2026 • 30 • 5

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-26-71

2026 • 30 • 5

Online ISSN 2500-3259

Online edition

VAVILOVSKII ZHURNAL GENETIKI I SELEKTSII

VAVILOV JOURNAL OF GENETICS AND BREEDING

Founded in 1997
Publication frequency: 8 issues a year
doi 10.18699/vjgb-26-71

СОДЕРЖАНИЕ • 2026 • 30 • 5

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

CONTENTS • 2026 • 30 • 5

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):713-724

doi 10.18699/vjgb-26-72

Pathogenic modulators of cellular senescence

E.V. Simoroz [1], Y.V. Antonov [1], G.S. Muravyov[1], J. Vasilevska [1] ✉, E.I. Rogaev [2]

[1] Sirius University of Science and Technology, Sirius Federal Territory, Krasnodar region, Russia
[2] UMass Chan Medical School, Department of Psychiatry, Shrewsbury, United States
✉ vasilevskaya.e@talantiuspeh.ru

Abstract. Cellular senescence is characterized by irreversible cell-cycle arrest, with cells remaining viable and metabolically active. This state features a proinflammatory senescence-associated secretory phenotype (SASP) that can harm neighboring tissues. Accumulation of senescent cells accelerates age-related physiological decline and associated pathologies. Furthermore, cellular senescence is implicated in various physiological processes, including embryonic development, wound healing, tumor progression, and immune response regulation. The complexity of the aging process arises from its diverse underlying mechanisms, potential reversibility, and intrinsic heterogeneity. Experimental gerontology focuses on identifying pathogenic modulators that regulate the formation and accumulation of senescent cells, as well as investigating their impact on tissue function. Within the field of experimental gerontology, considerable attention is devoted to identifying pathogenic modulators that regulate the formation and accumulation of senescent cells, as well as to investigating their impact on tissue function. Of particular interest is the characterization of novel molecular mechanisms that govern cellular aging. Key pathogenic molecular pathways include the p53-dependent senescence pathway, the p16INK4a/pRb pathway, and non-canonical pathways like IFIH1-MAVS, which contribute to oxidative stress, DNA damage, activation of SASP, and other cellular dysfunction. This study critically examines how viral and bacterial agents induce cellular senescence, particularly *in vitro*, reviewing the regulatory mechanisms involved. It discusses genetic variants affecting infection susceptibility and categorizes senescence markers. Investigating the molecular mechanisms underlying cellular aging presents promising avenues for the development of targeted and effective therapeutic interventions. Such strategies may include the selective induction of senescence in cancer cells, suppression of senescence to mitigate age-related diseases, or the comprehensive modulation of aging processes to optimize clinical outcomes.

Key words: senescence; viruses; bacteria; pathogenic modulators

For citation: Simoroz E.V., Antonov Y.V., Muravyov G.S., Vasilevska J., Rogaev E.I. Pathogenic modulators of cellular senescence. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(5):713-724. doi 10.18699/vjgb-26-72

Funding. The study was supported by the Sirius Federal Territory. Agreement 18-03, project MB-BFT-2403 (A.Y.V., M.G.S., V.J.).

Acknowledgements. The schematic figures were created using BioRender.com, while additional elements of the figures were produced by Illustrae Neuron.

Патогенные модуляторы клеточного старения

Е.В. Симороз [1], Е.В. Антонов [1], Г.С. Муравьев[1], Е. Василевская [1] ✉, Е.И. Рогаев [2]

[1] Научно-технологический университет «Сириус», федеральная территория «Сириус», Краснодарский край, Россия
[2] Медицинская школа Чан Массачусетского университета, департамент психиатрии, Шрусбери, США
✉ vasilevskaya.e@talantiuspeh.ru

Аннотация. Клеточное старение (сенесценция) представляет собой необратимую утрату способности к пролиферации при сохранении жизнеспособности и метаболической активности клетки. Этот процесс сопровождается формированием провоспалительного секреторного фенотипа, ассоциированного со старением (SASP), который способен оказывать повреждающее воздействие на окружающие ткани и организм в целом. Накопление сенесцентных клеток в тканях способствует ускорению возрастных изменений организма и развитию сопутствующих патологий, а также оказывает влияние на различные физиологические процессы, включая эмбриональное развитие, заживление ран, рост опухолей и регуляцию иммунного ответа. Сложность процесса старения обусловлена многообразием его механизмов, потенциальной обратимостью и гетерогенностью. В области экспериментальной геронтологии особое внимание уделяется выявлению патогенных модуляторов, регулирующих образование и накопление сенесцентных клеток, и исследованию их влияния на функциональное состояние тканей. Особый интерес представляет выявление новых молекулярных механизмов, участвующих в регуляции процессов клеточного старения. Основные патогенно опосредованные

молекулярные пути включают p53-опосредованное старение и путь p16INK4a/pRb, а также неканонические механизмы, такие как IFIH1-MAVS, которые вызывают окислительный стресс, повреждение ДНК, активацию SASP и другие клеточные нарушения. В данной работе проводится анализ влияния вирусов и бактерий на процессы клеточного старения с основным акцентом на индукцию сенесценции *in vitro*. Представлен всесторонний обзор регуляторных механизмов, посредством которых бактерии и вирусы инициируют старение клеток. Описаны генетические варианты, которые обусловливают предрасположенность к инфекциям, что, в свою очередь, влияет на патоген-индуцируемые процессы старения. Кроме того, систематизированы и проанализированы маркеры, применяемые для идентификации сенесцентных клеток. Исследование молекулярных механизмов клеточного старения открывает возможности для создания целенаправленных и эффективных терапий. Эти стратегии могут избирательно активировать старение в раковых клетках, подавлять его для замедления возрастных заболеваний или комплексно регулировать процессы старения для оптимальных результатов.
Ключевые слова: старение; вирусы; бактерии; патогенные модуляторы

Introduction

Infectious agents are factors that contribute to the aging process, exerting their effects through two interrelated mechanisms. First, they cause direct damage to cells and tissues through infection, leading to various age-related pathologies, including cardiovascular diseases (notably stroke) (Donato et al., 2018; Smolgovsky et al., 2021), neurodegenerative disorders (Alzheimer's disease and Parkinson's disease) (Pringsheim et al., 2014; Trevisan et al., 2019; Blackhurst, Funk, 2023), immune system dysfunctions (Sadighi Akha, 2018), as well as oncological diseases (DeSantis et al., 2019; Kannampuzha et al., 2023). Second, they indirectly promote the induction of aging and the accumulation of senescent cells, which, in turn, is recognized as one of the key mechanisms underlying the development of chronic diseases and overall decline in organismal viability. These changes lead to a progressive decline in physical and cognitive functions, as well as to an increased risk of disease and mortality (Aman et al., 2021).

Aging in mammals, including humans, is associated with progressive dysfunction of the immune system, termed "immunosenescence" (Arakelyan et al., 2025). This process is characterized by a quantitative decline and reduced functional activity of key immune cell populations, as well as disruption of immune homeostasis. Consequently, it leads to a weakened adaptive immune response, increasing susceptibility to infectious diseases and exacerbating their severity in old age. Immunosenescence is also considered one of the drivers of systemic inflammation and the acceleration of aging processes. Epidemiological data serve as clinical evidence of age-related vulnerability, demonstrating a higher incidence of SARS-CoV-2 infection and a greater risk of complications among elderly individuals compared to the 18–39 age cohort. In contrast, the immune response in childhood is characterized by high activity and adaptive potential, reflecting pronounced age-related dynamics in the immune system's response to pathogens (McGovern et al., 2023; Schmitt et al., 2023).

Cellular senescence is categorized into replicative and stress-induced senescence. Replicative senescence is associated with telomere shortening and the activation of pathways such as p53 and p21. Stress-induced senescence is driven by increased expression of the *CDKN2A* gene and the p16 protein, leading to cell cycle arrest (Yan et al., 2024).

For reliable identification of senescent cells, a combination of markers is used (Fig. 1): for example, β-galactosidase activity (SA-β-gal) and the presence of the senescence-associated secretory phenotype (SASP) (Fafián-Labora, O'Loghlen, 2020; Ogrodnik et al., 2024; Wyles et al., 2024).

There is evidence indicating that the senescent state represents a phenomenon with a high, yet not absolute, barrier to reversibility. A clear illustration of this is the induction of Yamanaka factors (Oct4, Sox2, Klf4, c-Myc), which can reprogram the cell and eliminate epigenetic markers of aging, reverting even senescent cells to a pluripotent state, with the subsequent possibility of their differentiation into various cell types. Under experimental conditions, reversal of senescence can be induced through genetic or epigenetic manipulations; however, this approach carries substantial risks, primarily the potential for oncogenesis (Aguirre et al., 2023; Jiang et al., 2026). Another example is provided by extracellular vesicles derived from embryonic stem cells. These vesicles have been shown to act as mediators modulating the activity of neighboring cells: they enhanced proliferative potential and reduced the expression of SA-β-gal as well as other markers of aging, such as mitochondrial membrane potential and reactive oxygen species levels. Furthermore, a decrease in the expression of classical markers of cellular senescence – p21WAF1/CIP1 (p21) and p16INK4a (p16) – was observed. This process has been found to be mediated by inhibition of the p38 MAPK signaling pathway (Liu Yu. et al., 2023).

The study by M. Reimann and colleagues (2024) demonstrates that the degree of reversibility of senescent states is more complex and context-dependent. Following exposure to a pathogen or a stressor, the development of the senescent phenotype can occur rapidly – within days or weeks. In the early stages, cells may retain a reversible state and the ability to restore function. However, as damage accumulates and corresponding signaling pathways

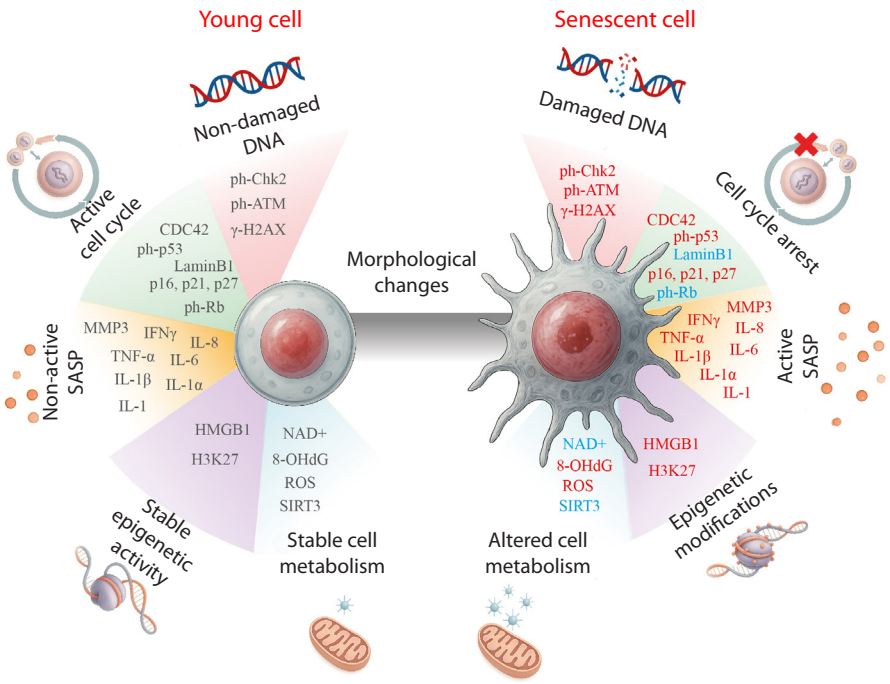


Fig. 1. Schematic representation comparing a young and a senescent cell, indicating the main changes and classical markers.

Markers, the expression of which is upregulated in senescent cells, are highlighted in red compared to the gray control, whereas markers with downregulated expression in senescent cells are indicated in blue.

(e. g. those mediated by p53 and p16INK4a) become activated, senescence stabilizes and acquires an irreversible character (Reimann et al., 2024).

In addition to age-related susceptibility to infections, it is important to note that individual aging trajectories can vary significantly depending on genetics. Genetic mutations may determine individual aging trajectories in response to infectious challenges. Such genes typically include those that maintain immune homeostasis (e. g. *PTPN2* and *FOXO3*) (Hale et al., 2020; Jeanpierre et al., 2024), as well as genes responsible for adaptive immunity, including HLA genes, the Toll-like receptor (TLR) family, and interferon regulatory factors (IRF) (Zhang et al., 2015; De Jong et al., 2018; Augusto et al., 2023). Furthermore, genes encoding receptors used by pathogens for cell entry also play a significant role.

It should be noted that research on aging processes in the context of infectious diseases is a rapidly evolving field of science, with a substantial number of new studies published monthly (Alam et al., 2021; Ryu et al., 2021; Masters et al., 2022; Aguado et al., 2023; Gu et al., 2025). This work reflects growing interest in this topic and opens prospects for identifying novel targets in combating infections and associated pathologies. The present review examines aging as a direct consequence of pathogen exposure, with a particular focus on *in vitro* studies that describe the activated signaling pathways and the identified markers of this process.

Bacteria as positive modulators of aging

There is growing interest in studying the role of the microbiome and bacterial toxins in regulating the aging processes of the host organism. Bacteria can either promote or delay cellular aging processes, depending on their species and metabolic activity. Beneficial probiotic microorganisms synthesize metabolites, such as short-chain fatty acids, which possess anti-inflammatory and antioxidant effects. These metabolites contribute to slowing down the aging processes by strengthening immunity and reducing levels of chronic inflammation.

In the study by S.-J. Han and colleagues (2023), it was demonstrated that different microbiota profiles shape distinct levels of immune optimization, which is particularly important in the context of combating pathogens and inflammation. Interventions aimed at modulating the microbiota, such as the use of probiotics, may represent effective strategies for preventing or delaying pathogen-induced aging. An optimal microbiome balance enhances immune responses, reduces inflammatory processes, and decreases the risk of developing age-related diseases (Han et al., 2023).

However, some pathogens, upon infecting cells, are capable of manipulating molecular pathways, which often leads to premature aging. This effect may manifest through DNA damage in host cells or the secretion of genotoxic proteins, thereby exacerbating cellular aging processes and negatively affecting the overall state of the

organism (Chumduri et al., 2016). For instance, the bacterium *Helicobacter pylori*, which causes chronic atrophic gastritis, acts as a factor inducing cellular senescence and contributing to cancer development, representing an evident phenomenon in the gastric mucosa. It has been demonstrated that this bacterium causes dysregulation of the Erk signaling pathway, leading to the accumulation of the cyclin-dependent kinase inhibitor p21Cip1/Waf1, which results in the arrest of cell proliferation characteristic of the aging process (Saito et al., 2010). Another study showed that *H. pylori* induces senescence of gastric epithelial cells both *in vitro* and *in vivo* through activation of the chemokine receptor CXCR2 signaling pathway (Cai et al., 2021).

