AUC) value is used as a metric that determines the variable importance; from the entire data spectrum, those wavelengths are left for which the AUC exceeds a certain threshold value;
– the derivative analysis (Savitzky, Golay, 1964); in this method, the most important wavelengths are selected as the high peaks and low valleys in the second derivative plot;
– partial least squares discriminant analysis (PLS-DA) (Mehmood et al., 2012); in this method, the wavelengths that correspond to the highest absolute values of β-coefficients are considered optimal wavelengths;
– ReliefF algorithm (Kononenko, 1994; Urbanowicz et al., 2018); it's a feature weighting algorithms that assigns different weights to features based on the category and correlation of each feature; features with weights below a certain threshold value are removed.

Vegetation indices extraction. Vegetation indices are algebraic combinations calculated from reflectance spectrum values for two or more selected wavelengths.

For our study, 16 vegetation indices (Table 1) were selected, characterizing the photochemical reflectance (PRI), physiology (NDVI, RENDVI, RVSI, PhRI), content of chlorophyll (MCARI, TVI, VOG1, VOG2, VOG3), pigments (PSSRa, PSSRb, CRI, ARI), nitrogen (NRI) and carbon (PSRI) in plant leaves (Wu et al., 2023).

Machine learning methods. Five machine learning methods (SVM, KNN, LDA, PLS-DA, RF), most commonly used in hyperspectral data classification (Singh et al., 2016; Benos et al., 2021), were considered in our study to differentiate healthy and diseased regions of strawberry leaves.

The Support Vector Machine (SVM) method. The main idea of the SVM method is to transfer the original vectors to a higher-dimensional space and search for the separating hyperplane with the largest gap in this space. The Gaussian radial basis function was taken as the classifier kernel.

The K-Nearest Neighbors (KNN) method. The classification is achieved by assigning the test object to the class that is most common among its K-nearest neighbors, the classes of which are already known. It applies the Euclidean distance in the multidimensional space as a similarity measurement to separate the test objects.

The Linear Discriminant Analysis (LDA) method. The high-dimensional data are projected into a lower-dimensional space to promote class separability. The optimal projection in classical LDA is obtained by maximizing the distance

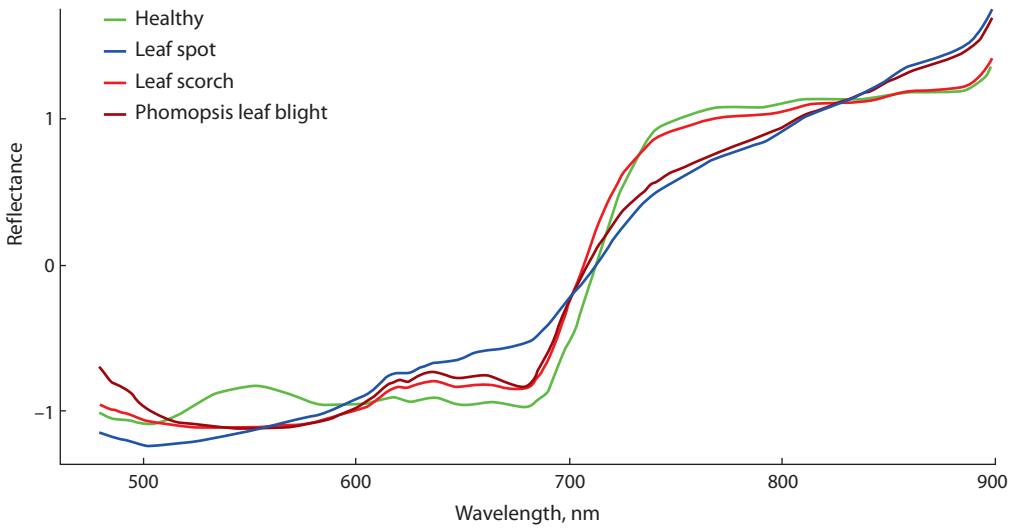


Fig. 2. Average reflectance spectrum of healthy and infected regions of strawberry leaves.

between different classes and minimizing the distance within a class.

The Partial Least Squares Discriminant Analysis (PLS-DA) method. It is a variant of combining Partial Least Squares regression (PLSR) and discriminant analysis (DA). Unlike classical discriminant analysis, which searches for hyperplanes of maximum variance of independent predictors, PLS-DA builds a linear regression model by projecting predicted and observed variables into a new reduced space.

The Random Forest (RF) method. It is a non-parametric method that uses multiple decision trees to classify data and is well suited to spectral data analysis.

The overall accuracy, calculated as the percentage of correctly classified objects to the total number of objects, was used as a metric to evaluate the quality of the models.

All calculations and data analysis were performed in the R software using the caret, kernlab, randomForest, klaR, pls, CORElearn, class, MASS and terra packages.

Results

Spectral behaviors

Figure 2 shows the averaged reflectance spectra of healthy and fungal disease-affected regions of strawberry leaves. The spectral curves are typical for plants (Mishra et al., 2017). A common feature of all spectral curves is a lower reflectance in the visible wavelength range, compared to the near-infrared range. At wavelengths around 670 nm, a decrease in reflectance is observed, which is due to the strong absorption of light by chlorophyll in the leaves. In the range of 670–760 nm, the reflectance of leaves increases sharply due to light scattering in the intercellular space. In the wavelength range of 760–900 nm, the reflectance remains high.

Certain differences between the spectra are observed. Thus, healthy green leaf regions have a characteristic peak at a wavelength of 550 nm (nitrogen absorption zone), while diseased regions have a decline in this area. In the range of 720–810 nm, healthy regions and regions affected by leaf scorch have a higher reflectance, compared to regions affected by leaf spot and phomopsis leaf blight. And in the range of 810–900 nm, on the contrary, it is lower. The reflectance of leaves affected by leaf spot disease increases uniformly over the entire wavelength range.

Optimal wavelengths selection

Analysis of variance (ANOVA) revealed the statistically significant differences between mean reflectance by disease type for each of the wavelengths, according to the F-criterion with a p-value < 0.001. In addition, a recursive feature elimination method was applied to each of the considered models, which also revealed that all wavelengths were significant for classification.

To reduce the dimensionality of the data, four different techniques were considered, which determined four different sets of optimal wavelengths (Figures S1–S4 in Supplementary Materials)[1].

Using ROC curve analysis, 23 wavelengths (nm) were identified for which the AUC exceeded the threshold value of 0.99: [541.39, 545.04, 548.92, 550.41, 553.99, 557.94, 561.3, 565.18, 568.58, 745.48, 748.98, 751.75, 756.45, 759.36, 763.0, 765.97, 769.44, 772.39, 775.92, 778.56, 781.11, 784.53, 787.2].

Using the second derivative analysis method with a threshold value of 1.0, the following 15 wavelengths (nm) were selected: [677.11, 680.47, 682.99, 685.28, 688.76, 691.62, 695.25, 697.97, 709.54, 712.19, 729.07, 732.25, 736.15, 739.20, 742.67].

Using the PLS-DA method, the following 16 wavelengths (nm) were selected for a threshold value of regression coefficients of 0.4: [498.68, 502.7, 505.97, 510.11, 513.5, 517.33, 522.39, 526.49, 529.98, 533.99, 541.39, 680.47, 682.99, 688.76, 691.62, 722.02].

[1] Figures S1–S5 are available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Cheshkova_Engl_29_2.pdf

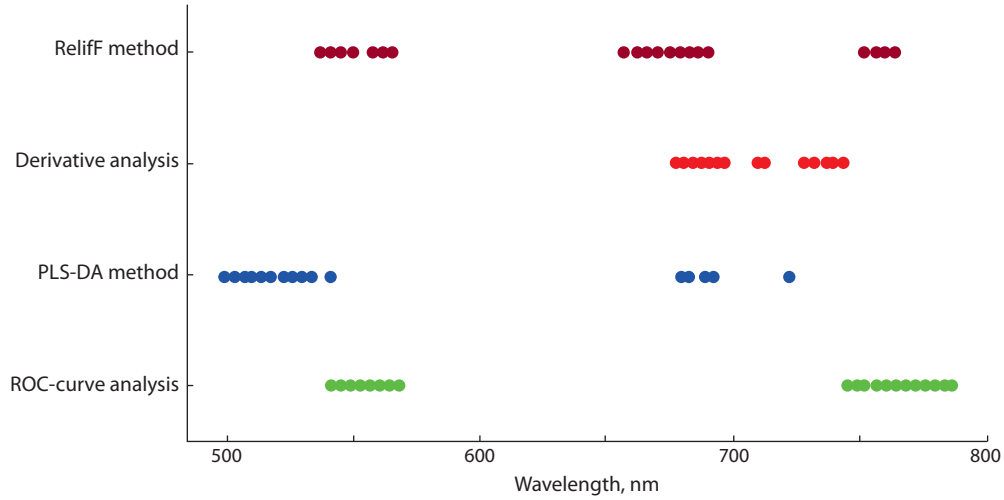


Fig. 3. Sets of optimal wavelengths determined by different methods.

Using the ReliefF method, the following 24 wavelengths (nm) were selected for the weight threshold of 0.5: [537.27, 541.39, 545.04, 548.92, 550.41, 557.94, 561.3, 565.18, 657.08, 662.06, 664.83, 668.35, 670.93, 674.46, 677.11, 680.47, 682.99, 685.28, 688.76, 691.62, 751.75, 756.45, 759.36, 763.00].

A comparison of the sets of optimal wavelengths selected by different methods (Fig. 3) allows us to conclude that the most informative wavelength ranges for classification are [542–565 nm] and [680–691 nm].

Vegetation indices calculation and analysis

Sixteen vegetation indices were calculated using the corresponding formulas (Table 1) for each pixel in the dataset. Analysis of variance (ANOVA) was performed for each index to determine the statistical significance of differences between mean values of indices by disease type. All 16 indices had statistically significant differences between means with a p-value < 0.001.