Another example is provided by the bacteria *Porphyromonas gingivalis* and *P. asaccharolytica*, identified in significant quantities in patients with colorectal cancer. Cocultivation of human fibroblasts in a medium containing culture supernatants of these bacteria led to irreversible cell cycle arrest, accompanied by increased expression of p16INK4a and p21Cip1/Waf1, as well as SASP factors (IL-1 and IL-6). Similar results were obtained in intestinal epithelial organoids. This effect is attributed to the high secretion of butyrate – a short-chain fatty acid – by these bacteria. The authors hypothesize that the impact of bacteria in inducing senescence processes contributes to the development and progression of colorectal cancer (Okumura et al., 2021).

Disorders of the hematopoietic system also represent a prominent sign of aging, accompanied by a reduced capacity for self-renewal, an increased number of myeloid cells, accumulation of DNA damage, and epigenetic and transcriptional changes. It has been suggested that the microbiota also plays a significant role in these processes. The study by L.V. Kovtonyuk et al. (2022) showed that in aged mice, the bone marrow releases more IL-1α and IL-1β, and patterns associated with microbes, such as TLR4 and TLR8 ligands, are detected in the blood. The authors suggest that the microbiome stimulates bone marrow cells to increase cytokine production. Supporting this hypothesis, it was found that bone marrow cells from aged mice, upon lipopolysaccharide stimulation, exhibit a more pronounced and sustained IL-1α/β production response both *in vitro* and *in vivo*. Moreover, antibiotic treatment reduced IL-1 levels in the bone marrow, confirming the role of the microbiome in shaping age-related changes in hematopoiesis via the IL-1R1 pathway (Kovtonyuk et al., 2022).

Similar to what has been described above, aging of the cardiovascular system, driven by endothelial cell senescence, may also be a consequence of bacterial activity. It has been shown that the bacterium *Clostridium* sp. ASF356, characterized by high production of the bacterial metabolite phenylacetic acid and its by-product phenylacetylglutamine, is significantly more abundant in aged mice and humans. In young mice, infection with these bacteria increases blood levels of phenylacetic acid and induces endothelial cell senescence, increasing mitochondrial H_2O_2. In aortic cells of infected mice, increased expression of the senescence marker p16INK4a, DNA damage (γ-H2A.X), elevated SASP factors, as well as increased β-galactosidase activity, are observed (Saeedi Saravi et al., 2025).

Another significant factor capable of inducing aging processes is *Mycobacterium tuberculosis*, the causative agent of tuberculosis. Infected patients exhibit accelerated epigenetic aging by approximately 12.7 years relative to their chronological age, accompanied by elevated levels of SASP components CXCL9, CXCL10, and TNFα (Bobak et al., 2022). Studies on murine alveolar-like macrophages have demonstrated that *M. tuberculosis* induces accelerated senescence *in vitro*. This results in reduced cell proliferation, morphological changes, activation of DNA damage pathways, and increased expression of senescence markers such as p21 and SA-β-galactosidase. Senescent macrophages secrete pro-inflammatory cytokines (SASP) and promote senescence of neighboring cells (Scarpa et al., 2024). It is known that *M. tuberculosis* infects approximately one-third of the world's population and causes about two million deaths annually. Nevertheless, only 10 % of infected individuals develop clinically manifest tuberculosis (Lin, Flynn, 2010). This phenomenon is likely due to genetic predisposition. A significant association of the disease with a single nucleotide polymorphism (SNP) in exon rs4331426, located on chromosomal region 18q11.2, has now been established (Thye et al., 2010). However, functional confirmation of this association is lacking.

Similar to the causative agent of tuberculosis, the malaria parasite *Plasmodium falciparum* is capable of inducing accelerated cellular senescence, where host genetic factors play a key role in the predisposition to developing severe malaria and its complications. Analysis of postmortem brain and peripheral blood samples from malaria patients revealed elevated levels of the p21 marker in the absence of changes in p16 expression (Hellani et al., 2024). Furthermore, this pathogen contributes to accelerated aging of erythrocytes, primarily due to alterations in the biochemical composition of the membrane, manifested by reduced phospholipid and cholesterol content. These changes play a significant role in the pathogenesis of malaria. The authors hypothesize that these alterations may result from peroxidative oxidative damage induced by the parasites (Omodeo-Salè et al., 2003). In a genome-wide association study (GWAS) comparing patients and healthy individuals, the most significant association signal was detected near the beta-hemoglobin gene, which contains the classic sickle hemoglobin S (*HbS*) polymorphism. It should be

noted that homozygotes for the rs334 allele develop life-threatening sickle cell anemia, whereas heterozygotes for this variant have an approximately tenfold reduced risk of severe malaria. Additionally, further genetic variants requiring additional study and validation have been identified, including mutations in the *SCO1* and *DDC* genes (Jallow et al., 2009).

One of the pressing issues in contemporary immunology is the study of immunosenescence induced by various bacterial agents. The study by S.L. Mathiasen and colleagues (2021) demonstrated that genotoxins from pathogenic Gram-negative bacteria (e. g., *Escherichia coli*, *Shigella dysenteriae*, *Campylobacter jejuni*, and others) induce premature senescence of CD4+ T lymphocytes. Cocultivation with these bacteria led to the acquisition of a senescent phenotype (SASP) in T cells, with increased levels of cytokines IL-13, IL-17, IFN-γ, and markers of DNA damage (ATM, γH2A.X). Molecular analysis of senescence-induction pathways revealed increased expression of phosphorylated p53 protein, accompanied by intense nuclear localization of p21CIP1/WAF1 (Mathiasen et al., 2021). It has previously been shown that the microbiota initiates neutrophil aging through activation of signaling pathways mediated by Toll-like receptors TLR2 and TLR4, as well as the adaptor protein myeloid differentiation factor 88 (Myd88). Notably, senescent neutrophils do not express classical markers of cellular senescence but exhibit elevated CXCR4 levels, which contributes to their reduced numbers in the bone marrow. Consequently, neutrophil production is inhibited through signaling cascades involving IL-17 and granulocyte colony-stimulating factor (G-CSF). Moreover, aging neutrophils are characterized by decreased L-selectin expression. Functional analysis showed that these cells have impaired migratory capacity and attenuated pro-inflammatory functions (Zhang et al., 2015).

In the context of B-lymphocyte senescence, it has been demonstrated that this process is also induced by the bacterial microbiota, leading to reduced production of immunoglobulin A (IgA) in the intestine and, as a result, exerting a modulatory effect on the composition and functional state of the microbiota (Kawamoto et al., 2023).

Bacteria as negative modulators of aging

A number of studies have demonstrated the positive influence of certain bacterial strains on processes associated with delayed aging. In 2025, L. Liu and colleagues investigated the effects of bacteria from plant roots on the lifespan of the nematode *Caenorhabditis elegans*. The most notable was the strain *Mycobacterium* sp. Root265, which synthesizes polysaccharides and arabinogalactan peptidoglycan, enhancing protein homeostasis and increasing lifespan through DAF-16-dependent and -independent pathways (Liu L. et al., 2025). Using the same nematode model, it was demonstrated that administration of live lactic acid bacteria, in particular *Lactobacillus rhamnosus* CNCM I-3690, increases their viability by 30 % and also extends the mean lifespan of the worms by 20 %. This effect is achieved through activation of DAF-16, which is associated with insulin-like signaling (Grompone et al., 2012).

Research using mouse models has shown that the bacterium *Parabacteroides merdae* can inhibit aging processes. It produces branched-chain amino acids that are converted into isovalerate, 2-methylbutyrate, and isobutyrate. In a mouse model of induced accelerated aging (under D-galactose administration), treatment with these metabolites reduced oxidative stress and inflammation, improved muscle function, increased acetylcholine levels in the brain, and normalized blood glucose levels (Azman, Zakaria, 2019). Microbiota analysis revealed an increase in beneficial bacteria, such as *Bifidobacterium pseudolongum* (Wang H. et al., 2025).

In a study conducted on healthy elderly individuals, it was found that administration of the probiotic BB68S – a specific strain of *Bifidobacterium longum* – significantly improved the cognitive functions of the participants. Specifically, improvements were observed in immediate memory, visuospatial and constructive memory, attention, as well as delayed memory. Furthermore, BB68S administration had a beneficial effect on the gut microbiota, promoting an increase in the relative abundance of beneficial bacteria such as *Lachnospira*, *Bifidobacterium*, *Dorea*, and *Cellulosilyticum*, while simultaneously reducing the abundance of bacteria associated with cognitive impairment, including *Collinsella*, *Parabacteroides*, *Tyzzerella*, *Bilophila*, and others (Shi et al., 2022).

Accumulated scientific evidence convincingly indicates that chronic and recurrent bacterial infections represent a significant factor contributing to the acceleration of systemic aging and the development of age-associated pathologies. The primary mechanism underlying this effect is the persistent activation of classical pro-inflammatory and cellular stress molecular pathways, the key ones of which are summarized in Figure 2*a*.

Despite the current understanding of general mechanisms, the fundamental question of genetic predisposition to frequent or severe bacterial infections capable of inducing premature aging remains largely underexplored. For instance, although researchers have identified monogenic susceptibility variants for *M. tuberculosis*, data for the vast majority of other bacteria remain extremely fragmentary. Groundbreaking advances in microbiome research, especially the study of interactions between both commensal and pathogenic bacteria and host cells, are crucial for identifying novel molecular signaling pathways involved in lifespan regulation. These insights are essential for advancing this field further.

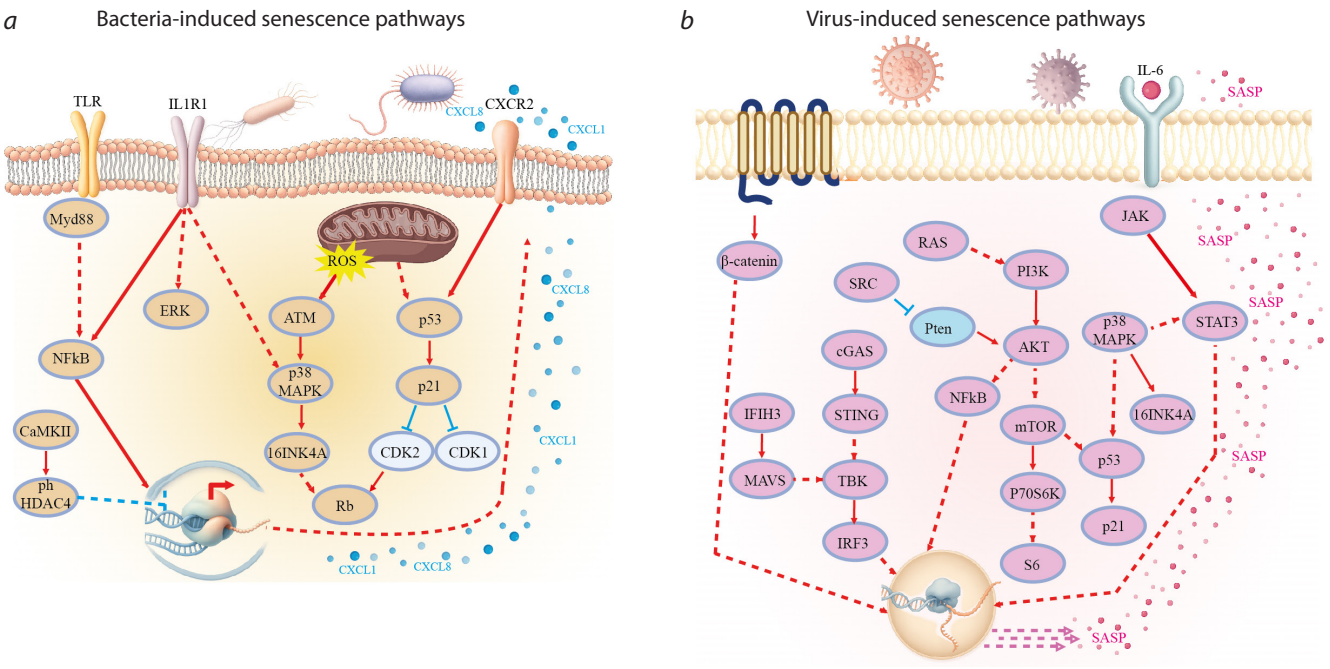


Fig. 2. Illustration of molecular signaling pathways involved in senescence processes activated by (*a*) bacteria and (*b*) viruses.
Solid arrows indicate the molecular connections discussed in this review, while dashed arrows represent established, known molecular connections.

Viruses as modulators of aging

In contrast to bacterial infections, viral infections exert an exclusively stimulatory effect on cellular senescence processes. Additionally, increasing evidence points to a genetic predisposition that worsens the course and consequences of viral infections. Viruses represent complex regulators of aging processes. Viruses act as complex regulators of cellular senescence: some trigger senescence as a protective measure to limit their replication, while others exploit senescent cells to facilitate their propagation. This interaction contributes to pathological conditions during viral infections, especially in elderly individuals, through mechanisms such as cytokine storms and immune response dysfunction (Seoane et al., 2020).

Although there is no evidence supporting a rejuvenating effect of viral infections, the role of viruses in inducing senescence remains unclear. Some studies report that certain viruses, such as herpesviruses, increase cellular senescence and accelerate tissue aging. However, other research either does not confirm these effects or finds them only in specific populations or after prolonged infection. These discrepancies may result from differences in research models (cell cultures, animal studies, or clinical data), experimental conditions, virus types, and individual genetic predispositions. Additionally, the impact of viruses may vary depending on the infection stage, the patient's immune status, and comorbidities. This highlights the need for further research to clarify viruses' exact role in aging and to identify potential therapeutic targets.

Recently, significant attention has focused on the SARS-CoV-2 virus. Studies indicate that it induces cellular senescence in infected cells as a stress response. In COVID-19 patients, senescence markers have been detected in the airways, alongside elevated levels of SASP factors in the blood. Senolytic drugs, such as navitoclax and the combination of dasatinib with quercetin, have demonstrated potential in eliminating virus-induced senescent cells and reducing lung damage in animal models (Lee et al., 2021). Additionally, the viral spike protein accelerates cellular senescence by increasing Cdc42 expression, which activates the Wnt/β-catenin pathway. Treatment with Cdc42 inhibitors has been shown to reduce lung damage and inflammation (Nong et al., 2024).

Genetics also plays a significant role in the risk of SARS-CoV-2 infection. Specific genetic markers associated with increased susceptibility have been identified, including variants rs11246068-CC, rs12252-CC, rs1990760, and rs2111485 in the *IFITM3* gene; rs5742933-GG in the *ORMDL1* gene; rs35337543-CG in the *IFIH1* gene; and the GGGCT haplotype, which includes polymorphisms rs2070788, rs2298659, rs17854725, rs12329760, and rs3787950 in the *TMPRSS2* gene. These genes are directly involved in antiviral immune responses and viral replication. These genetic variants are important markers of infection susceptibility and may also play a role in accelerating virus-induced aging processes (Beckmann, Becker, 2021; Xu et al., 2022; Majeed et al., 2024; Azcarate et al., 2025).

Similarly, infection with influenza A virus (IAV) leads to the development of both acute and chronic lung tissue damage. In a mouse study using a sublethal dose of the H1N1p2009 virus, cellular senescence was detected, marked by increased expression of p16, p21, β-galactosidase, and the DNA damage marker γ-H2A.X in lung tissue. The process of cellular senescence began as early as day four post-infection in the bronchial epithelium and subsequently spread to the lung parenchyma. Thus, virus-induced cellular senescence contributes to the development of influenza-related pulmonary complications (Lipskaia et al., 2025). Genetic factors also affect susceptibility and severity of influenza A infections. Analysis of patients with severe pneumonia identified polymorphisms on chromosomes 1 and 17 that may influence risk, including variants in *FCGR2A* (rs1801274), *RPAIN* (rs8070740), and *C1QBP* (rs3786054) genes (Zúñiga et al., 2012). Additional studies found polymorphisms in *IFITM3* (rs12252), and Toll-like receptor genes *TLR2*, *TLR3*, and *TLR4* (rs5743708, rs5743313, rs5743315, rs4986790, and rs4986791) (Hidaka et al., 2006; Esposito et al., 2012; Chen C. et al., 2016). Furthermore, polymorphisms in *CD55* (rs2564978) and *C1QBP* (rs3786054) have been linked to increased risk of fatal outcomes of influenza infection (Chatzopoulou et al., 2019).

Human respiratory syncytial virus (HRSV) is a leading cause of lower respiratory tract infections. Infection with HRSV induces cell cycle arrest at the G1 or G2/M phases, depending on the cell type, through accumulation of proteins p53 and p21 (Bian et al., 2012). It also triggers the expression of DNA damage markers and proliferation arrest markers, such as phosphorylated ATM (ph-ATM) and γH2A.X (Martínez et al., 2016). Regarding genetic susceptibility, a single nucleotide polymorphism in the olfactory receptor gene *OR13C5*, rs199665292, has been identified as a promising candidate variant. Additionally, variants in the major histocompatibility complex HLA genes (*HLA-DQA1*, *HLA-DPB1*) and the mucin 4 gene (*MUC4*) have been implicated (Salas et al., 2017). Mutations in *TLR4*, specifically Asp299Gly and Thr399Ile, either alone or combined, occur more frequently in patients with severe disease forms (Tulic et al., 2007).

Human cytomegalovirus (HCMV) is the causative agent of cytomegaly. To replicate its double-stranded DNA genome, HCMV induces significant alterations in cellular homeostasis, analogous to cellular senescence. It promotes decreased expression of Lamin B1 and KI67, alongside increased secretion of pro-inflammatory cytokines IL-6 and IL-8 in various cell types, which are well-established markers of cellular senescence (Raviola et al., 2024). Supporting this, T. Hasegawa et al. (2023) found HCMV DNA and RNA in normal human skin, with viral levels significantly higher in samples from elderly individuals, suggesting enhanced viral replication in senescent cells (Hasegawa et al., 2023). While age-related predisposition to HCMV infection is recognized, several genetic variants have been proposed as contributors to infection progression, though their clinical relevance remains uncertain. The rs11686168 polymorphism in the *CDC42EP3* gene is the most promising candidate, although its functional role is not yet understood (Casto et al., 2021). It is hypothesized that *CDC42EP3* may not directly influence the virus but could affect cellular processes that viruses exploit.

Human immunodeficiency virus type 1 (HIV-1), belonging to the retrovirus family, promotes premature aging in patients with chronic infection. Individuals with HIV infection exhibit decreased bone mineral density, as well as an increased prevalence of osteopenia and osteoporosis. This phenomenon is attributed to the viral proteins Tat and Nef, which induce premature aging of bone marrow mesenchymal stem cells, reducing their proliferation due to oxidative stress and mitochondrial dysfunction. Consequently, their ability to differentiate into osteoblasts is diminished. The Tat protein additionally activates NF-κB and stimulates the secretion of pro-inflammatory cytokines, contributing to the inflammatory response (Beaupere et al., 2015). HIV-1 also induces senescence in microglial cells, characterized by increased β-galactosidase activity, elevated p21, and higher levels of cytokines IL-6 and IL-8, alongside decreased expression of SIRT3, a mitochondrial enzyme essential for cellular homeostasis and metabolism. These changes coincide with increased mitochondrial oxidative stress and the development of a SASP (Chen N.C. et al., 2018; Thangaraj et al., 2021).

In addition to viral infections that accelerate aging (Lee et al., 2021), the host genome contains retroelements – highly mobile DNA sequences integrated throughout the genome, many of which also exist as extrachromosomal DNA. Reactivation of endogenous retroviruses (ERVs), a subtype of these retroelements, has been linked to the aging process and is driven by an epigenetic switch. A class of ERVs is epigenetically silenced in proliferating cells but becomes activated during replicative senescence, RAS- and Akt-induced senescence, or after *HDAC4* gene knockout (Di Giorgio et al., 2024). Moreover, ERVs are activated by the transcription factor ATF3, leading to the formation of double-stranded RNAs that trigger the RIG-I/MDA5-MAVS signaling pathway and stimulate type I interferon (IFN-I) production in senescent fibroblasts. This aging-associated molecular mechanism has been observed in tissues from healthy elderly individuals as well as patients with Hutchinson–Gilford progeria syndrome (Mao et al., 2024).

HIV-1 is among the most extensively studied viruses, with recent research showing that susceptibility to HIV-1 acquisition is a polygenic and heritable trait. Heritability estimates based on single nucleotide polymorphisms (SNPs) account for 28 to 42 % of the variation in sus-

ceptibility across the population. Genetic analyses have identified *EFCAB14* as a novel gene linked to HIV-1 susceptibility. Additionally, a higher genetic risk for HIV-1 acquisition correlates with lower levels of the chemokine ligand *CCL17* (Powell et al., 2020). It is known that HIV-1 uses the CCR5 co-receptor for infection, and in some cases, the CCR2 and CXCR4 co-receptors. Consequently, genetic mutations in these receptors represent an important factor influencing susceptibility to infection. The frequency of heterozygous mutations in the *CCR5* gene is significantly elevated among individuals who are able to tolerate infection for many years. Moreover, the CCR5delta32 and CCR2-64I mutations may act as factors limiting HIV-1 infection and also confer a dominant phenotype of delayed disease progression in infected patients (Dean et al., 1996; Smith et al., 1997; Martin, 1998).

Thus, numerous studies confirm that viruses act as potent inducers of cellular senescence. They initiate this process through various molecular pathways, as schematically represented in Figure 2*b*. It is known that individual susceptibility to viral infections and the severity of their course have a genetic basis. However, most of the identified predisposition genes encode proteins directly involved in virus recognition and the formation of innate or adaptive immune responses (e. g. interferon pathway genes, pattern recognition receptors) (Table S1 in Supplementary Material)[1]. However, functional studies directly linking these genetic variants to the modulation of virus-induced cellular senescence are currently lacking. Consequently, a critical question remains: does genetic predisposition influence the intensity of the senescent response to infection and what are the underlying molecular mechanisms? Addressing this could open promising avenues for targeted control and therapeutic regulation of virus-induced cellular senescence.

Conclusion

This review summarizes data on various pathogenic modulators that can both positively and negatively influence aging processes. It provides a comprehensive analysis of how pathogenic agents affect classical molecular senescence pathways and trigger novel cascades leading to age-related changes. The review also highlights candidate genes that contribute to genetic predisposition to infections, which in turn can accelerate or delay aging processes (Table S1). Studies conducted on model organisms and cell cultures demonstrate that exposure to even a single such factor can induce irreversible structural and functional changes. Investigating the molecular mechanisms of pathogen-induced senescence is a vital interdisciplinary field that integrates infectious disease research, gerontology, genetics, and the study of age-related diseases. Understanding these mechanisms is essential not only for deepening our knowledge of aging but also for opening new avenues in the prevention and treatment of age-related conditions, as well as for promoting healthy lifespan extension. This approach frames aging as the combined outcome of external factors, such as infections, and internal biological processes. It lays the groundwork for a new generation of preventive medicine focused not on treating individual diseases but on systemically slowing aging by addressing its root causes, including pathogenic influences. Such a strategy has the potential to prolong the active, healthy years of life, lessen the impact of age-related diseases, and enhance quality of life for older adults.

The epidemiological significance of the bacteria discussed is extremely high: for example, *H. pylori* infects approximately half of the world's population, with the highest prevalence in low-income countries (Li et al., 2023). Representatives of periodontopathogens, such as *P. gingivalis*, are widespread among the adult population, thus linking chronic infections with an increased risk of age-associated diseases (Mei et al., 2020). The key conclusion concerns the dualistic role of bacteria. On the one hand, certain pathogens can act as inducers of aging through the activation of SASP in cells, cell cycle arrest, and immunosenescence. On the other hand, commensal bacteria, such as *P. merdae*, exhibit geroprotective potential, counteracting oxidative stress and inflammation through the production of specific metabolites. So, aging can be viewed not only as an internal biological process but also as a phenomenon influenced by the human microbiota. Aging processes are driven by the interplay of cell-autonomous and non-cell-autonomous mechanisms. In particular, senescence induction caused by bacterial infection is primarily regulated by classical molecular pathways such as Erk, p53-p21, and ATM-p38. Furthermore, additional signaling cascades have been identified, mediated through the CXCR2, IL-1R1, CaMKII-HDAC4, and TLR-MyD88 mechanisms. It should be noted that activation of the CXCR2 signaling pathway has been demonstrated in the context of oncological aging, whereas the TLR-MyD88 pathway is associated with immune aging. At the same time, the mechanisms of so-called anti-aging induced by probiotics remain, in most cases, insufficiently understood and require further investigation.

Unlike bacteria, viral infections have an exclusively negative impact on cellular senescence, with no evidence to date of any rejuvenating effects. Despite this overall detrimental influence, viruses regulate aging processes through diverse molecular pathways, including PI3K–AKT–p70S6K, p38MAPK, SRC, Wnt/β-catenin, p53, IL-6-JAK-STAT3, and NF-κB. Retroviruses specifically engage signaling cascades such as IRF3, cGAS–STING, RAS, IFIH1-MAVS, and ATF3–RIG-I/MDA5–MAVS. Understanding these virus-induced aging mechanisms is

[1] Table S1 is available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Simoroz_Engl_30_5.pdf

critically important for developing therapeutic strategies to treat viral infections and promote healthy aging. However, research in this area remains limited, and there is a pressing need to identify therapeutic targets and develop effective pharmaceutical interventions.

Two main types of cellular senescence are recognized: senescence of immune cells (immunosenescence) and senescence of non-immune cells. Growing evidence shows that senescent non-immune cells, especially tumor cells, can induce senescence in immune cells. For example, senescent cancer cells increase expression of programmed death-ligand 1 (PD-L1) on their surface or release extracellular vesicles containing PD-L1, which trigger T-lymphocyte senescence (Majewska et al., 2024; Hwang et al., 2025; Ma et al., 2025). Conversely, immune cell senescence exerts systemic effects on the organism (Liu Z. et al., 2023). Notably, senescent cells within a single population exhibit significant heterogeneity, complicating the development of effective senolytic agents targeting these cells. Current approaches leverage machine learning for subclassification based on morphological features alongside omics technologies for molecular profiling (Kamat et al., 2025; Wang Z. et al., 2025). However, existing data remain insufficient to define a unique senescence phenotype. It remains unresolved whether the observed heterogeneity reflects distinct subpopulations or temporal phenotypic evolution within the same cell population (Bitencourt et al., 2024).

It is important to recognize that research on senescence mechanisms often faces certain biases, as it tends to focus primarily on well-established signaling pathways and markers, potentially overlooking many significant but less studied mechanisms. Additionally, there is a methodological challenge stemming from discrepancies between *in vitro* and *in vivo* study results. Model systems frequently fail to fully replicate the complexity of human tissues and whole organisms, limiting the applicability of findings. The heterogeneity of senescent cells, in morphology and function, also complicates their accurate identification and evaluation within tissues. These limitations must be carefully considered when interpreting results and it is important to highlight the need for further research to improve the translational relevance of findings and to develop effective preventive strategies.

Understanding the molecular mechanisms of pathogen-induced senescence holds great potential for developing novel diagnostic tools that enable early detection of infection-related pathological changes and timely preventive interventions. Moreover, elucidating the role of pathogens in driving aging paves the way for therapeutic strategies aimed at mitigating inflammatory responses and delaying aging processes in patients with chronic or prolonged infections. Incorporating these approaches into clinical practice could significantly enhance the quality of life for elderly individuals, lower the risk of age-related diseases, and broaden the scope of comprehensive management for age-associated conditions. Therefore, this research area not only advances academic knowledge but also lays the groundwork for future medical breakthroughs and the extension of a healthy human lifespan.

References

Aguado J., Amarilla A.A., Taherian Fard A., Albornoz E.A., Tyshkovskiy A., Schwabenland M., Chaggar H.K., … Gladyshev V.N., Woodruff T.M., Mar J.C., Watterson D., Wolvetang E.J. Senolytic therapy alleviates physiological human brain aging and COVID-19 neuropathology. *Nat Aging*. 2023;3(12):1561-1575. doi 10.1038/s43587-023-00519-6

Aguirre M., Escobar M., Forero Amézquita S., Cubillos D., Rincón C., Vanegas P., Tarazona M.P., Atuesta Escobar S., Blanco J.C., Celis L.G. Application of the Yamanaka transcription factors *Oct4*, *Sox2*, *Klf4*, and *c-Myc* from the laboratory to the clinic. *Genes*. 2023;14(9):1697. doi 10.3390/genes14091697

Alam M.S., Gangiredla J., Hasan N.A., Barnaba T., Tartera C. Aging-induced dysbiosis of gut microbiota as a risk factor for increased *Listeria monocytogenes* infection. *Front Immunol.* 2021;12:672353. doi 10.3389/fimmu.2021.672353

Aman Y., Schmauck-Medina T., Hansen M., Morimoto R.I., Simon A.K., Bjedov I., Palikaras K., … Rubinsztein D.C., Partridge L., Kroemer G., Labbadia J., Fang E.F. Autophagy in healthy aging and disease. *Nat Aging*. 2021;1(8):634-650. doi 10.1038/s43587-021-00098-4

Arakelyan N.A., Kupriyanova D.A., Vasilevska J., Rogaev E.I. Sexual dimorphism in immunity and longevity among the oldest old. *Front Immunol.* 2025;16:1525948. doi 10.3389/fimmu.2025.1525948

Augusto D.G., Murdolo L.D., Chatzileontiadou D.S.M., Sabatino J.J., Yusufali T., Peyser N.D., Butcher X., … Olgin J.E., Pletcher M.J., Maiers M., Gras S., Hollenbach J.A. A common allele of *HLA* is associated with asymptomatic SARS-CoV-2 infection. *Nature*. 2023;620(7972):128-136. doi 10.1038/s41586-023-06331-x

Azcarate D., Olasagasti Arsuaga F., Granizo Rodriguez E., Arana-Arri E., España P.P., Intxausti M., Sancho C., García De Vicuña Meléndez A., Ibarrondo O., De Pancorbo M.M. Human-genetic variants associated with susceptibility to SARS-CoV-2 infection. *Gene*. 2025;953:149423. doi 10.1016/j.gene.2025.149423

Azman K.F., Zakaria R. d-Galactose-induced accelerated aging model: an overview. *Biogerontology*. 2019;20(6):763-782. doi 10.1007/s10522-019-09837-y