Classification results based on the full range of wavelengths

In our study, five different models (SVM, KNN, LDA, PLS-DA, RF) were applied to differentiate healthy and fungal disease-affected strawberry leaves. First, models were built for the full spectrum of wavelengths (131 wavelengths in the range 470–900 nm). The following optimal hyperparameters were selected using the cross-validation: SVM (sigma = 0.03, C = 6), KNN (K = 9), RF (mtry = 11), PLS-DA (ncomp = 38). The classification results are shown in Table 2. Analysis of the results allows us to conclude that the main errors in classification occur when differentiating between leaf scorch and phomopsis leaf blight, since these areas have a similar reflectance spectrum.

The support vector machine method on the full range of wavelengths demonstrated the highest classification accuracy (90 %), while the K-nearest neighbors method showed the lowest accuracy (85 %).

Classification results based on sets of optimal wavelengths and vegetation indices

Each of the five classification models (SVM, KNN, LDA, PLS-DA, RF) was trained on sets of optimal wavelengths obtained by applying four different dimensionality reduction methods (ROC curve analysis, derivative analysis, PLS-DA, ReliefF), as well as on a set of 16 vegetation indices (Table 3).

As can be seen from the results presented in Table 3, the overall classification accuracy of all methods decreased compared to the classification using the full spectrum. The highest classification accuracy for all models was obtained for the set of vegetation indices and for the set of wavelengths selected by the PLS-DA method.

Identification of fungal diseases of strawberry

Trained and optimized models can be used to detect and differentiate fungal diseases of strawberry. Figure 4 shows an example of the application of the SVM model for the diagnosis of different types of fungal diseases.

Discussion

The analysis of hyperspectral images by machine learning methods has already been successfully applied in scientific research to detect strawberry diseases. For example, G. Wu et al. (2023) focused on the potential of using hyperspectral imaging (HSI) combined with spectral features, vegetation indices (VIs), and textural features (TFs) for the detection of gray mold on strawberry leaves under laboratory conditions. Three machine learning models (ELM, KNN, SVM) were trained and optimized. The overall classification accuracy of the models reached 96 %.

In (Jiang et al., 2021) six machine learning methods (SVM, ELM, KNN, PLS-DA, RF, NB) were developed based on the selected spectral fingerprint features for early identification of anthracnose and gray mold in strawberries using a hyperspectral imaging system. Most classification models obtain relatively good accuracy (100 %) and robust performance, recognizing asymptomatic fungus infections classes before

Table 2. Confusion matrices for hyperspectral image classification by different methods using the full spectrum

Predicted	Actual				Overall accuracy, %
	Healthy	Leaf spot	Leaf scorch	Phomopsis leaf blight	
SVM					89.90
Healthy	98.07	0.73	0.07	0.00	
Leaf spot	0.73	95.13	1.07	1.27	
Leaf scorch	0.33	3.20	92.67	25.00	
Phomopsis leaf blight	0.87	0.93	6.20	73.73	
KNN					85.17
Healthy	98.73	1.13	0.33	0.00	
Leaf spot	0.27	90.80	1.27	1.73	
Leaf scorch	0.27	5.33	92.13	39.27	
Phomopsis leaf blight	0.73	2.73	6.27	59.00	
RF					86.93
Healthy	98.20	0.87	0.20	0.00	
Leaf spot	0.40	93.27	0.67	0.47	
Leaf scorch	0.27	3.53	90.40	33.67	
Phomopsis leaf blight	1.13	2.33	8.73	65.87	
LDA					89.15
Healthy	99.00	1.00	0.07	0.0	
Leaf spot	0.07	93.07	0.93	0.4	
Leaf scorch	0.93	3.53	89.53	24.6	
Phomopsis leaf blight	0.00	2.40	9.47	75.0	
PLS-DA					87.63
Healthy	99.93	1.87	1.07	0.00	
Leaf spot	0.00	94.60	1.87	0.33	
Leaf scorch	0.00	1.47	81.13	24.80	
Phomopsis leaf blight	0.07	2.07	15.93	74.87	

Note. SVM – the support vector machine method; KNN – the K-nearest neighbors method; LDA – the linear discriminant analysis method; PLS-DA – the partial least squares discriminant analysis method; RF – the random forest method.

Table 3. Overall classification accuracy (%) for different models using selected wavelengths and vegetation indices

Model	Dimensionality reduction method				16 vegetation indices
	ROC curve analysis	Derivative analysis	PLS-DA	ReliefF	
SVM	77.37	83.03	85.13	78.32	89.75
KNN	76.40	81.60	83.27	77.48	82.65
RF	77.55	82.67	83.50	78.57	84.25
LDA	74.32	80.55	85.88	76.93	84.62
PLS-DA	69.00	75.45	85.60	77.15	84.70

Note. SVM – the support vector machine method; KNN – the K-nearest neighbors method; RF – the random forest method; LDA – the linear discriminant analysis method; PLS-DA – the partial least squares discriminant analysis method.

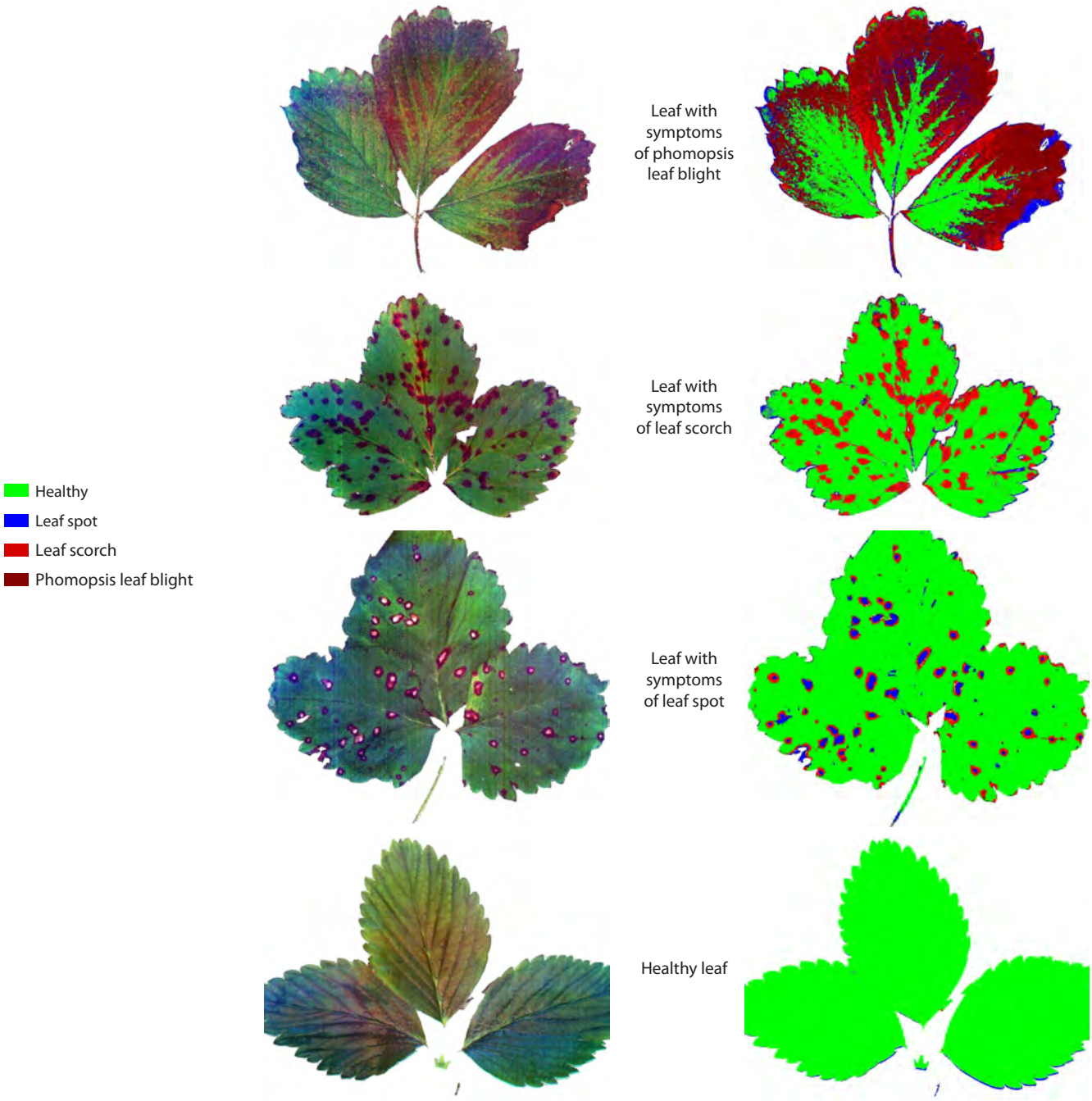


Fig. 4. Visualization of strawberry disease classification using support vector machines on a full range of values.

The left column shows the original color images of the strawberry leaves and the right column visualizes the classification results after applying the SVM model to the full spectrum in every single pixel in the images.

the obvious signs of disease appear notably in the strawberry. In our study, the obtained accuracy of disease classification did not exceed 90 %. This result can be explained by several reasons. First, three types of disease were considered at once, rather than one or two as in other studies. Secondly, successful differentiation of diseases requires a difference in the spectral characteristics of plant leaves affected by pathogens. Our study revealed that the main errors in classification occur when differentiating leaf scorch and phomopsis leaf blight, since these diseases have a similar reflectance spectrum. A possible way to improve classification accuracy is to use convolutional neural networks that take into account not only spectral but also textural characteristics of the affected leaves, such as shape and location of spots.

The choice of classification method depends on the diseases under study. Among the five popular machine learning models we considered (SVM, KNN, LDA, PLS-DA, RF), the support vector machine (SVM) demonstrated the best classification

accuracy, which is in agreement with the results of other studies (Benos et al., 2021).

In order to reduce dimensionality and select optimal wavelengths for model building, researchers have applied various methods. Thus, the CARS, CARS-RF, ReliefF, and ROC algorithms were used in (Luo et al., 2012; Jiang et al., 2021; Wu et al., 2023). In many studies, dimensionality reduction does not reduce the accuracy of the models, but in our case, all wavelengths were significant and the classification accuracy decreased slightly compared to the full spectrum.