Beaupere C., Garcia M., Larghero J., Fève B., Capeau J., Lagathu C. The HIV proteins Tat and Nef promote human bone marrow mesenchymal stem cell senescence and alter osteoblastic differentiation. *Aging Cell*. 2015;14(4):534-546. doi 10.1111/acel.12308

Beckmann N., Becker K.A. Ceramide and related molecules in viral infections. *Int J Mol Sci.* 2021;22(11):5676. doi 10.3390/ijms22115676

Bian T., Gibbs J.D., Örvell C., Imani F. Respiratory syncytial virus matrix protein induces lung epithelial cell cycle arrest through a p53 dependent pathway. *PLoS One*. 2012;7(5):e38052. doi 10.1371/journal.pone.0038052

Bitencourt T.C., Vargas J.E., Silva A.O., Fraga L.R., Filippi-Chiela E. Subcellular structure, heterogeneity, and plasticity of senescent cells. *Aging Cell*. 2024;23(4):e14154. doi 10.1111/acel.14154

Blackhurst B.M., Funk K.E. Viral pathogens increase risk of neurodegenerative disease. *Nat Rev Neurol*. 2023;19(5):259-260. doi 10.1038/s41582-023-00790-6

Bobak C.A., Abhimanyu, Natarajan H., Gandhi T., Grimm S.L., Nishiguchi T., Koster K., … Cirillo J.D., O'Malley J., Hill J.E., Coarfa C., DiNardo A.R. Increased DNA methylation, cellular senescence and premature epigenetic aging in guinea pigs and humans with tuberculosis. *Aging*. 2022;14(5):2174-2193. doi 10.18632/aging.203936

Cai Q., Shi P., Yuan Y., Peng J., Ou X., Zhou W., Li J., Su T., Lin L., Cai S., He Y., Xu J. Inflammation-associated senescence promotes *Helicobacter pylori* – induced atrophic gastritis. *Cell Mol Gastroenterol Hepatol*. 2021;11(3):857-880. doi 10.1016/j.jcmgh.2020.10.015

Casto A.M., Seo S., Levine D.M., Storer B.E., Dong X., Hansen J.A., Boeckh M., Martin P.J. Genetic variants associated with cytomegalovirus infection after allogeneic hematopoietic cell transplantation. *Blood*. 2021;138(17):1628-1636. doi 10.1182/blood.2021012153

Chatzopoulou F., Gioula G., Kioumis I., Chatzidimitriou D., Exindari M. Identification of complement-related host genetic risk factors associated with influenza A(H1N1)pdm09 outcome: challenges ahead. *Med Microbiol Immunol*. 2019;208(5):631-640. doi 10.1007/s00430-018-0567-9

Chen C., Wang M., Zhu Z., Qu J., Xi X., Tang X., Lao X., … Bai C., Zhang Z., Lu S., Song Y., Zhou W. Multiple gene mutations identified in patients infected with influenza A (H7N9) virus. *Sci Rep*. 2016;6(1):25614. doi 10.1038/srep25614

Chen N.C., Partridge A.T., Tuzer F., Cohen J., Nacarelli T., Navas-Martín S., Sell C., Torres C., Martín-García J. Induction of a senescence-like phenotype in cultured human fetal microglia during HIV-1 infection. *J Gerontol A Biol Sci Med Sci*. 2018;73(9):1187-1196. doi 10.1093/gerona/gly022

Chumduri C., Gurumurthy R.K., Zietlow R., Meyer T.F. Subversion of host genome integrity by bacterial pathogens. *Nat Rev Mol Cell Biol*. 2016;17(10):659-673. doi 10.1038/nrm.2016.100

De Jong S.J., Créquer A., Matos I., Hum D., Gunasekharan V., Lorenzo L., Jabot-Hanin F., … Franco J.L., Burger B., Orth G., Jouanguy E., Casanova J.-L. The human CIB1–EVER1–EVER2 complex governs keratinocyte-intrinsic immunity to β-papillomaviruses. *J Exp Med*. 2018;215(9):2289-2310. doi 10.1084/jem.20170308

Dean M., Carrington M., Winkler C., Huttley G.A., Smith M.W., Allikmets R., Goedert J.J., … Kaslow R., Saah A., Rinaldo C., Detels R., O'Brien S.J. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the *CKR5* structural gene. *Science*. 1996;273(5283):1856-1862. doi 10.1126/science.273.5283.1856

DeSantis C.E., Miller K.D., Dale W., Mohile S.G., Cohen H.J., Leach C.R., Goding Sauer A., Jemal A., Siegel R.L. Cancer statistics for adults aged 85 years and older, 2019. *CA Cancer J Clin*. 2019;69(6):452-467. doi 10.3322/caac.21577

Di Giorgio E., Ranzino L., Tolotto V., Dalla E., Burelli M., Gualandi N., Brancolini C. Transcription of endogenous retroviruses in senescent cells contributes to the accumulation of double-stranded RNAs that trigger an anti-viral response that reinforces senescence. *Cell Death Dis*. 2024;15(2):157. doi 10.1038/s41419-024-06548-2

Donato A.J., Machin D.R., Lesniewski L.A. Mechanisms of dysfunction in the aging vasculature and role in age-related disease. *Circ Res*. 2018;123(7):825-848. doi 10.1161/CIRCRESAHA.118.312563

Esposito S., Molteni C.G., Giliani S., Mazza C., Scala A., Tagliaferri L., Pelucchi C., Fossali E., Plebani A., Principi N. Toll-like receptor 3 gene polymorphisms and severity of pandemic A/H1N1/2009 influenza in otherwise healthy children. *Virol J*. 2012;9(1):270. doi 10.1186/1743-422X-9-270

Fafián-Labora J.A., O'Loghlen A. Classical and nonclassical intercellular communication in senescence and ageing. *Trends Cell Biol*. 2020;30(8):628-639. doi 10.1016/j.tcb.2020.05.003

Grompone G., Martorell P., Llopis S., González N., Genovés S., Mulet A.P., Fernández-Calero T., Tiscornia I., Bollati-Fogolín M., Chambaud I., Foligné B., Montserrat A., Ramón D. Anti-inflammatory *Lactobacillus rhamnosus* CNCM I-3690 strain protects against oxidative stress and increases lifespan in *Caenorhabditis elegans*. *PLoS One*. 2012;7(12):e52493. doi 10.1371/journal.pone.0052493

Gu P., Wei R., Liu R., Yang Q., He Y., Guan J., He W., … Zeng Z., Chen Z., Jiang Y., Liu Z., Chen P. Aging-induced alternation in the gut microbiota impairs host antibacterial defense. *Adv Sci*. 2025;12(12):2411008. doi 10.1002/advs.202411008

Hale A.E., Collins-McMillen D., Lenarcic E.M., Igarashi S., Kamil J.P., Goodrum F., Moorman N.J. FOXO transcription factors activate alternative major immediate early promoters to induce human cytomegalovirus reactivation. *Proc Natl Acad Sci USA*. 2020;117(31):18764-18770. doi 10.1073/pnas.2002651117

Han S.-J., Stacy A., Corral D., Link V.M., De Siqueira M.K., Chi L., Teijeiro A., Yong D.S., Perez-Chaparro P.J., Bouladoux N., Lim A.I., Enamorado M., Belkaid Y., Collins N. Microbiota configuration determines nutritional immune optimization. *Proc Natl Acad Sci USA*. 2023;120(49):e2304905120. doi 10.1073/pnas.2304905120

Hasegawa T., Oka T., Son H.G., Oliver-García V.S., Azin M., Eisenhaure T.M., Lieb D.J., Hacohen N., Demehri S. Cytotoxic CD4+ T cells eliminate senescent cells by targeting cytomegalovirus antigen. *Cell*. 2023;186(7):1417-1431.e20. doi 10.1016/j.cell.2023.02.033

Hellani F., Leleu I., Saidi N., Martin N., Lecoeur C., Werkmeister E., Koffi D., Trottein F., Yapo-Etté H., Das B., Abbadie C., Pied S. Role of astrocyte senescence regulated by the non-canonical autophagy in the neuroinflammation associated to cerebral malaria. *Brain Behav Immun*. 2024;117:20-35. doi 10.1016/j.bbi.2023.12.030

Hidaka F., Matsuo S., Muta T., Takeshige K., Mizukami T., Nunoi H. A missense mutation of the Toll-like receptor 3 gene in a patient with influenza-associated encephalopathy. *Clin Immunol*. 2006;119(2):188-194. doi 10.1016/j.clim.2006.01.005

Hwang H.J., Kang D., Shin J., Jung J., Ko S., Jung K.H., Hong S.-S., Park J.E., Oh M.J., An H.J., Yang W.-H., Ko Y.-G., Cha J.-H., Lee J.-S. Therapy-induced senescent cancer cells contribute to cancer progression by promoting ribophorin 1-dependent PD-L1 upregulation. *Nat Commun*. 2025;16(1):353. doi 10.1038/s41467-024-54132-1

Jallow M., Teo Y.Y., Small K.S., Rockett K.A., Deloukas P., Clark T.G., Kivinen K., … Williams T.N., Wilson M., Kwiatkowski D.P.; Wellcome Trust Case Control Consortium; Malaria Genomic Epidemiology Network. Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat Genet*. 2009;41(6):657-665. doi 10.1038/ng.388

Jeanpierre M., Cognard J., Tusseau M., Riller Q., Bui L.-C., Berthelet J., Laurent A., … Rodrigues Lima F., Walzer T., Rieux-Laucat F., Belot A., Mathieu A.-L. Haploinsufficiency in PTPN2 leads to early-onset systemic autoimmunity from Evans syndrome to lupus. *J Exp Med*. 2024;221(9):e20232337. doi 10.1084/jem.20232337

Jiang M., Xu Q., Wu Z. Optimized Yamanaka factors combined with TERT gene therapy for enhanced anti-aging effects. *Genes Dis*. 2026;13(2):101669. doi 10.1016/j.gendis.2025.101669

Kamat P., Macaluso N., Li Y., Agrawal A., Winston A., Pan L., Stewart T., Starich B., Milcik N., Min C., Wu P.-H., Walston J., Fan J., Phillip J.M. Single-cell morphology encodes functional subtypes of senescence in aging human dermal fibroblasts. *Sci Adv*. 2025;11(17):eads1875. doi 10.1126/sciadv.ads1875

Kannampuzha S., Gopalakrishnan A.V., Padinharayil H., Alappat R.R., Anilkumar K.V., George A., Dey A., Vellingiri B., Madhyastha H., Ganesan R., Ramesh T., Jayaraj R., Prabakaran D.S. Onco-pathogen mediated cancer progression and associated signaling pathways in cancer development. *Pathogens*. 2023;12(6):770. doi 10.3390/pathogens12060770

Kawamoto S., Uemura K., Hori N., Takayasu L., Konishi Y., Katoh K., Matsumoto T., … Ohtani N., Standley D.M., Suda W., Fukuda S., Hara E. Bacterial induction of B cell senescence promotes age-related changes in the gut microbiota. *Nat Cell Biol*. 2023;25(6):865-876. doi 10.1038/s41556-023-01145-5

Kovtonyuk L.V., Caiado F., Garcia-Martin S., Manz E.-M., Helbling P., Takizawa H., Boettcher S., Al-Shahrour F., Nombela-Arrieta C., Slack E., Manz M.G. IL-1 mediates microbiome-induced inflammaging of hematopoietic stem cells in mice. *Blood*. 2022;139(1):44-58. doi 10.1182/blood.2021011570

Lee S., Yu Y., Trimpert J., Benthani F., Mairhofer M., Richter-Pechanska P., Wyler E., … Ralser M., Eils R., Reimann M., Fan D.N.Y., Schmitt C.A. Virus-induced senescence is a driver and therapeutic

target in COVID-19. *Nature*. 2021;599(7884):283-289. doi 10.1038/s41586-021-03995-1

Li Y., Choi H., Leung K., Jiang F., Graham D.Y., Leung W.K. Global prevalence of *Helicobacter pylori* infection between 1980 and 2022: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol*. 2023;8(6):553-564. doi 10.1016/S2468-1253(23)00070-5

Lin P.L., Flynn J.L. Understanding latent tuberculosis: a moving target. *J Immunol*. 2010;185(1):15-22. doi 10.4049/jimmunol.0903856

Lipskaia L., Delval L., Sencio V., Born E., Heumel S., Houssaini A., Valentin C., … Le Grand R., Le Goffic R., Bernard D., Adnot S., Trottein F. Virus-induced cellular senescence causes pulmonary sequelae post-influenza infection. *Aging Cell*. 2025;24(9):e70140. doi 10.1111/acel.70140

Liu L., Hao X., Bai Y., Tian Y. The soil *Mycobacterium* sp. promotes health and longevity through different bacteria-derived molecules in *Caenorhabditis elegans*. *Aging Cell*. 2025;24(3):e14416. doi 10.1111/acel.14416

Liu Yu., Gu S., Su Y., Wang S., Cheng Y., Sang X., Jin L., Liu Yi., Li C., Liu W., Chen M., Wang X., Wang Z. Embryonic stem cell extracellular vesicles reverse the senescence of retinal pigment epithelial cells by the p38MAPK pathway. *Exp Eye Res*. 2023;227:109365. doi 10.1016/j.exer.2022.109365

Liu Z., Liang Q., Ren Y., Guo C., Ge X., Wang L., Cheng Q., Luo P., Zhang Y., Han X. Immunosenescence: molecular mechanisms and diseases. *Sig Transduct Target Ther*. 2023;8(1):200. doi 10.1038/s41392-023-01451-2

Ma F., Liu X., Zhang Y., Tao Y., Zhao L., Abusalamah H., Huffman C., Harbison R.A., Puram S.V., Wang Y., Peng G. Tumor extracellular vesicle-derived PD-L1 promotes T cell senescence through lipid metabolism reprogramming. *Sci Transl Med.* 2025;17(785):eadm7269. doi 10.1126/scitranslmed.adm7269

Majeed A.Y., Zulkafli N.E.S., Ad'hiah A.H. The *IFIH1* gene variants rs1990760 and rs2111485 are associated with COVID-19 susceptibility and affect IFIH1 protein levels in Iraqis. *Egypt J Med Hum Genet*. 2024;25(1):132. doi 10.1186/s43042-024-00601-7

Majewska J., Agrawal A., Mayo A., Roitman L., Chatterjee R., Sekeresova Kralova J., Landsberger T., … Wagner W., Maimon A., Amit I., Alon U., Krizhanovsky V. p16-dependent increase of PD-L1 stability regulates immunosurveillance of senescent cells. *Nat Cell Biol*. 2024;26(8):1336-1345. doi 10.1038/s41556-024-01465-0

Mao J., Zhang Q., Zhuang Y., Zhang Y., Li L., Pan J., Xu L., Ding Y., Wang M., Cong Y.-S. Reactivation of senescence-associated endogenous retroviruses by ATF3 drives interferon signaling in aging. *Nat Aging*. 2024;4(12):1794-1812. doi 10.1038/s43587-024-00745-6

Martin M.P. Genetic acceleration of AIDS progression by a promoter variant of CCR5. *Science*. 1998;282(5395):1907-1911. doi 10.1126/science.282.5395.1907

Martínez I., García-Carpizo V., Guijarro T., García-Gomez A., Navarro D., Aranda A., Zambrano A. Induction of DNA double-strand breaks and cellular senescence by human respiratory syncytial virus. *Virulence*. 2016;7(4):427-442. doi 10.1080/21505594.2016. 1144001

Masters M.C., Landay A.L., Robbins P.D., Tchkonia T., Kirkland J.L., Kuchel G.A., Niedernhofer L.J., Palella F.J. Chronic HIV infection and aging: application of a geroscience-guided approach. *J Acquir Immune Defic Syndr*. 2022;89(Suppl.1):S34-S46. doi 10.1097/QAI.0000000000002858

Mathiasen S.L., Gall-Mas L., Pateras I.S., Theodorou S.D.P., Namini M.R.J., Hansen M.B., Martin O.C.B., … Akbar A.N., Gorgoulis V.G., Frisan T., Ødum N., Krejsgaard T. Bacterial genotoxins induce T cell senescence. *Cell Rep*. 2021;35(10):109220. doi 10.1016/j.celrep.2021.109220

McGovern K.E., Sonar S.A., Watanabe M., Coplen C.P., Bradshaw C.M., Nikolich J.Ž. The aging of the immune system and its implications for transplantation. *GeroScience*. 2023;45(3):1383-1400. doi 10.1007/s11357-022-00720-2

Mei F., Xie M., Huang X., Long Y., Lu X., Wang X., Chen L. *Porphyromonas gingivalis* and its systemic impact: current status. *Pathogens*. 2020;9(11):944. doi 10.3390/pathogens9110944

Nong C., Wu Z., Yang C., Xu W., Luo L., Zhou J., Shen L., Chen Y., Yuan Y., Hu G. Cdc42 improve SARS-CoV-2 spike protein-induced cellular senescence through activating of Wnt/β-Catenin signaling pathway. *Front Cell Infect Microbiol.* 2024;14:1449423. doi 10.3389/fcimb.2024.1449423

Ogrodnik M., Carlos Acosta J., Adams P.D., d'Adda Di Fagagna F., Baker D.J., Bishop C.L., Chandra T., … Witwer K., Von Zglinicki T., Yun M.H., Grillari J., Demaria M. Guidelines for minimal information on cellular senescence experimentation *in vivo*. *Cell*. 2024;187(16):4150-4175. doi 10.1016/j.cell.2024.05.059

Okumura S., Konishi Y., Narukawa M., Sugiura Y., Yoshimoto S., Arai Y., Sato S., … Fukunaga Y., Ueno M., Sakai Y., Nagayama S., Hara E. Gut bacteria identified in colorectal cancer patients promote tumourigenesis via butyrate secretion. *Nat Commun*. 2021;12(1):5674. doi 10.1038/s41467-021-25965-x

Omodeo-Salè F., Motti A., Basilico N., Parapini S., Olliaro P., Taramelli D. Accelerated senescence of human erythrocytes cultured with *Plasmodium falciparum*. *Blood*. 2003;102(2):705-711. doi 10.1182/blood-2002-08-2437

Powell T.R., Duarte R.R.R., Hotopf M., Hatch S.L., De Mulder Rougvie M., Breen G.D., Lewis C.M., Nixon D.F. The behavioral, cellular and immune mediators of HIV-1 acquisition: new insights from population genetics. *Sci Rep*. 2020;10(1):3304. doi 10.1038/s41598-020-59256-0

Pringsheim T., Jette N., Frolkis A., Steeves T.D.L. The prevalence of Parkinson's disease: a systematic review and meta-analysis. *Mov Disord*. 2014;29(13):1583-1590. doi 10.1002/mds.25945

Raviola S., Griffante G., Iannucci A., Chandel S., Lo Cigno I., Lacarbonara D., Caneparo V., … Boldorini R., Cantaluppi V., Landolfo S., Gariglio M., De Andrea M. Human cytomegalovirus infection triggers a paracrine senescence loop in renal epithelial cells. *Commun Biol*. 2024;7(1):292. doi 10.1038/s42003-024-05957-5

Reimann M., Lee S., Schmitt C.A. Cellular senescence: neither irreversible nor reversible. *J Exp Med*. 2024;221(4):e20232136. doi 10.1084/jem.20232136

Ryu S., Shchukina I., Youm Y.-H., Qing H., Hilliard B., Dlugos T., Zhang X., … Horvath T.L., Dietrich M.O., Artyomov M., Wang A., Dixit V.D. Ketogenic diet restrains aging-induced exacerbation of coronavirus infection in mice. *eLife*. 2021;10:e66522. doi 10.7554/eLife.66522

Sadighi Akha A.A. Aging and the immune system: an overview. *J Immunol Methods*. 2018;463:21-26. doi 10.1016/j.jim.2018.08.005

Saeedi Saravi S.S., Pugin B., Constancias F., Shabanian K., Spalinger M., Thomas A., Le Gludic S., … Falchi M., Alimonti A., Ruschitzka F., Paneni F., Beer J.H. Gut microbiota-dependent increase in phenylacetic acid induces endothelial cell senescence during aging. *Nat Aging*. 2025;5(6):1025-1045. doi 10.1038/s43587-025-00864-8

Saito Y., Murata-Kamiya N., Hirayama T., Ohba Y., Hatakeyama M. Conversion of *Helicobacter pylori* CagA from senescence inducer to oncogenic driver through polarity-dependent regulation of p21. *J Exp Med*. 2010;207(10):2157-2174. doi 10.1084/jem.20100602

Salas A., Pardo-Seco J., Cebey-López M., Gómez-Carballa A., Obando-Pacheco P., Rivero-Calle I., Currás-Tuala M.-J., … Martínez-Padilla M.D.C., Baca-Cots M., Moreno-Pérez D., Beatriz-Reyes S., León-León M.C. Whole exome sequencing reveals new candidate genes in host genomic susceptibility to respiratory syncytial virus disease. *Sci Rep*. 2017;7(1):15888. doi 10.1038/s41598-017-15752-4

Scarpa E., Oliveri D., Moschetti G., Griego A., Vianello A., Rondelli D., Antinori B., … Franzè S., Cidonio G., Manganaro L., Lorè N.I., Rizzello L. Chronic mycobacteria infection triggers macrophage senescence. *bioRxiv*. 2024. doi 10.1101/2024.11.05.622137

Schmitt C.A., Tchkonia T., Niedernhofer L.J., Robbins P.D., Kirkland J.L., Lee S. COVID-19 and cellular senescence. *Nat Rev Immunol*. 2023;23(4):251-263. doi 10.1038/s41577-022-00785-2

Seoane R., Vidal S., Bouzaher Y.H., El Motiam A., Rivas C. The interaction of viruses with the cellular senescence response. *Biology*. 2020;9(12):455. doi 10.3390/biology9120455

Shi S., Zhang Q., Sang Y., Ge S., Wang Q., Wang R., He J. Probiotic *Bifidobacterium* longum BB68S improves cognitive functions in healthy older adults: a randomized, double-blind, placebo-controlled trial. *Nutrients*. 2022;15(1):51. doi 10.3390/nu15010051

Smith M.W., Dean M., Carrington M., Winkler C., Huttley G.A., Lomb D.A., Goedert J.J., … Multicenter AIDS Cohort Study (MACS); Multicenter Hemophilia Cohort Study (MHCS); San Francisco City Cohort (SFCC); ALIVE Study; O'Brien S.J. Contrasting genetic influence of *CCR2* and *CCR5* variants on HIV-1 infection and disease progression. *Science*. 1997;277(5328):959-965. doi 10.1126/science.277.5328.959

Smolgovsky S., Ibeh U., Tamayo T.P., Alcaide P. Adding insult to injury – inflammation at the heart of cardiac fibrosis. *Cell Signal*. 2021; 77:109828. doi 10.1016/j.cellsig.2020.109828

Thangaraj A., Chivero E.T., Tripathi A., Singh S., Niu F., Guo M.-L., Pillai P., Periyasamy P., Buch S. HIV TAT-mediated microglial senescence: role of SIRT3-dependent mitochondrial oxidative stress. *Redox Biol*. 2021;40:101843. doi 10.1016/j.redox.2020.101843

Thye T., Vannberg F.O., Wong S.H., Owusu-Dabo E., Osei I., Gyapong J., Sirugo G., … Wellcome Trust Case Control Consortium; Morris A.P., Meyer C.G., Horstmann R.D., Hill A.V.S. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat Genet*. 2010;42(9):739-741. doi 10.1038/ng.639

Trevisan K., Cristina-Pereira R., Silva-Amaral D., Aversi-Ferreira T.A. Theories of aging and the prevalence of Alzheimer's disease. *BioMed Res Int*. 2019;2019:9171424. doi 10.1155/2019/9171424

Tulic M.K., Hurrelbrink R.J., Prêle C.M., Laing I.A., Upham J.W., Le Souef P., Sly P.D., Holt P.G. TLR4 polymorphisms mediate impaired responses to respiratory syncytial virus and lipopolysaccharide. *J Immunol*. 2007;179(1):132-140. doi 10.4049/jimmunol.179.1.132

Wang H., Feng L., Pei Z., Zhao J., Lu S., Lu W. Gut microbiota metabolism of branched-chain amino acids and their metabolites can improve the physiological function of aging mice. *Aging Cell*. 2025; 24(4):e14434. doi 10.1111/acel.14434

Wang Z., Ge S., Liao T., Yuan M., Qian W., Chen Q., Liang W., Cheng X., Zhou Q., Ju Z., Zhu H., Xiong W. Integrative single-cell metabolomics and phenotypic profiling reveals metabolic heterogeneity of cellular oxidation and senescence. *Nat Commun*. 2025;16(1):2740. doi 10.1038/s41467-025-57992-3

Wyles S.P., Yu G.T., Gold M., Behfar A. Topical platelet exosomes reduce senescence signaling in human skin: an exploratory prospective trial. *Dermatol Surg*. 2024;50(11S):S160-S165. doi 10.1097/DSS.0000000000004426

Xu F., Wang G., Zhao F., Huang Y., Fan Z., Mei S., Xie Y., … Cen S., Liang C., Ren L., Guo F., Wang J. IFITM3 inhibits SARS-CoV-2 infection and is associated with COVID-19 susceptibility. *Viruses*. 2022;14(11):2553. doi 10.3390/v14112553

Yan J., Chen S., Yi Z., Zhao R., Zhu J., Ding S., Wu J. The role of p21 in cellular senescence and aging-related diseases. *Mol Cells*. 2024; 47(11):100113. doi 10.1016/j.mocell.2024.100113

Zhang D., Chen G., Manwani D., Mortha A., Xu C., Faith J.J., Burk R.D., Kunisaki Y., Jang J.-E., Scheiermann C., Merad M., Frenette P.S. Neutrophil ageing is regulated by the microbiome. *Nature*. 2015;525(7570):528-532. doi 10.1038/nature15367

Zúñiga J., Buendía-Roldán I., Zhao Y., Jiménez L., Torres D., Romo J., Ramírez G., … Su L., Tager A.M., Pardo A., Selman M., Christiani D.C. Genetic variants associated with severe pneumonia in A/H1N1 influenza infection. *Eur Respir J*. 2012;39(3):604-610. doi 10.1183/09031936.00020611

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

Received December 23, 2025. Revised April 28, 2026. Accepted May 4, 2026.

PLANT GENETICS
Original article

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):725-734

doi 10.18699/vjgb-26-73

Changes in telomere length in *Arabidopsis thaliana* plants infected with the phytopathogenic bacteria *Pseudomonas syringae* DC3000

A.V. Sannikova, L.R. Abdulkina, M.R. Sharipova, L.R. Valeeva

Institute of Fundamental Medicine and Biology, Kazan (Volga Region) Federal University, Kazan, Russia
liarvaleeva@gmail.com

Abstract. Telomeres are specialized nucleoprotein complexes at the physical ends of linear chromosomes of most eukaryotes that represent a key structural element in genome stability regulation. One of the most important characteristics determining the protective capacity of telomeres is telomere length (TL). In *Arabidopsis thaliana*, telomere length varies among ecotypes, which may be considered as an adaptive mechanism to various environmental and stress conditions, including phytopathogenic infections. However, the mechanisms of TL regulation under biotic stress remain unclear. Here, we show the results of our study of the TL changes in *A. thaliana* plants under the *Pseudomonas syringae* pv. *tomato* DC3000 infection. In the study, we used natural ecotypes Col-0 (medium telomere length) and Sf-2 (long telomeres), as well as mutant plant lines with disordered and wild-type telomere length phenotypes, *oli5-2*$^{-/-}$ and *nop2c-2*$^{-/-}$, respectively. It was found that susceptibility to the phytopathogen varies in plants: the line with short telomeres, *oli5-2*$^{-/-}$, showed the highest sensitivity compared to the WT and *nop2c-2*$^{-/-}$. Sf-2 plants possessed the highest resistance to the infection. Significant shortening of TL on the third day after infection was found on the short arm of chromosome 4 of infected *nop2c-2*$^{-/-}$ plants, whereas TL of infected *oli5-2*$^{-/-}$ plants slightly decreased on the third day and increased on the seventh day after infection. The most resistant line Sf-2 retained TL on the third and seventh days after infection but significantly lost it on the 18th day. So, it was shown for the first time that phytopathogenic infection affects telomere length in plants, which opens a new direction in the study of plant telomere regulation under biotic stress.

Key words: plant telomeres; DNA stability; biotic stress; bacterial infection; ribosomal proteins; rRNA methyltransferase; phytopathogens

For citation: Sannikova A.V., Abdulkina L.R., Sharipova M.R., Valeeva L.R. Changes in telomere length in *Arabidopsis thaliana* plants infected with the phytopathogenic bacteria *Pseudomonas syringae* DC3000. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(5):725-734. doi 10.18699/vjgb-26-73

Funding. The study was supported by the Strategic Academic Leadership Program (PRIORITY-2030).