The obtained results of laboratory studies indicate the potential of using hyperspectral imaging methods for diagnosing fungal diseases of strawberries in agricultural production. Scientific publications have already described examples of successful application of hyperspectral sensors mounted on UAVs for diagnostics of biotic and abiotic plant stresses (Yang et al., 2017).

In our further research, we plan to test the application of hyperspectral imaging methods in field conditions to automate the diagnosis of fungal diseases of strawberries.

Conclusion

This study explored the feasibility of using hyperspectral imaging technique combined with machine learning for the detection and identification of leaf spot, leaf scorch and phomopsis leaf blight diseases on strawberry leaves in the presence of visible symptoms. In order to identify the strawberry leaves disease effectively, diverse classifiers (SVM, KNN, LDA, PLS-DA, RF) were developed and evaluated using the full spectrum. The Support Vector Machine (SVM) demonstrated the highest classification accuracy of 89.9 % on 131 wavelengths in the range of 470–900 nm. In order to simplify the models and increase the speed of data processing, four different dimensionality reduction methods were considered (ROC curve analysis, derivative analysis, PLS-DA, ReliefF). Moreover, 16 vegetation indices were used as features. The overall classification accuracy of all methods decreased slightly compared to classification using the full spectrum. The set of 16 optimal wavelengths obtained by the PLS-DA method and the set of 16 vegetation indices had higher classification accuracy than the other wavelength sets.

References

Benos L., Tagarakis A., Dolias G., Berruto R., Kateris D., Bochtis D. Machine learning in agriculture: a comprehensive updated review. *Sensors*. 2021;21:3758. doi 10.3390/s21113758

Cheshkova A.F. A review of hyperspectral image analysis techniques for plant disease detection and identification. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2022;26(2):202-213. doi 10.18699/VJGB-22-25

Cheshkova A.F. Application of machine learning methods for the differentiation of fungal diseases in strawberry based on hyperspectral image analysis. *E3S Web Conf*. 2023;390:03018. doi 10.1051/e3sconf/202339003018

Cockerton H.M., Li B., Vickerstaff R.J., Eyre C.A., Sargent D.J., Armitage A.D., Marina-Montes C., Garcia-Cruz A., Passey A.J., Simpson D.W., Harrison R.J. Identifying *Verticillium dahlia* resistance in strawberry through disease screening of multiple populations and image based phenotyping. *Front Plant Sci*. 2019;10:924. doi 10.3389/fpls.2019.00924

Garrido C., Carbu M., Fernandez-Acero J.F., Gonzalez-Rodriguez V.E., Cantoral J.M., New insights in the study of strawberry fungal pathogens. In: Husaini A.M., Mercado J.A. (Eds) Genomics, Transgenics, Molecular Breeding and Biotechnology of Strawberry. Global Science Books, UK, 2011;24-39

Govorova G.F., Govorov D.N. Fungal Diseases of Garden Strawberries: Breeding for Immunity and Other Methods of Protection. Moscow, 2015 (in Russian)

Jiang Q., Wu G., Tian C., Li N., Yang H., Bai Y., Zhang B. Hyperspectral imaging for early identification of strawberry leaves diseases with machine learning and spectral fingerprint features. *Infrared Phys Technol*. 2021;118:103898. doi 10.1016/j.infrared.2021.103898

Kononenko I. Estimating attributes: analysis and extensions of RELIEF. In: Bergadano F., De Raedt L. (Eds) Machine Learning: ECML-94. Lecture Notes in Computer Science. Vol. 784. Berlin: Springer, 1994;171-182. doi 10.1007/3-540-57868-4_57

Liu D., Sun D.W., Zeng X.A. Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry. *Food Bioprocess Technol*. 2014;7:307-323. doi 10.1007/s11947-013-1193-6

Lu J., Ehsani R., Shi Y., Abdulridha J., de Castro A.I., Xu Y. Field detection of anthracnose crown rot in strawberry using spectroscopy technology. *Comput Electron Agric*. 2017;135:289-299. doi 10.1016/j.compag.2017.01.017

Luo X., Takahashi T., Kyo K., Zhang S. Wavelength selection in vis/NIR spectra for detection of bruises on apples by ROC analysis. *J Food Eng*. 2012;109(3):457-466. doi 10.1016/j.jfoodeng.2011.10.035

Mahlein A.-K., Rumpf T., Welke P., Dehne H.-W., Plümer L., Steiner U., Oerke E.-C. Development of spectral indices for detecting and identifying plant diseases. *Remote Sens Environ*. 2013;128:21-30. doi 10.1016/j.rse.2012.09.019

Mahlein A.-K., Kuska M.T., Behmann J., Polder G., Walter A. Hyperspectral sensors and imaging technologies in phytopathology: state of the art. *Annu Rev Phytopathol*. 2018;56:535-558. doi 10.1146/annurev-phyto-080417-050100

Mahmud M.S., Zaman Q.U., Esau T.J., Chang Y.K., Price G.W., Prithiviraj B. Real-time detection of strawberry powdery mildew disease using a mobile machine vision system. *Agronomy*. 2020;10:1027. doi 10.3390/agronomy10071027

Maximov L.V., Gurova T.A., Elkin O.V. Development of software and hardware complex for phenotyping of grain crop stress by hyperspectral images. In: Spatial Data Processing for Monitoring of Natural and Anthropogenic Processes (SDM-2023). Novosibirsk, 2023;289-293. doi 10.25743/sdm.2023.42.73.049 (in Russian)

Mehmood T., Liland K.H., Snipen L., Sæbø S. A review of variable selection methods in Partial Least Squares Regression. *Chemom Intell Lab Syst*. 2012;118:62-69. doi 10.1016/j.chemolab.2012.07.010

Mishra P., Asaari M., Herrero-Langreo A., Lohumi S., Diezma B., Scheunders P. Close range hyperspectral imaging of plants: a review. *Biosyst Eng*. 2017;164:49-67. doi 10.1016/j.biosystemseng.2017.09.009

Nagaraju M., Chawla P. Systematic review of deep learning techniques in plant disease detection. *Int J Syst Assur Eng Manag*. 2020;11(3):547-560. doi 10.1007/s13198-020-00972-1

Savitzky A., Golay M.J.E. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem*. 1964;36(8):1627-1639

Singh A., Ganapathysubramanian B., Singh A.K., Sarkar S. Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci*. 2016;21(2):110-124. doi 10.1016/j.tplants.2015.10.015

Sun W., Du Q. Hyperspectral band selection: a review. *IEEE Geosci Remote Sens Mag*. 2019;7(2):118-139. doi 10.1109/MGRS.2019.2911100

Urbanowicz R.J., Meeker M., Cava W., Olson R.S., Moore J.H. Relief-based feature selection: introduction and review. *J Biomed Inf*. 2018;85:189-203. doi 10.1016/j.jbi.2018.07.014

Vidal M., Amigo J.M. Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemom Intell Lab*. 2012;117:138-148. doi 10.1016/j.chemolab.2012.05.009

Wu G., Fang Y., Jiang Q., Cui M., Li N., Ou Y., Diao Z., Zhang B. Early identification of strawberry leaves disease utilizing hyperspectral imaging combing with spectral features, multiple vegetation indices and textural features. *Comput Electron Agric*. 2023;204:107553. doi 10.1016/j.compag.2022.107553

Yang G., Liu J., Zhao C., Li Z., Huang Y., Yu H., Xu B., Yang X., Zhu D., Zhang X., Zhang R., Feng H., Zhao X., Li Z., Li H., Yang H. Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. *Front Plant Sci*. 2017;8:1111. doi 10.3389/fpls.2017.01111

Zheng C., Abd-Elrahman A., Whitaker V. Remote sensing and machine learning in crop phenotyping and management, with an emphasis on applications in strawberry farming. *Remote Sens*. 2021;13:531. doi 10.3390/rs13030531

Conflict of interest. The author declares no conflict of interest.

Received April 27, 2024. Revised November 7, 2024. Accepted November 7, 2024.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(2):320-329

doi 10.18699/vjgb-25-35

CropGene: a software package for the analysis of genomic and transcriptomic data of agricultural plants

A.Yu. Pronozin [1, 2] ✉, D.I. Karetnikov [1, 2], N.A. Shmakov [1, 2], M.E. Bocharnikova [1, 2], S.D. Afonnikova [1, 2], D.A. Afonnikov [1, 2], N.A. Kolchanov [1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia
✉ PronozinAU@bionet.nsc.ru

Abstract. Currently, the breeding of agricultural plants is increasingly based on the use of molecular biological data on genetic sequences, which makes it possible to significantly accelerate the breeding process, create new plant varieties through genomic editing. These data have a large volume, variety and require a large amount of resources, both labor and computing, to analyze the costs. Data analysis of such volume and complexity can be effective only when using modern bioinformatics methods, which include algorithms for identifying genes, predicting their function, and evaluating the effect of mutation on plant phenotype. Such an analysis has recently become impossible without the use of integrated software systems that solve problems of different levels by executing computational pipelines. The paper describes the CropGene software package developed for the comprehensive analysis of genomic and transcriptomic data of agricultural plants. CropGene includes several blocks of bioinformatic analysis, such as analysis of gene variations, assembly of genomes and transcriptomes, as well as annotation of genes and proteins. CropGene implements new methods for analyzing long non-coding RNAs, protein domains, searching and analyzing polymorphisms, and genome-wide association research. CropGene has a user-friendly interface and supports working with various types of data, which greatly simplifies its use for researchers who do not have deep knowledge in the field of bioinformatics. The paper provides examples of the use of CropGene for the analysis of agricultural organisms such as *Solanum tuberosum* and *Zea mays*. With CropGene, genetic markers have been identified that explain up to 50 % of the variability in seed color parameters; potential genes that may become promising material for producing potato varieties; more than 100 thousand new long non-coding RNAs. Orthogroups were also found, the domain structure of which shows a marked similarity with the domain architecture of characteristic secreted A2 phospholipases. Thus, CropGene is an important tool for scientists and practitioners working in the field of agrobiotechnology and plant genetics.