Изменение длины теломер растений *Arabidopsis thaliana* в результате инфицирования фитопатогенными бактериями *Pseudomonas syringae* DC3000

А.В. Санникова, Л.Р. Абдулкина, М.Р. Шарипова, Л.Р. Валеева

Институт фундаментальной медицины и биологии, Казанский (Приволжский) федеральный университет, Казань, Россия
liarvaleeva@gmail.com

Аннотация. Ключевым структурно-функциональным элементом, обеспечивающим стабильность генома большинства эукариот, являются теломеры – специализированные нуклеопротеиновые комплексы на физических концах линейных хромосом. Одна из важнейших характеристик теломер, определяющих их защитный потенциал, – длина теломерной ДНК. У растений *Arabidopsis thaliana* она варьирует среди экотипов, что может указывать на роль теломер в адаптации к различным условиям окружающей среды и стрессу, включая воздействие фитопатогенов. Однако механизмы регуляции длины теломер при биотическом стрессе остаются неизученными. В статье представлены результаты исследования зависимости длины теломер *A. thaliana* от инфицирования фитопатогенным штаммом *Pseudomonas syringae* pv. *tomato* DC3000. В работе использовались природные экотипы растений Col-0 (средняя длина теломер) и Sf-2 (длинные теломеры), а также мутантные линии с нарушением регуляции длины теломер и с теломерным фенотипом дикого типа, *oli5-2*$^{-/-}$ и *nop2c-2*$^{-/-}$ соответственно. Установлено, что восприимчивость к фитопатогену различается у растений разных генотипов: линия *oli5-2*$^{-/-}$ с короткими теломерами проявляла наибольшую чувствительность, в том числе по сравнению

с диким типом и линией *nop2c-2*$^{-/-}$, в то время как растения экотипа с длинными теломерами Sf-2 отличались наибольшей устойчивостью. Значимое укорочение длины теломер на третьи сутки после инфицирования было обнаружено на коротком плече хромосомы 4 у инфицированных растений линии *nop2c-2*$^{-/-}$, тогда как длина теломер инфицированных растений *oli5-2*$^{-/-}$ незначительно сокращалась на третьи сутки и возрастала на седьмые сутки после инфицирования, что указывает на динамичность регуляции длины теломер при стрессе. Длина теломер растений дикого типа значительно не изменялась при инфицировании. Наиболее устойчивая линия Sf-2 сохраняла длину теломер на третьи и седьмые сутки после инфицирования, но на 18-е сутки происходило значительное укорочение теломер. Таким образом, впервые показано, что фитопатогенная инфекция влияет на длину теломер у растений, что открывает новое направление в изучении их регуляции при биотическом стрессе.
Ключевые слова: теломеры растений; стабильность ДНК; биотический стресс; бактериальная инфекция; рибосомные белки; метилтрансфераза рРНК; фитопатогены

Introduction

The interaction of living organisms with the environment requires diverse and finely regulated adaptive mechanisms to maintain homeostasis, among which genome protection plays a crucial role. Biotic stress can enforce enormous pressure on the maintenance of genome integrity. In animal cells, microbial metabolites such as colibactin E have been shown to exert genotoxic effects on host DNA, including through alkylation, leading to its degradation (Wilson et al., 2019; Xue et al., 2019; Puschhof, Sears, 2022). Furthermore, pathogens such as *Helicobacter pylori* and *Staphylococcus aureus* can induce double-strand breaks and fragmentation of human DNA (Chakraborty et al., 2011; Toller et al., 2011). Biotic stress, particularly that induced by phytopathogenic microorganisms, also causes complex and multilevel effects on plant organisms (Kong, Yang, 2023; Solomon et al., 2024). At the genomic level, microbial infections can induce double-strand DNA breaks in plants, alter chromatin structure and architecture, and trigger changes in DNA and histone methylation as part of the stress response (Pavet et al., 2006; Dowen et al., 2012; López Sánchez et al., 2016; Alonso et al., 2018; Hannan Parker et al., 2022; Corrêa et al., 2024).

In turn, telomeres are among the fundamental structural and functional elements responsible for maintaining chromatin integrity in plants, as in most other eukaryotes (McKnight et al., 2002; Riha, Shippen, 2003; Zakian, 2012; Peska, Garcia, 2020; Shakirov et al., 2022). Telomeres are specialized nucleoprotein structures located at the ends of the linear chromosomes of most eukaryotes that consist of tandem DNA repeats together with associated protein complexes. They maintain linear genome stability by preventing chromosome ends from being recognized by endogenous exonucleases and DNA repair machinery and by avoiding pathological chromosome fusions (Riha, Shippen, 2003). In addition, telomeres are involved in controlling unprogrammed genome degradation during cell division, which contributes to replicative senescence (Watson, Riha, 2010). Therefore, telomere shortening is considered an important marker of genome instability (Gan, 2003; Amiard et al., 2014; Shakirov et al., 2022; Agabekian et al., 2024; Di Pietro et al., 2024). Stress can also affect telomere structure, promote telomeric DNA shortening, and reduce telomerase activity. However, the relationship between telomere regulation and infectious diseases under both laboratory and natural conditions remains poorly understood, and the underlying mechanisms remain controversial (Karell et al., 2017; Giraudeau et al., 2019; Sudyka et al., 2019; Wolf et al., 2023; Ravindran et al., 2024). Moreover, reliable information regarding the effects of biotic stress on the regulation of telomere length in plants is currently lacking.

Telomere length in *A. thaliana* is known to vary widely among hundreds of characterized ecotypes (Riha et al., 2002; Shakirov, Shippen, 2004; Abdulkina et al., 2019; Choi et al., 2021). Moreover, it has previously been shown that the molecular mechanisms regulating telomeres are closely linked to the maintenance of the structure and function of another important subcellular structure – the ribosome (Abdulkina et al., 2019; Valeeva et al., 2023). Accordingly, several proteins associated with ribosome biogenesis have been identified as positive regulators of telomere length in *Arabidopsis* (Abdulkina et al., 2019; Agabekian et al., 2024). The *NOP2A/OLI2* gene of *A. thaliana* encodes a key enzyme involved in 25S rRNA methylation and maturation of the 60S ribosomal subunit, and mutation of this gene is followed by severe defects in plant development and metabolism (Maekawa, Yanagisawa, 2021), as well as by a 27 % reduction in telomere length relative to wild-type plants (Abdulkina et al., 2019). The NOP2A paralog, NOP2C/NSUN5, also belongs to the NOP2/NSUN family of 25S rRNA methyltransferases (Burgess et al., 2015). However, mutations in the gene encoding this enzyme do not lead to changes in telomere length (Kobayashi, 2018), indicating functional divergence between these paralogs. Telomere shortening in *A. thaliana* was also observed in mutants of the ribosomal protein L5 paralog genes *OLI5/RPL5A* and *OLI7/RPL5B* (Abdulkina et al., 2019). In addition, more data indicate that ribosomal proteins, as well as ribosome-binding and ribosome-modifying proteins, play an important role in stress adaptation (Schosserer et al., 2015; Tieu Ngoc et al., 2021; Fakih et al., 2023; Siodmak et al., 2023; Wang et al., 2024; Cai et al., 2025; Fakih, Germain, 2025).

In this work, we studied the effect of bacterial infection on telomere length in *A. thaliana*. In our study, we used plant lines differing both in genotype and in telomere length, including two natural *A. thaliana* Col-0 and Sf-2 ecotypes, characterized by intermediate and long telomeres, respectively, as well as two mutant lines: *oli5-2*$^{-/-}$, with short telomeres, and *nop2c-2*$^{-/-}$, with intact telomere length. Therefore, our findings represent an important first step toward understanding the regulation of telomere length in plants under biotic stress.

Materials and methods

Plant material and bacterial strains. *Arabidopsis thaliana* ecotypes Columbia-0 (Col-0) and San Feliu-2 (Sf-2), as well as the mutant lines *nop2c-2*$^{-/-}$ (At1g06560; encoding the rRNA S-adenosyl-L-methionine-dependent methyltransferase *NOP2C*) and *oli5-2*$^{-/-}$ (At3g25520; encoding the large ribosomal subunit protein *OLI5/RPL5A*), were used in this study. Second-generation (G2) mutant plants were homozygous for the T-DNA insertion in the target gene. Seeds were obtained from the Arabidopsis Biological Resource Center (ABRC; Ohio State University, USA).

Seeds were surface-sterilized in sodium hypochlorite solution according to a standard protocol (Lindsey et al., 2017). After sterilization, seeds were stratified at 4 °C for 24 h. Subsequently, 10 seeds were plated on solid MS medium (Murashige and Skoog basal medium: 1/2 MS, 1 % sucrose, pH 5.6) in Petri dishes and grown for 14 days under standard conditions (16 h light/8 h dark photoperiod, 20–22 °C, 45–60 % relative humidity, light intensity 7,000–9,000 lux) in a growth chamber (Sanyo, Japan). The phytopathogenic strain *Pseudomonas syringae* pv. *tomato* DC3000 (Pst DC3000) from the laboratory collection was used for plant infection experiments. Pst DC3000 cultures were grown at 28 °C in liquid MG medium (mannitol-glutamate medium; mannitol, 54.9 mM; glutamic acid, 6.8 mM; KH_2PO_4, 0.735 mM; $CaCl_2$, 1.80 mM; NaCl, 3.42 mM; $MgSO_4 \cdot 7H_2O$, 0.812 mM; $FeSO_4$, 1.78 mM; Na-EDTA, 0.994 mM; pH 7.0) supplemented with rifampicin (50 μg/mL) under shaking at 250 rpm.

Plant infection was performed using an optimized flood-inoculation method developed for *A. thaliana* seedlings grown in Petri dishes (Ishiga et al., 2011). Prior to infection, all plant lines were cultivated for 14 days under the standard conditions described in section "Plant material and bacterial strains". Pst DC3000 cultures were grown for 12–14 h to an optical density of $OD_{600} = 0.4$, also as described above. Before plant inoculation, bacterial cells were harvested by centrifugation at 5,000 rpm, washed twice to remove residual medium, and resuspended in 2.5 % sucrose solution supplemented with 0.025 % Silwet L-77. Control plants were treated with sterile 2.5 % sucrose solution containing 0.025 % Silwet L-77 without bacteria. Plant stress responses and visible disease symptoms were monitored and photographed at 1, 3, 5, and 7 days post infection (dpi); plants of the Sf-2 line were additionally monitored for up to 18 dpi. For each plant line, experiments were performed in at least three biological replicates, including three control groups and three infected groups for each time point. Plant tissue samples for telomere length analysis were collected at 3, 5, and 7 dpi, as well as at an additional 18 dpi for the Sf-2 line.

Analysis of bacterial infection in plant tissues. To evaluate bacterial colonization of *A. thaliana* tissues by *P. syringae* DC3000, bacteria were reinoculated from both the plant surface and homogenized plant tissues as described by Y. Ishiga et al. (2011). The analysis was conducted at six days post infection.

Analysis of plant telomere length. Genomic DNA for telomere length analysis was isolated from plant tissues using a standard CTAB-based extraction method (Castillo-González et al., 2022). Telomere lengths at specific chromosome arms were analyzed using the PETRA (Primer Extension Telomere Repeat Amplification) assay with primers specific to the subtelomeric regions of chromosome arms 4 and 5: primer 4R for the short arm of chromosome 4, and primers 5L and 5R for the short and long arms of chromosome 5, respectively (Mayer et al., 1999; Tabata et al., 2000; Heacock et al., 2004; Watson, Shippen, 2007). PCR products were analyzed by non-radioactive Southern blot hybridization using a DIG-labeled telomeric DNA probe (Sigma-Aldrich, USA) and DIG-specific antibodies (Roche, USA). Mean telomere lengths were calculated using the WALTER online tool (https://www.ceitec.eu/chromatin-molecular-complexes/rg51/tab?tabId=125#WALTER) (Lyčka et al., 2021).

Statistical analyses were performed using GraphPad Prism version 8.0 (GraphPad Software, USA). Mean telomere length values and standard deviations were used for data visualization and graph generation. Statistical significance was assessed using a two-sided *t*-test, and differences were considered statistically significant at $p < 0.05$.

Results and discussion

Differences in susceptibility to phytopathogenic infection among plants with different genotypes

To identify an association between the biotic stress responses and telomere length regulation, 14-day-old *A. thaliana* plants of four genotypes were infected with the phytopathogenic strain *P. syringae* DC3000 (Supplementary Fig. S1)[1]:

1) the control ecotype Col-0, with a telomere length of 2.0–5.0 kb (Shakirov, Shippen, 2004);
2) the Sf-2 ecotype, with long telomeres of 4.5–7.0 kb (Abdulkina et al., 2019);
3) the *oli5-2*$^{-/-}$ mutant line, deficient in the ribosomal protein gene *OLI5/RPL5A* (telomeres are shorter compared to the Col-0 wild type, 1.8–2.0 kb);
4) the *nop2c-2*$^{-/-}$ mutant line, deficient in the gene *NOP2C* of the S-acetyl-L-methionine-dependent rRNA methyltransferase, a regulator of ribosome biogenesis (the average telomere length across all chromosomes remains intact compared to the Col-0 wild type) (Kobayashi, 2018).

To confirm infection and the invasion of the phytopathogen into plant cells, we performed bacterial plating both from surface washes and from homogenates derived from surface-sterilized tissues of control and infected plants. It was found that the phytopathogenic strain is present within both the intracellular and intercellular tissue compartments, thereby confirming the pathogen's invasion into the intercellular spaces of the tissues as well as into the interior of the plant cells (Fig. S2). The presence of Pst DC3000 cells was also detected on the plant surfaces (Fig. S2). In the control group (uninfected plants), no colonies of *P. syringae* DC3000 were detected, neither on the surface nor within the plants, indicating an absence of contamination during the experiment (Fig. S2).

During the experiment, all plants in the uninfected control groups showed normal growth and no visible signs of

[1] Supplementary Figures S1–S3 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_San_Engl_30_5.pdf

stress (Fig. 1). In contrast, infected plants displayed distinct responses to the phytopathogen depending on the genotype. No noticeable morphological or phenotypic changes were observed in any of the infected plant lines during the first two days after infection. However, by three days post infection, symptoms of leaf pathology became evident, with the most pronounced effects observed in the mutant line *oli5-2*$^{-/-}$. In *oli5-2*$^{-/-}$ plants, changes in leaf pigmentation were observed, indicating the accumulation of anthocyanins, protective pigments that are produced in response to various stress factors and contribute to protection against increased levels of reactive oxygen species (ROS) (Gould, 2004; Li, Ahammed, 2023). In addition, chlorosis was detected, primarily in the cotyledons. Chlorosis is a characteristic symptom of DC3000 infection in *Arabidopsis* (Katagiri et al., 2002) and is associated with the progression of leaf senescence. Symptoms of tissue damage, including chlorosis, necrosis, and leaf wilting, progressively increased in *oli5-2*$^{-/-}$ plants throughout the experiment. By seven dpi, tissue damage had reached a critical level, and the experiment was terminated for this line (Fig. 1). The *nop2c-2*$^{-/-}$ mutant and wild-type Col-0 plants exhibited higher resistance to infection. At three dpi, plants of both lines retained a darker green coloration, showed no obvious anthocyanin accumulation, and displayed fewer chlorotic leaves than *oli5-2*$^{-/-}$ plants. Disease symptoms became more pronounced at five dpi and reached severe levels by seven dpi (Fig. 1).

Notably, the long-telomere ecotype Sf-2 showed the highest resistance to infection among all plant lines analyzed. Beginning at five dpi, infected plants demonstrated reduced growth, shown as smaller leaves comparing to the control group (Fig. 1). Symptoms of tissue damage, including chlorosis and leaf necrosis, were first observed after 14 dpi. The major part of shoots showed visible signs of infection only at 18 dpi.

Thus, our results indicate that different *A. thaliana* genotypes exhibit distinct levels of resistance and stress responses to biotic stress caused by phytopathogenic bacteria. Among the lines examined, the *oli5-2*$^{-/-}$ mutant, which displays a telomere-shortening phenotype, was the most susceptible to infection, whereas the long-telomere ecotype Sf-2 exhibited the highest level of resistance. Taken together, these observations suggest a possible association between telomere length and resistance to phytopathogenic bacteria.

Effect of bacterial infection on telomere length in *A. thaliana*

The plant infection experiment revealed that the *oli5-2*$^{-/-}$ mutant line, characterized by shortened telomeres, was the most susceptible to infection by *P. syringae* DC3000. More generally, plant lines with shorter telomeres tended to exhibit lower resistance to this phytopathogen. To further investigate the relationship between infection and telomere dynamics, telomere lengths were analyzed in infected and control plants. Telomere length was measured using the PCR-based PETRA (Primer Extension Telomere Repeat Amplification) assay, which enables accurate analysis of telomere length using small amounts of plant material and low DNA input (Heacock et al., 2004; Watson, Shippen, 2007). In addition, PETRA allows determination of telomere length at individual chromosome arms, facilitating the detection of statistically significant differences even when absolute changes in telomere length are relatively small.

Telomere lengths were analyzed at the short arm of chromosome 4 (4R) and at the short and long arms of chromosome 5 (5L and 5R, respectively). Measurements were performed at three days post infection, when differences in disease symptoms and stress responses among the plant lines were clearly apparent. Telomere shortening at the short arm of chromosome 4 was observed in infected plants compared to controls in all three plant lines (see the Table, Fig. 2). However, statistically significant shortening was detected only in the *nop2c-2*$^{-/-}$ mutant line, in which mean telomere length decreased from

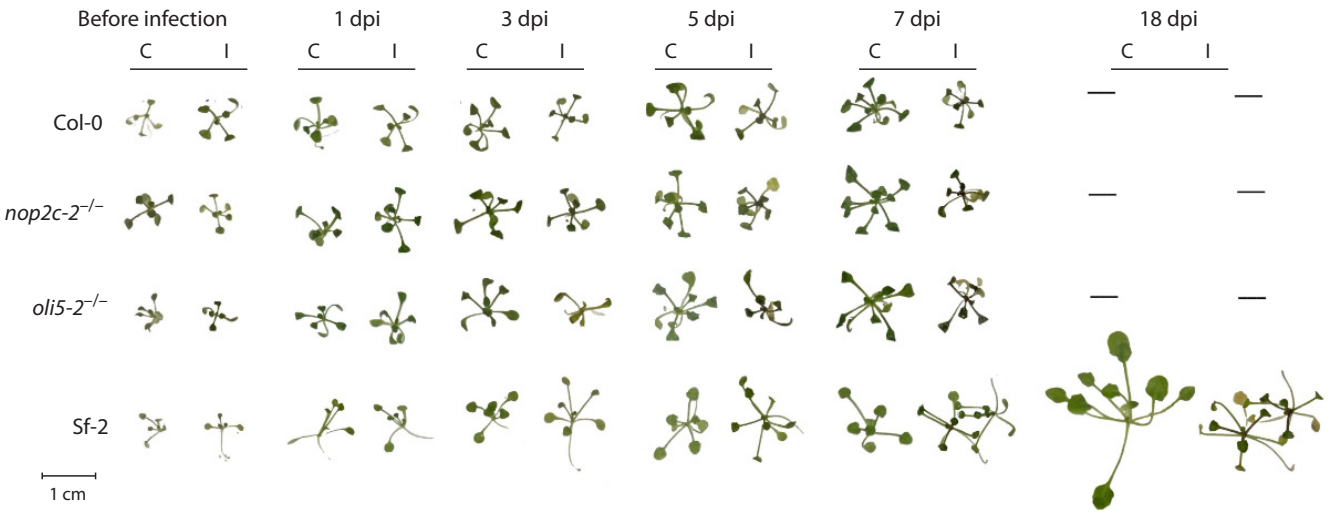


Fig. 1. Progression of disease symptoms in *A. thaliana* seedlings following infection (I) and in control plants (C).
Col-0, wild type; *oli5-2*$^{-/-}$ and *nop2c-2*$^{-/-}$, mutant lines in the Col-0 genetic background; Sf-2, a long-telomere ecotype.

6.14 ± 0.39 kb in control plants to 5.19 ± 0.16 kb in infected plants, with an average shortening of 0.95 kb ($p = 0.0042$) (see the Table, Fig. 2). In the *oli5-2*$^{-/-}$ line, which displayed the most severe visible disease symptoms at this time point, telomere shortening was also observed, with mean telomere length decreasing from 3.24 ± 0.17 kb in control plants to 3.11 ± 0.16 kb in infected plants (average shortening of 0.13 kb). However, this difference was not statistically significant ($p = 0.20$). A similar pattern was observed in the wild-type Col-0 line, in which telomere length decreased from 4.62 ± 0.49 kb in control plants to 4.10 ± 0.11 kb in infected plants (average shortening of 0.52 kb; $p = 0.24$) (see the Table, Fig. 2).

To assess telomere dynamics at later stages of infection, an additional analysis of telomere length at chromosome arm 4R was performed in *oli5-2*$^{-/-}$ plants at seven dpi. At this later stage, when disease symptoms were substantially more visible, telomeres in infected plants were significantly longer than in control plants (3.29 ± 0.01 kb *vs* 3.15 ± 0.06 kb, respectively; $p = 0.025$) (see the Table, Fig. S3).

The dynamics of telomere length changes at chromosome arm 4R in the long-telomere ecotype Sf-2 closely paralleled the progression of infection in this plant line (see the Table). At three and seven dpi, when no obvious signs of disease or growth retardation were observed, telomere length did not differ from that of control plants. In contrast, by 18 dpi, when plants displayed pronounced symptoms of biotic stress, a significant reduction in telomere length at chromosome arm 4R was detected.

To determine whether the observed changes in telomere length were common to multiple chromosome arms, we additionally analyzed telomeric loci on chromosome 5 using primers specific to long and short chromosome arms (5R and 5L, respectively). However, a controversial trend in telomere length regulation was observed at chromosome arms 5R and 5L. At the 5L arm, telomeres tended to be longer in infected plants than in the corresponding controls across all plant lines examined, although these differences were not statistically significant. In contrast, no differences in telomere length

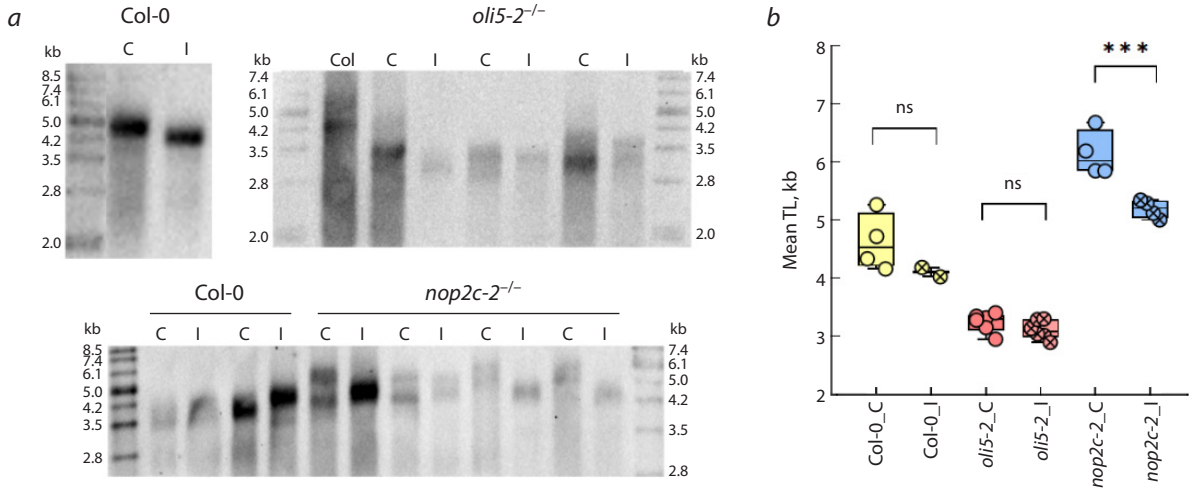


Fig. 2. Telomere length at the short arm of chromosome 4 (4R) in Col-0, *oli5-2*$^{-/-}$, and *nop2c-2*$^{-/-}$ plants at three days post infection.

a – Southern blot analysis of PETRA products; *b* – mean telomere length (kb). Statistical significance was assessed using Student's *t*-test. ns, not significant; *** $p \leq 0.001$.

Telomere lengths of *A. thaliana* lines under control conditions and following infection with *Pseudomonas syringae* pv. *tomato* DC3000

Chromosome arm	Col-0		*oli5-2*$^{-/-}$		*nop2c-2*$^{-/-}$		Sf-2	
	C	I	C	I	C	I	C	I
3 days post infection								
4R	4.62 ± 0.49	4.10 ± 0.11	3.24 ± 0.17	3.11 ± 0.16	6.14 ± 0.39	5.19 ± 0.16	7.11	6.49
5L	3.67 ± 0.14	4.29	3.42 ± 0.46	3.68 ± 0.26	4.21 ± 0.46	4.35 ± 0.72		nd
5R	3.13 ± 0.20	2.76 ± 0.26	1.97 ± 0.08	2.16 ± 0.14	3.83 ± 0.11	3.87 ± 0.25		nd
7 days post infection								
4R		nd	3.15 ± 0.06	3.29 ± 0.01		nd	7.27 ± 0.14	6.96 ± 0.04
18 days post infection								
4R		nd		nd		nd	6.63 ± 0.70	3.71 ± 1.14

Note. C – control plants; I – infected plants; nd – no data.

were detected at the 5R arm (see the Table). The difference between the results obtained for chromosome arm 4R and those observed at chromosome 5 may be explained by the fact that telomere length dynamics can be regulated independently at different chromosome arms within the same organism (Graakjaer et al., 2003; Perner et al., 2003; Sholes et al., 2022; Karimian et al., 2024).

Taken together, these results indicate that infection with the phytopathogen *P. syringae* DC3000 is associated with alterations in telomere length in *A. thaliana*, and that these changes depend on plant genotype.

Despite the long history of telomere research, relatively little is known about telomere length regulation in response to viral or bacterial infections and parasitic interactions. Moreover, the limited number of available studies have been conducted primarily in animals or clinical human samples. To date, reliable information regarding the effects of biotic stress on plant telomeres remains lacking. In the present study, we used infection of *A. thaliana* plants of different genotypes with the phytopathogenic bacteria *P. syringae* DC3000 as a model system to investigate changes in telomere length. Our principal finding is the identification of an association between infection and telomere length dynamics, as well as evidence that telomere length changes depend on plant genotype. Although only a limited number of natural ecotypes and mutant lines were examined, the present results lay ground for future studies of telomere regulation and chromatin stability in the context of plant disease and plant-microorganism interactions.

Evidence for an association between biotic stress and telomere length regulation is particularly well established in virus-derived diseases of humans and animals. On the one hand, viral infections can induce genomic DNA damage, telomere shortening, and telomere dysfunction. Several studies have demonstrated that chronic viral infections are associated with damage to host telomeric DNA and progressive telomere shortening (Bellon, Nicot, 2017; Dowd et al., 2017; Ji et al., 2019; Lim et al., 2022). On the other hand, viral effects on telomere biology may also be mediated through increased telomerase activity and subsequent oncogenic transformation of infected cells. Both DNA and RNA oncogenic viruses have been shown to interact with telomerase, either by integrating into host telomeric regions or by activating telomerase expression through promoter regulation (Salimi-Jeda et al., 2021). Viruses such as Epstein–Barr virus (Kamranvar, Masucci, 2017), human papillomavirus (Liu X. et al., 2008), and hepatitis B virus (Salimi-Jeda et al., 2021) can disrupt telomere and telomerase regulation, thereby contributing to cellular immortalization and tumor development. In addition, several viruses have been reported to induce alternative lengthening of telomeres (ALT) pathways (Zhu et al., 2017; Lippert et al., 2021; Tornesello et al., 2022).

Beyond studies of viral infections in humans, several investigations have examined the effects of other forms of biotic stress on telomere length in animals. For example, malaria infection has been shown to promote telomere shortening in bird species such as tawny owls and blue tits and to affect their lifespan (Karell et al., 2017; Sudyka et al., 2019). In contrast, a study of Soay sheep found no association between leukocyte telomere length and infection by parasitic nematodes, leading the authors to conclude that the costs of immune maintenance are not directly linked to telomere length in this species (Ravindran et al., 2024).

Our results indicate that plant lines with shorter telomeres were more susceptible to phytopathogenic infection and exhibited telomere shortening at earlier stages following infection. These observations raise the question of whether initial telomere length is associated with resistance to infection. Could telomere length serve as a reliable prognostic biomarker for predicting susceptibility or, conversely, enhanced resistance to biotic stress? Addressing this question will require studies involving larger and more diverse populations, including natural *A. thaliana* ecotypes spanning a broad range of telomere lengths, as well as additional mutants affected in known regulators of plant telomere maintenance.

To date, large-scale investigations of the relationship between telomere length and infection have been conducted primarily in the context of human viral diseases. Interest in this topic increased substantially following the COVID-19 pandemic, when several studies reported an inverse association between telomere length and the severity of clinical outcomes in hospitalized patients. These findings suggest that longer telomeres may be associated with enhanced immune function and reduced susceptibility to severe COVID-19 outcomes (Dos Santos et al., 2021; Vos et al., 2025). However, large-scale studies examining the relationship between telomere length and infectious diseases in both natural animal populations and human cohorts are complicated by numerous confounding factors, including differences in sample size, age structure, infection type and severity, telomere measurement methodology, and the tissues analyzed. Furthermore, establishing a causal relationship between telomere length and susceptibility to infection or disease severity remains challenging (Tunnicliffe et al., 2024).

Another important question raised by our findings concerns the mechanism through which phytopathogenic infection influences telomere length regulation in plants. We hypothesize that one possible mechanism involves the detrimental effects of reactive oxygen species on cellular homeostasis and chromatin stability. ROS levels are known to increase during plant immune responses to biotic stress. Under non-stress conditions, ROS produced by plant cells function as important signaling molecules involved in the regulation of numerous cellular processes, and their production is tightly controlled (Guo et al., 2023; Sood, 2025). However, during biotic stress, including phytopathogen recognition and infection, ROS production increases substantially. Elevated ROS levels contribute to several key components of plant immunity, including direct antimicrobial activity against pathogen membranes, proteins, and nucleic acids; activation of signaling pathways and immune-response cascades; and induction of genes associated with stress responses and systemic acquired resistance (SAR) (Sewelam et al., 2016; Zhang M., Zhang S., 2022; Liu C. et al., 2024; Haghpanah et al., 2025; Sood, 2025). At the same time, intracellular ROS homeostasis must remain tightly regulated, as excessive ROS accumulation can cause oxidative damage to cellular proteins, lipids, and nucleic acids, ultimately leading

to programmed cell death (Vanderauwera et al., 2011; Dvorak et al., 2021; Sahu et al., 2022; Guo et al., 2023; Sood, 2025).