Key words: bioinformatics pipeline; software package; SNP; analyzing polymorphisms; identification of genes

For citation: Pronozin A.Yu., Karetnikov D.I., Shmakov N.A., Bocharnikova M.E., Afonnikova S.D., Afonnikov D.A., Kolchanov N.A. CropGene: a software package for the analysis of genomic and transcriptomic data of agricultural plants. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(2):320-329. doi 10.18699/vjgb-25-35

Funding. The work on the creation of the CropGene software package was carried out with the support of budget project No. FWNR-2022-0020.

CropGene: программный комплекс анализа геномных и транскриптомных данных сельскохозяйственных растений

А.Ю. Пронозин [1, 2] ✉, Д.И. Каретников [1, 2], Н.А. Шмаков [1, 2], М.Е. Бочарникова [1, 2], С.Д. Афонникова [1, 2], Д.А. Афонников [1, 2], Н.А. Колчанов [1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия
✉ PronozinAU@bionet.nsc.ru

Аннотация. В настоящее время селекция сельскохозяйственных растений все больше опирается на использование молекулярно-биологических данных о генетических последовательностях, что позволяет существенно ускорить селекционный процесс создания новых сортов растений за счет геномного редактирования. Эти данные имеют большой объем, разнообразны и требуют для анализа затрат большого количества ресурсов, как трудовых, так и вычислительных. Анализ данных с такими объемом и сложностью может быть эффективным лишь с применением современных методов биоинформатики, включающих алгоритмы идентификации генов, предсказания их функции, оценку влияния эффекта мутации на фенотип растений. Такой анализ в последнее время стал невозможным без использования интегрированных программных комплексов, решающих задачи разного уровня за счет выполнения вычислительных конвейеров. В статье описан программный комплекс CropGene, раз-

работанный для комплексного анализа геномных и транскриптомных данных сельскохозяйственных растений. Система включает в себя несколько блоков биоинформатического анализа, таких как анализ вариаций генов, сборка геномов и транскриптомов, а также аннотация генов и белков. В комплексе реализованы новые методы анализа длинных некодирующих РНК, белковых доменов, поиска и анализа полиморфизмов и полногеномного исследования ассоциаций. В работе представлены примеры применения CropGene для анализа сельскохозяйственных организмов, таких как *Solanum tuberosum*, *Zea mays*. С помощью данного программного пакета найдены: генетические маркеры, объясняющие до 50 % изменчивости параметров окраски семян; потенциальные гены, которые могут стать перспективным материалом для получения сортов картофеля; более 100 тыс. новых длинных некодирующих РНК. Также обнаружены ортогруппы, доменная структура которых проявляет заметное сходство с доменной архитектурой характерных секретируемых фосфолипаз А2. Таким образом, CropGene представляет собой важный инструмент для ученых и практиков, работающих в области агробиотехнологий и генетики растений.
Ключевые слова: биоинформатический конвейер; программный пакет; SNP; анализ полиморфизмов; идентификация генов

Introduction

In contemporary agricultural science, the development of plant breeding strategies is increasingly dependent on the utilization of molecular biological data, particularly genetic sequence information. Genetic sequence information facilitates a significant acceleration of the breeding process (Khlestkina, 2014) and enables the creation of novel plant varieties through advanced genomic editing techniques. The extensive size, high dimension, and inherent complexity of these data sets demand substantial computational and labor resources for thorough investigation. The effective interpretation of such large-scale and intricate data is achievable only through the application of modern bioinformatics methodologies, which encompass algorithms for gene identification, functional annotation, and the assessment of mutational impacts on phenotypic expression. In recent years, the integration of computational modeling and deep learning algorithms has become indispensable for such analyses. Furthermore, the development of automated computational pipeline technologies is advancing to streamline and optimize data processing workflows within the field of bioinformatics.

Key terms

Intronic lncRNAs – overlap with the intron of a gene

Antisense lncRNAs – oriented against the direction of transcription of a protein-coding gene

Intergenic lncRNAs – located between two gene loci

Genome-Wide Association Studies (GWAS) – a method of genome research aimed at finding statistical relationships between genetic variations and certain phenotype traits

Transcriptome – a set of all transcripts present in a cell at a certain stage of development or under certain physiological conditions

Gene network – a group of coordinated functioning genes interacting with each other both through their primary products (RNA and proteins) and through a variety of metabolites and other secondary products of gene network functioning

The investigation of genetic and transcriptomic information in plant species involves numerous crucial endeavors, notably the examination of genetic variety. Genetic diversity is an important basis for identifying genes associated with resistance to biotic and abiotic stresses, as well as for developing novel, highly adaptive, and high-yielding crop varieties. The assessment of genetic diversity is conducted through a variety of genetic analysis methodologies. Notably, genetic markers play a pivotal role in such studies (Khlestkina, 2014). Among these markers, single-nucleotide polymorphisms (SNPs), which represent single-nucleotide substitutions occurring at varying frequencies within plant populations, are of particular significance (Sukhareva, Kuluev, 2018). SNP analysis is extensively employed to examine allelic polymorphism, analyze haplotypes and pedigrees, and facilitate genotyping and the construction of genetic maps.

In addition to SNP analysis, copy number variation (CNV) is employed to investigate genetic diversity. CNV represents a form of genetic polymorphism characterized by differences in the number of copies of specific genomic regions among individuals. These variations encompass deletions or duplications of individual genes or clusters of linked genes. CNVs can span extensive genomic regions, ranging from several kilobases to millions of base pairs, and play a significant role in contributing to genomic variability and phenotypic diversity.

Genome-wide information on SNPs across hundreds of samples can be obtained through next-generation high-throughput sequencing technologies. SNP identification is achievable using two primary strategies: whole-genome sequencing (WGS) and genotyping by sequencing (GBS) (Scheben et al., 2017). The GBS approach is notably faster and more cost-effective compared to WGS. This efficiency is achieved by sequencing genomic DNA fragments only in proximity to restriction enzyme recognition sites, thereby reducing the overall sequencing cost. However, this method results in fragmented genome coverage and yields a lower density of SNPs compared to comprehensive whole-genome sequencing. Despite these limitations, the data generated

through GBS are sufficiently robust to characterize the genetic diversity of agricultural plant populations with acceptable accuracy. Furthermore, GBS data are widely utilized in genome-wide association studies (GWAS), a powerful tool for identifying genes associated with complex quantitative traits (Burghardt et al., 2017).

In addition to providing fundamental insights into the genetic mechanisms underlying traits of interest, GWAS also facilitate the discovery of genetic markers that can be directly applied to breeding programs (Tsai et al., 2010; Zatybekov et al., 2017; Larkin et al., 2019; Muqaddasi et al., 2020).

Another area of bioinformatics research in agricultural plants involves the assembly of genomes and transcriptomes. Genome assembly represents a foundational step in genomic analysis, providing essential insights into the organization of protein-coding genes, regulatory elements, and mobile genetic elements. The transcriptome, on the other hand, serves as a crucial link between an organism's genome and its phenotypic expression (Velculescu et al., 1997). Currently, the most widely used method for transcriptomic analysis is RNA sequencing (RNA-seq), a high-throughput technology that enables comprehensive profiling of the transcriptome using next-generation sequencing platforms (Shendure, 2008).

The most widely recognized application of RNA-seq is the identification of differentially expressed genes in comparative experiments, such as those involving experimental and control conditions (Drewe et al., 2013). However, beyond this, RNA-seq technology has several other critical applications, including *de novo* transcriptome assembly (Cardoso-Silva et al., 2014), detection of genetic polymorphisms (Piskol et al., 2013), and the discovery of novel splicing variants. When sequencing and reconstructing the genomes of non-model organisms, transcriptome sequencing is often performed in parallel, as it significantly aids in genome annotation, prediction, and functional characterization of protein-coding genes. Nevertheless, due to the extensive genomic and morphological diversity within species, driven by structural variations, a single reference genome is insufficient to capture the complete gene repertoire of a species. To address this limitation, the concepts of pan-genome and pan-transcriptome have been introduced.

Reconstructing genomes and transcriptomes across a population enables the generation and analysis of pan-genomes and pan-transcriptomes in plants (Pronozin et al., 2021). The pan-genome concept encompasses sequences that are subject to structural variation and may be absent from the reference genome of a single representative of the species (Vernikos et al., 2015). Numerous studies have demonstrated that analyzing pan-genomes and pan-transcriptomes enhances the efficiency of research and increases the total number of predicted genes, compared to relying solely on the genome of a single representative (Jin et al., 2016). This approach improves the accuracy and completeness of the gene set under investigation.

Another area of bioinformatic analysis is the annotation of the genome and transcriptome. For protein-coding genes, an important part of their annotation is the identification of protein domains, a structural fragment of a protein that acts as an independent functional unit. It can form a unique structure or be part of multi-domain proteins, functioning both independently and in combination with other domains. For the functional identification of proteins, it is also significant to search already known genomes for orthologs, proteins that perform the same functions in different organisms.

Note also that more than 90 % of all transcripts are not translated into proteins (Carninci et al., 2005) and are noncoding sequences. Noncoding RNAs (ncRNAs) perform a number of important functions in plant genomes related to the regulation of gene expression and homeostasis of plant physiological parameters. One essential class of ncRNAs is long noncoding RNAs (lncRNAs) (Nazipova, 2021). The lncRNAs are a class of linear or circular RNA molecules 200 nucleotides or more in length that do not code proteins (Kim, Sung, 2012). The participation of lncRNAs has been revealed in the regulation of gene expression, formation of the structure of macromolecular complexes, interaction with proteins, and pathogenesis. To date, more than half a million lncRNA sequences have been identified for various organisms.

Data on gene expression levels obtained from transcriptomic experiments are widely used to reconstruct gene networks (Johnson, Krishnan, 2022). Gene networks, in turn, make it possible to model the dynamics of specific processes in an organism and predict its behavior under various conditions.

This paper presents the CropGene system for complex analysis of genomic, transcriptomic data, features of molecular evolution of agricultural plant genes. The system includes blocks of bioinformatic data analysis: analysis of gene variations, assembly of genomes and transcriptomes, annotation of genes and proteins.

Materials and methods

The CropGene software package includes the software packages shown in Figure 1.

The structure of the software package includes the following blocks for solving problems:

Module for the analysis of genome-wide associations. This module implements the following analysis steps:

- analysis of phenotyping data. The phenotyping data is processed using the R, pastecs, and psych packages (Grosjean et al., 2018),
- processing of genotyping data. The step is aimed at processing genotyping data obtained by the microarray genotyping method and the GBS method. Processing includes the evaluation of raw data quality, mapping the reads to the reference genome using BWA-MEM (Li, 2013), and searching for genetic variants using vcftools (Danecek et al., 2011). The variants identified by the above genotyping methods are filtered by quality, minor allele frequency,

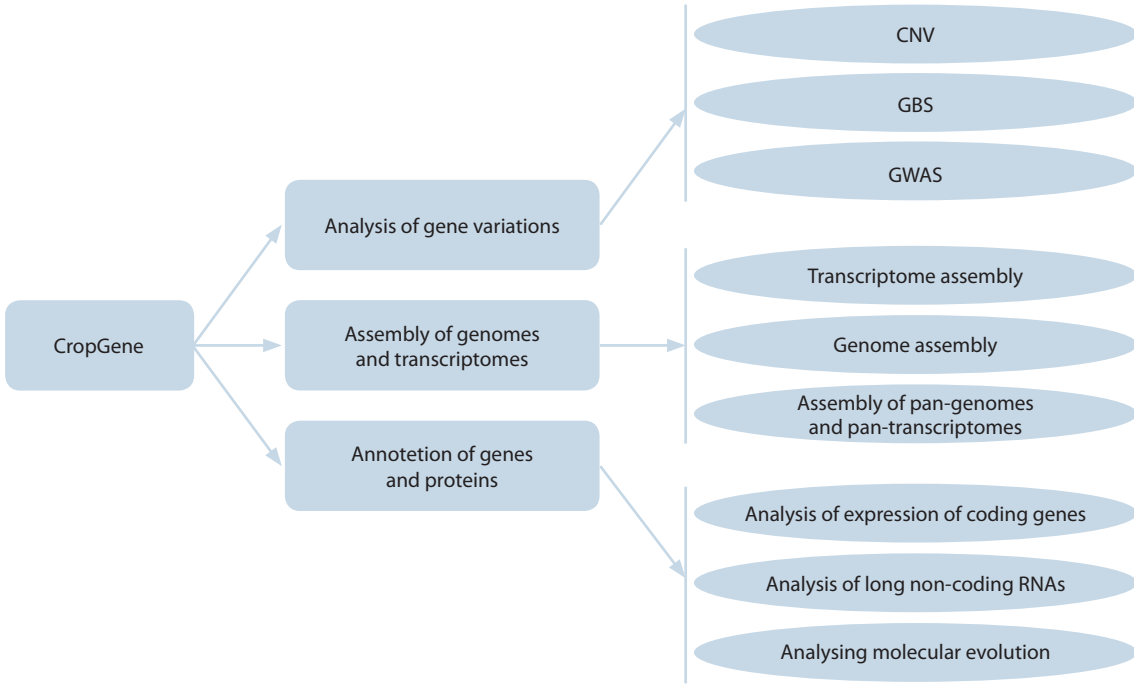


Fig. 1. Diagram of the CropGene software package, with an indication of the main blocks of analysis (rounded rectangles in the center) and specific tasks to be solved (ovals on the right).

heterozygosity, and the amount of missing data. This stage is performed by the bcftools instrument (Danecek et al., 2021). BEAGLE 5.2 is used to impute missing genotyping data (Browning et al., 2018),

- genome-wide analysis of associations. At this stage, genome-wide association analysis is carried out. It is implemented in the R programming language using the functions of the GAPIT3 package (Wang, Zhang, 2021),
- prioritization of genes at the identified loci. This module of genome-wide association analysis is aimed at identifying candidate genes associated with traits of interest. First of all, using the functions of the R "genetics" package, the boundaries of loci are determined, which include variants significantly associated with the phenotype. Further, based on published data on gene expression in the studied organism and on the resources of the Knetminer platform (Hassani-Pak et al., 2021), the genes are prioritized among the detected loci.

Module for the CNV analysis. This module is aimed at solving the tasks of estimating and analyzing variations in the number of copies in the genome. It implements several stages of analysis:

- the sets of raw reads are filtered by quality and length using the fastp program (Chen et al., 2018). Then the filtered and processed reads are mapped to the reference genome of potato using the BWA program (Li, Durbin, 2009). Duplicates in the mapped reads are marked and deleted, after which the reads are sorted and indexed using SAMtools (Li et al., 2009),
- the BAM files are used as input in CNVpytor (Suvakov et al., 2021). Copy number variations are detected on all chromosomes of the reference genome. The detected CNVs are filtered as follows: length greater than 1,000 bp, p-value < 0.01, q0 < 50 %, pN < 50 %. The intansv R package is used to compare the identified CNVs with the genes of the reference genome (Jia et al., 2020),
- for subsequent processing, the CNV list was presented in the form of a matrix in which the rows correspond to a specific genotype, and the columns correspond to the gene of the reference genome. Each element of the matrix is represented in three variants: +1 (potential duplication), –1 (potential deletion) and 0 (absence of a significant CNV). Next, principal component analysis (PCA) is performed using the Scikit-learn v1.1.2 package, which makes it possible to assess genetic diversity (Pedregosa et al., 2011).

Bioinformatic pipeline GBS-DP. This software module is aimed at analyzing the data obtained by the GBS method and consists of three main stages (Pronozin et al., 2023):

- data preprocessing includes checking the quality of raw FastQC readings, removing fast adapters (Chen et al., 2018), and building a reference genome index,
- the search for polymorphisms consists of mapping preprocessed reads to the Bwa-Mem2 reference genome (Li, Durbin, 2009), sorting mapped SAMtools reads (Li et al., 2009), and searching for single-nucleotide polymorphisms Bcftools (Li, 2011),
- the analysis of genetic diversity is divided into two data processing options: if the data obtained exceed the occupied memory capacity of 1 TB and if the data obtained do not exceed the occupied memory capacity of 1 TB. The appropriate option is selected automatically and is

associated with an increased load on the computer's RAM when working with big data. The R – SNPrelate package is used to analyze the main components filtered by SNPs (Zheng, 2013), and the SNPrelate package is used to build a phylogenetic tree.

Transcriptome reconstruction module. This module includes realisation of the following analysis stages:

- contig reconstruction from RNA-seq libraries. Several programs are implemented during this stage: Trinity (Grabherr et al., 2011), Trans-ABySS (Robertson et al., 2010), rnaSpades (Bushmanova et al., 2019),
- aggregation of contig sets obtained during the previous stage and redundancy removal with the tr2aacds.pl tool from the EvidentialGene toolbox,
- quality control of the resulting sequences; the BUSCO software (Simão et al., 2015) assesses transcriptome completeness; kallisto (Bray et al., 2016) shows percentage of initial RNA-seq libraries used in transcriptome reconstruction; rnaQUAST (Bushmanova et al., 2016) evaluates several metrics, including homology with genome sequence of reference organism or closely related organism in case the study is performed on non-model species.

Pangenome reconstruction and analysis module. This module implements the following analysis steps:

- reconstruction of each genome based on paired short reads using the MaSuRCA genome assembler (Zimin et al., 2013),
- masking of mobile genetic elements using RepeatMasker and further *de novo* annotation of reconstructed masked genomes with further translation of open reading frames using the AUGUSTUS program (Stanke et al., 2004),
- identification of orthologous groups in a set of amino acid sequences obtained on the basis of open reading frames using OrthoFinder (Emms, Kelly, 2019).

Gene expression evaluation module. The estimation of gene expression in this module can be performed either based on the reference genome or based on the *de novo* reconsructed transcriptome:

- to quantify the expression of reference genome genes, short read alignment to the genome sequence is performed using the Dart (Lin, Hsu, 2018) software. Next, based on genome annotation and known positions of genes, the number of reads mapped to each gene is estimated using the featureCounts (Liao et al., 2014) software,
- to evaluate expression based on the previously reconstructed transcriptome, the kallisto software is used, which performs so-called pseudoalignment of reads to determine to which transcript they belong to; this allows for quantification of expression levels of these transcripts.

Bioinformatic pipeline ICAnnoLncRNA. This module, aimed at identifying and annotating lncRNAs, implements three stages of processing transcriptomic sequences (Pronozin, Afonnikov, 2023):

1) quality control. This stage includes two operations: the construction of an index file for the genomic sequence by the gmap program (Wu, Watanabe, 2005) and the training of the lncRNA recognition model by the LncFinder v1.1.4 program (Han et al., 2019).
2) lncRNA identification. This block consists of three stages: prediction of lncRNA candidates from the input set of transcripts using the LncFinder method; filtering of the obtained candidate sequences based on the identification of transmembrane segments in the OPC; alignment of filtered lncRNA candidate sequences to the reference genome,
3) annotation. The annotation includes the determination of lncRNA sequence types by alignment to protein-coding genes, identification of conserved lncRNAs, and analysis of the structural features of lncRNAs and their expression.

Module for analyzing protein evolution OrthoDOM. The module implements four key stages of protein sequence analysis:

1) the input data is validated and the presence of functional domains specified by the user for reference proteins is checked for,
2) the presence of key domains in the reference sequences is checked for,
3) the Orthofinder program runs for the studied proteomes,
4) the identified orthologs are checked for the presence of sets of specified domains in their sequence.

Results and discussion

The modules of the CropGene software package have been used to solve various problems of bioinformatic analysis of genomes and transcriptomes of agricultural plants.

A software pipeline that detects CNVs based on genome-wide data was previously used in the analysis of the structure of potato genomes of domestic varieties (Karetnikov et al., 2023). It allowed us to identify all the copy number variations in potato genomes and to conduct a comparative analysis of the number of copies of genes with South American potatoes. The analysis revealed that the frequency of CNV occurrence in four of the 48 known genes associated with tuber formation and photoperiod response differs between the genomes of Russian varieties adapted to long daylight hours in northern latitudes and local Andean cultivars adapted to short daylight hours.

This work used GBS-DP to analyze 219 varieties of barley. 61,620 SNPs were identified. Based on the identified polymorphisms, clustering was performed using the principal component method (Fig. 2) and a dendrogram constructed using the hierarchical clustering method (Fig. 3).

The genome-wide association analysis module was used in the search for candidate genes of common winter wheat associated with pre-harvest sprouting and red grain color (Afonnikova et al., 2024). In addition to the discovery of genetic markers that explain up to 50 % of variability in grain lightness, red and blue color, the work has identified two candidate genes associated with the formation of grain color. The first gene, TraesCS1D02G319700, is located on chromosome 1D and participates in the synthesis of flavonols in the biosynthesis of flavonoids. The other gene,

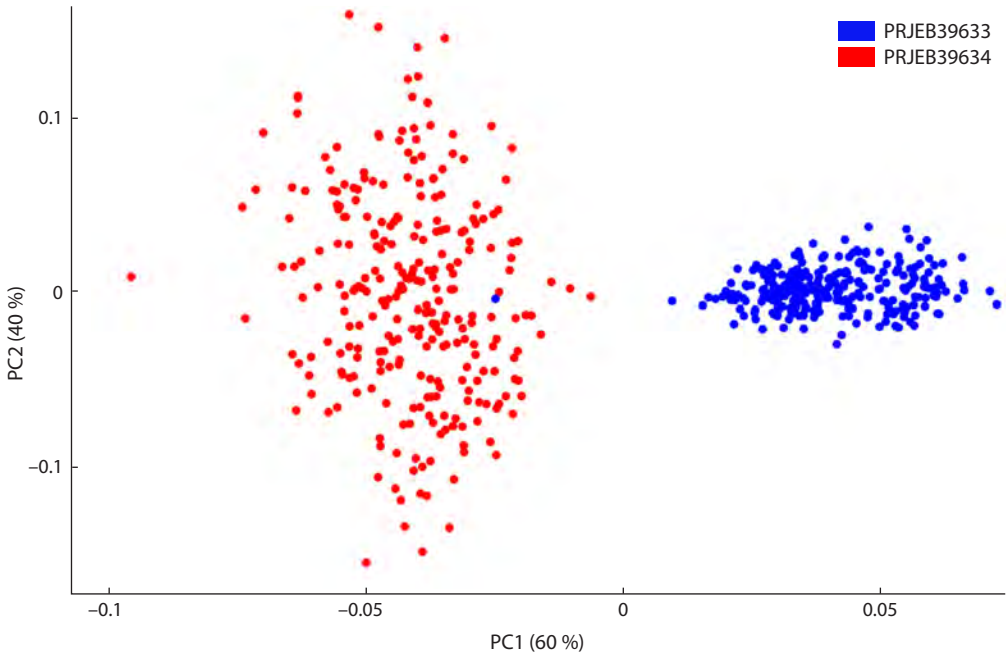


Fig. 2. Visualization of the genetic diversity of 219 barley libraries using the PCA method.
The first and second main components are directed along the *X* and *Y* axes, respectively.

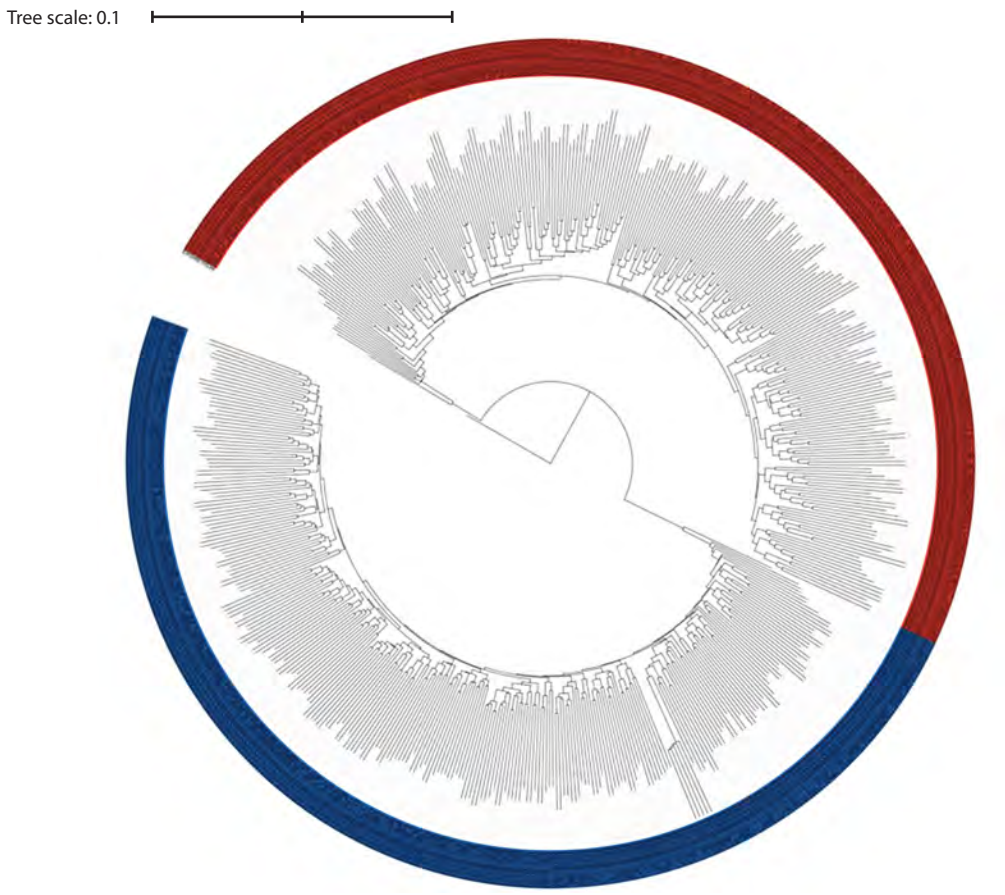


Fig. 3. A dendrogram characterizing the genetic diversity of 219 barley libraries constructed by hierarchical clustering based on GBS data.
The dendrogram is constructed on the basis of the found single nucleotide polymorphisms.

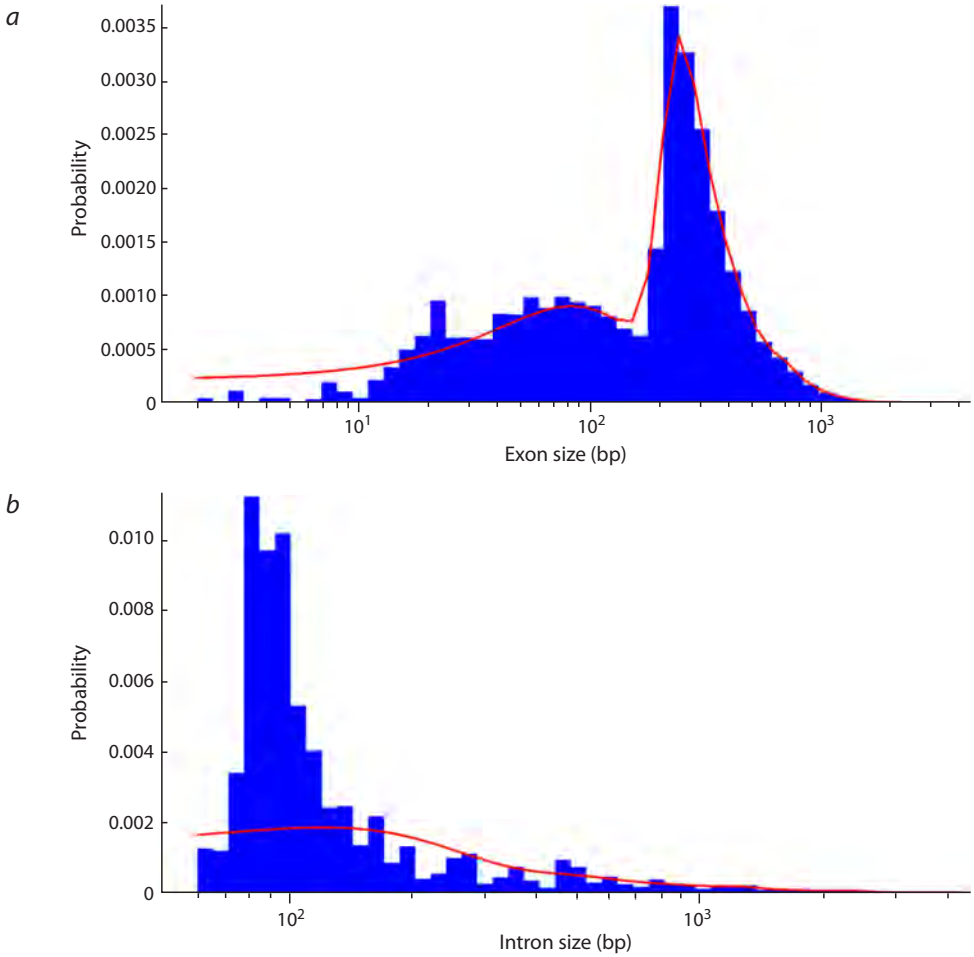


Fig. 4. The ratio of the number of exons per lncRNA (*a*) and the distribution of the size of introns relative to lncRNA, respectively (*b*).

TraesCS7B02G482000, is localized on chromosome 7B and encodes phytoene synthase involved in one of the initial stages of carotenoid synthesis. The main candidate gene for the resistance to pre-harvest sprouting is the TraesCS6B02G147900 gene encoding the aleurone layer morphogenesis protein. Genetic markers were also identified that explain up to 25.3 % of the variability of pre-harvest sprouting traits – the germination index at the milk/hard dough stage of grain development.

Based on the transcriptome analysis module, the transcriptome of four potato varieties of *Solanum tuberosum* group *phureja* (Bintier, Siverskij, Sudarynya, Evraziya) and wild-growing *S. stoloniferum* L. was constructed. Genes encoding proteins of the Nucleotide-binding site – Leucine rich repeats (NBS-LRR) family involved in the formation of the plant immune response were detected (Kochetov et al., 2021). It was found that the repertoires of these genes in the studied potato varieties and in wild nightshade differ significantly, which is consistent with the available data on the rapid evolution of these genes. Some of the NBS-LRR family genes observed in this work had not previously been detected in Solanaceae and potatoes in particular. These genes may become promising material for producing potato varieties that are more resistant to various pathogens and parasites.

The ICAnnoLncRNA pipeline was used to investigate 54 barley transcriptomes. 143,279 new lncRNAs were identified. Of these, 29,987 belong to the class of intronic lncRNAs, 48,369, to intergenic lncRNAs, 64,923, to antisense lncRNAs. Analysis of the lncRNA structure showed that the majority (60 %) contain only one exon. At the same time, the average exon length is 371 nucleotides, a small proportion of exons are up to 10 bp long, the vast majority are from 10 to 1,000 bp long, and their distribution has two characteristic peaks, one wide, with a maximum in the region of 100 bp, and the other narrow, in the region of 250–300 bp (Fig. 4). Tissue specificity analysis showed that the majority of lncRNAs are expressed in the tissues of barley sprouts (Fig. 5*a*). This is observed for both conservative and non-conservative lncRNAs. The same is true for mRNAs (Fig. 5*a*).

The use of the OrthoDOM conveyor to detect phospholipase A2 family proteins in barley and wheat allowed us to confirm their presence in the genomes of these plants.

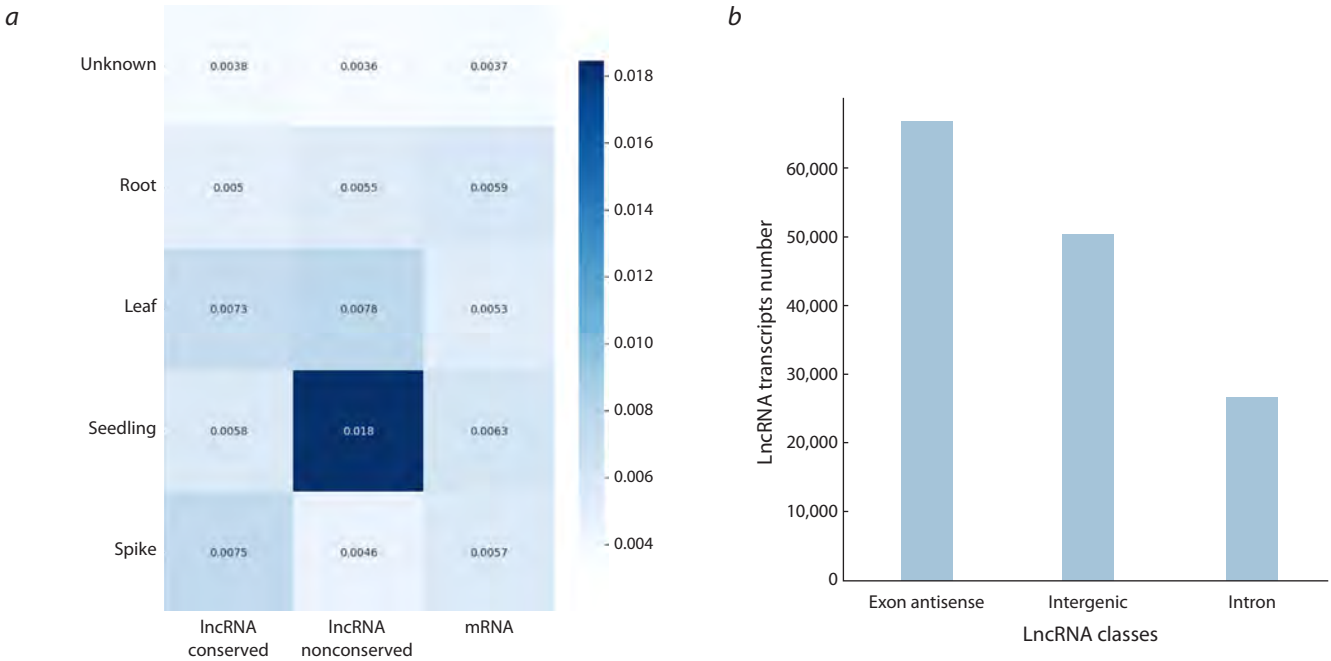


Fig. 5. Specificity of mRNA expression in relation to various barley tissues, shown as a heat diagram.

The *X* axis shows data for two classes of lncRNAs (conservative and non-conservative) and mRNAs. The correspondence of cell color and specificity value is shown by the scale to the right of the diagram (the higher the value in the cell, the more transcripts are specific to that tissue) (*a*), and the distribution of classes of barley lncRNAs (*b*).

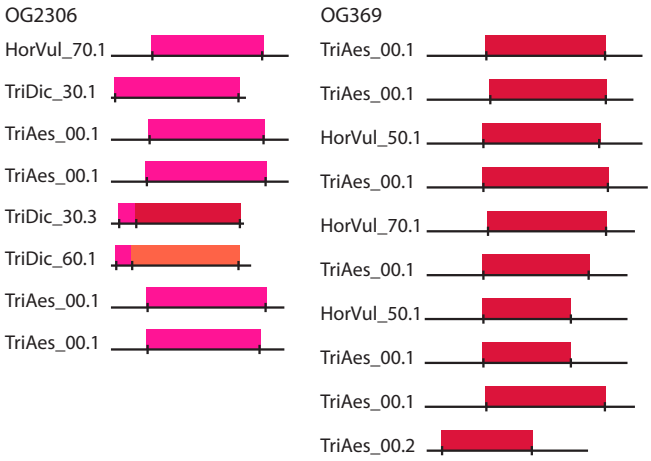


Fig. 6. Domain structure of orthogroup sequences – 2306, 369, from left to right.

The PLA2 beta domain is marked in red, PLA2 alpha is pink, and PLA2G12 is orange.

During the study, two orthogroups were found. The domain structure (Fig. 6) of these groups shows marked similarity to the domain architecture of characteristic secreted A2 phospholipases (Larkin et al., 2019). The length of phospholipase A2 sequences in the orthogroups can be estimated as approximately 150 amino acids, with the PLA2 domain being the predominant part of the sequences, which corresponds to the known structure of secreted PLA2 forms.

Conclusion

The developed CropGene software package includes the main blocks of programs necessary for the analysis of genomic and transcriptomic data of agricultural plants. These are blocks related to the assembly and analysis of the genome and transcriptome, including the formation of a pan-genome and a pan-transcriptome, analysis of GBS data, analysis of gene expression, recognition of long non-coding RNAs in plant transcriptomes, necessary for a comprehensive analysis of genomic, transcriptomic data, and features of the molecular evolution of agricultural plant genes. The use of these modules has made it possible to solve a number of important tasks in the analysis of genomic and transcriptomic data for crops such as potatoes, wheat, and barley.

References

Afonnikova S.D., Kiseleva A.A., Fedyaeva A.V., Komyshev E.G., Koval V.S., Afonnikov D.A., Salina E.A. Identification of novel loci precisely modulating pre-harvest sprouting resistance and red color components of the seed coat in *T. aestivum* L. *Plants*. 2024;13(10):1309. doi 10.3390/plants13101309

Bray N.L., Pimentel H., Melsted P., Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34(5):525-527. doi 10.1038/nbt.3519

Browning B.L., Zhou Y., Browning S.R. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet.* 2018;103(3):338-348. doi 10.1016/j.ajhg.2018.07.015

Burghardt L.T., Young N.D., Tiffin P. A guide to genome-wide association mapping in plants. *Curr Protoc Plant Biol.* 2017;2(1):22-38. doi 10.1002/cppb.20041

Bushmanova E., Antipov D., Lapidus A., Suvorov V., Prjibelski A.D. rnaQUAST: a quality assessment tool for *de novo* transcriptome

assemblies. *Bioinformatics*. 2016;32(14):2210-2212. doi 10.1093/bioinformatics/btw218

Bushmanova E., Antipov D., Lapidus A., Prjibelski A.D. rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. *GigaScience*. 2019;8(9):giz100. doi 10.1093/gigascience/giz100

Cardoso-Silva C.B., Costa E.A., Mancini M.C., Balsalobre T.W.A., Canesin L.E.C., Pinto L.R., Carneiro M.S., Garcia A.A.F., de Souza A.P., Vicentini R. *De novo* assembly and transcriptome analysis of contrasting sugarcane varieties. *PloS One*. 2014;9(2):e88462. doi 10.1371/journal.pone.0088462

Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N., Oyama R., … Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J., Hayashizaki Y. The transcriptional landscape of the mammalian genome. *Science*. 2005;309(5740):1559-1563. doi 10.1126/science.1112014

Chen S., Zhou Y., Chen Y., Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884-i890. doi 10.1093/bioinformatics/bty560

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R.; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156-2158. doi 10.1093/bioinformatics/btr330

Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., Whitwham A., Keane T., McCarthy S.A., Davies R.M., Li H. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10(2):giab008. doi 10.1093/gigascience/giab008

Drewe P., Stegle O., Hartmann L., Kahles A., Bohnert R., Wachter A., Borgwardt K., Rätsch G. Accurate detection of differential RNA processing. *Nucleic Acids Res.* 2013;41(10):5189-5198. doi 10.1093/nar/gkt211

Emms D.M., Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 2019;20(1):238. doi 10.1186/s13059-019-1832-y

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., … Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29(7):644-652. doi 10.1038/nbt.1883

Grosjean P., Ibanez F., Etienne M., Grosjean M.P. Package 'Pastecs'. 2018. Available online: http://masterdistfiles.gentoo.org/pub/cran/web/packages/pastecs/pastecs.pdf

Han S., Liang Y., Ma Q., Xu Y., Zhang Y., Du W., Wang C., Li Y. LncFinder: an integrated platform for long non-coding RNA identification utilizing sequence intrinsic composition, structural information and physicochemical property. *Brief Bioinform.* 2019;20(6):2009-2027. doi 10.1093/bib/bby065

Hassani-Pak K., Singh A., Brandizi M., Hearnshaw J., Parsons J.D., Amberkar S., Phillips A.L., Doonan J.H., Rawlings C. KnetMiner: a comprehensive approach for supporting evidence-based gene discovery and complex trait analysis across species. *Plant Biotechnol J.* 2021;19(8):1670-1678. doi 10.1111/pbi.13583

Jia L., Liu N., Huang F., Zhou Z., He X., Li H., Wang Z., Yao W. intansv: an R package for integrative analysis of structural variations. *PeerJ*. 2020;8:e8867. doi 10.7717/peerj.8867

Jin M., Liu H., He C., Fu J., Xiao Y., Wang Y., Xie W., Wang G., Yan J. Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation. *Sci Rep.* 2016;6(1):18936. doi 10.1038/srep18936

Johnson K.A., Krishnan A. Robust normalization and transformation techniques for constructing gene coexpression networks from RNA-seq data. *Genome Biol.* 2022;23(1):1. doi 10.1186/s13059-021-02568-9

Karetnikov D.I., Vasiliev G.V., Toshchakov S.V., Shmakov N.A., Genaev M.A., Nesterov M.A., Ibragimova S.M., Rybakov D.A., Gavrilenko T.A., Salina E.A., Patrushev M.V., Kochetov A.V., Afonnikov D.A. Analysis of genome structure and its variations in potato cultivars grown in Russia. *Int J Mol Sci.* 2023;24(6):5713. doi 10.3390/ijms24065713

Khlestkina E.K. Molecular markers in genetic studies and breeding. *Russ J Genet Appl Res.* 2014;4:236-244. doi 10.1134/S2079059714030022

Kim E.-D., Sung S. Long noncoding RNA: unveiling hidden layer of gene regulatory networks. *Trends Plant Sci.* 2012;17(1):16-21. doi 10.1016/j.tplants.2011.10.008

Kochetov A.V., Afonnikov D.A., Shmakov N., Vasiliev G.V., Antonova O.Y., Shatskaya N.V., Glagoleva A.Y., Ibragimova S.M., Khiutti A., Afanasenko O.S., Gavrilenko T.A. NLR genes related transcript sets in potato cultivars bearing genetic material of wild Mexican Solanum species. *Agronomy*. 2021;11(12):2426. doi 10.3390/agronomy11122426

Larkin D.L., Lozada D.N., Mason R.E. Genomic selection – considerations for successful implementation in wheat breeding programs. *Agronomy*. 2019;9(9):479. doi 10.3390/agronomy9090479

Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993. doi 10.1093/bioinformatics/btr509

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv.* 2013;1303.3997

Li H., Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi 10.1093/bioinformatics/btp324

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi 10.1093/bioinformatics/btp352

Liao Y., Smyth G.K., Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930. doi 10.1093/bioinformatics/btt656

Lin H.-N., Hsu W.-L. DART: a fast and accurate RNA-seq mapper with a partitioning strategy. *Bioinformatics*. 2018;34(2):190-197. doi 10.1093/bioinformatics/btx558

Muqaddasi Q.H., Brassac J., Ebmeyer E., Kollers S., Korzun V., Argillier O., Stiewe G., Plieske J., Ganal M.W., Röder M.S. Prospects of GWAS and predictive breeding for European winter wheat's grain protein content, grain starch content, and grain hardness. *Sci Rep.* 2020;10(1):12541. doi 10.1038/s41598-020-69381-5

Nazipova N.N. Variety of non-coding RNAs in eukaryotic genomes. *Matematicheskaya Biologiya i Bioinformatika = Mathematical Biology Bioinformatics*. 2021;16(2):256-298. doi 10.17537/2021.16.256 (in Russian)

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., … Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12:2825-2830

Piskol R., Ramaswami G., Li J.B. Reliable identification of genomic variants from RNA-seq data. *Am J Hum Genet.* 2013;93(4):641-651. doi 10.1016/j.ajhg.2013.08.008

Pronozin A.Yu., Afonnikov D.A. ICAnnoLncRNA: A Snakemake pipeline for a long non-coding-RNA search and annotation in transcriptomic sequences. *Genes*. 2023;14(7):1331. doi 10.3390/genes14071331

Pronozin A.Yu., Bragina M.K., Salina E.A. Crop pangenomes. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2021;25(1):57-63. DOI 10.18699/VJ21.007

Pronozin A.Yu., Salina E.A., Afonnikov D.A. GBS-DP: a bioinformatics pipeline for processing data coming from genotyping by sequencing. *Vavilov J Genet Breed.* 2023;27(7):737-745. doi 10.18699/VJGB-23-86

Robertson G., Schein J., Chiu R., Corbett R., Field M., Jackman S.D., Mungall K., … Hirst M., Marra M.A., Jones S.J., Hoodless P.A., Bi-

rol I. *De novo* assembly and analysis of RNA-seq data. *Nat Methods*. 2010;7(11):909-912. doi 10.1038/nmeth.1517

Scheben A., Batley J., Edwards D. Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnol J*. 2017;15(2):149-161. doi 10.1111/pbi.12645

Shendure J. The beginning of the end for microarrays? *Nat Methods*. 2008;5(7):585-587. doi 10.1038/nmeth0708-585

Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19): 3210-3212. doi 10.1093/bioinformatics/btv351

Stanke M., Steinkamp R., Waack S., Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res*. 2004;32(Suppl. 2):W309-W312. doi 10.1093/nar/gkh379

Sukhareva A.S., Kuluev B.R. DNA markers for genetic analysis of crops. *Biomika = Biomics*. 2018;10(1):69-84. doi 10.31301/2221-6197.bmcs.2018-15 (in Russian)

Suvakov M., Panda A., Diesh C., Holmes I., Abyzov A. CNVpytor: a tool for copy number variation detection and analysis from read depth and allele imbalance in whole-genome sequencing. *GigaScience*. 2021;10(11):giab074. doi 10.1093/gigascience/giab074

Tsai M.-C., Manor O., Wan Y., Mosammaparast N., Wang J.K., Lan F., Shi Y., Segal E., Chang H.Y. Long noncoding RNA as modular scaffold of histone modification complexes. *Science*. 2010;329(5992): 689-693. doi 10.1126/science.1192002

Velculescu V.E., Zhang L., Zhou W., Vogelstein J., Basrai M.A., Bassett D.E., Hieter P., Vogelstein B., Kinzler K.W. Characterization of the yeast transcriptome. *Cell*. 1997;88(2):243-251. doi 10.1016/S0092-8674(00)81845-0

Vernikos G., Medini D., Riley D.R., Tettelin H. Ten years of pan-genome analyses. *Curr Opin Microbiol*. 2015;23:148-154. doi 10.1016/j.mib.2014.11.016

Wang J., Zhang Z. GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinformatics*. 2021;19(4):629-640. doi 10.1016/j.gpb.2021.08.005

Wu T.D., Watanabe C.K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics*. 2005; 21(9):1859-1875. doi 10.1093/bioinformatics/bti310

Zatybekov A., Abugalieva S., Didorenko S., Gerasimova Y., Sidorik I., Anuarbek S., Turuspekov Y. GWAS of agronomic traits in soybean collection included in breeding pool in Kazakhstan. *BMC Plant Biol*. 2017;17(S1):179. doi 10.1186/s12870-017-1125-0

Zheng X. A tutorial for the R Package SNPRelate. Washington, USA: University of Washington, 2013

Zimin A.V., Marçais G., Puiu D., Roberts M., Salzberg S.L., Yorke J.A. The MaSuRCA genome assembler. *Bioinformatics*. 2013;29(21): 2669-2677. doi 10.1093/bioinformatics/btt476

Conflict of interest. The authors declare no conflict of interest.

Received November 27, 2024. Revised January 15, 2025. Accepted January 15, 2025.

Прием статей через электронную редакцию на сайте http://vavilov.elpub.ru/index.php/jour
Предварительно нужно зарегистрироваться как автору, затем в правом верхнем углу страницы выбрать «Отправить рукопись». После завершения загрузки материалов обязательно выбрать опцию «Отправить письмо», в этом случае редакция автоматически будет уведомлена о получении новой рукописи.

«Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» до 2011 г. выходил под названием «Информационный вестник ВОГиС»/ "The Herald of Vavilov Society for Geneticists and Breeding Scientists".

Сетевое издание «Вавиловский журнал генетики и селекции (Vavilov Journal of Genetics and Breeding)» – реестровая запись СМИ Эл № ФС77-85772, зарегистрировано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций 14 августа 2023 г.

Издание включено ВАК Минобрнауки России в Перечень рецензируемых научных изданий, в которых должны быть опубликованы основные результаты диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук, Russian Science Citation Index, Российский индекс научного цитирования, ВИНИТИ, Web of Science CC, Scopus, PubMed Central, DOAJ, ROAD, Ulrich's Periodicals Directory, Google Scholar.

Открытый доступ к полным текстам:
русскоязычная версия – на сайте https://vavilovj-icg.ru/
и платформе Научной электронной библиотеки, elibrary.ru/title_about.asp?id=32440

англоязычная версия – на сайте vavilov.elpub.ru/index.php/jour
и платформе PubMed Central, https://www.ncbi.nlm.nih.gov/pmc/journals/3805/

✉ email: vavilov_journal@bionet.nsc.ru

Издатель: Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук»,
проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Адрес редакции: проспект Академика Лаврентьева, 10, Новосибирск, 630090.
Секретарь по организационным вопросам С.В. Зубова. Тел.: (383)3634977.
Издание подготовлено информационно-издательским отделом ИЦиГ СО РАН. Тел.: (383)3634963*5218.
Начальник отдела: Т.Ф. Чалкова. Редакторы: В.Д. Ахметова, И.Ю. Ануфриева. Дизайн: А.В. Харкевич.
Компьютерная графика и верстка: Т.Б. Коняхина, О.Н. Савватеева.
Дата публикации 31.03.2025. Формат 60 × 84 $^{1}/_{8}$. Уч.-изд. л. 21.5.