Telomeres may be particularly vulnerable to oxidative damage. Because telomeric DNA is enriched in guanine, the most readily oxidized of the canonical DNA bases, telomeric regions represent preferential targets of ROS-mediated damage. Oxidation of guanine residues results in the formation of lesions such as 8-oxoguanine (8-oxo-G), contributing to genomic instability (Castillo-González et al., 2022). Indeed, telomeres have been reported to accumulate disproportionately high levels of oxidative damage. Under oxidative stress conditions, telomeric DNA accounts for approximately 0.2–15 % of the total cellular 8-oxo-G content, despite representing only 0.02–0.07 % of the genome. Consequently, telomeres may accumulate 8-oxo-G at frequencies up to 100-fold higher than the remainder of the genome (Castillo-González et al., 2022). In addition, reduced telomerase activity may further exacerbate the detrimental effects of oxidative stress on telomere maintenance. Consistently with this hypothesis, decreased telomerase activity has been associated with telomere dysfunction in *A. thaliana* plants exposed to spaceflight conditions, a model characterized by elevated oxidative stress (Barcenilla et al., 2023).

The use of mutant lines remains an important approach for investigating telomere biology and telomere responses to environmental stress in model organisms (Campitelli et al., 2022; Castillo-González et al., 2022; Barcenilla et al., 2023). In the present study, we analyzed two *A. thaliana* mutant lines carrying mutations in the *OLI5/RPL5A* and *NOP2C* genes. Both gene products are known components of the ribosome biogenesis pathway (Fujikura et al., 2009; Burgess et al., 2015; Kobayashi, 2018; Abdulkina et al., 2019). Our results showed that *oli5-2*$^{-/-}$ plants were more susceptible to infection than *nop2c-2*$^{-/-}$ plants. However, significant telomere shortening at three dpi was detected only in the *nop2c-2*$^{-/-}$ line. In contrast, telomere length in *oli5-2*$^{-/-}$ plants decreased slightly at three dpi and increased at seven dpi, suggesting dynamic regulation of telomere length during the progression of stress.

The observed relationship between mutations in ribosome-associated genes, susceptibility to infection, and telomere dynamics may reflect the central role of ribosomes in plant adaptation to environmental stress, including pathogen challenge. Under both biotic and abiotic stress conditions, protein translation represents a major regulatory target through which cells reduce energy consumption by selective synthesis of proteins involved in stress adaptation. For example, studies of plant ribosomes have revealed stress-dependent expression patterns of genes encoding proteins of both the large (RPL) and small (RPS) ribosomal subunits (Fakih, Germain, 2025). Furthermore, silencing of the ribosomal protein genes *NbRPSaA*, *NbRPS5A*, and *NbRPS24A* in *Nicotiana benthamiana* was shown to reduce the translation of defense-related transcripts, whereas silencing of *NbRACK1A* impaired the translation of specific antioxidant enzymes (Fakih et al., 2023). In addition, proteins involved in ribosome biogenesis, including rRNA methyltransferases, have been implicated in stress resistance. M. Schosserer et al. (2015) demonstrated that reduced expression of NSUN5-family rRNA methyltransferases, which include the NOP2C homolog, increases lifespan and stress resistance in yeast, nematodes, and flies. Notably, loss of NSUN5 alters ribosome conformation in yeast, affecting translational fidelity and promoting selective recruitment of mRNAs involved in oxidative stress responses (Schosserer et al., 2015).

Conclusion

In summary, the present study demonstrates that infection of *A. thaliana* with the phytopathogenic bacterium *P. syringae* DC3000 is associated not only with the development of disease symptoms but also with changes in telomere length regulation. Since telomeres play a central role in maintaining genome stability, these findings suggest that telomere dynamics may represent an important component of plant responses to biotic stress. Our results further indicate that infection-associated changes in telomere length depend on plant genotype. In particular, the *nop2c-2*$^{-/-}$ mutant line exhibited significant telomere shortening at an early stage of infection, whereas the long-telomere ecotype Sf-2 showed both the highest resistance to infection and stable telomere length during the early stages of disease development. Taken together, these findings provide the first evidence of an association between phytopathogenic infection and telomere length dynamics in plants. The results establish a foundation for future studies aimed at understanding the molecular mechanisms linking biotic stress, genome stability, and telomere regulation. In the longer term, telomere length and telomere-associated pathways may have potential as biomarkers of plant stress resistance and as targets for the development of crop varieties with enhanced tolerance to pathogens and other environmental stresses.

References

Abdulkina L.R., Kobayashi C., Lovell J.T., Chastukhina I.B., Aklilu B.B., Agabekian I.A., Suescún A.V., Valeeva L.R., Nyamsuren C., Aglyamova G.V., Sharipova M.R., Shippen D.E., Juenger T.E., Shakirov E.V. Components of the ribosome biogenesis pathway underlie establishment of telomere length set point in *Arabidopsis*. *Nat Commun.* 2019;10(1):5479. doi 10.1038/s41467-019-13448-z

Agabekian I.A., Abdulkina L.R., Lushnenko A.Y., Young P.G., Valeeva L.R., Boskovic O., Lilly E.G., Sharipova M.R., Shippen D.E., Juenger T.E., Shakirov E.V. Arabidopsis *AN3* and *OLIGOCELLULA* genes link telomere maintenance mechanisms with cell division and expansion control. *Plant Mol Biol.* 2024;114(3):65. doi 10.1007/s11103-024-01457-6

Alonso C., Ramos-Cruz D., Becker C. The role of plant epigenetics in biotic interactions. *New Phytol.* 2018;221(2):731-737. doi 10.1111/nph.15408

Amiard S., Olivier M., Allain E., Choi K., Smith-Unna R., Henderson I.R., White C.I., Gallego M.E. Telomere stability and development of *ctc1* mutants are rescued by inhibition of EJ recombination pathways in a telomerase-dependent manner. *Nucleic Acids Res.* 2014;42(19):11979-11991. doi 10.1093/nar/gku897

Barcenilla B.B., Meyers A.D., Castillo-González C., Young P., Min J.H., Song J., Phadke C., Land E., Canaday E., Perera I.Y., Bailey S.M., Aquilano R., Wyatt S.E., Shippen D.E. Arabidopsis telomerase takes off by uncoupling enzyme activity from telomere length maintenance in space. *Nat Commun.* 2023;14(1):7854. doi 10.1038/s41467-023-41510-4

Bellon M., Nicot C. Telomere dynamics in immune senescence and exhaustion triggered by chronic viral infection. *Viruses*. 2017;9(10):289. doi 10.3390/v9100289

Burgess A.L., David R., Searle I.R. Conservation of tRNA and rRNA 5-methylcytosine in the kingdom Plantae. *BMC Plant Biol.* 2015; 15:199. doi 10.1186/s12870-015-0580-8

Cai J., Shen L., Kang H., Xu T. RNA modifications in plant adaptation to abiotic stresses. *Plant Commun.* 2025;6(2):101229. doi 10.1016/j.xplc.2024.101229

Campitelli B.E., Razzaque S., Barbero B., Abdulkina L.R., Hall M.H., Shippen D.E., Juenger T.E., Shakirov E.V. Plasticity, pleiotropy and fitness trade-offs in Arabidopsis genotypes with different telomere lengths. *New Phytol.* 2022;233(4):1939-1952. doi 10.1111/nph.17880

Castillo-González C., Barbero Barcenilla B., Young P.G., Hall E., Shippen D.E. Quantification of 8-*oxoG* in plant telomeres. *Int J Mol Sci.* 2022;23(9):4990. doi 10.3390/ijms23094990

Chakraborty S.P., Kar Mahapatra S., Sahu S.K., Das S., Tripathy S., Dash S., Pramanik P., Roy S. Internalization of *Staphylococcus aureus* in lymphocytes induces oxidative stress and DNA fragmentation: possible ameliorative role of nanoconjugated vancomycin. *Oxid Med Cell Longev.* 2011;2011:942123. doi 10.1155/2011/942123

Choi J.Y., Abdulkina L.R., Yin J., Chastukhina I.B., Lovell J.T., Agabekian I.A., Young P.G., Razzaque S., Shippen D.E., Juenger T.E., Shakirov E.V., Purugganan M.D. Natural variation in plant telomere length is associated with flowering time. *Plant Cell.* 2021;33(4): 1118-1134. doi 10.1093/plcell/koab022

Corrêa R.L., Kutnjak D., Ambrós S., Bustos M., Elena S.F. Identification of epigenetically regulated genes involved in plant-virus interaction and their role in virus-triggered induced resistance. *BMC Plant Biol.* 2024;24(1):172. doi 10.1186/s12870-024-04866-3

Di Pietro E., Burla R., La Torre M., González-García M.P., Dello Ioio R., Saggio I. Telomeres: an organized string linking plants and mammals. *Biol Direct.* 2024;19(1):119. doi 10.1186/s13062-024-00558-y

Dos Santos G.A., Pimenta R., Viana N.I., Guimarães V.R., Romão P., Candido P., de Camargo J.A., Hatanaka D.M., Queiroz P.G.S., Teruya A., Leite K.R.M., Srougi V., Srougi M., Reis S.T. Shorter leukocyte telomere length is associated with severity of COVID-19 infection. *Biochem Biophys Rep.* 2021;27:101056. doi 10.1016/j.bbrep.2021.101056

Dowd J.B., Bosch J.A., Steptoe A., Jayabalasingham B., Lin J., Yolken R., Aiello A.E. Persistent herpesvirus infections and telomere attrition over 3 years in the Whitehall II cohort. *J Infect Dis.* 2017; 216(5):565-572. doi 10.1093/infdis/jix255

Dowen R.H., Pelizzola M., Schmitz R.J., Lister R., Dowen J.M., Nery J.R., Dixon J.E., Ecker J.R. Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci USA.* 2012; 109(32):E2183-E2191. doi 10.1073/pnas.1209329109

Dvorak P., Krasylenko Y., Zeiner A., Šamaj J., Takáč T. Signaling toward reactive oxygen species-scavenging enzymes in plants. *Front Plant Sci.* 2021;11:618835. doi 10.3389/fpls.2020.618835

Fakih Z., Germain H. Implication of ribosomal protein in abiotic and biotic stress. *Planta.* 2025;261(4):85. doi 10.1007/s00425-025-04665-6

Fakih Z., Plourde M.B., Germain H. Differential participation of plant ribosomal proteins from the small ribosomal subunit in protein translation under stress. *Biomolecules*. 2023;13(7):1160. doi 10.3390/biom13071160

Fujikura U., Horiguchi G., Ponce M.R., Micol J.L., Tsukaya H. Coordination of cell proliferation and cell expansion mediated by ribosome-related processes in the leaves of *Arabidopsis thaliana*. *Plant J.* 2009;59(3):499-508. doi 10.1111/j.1365-313X.2009.03886.x

Gan S. Mitotic and postmitotic senescence in plants. *Sci Aging Knowledge Environ.* 2003;2003(38):re7. doi 10.1126/sageke.2003.38.re7

Giraudeau M., Heidinger B., Bonneaud C., Sepp T. Telomere shortening as a mechanism of long-term cost of infectious diseases in natural animal populations. *Biol Lett.* 2019;15(5):20190190. doi 10.1098/rsbl.2019.0190

Gould K.S. Nature's Swiss army knife: the diverse protective roles of anthocyanins in leaves. *J Biomed Biotechnol.* 2004;2004(5):314-320. doi 10.1155/S1110724304406147

Graakjaer J., Bischoff C., Korsholm L., Holstebroe S., Vach W., Bohr V.A., Christensen K., Kølvraa S. The pattern of chromosome-specific variations in telomere length in humans is determined by inherited, telomere-near factors and is maintained throughout life. *Mech Ageing Dev.* 2003;124(5):629-640. doi 10.1016/s0047-6374(03)00081-2

Guo W., Xing Y., Luo X., Li F., Ren M., Liang Y. Reactive oxygen species: a crosslink between plant and human eukaryotic cell systems. *Int J Mol Sci.* 2023;24(17):13052. doi 10.3390/ijms241713052

Haghpanah M., Namdari A., Kaleji M.K., Nikbakht-Dehkordi A., Arzani A., Araniti F. Interplay between ROS and hormones in plant defense against pathogens. *Plants*. 2025;14(9):1297. doi 10.3390/plants14091297

Hannan Parker A., Wilkinson S.W., Ton J. Epigenetics: a catalyst of plant immunity against pathogens. *New Phytol.* 2022;233(1):66-83. doi 10.1111/nph.17699

Heacock M., Spangler E., Riha K., Puizina J., Shippen D.E. Molecular analysis of telomere fusions in Arabidopsis: multiple pathways for chromosome end-joining. *EMBO J.* 2004;23(11):2304-2313. doi 10.1038/sj.emboj.7600236

Ishiga Y., Ishiga T., Uppalapati S.R., Mysore K.S. Arabidopsis seedling flood-inoculation technique: a rapid and reliable assay for studying plant-bacterial interactions. *Plant Methods.* 2011;7:32. doi 10.1186/1746-4811-7-32

Ji Y., Dang X., Nguyen L.N.T., Nguyen L.N., Zhao J., Cao D., Khanal S., … El Gazzar M., Ning S., Wang L., Moorman J.P., Yao Z.Q. Topological DNA damage, telomere attrition and T cell senescence during chronic viral infections. *Immun Ageing*. 2019;16:12. doi 10.1186/s12979-019-0153-z

Kamranvar S.A., Masucci M.G. Regulation of telomere homeostasis during Epstein–Barr virus infection and immortalization. *Viruses*. 2017;9(10):217. doi 10.3390/v9080217

Karell P., Bensch S., Ahola K., Asghar M. Pale and dark morphs of tawny owls show different patterns of telomere dynamics in relation to disease status. *Proc Biol Sci.* 2017;284(1859):20171127. doi 10.1098/rspb.2017.1127

Karimian K., Groot A., Huso V., Kahidi R., Tan K.T., Sholes S., Keener R., McDyer J.F., Alder J.K., Li H., Rechtsteiner A., Greider C.W. Human telomere length is chromosome end-specific and conserved across individuals. *Science.* 2024;384(6695):533-539. doi 10.1126/science.ado0431

Katagiri F., Thilmony R., He S.Y. The *Arabidopsis thaliana–Pseudomonas syringae* interaction. *Arabidopsis Book.* 2002;1:e0039. doi 10.1199/tab.0039

Kobayashi C.R. Gene Evolution and Function in *Arabidopsis* Telomere Biology System. PhD Dissertation. College Station, TX, USA: Texas A&M University, 2018

Kong F., Yang L. Pathogen-triggered changes in plant development: virulence strategies or host defense mechanism? *Front Microbiol*. 2023;14:1122947. doi 10.3389/fmicb.2023.1122947

Li Z., Ahammed G.J. Hormonal regulation of anthocyanin biosynthesis for improved stress tolerance in plants. *Plant Physiol Biochem.* 2023;201:107835. doi 10.1016/j.plaphy.2023.107835

Lim Y.S., Nguyen M.T.N., Pham T.X., Huynh T.T.X., Park E.M., Choi D.H., Kang S.M., Tark D., Hwang S.B. Hepatitis C virus nonstructural 5A protein interacts with telomere length regulation protein: implications for telomere shortening in patients infected with HCV. *Mol Cells*. 2022;45(3):148-157. doi 10.14348/molcells.2021.0167

Lindsey B.E. 3rd, Rivero L., Calhoun C.S., Grotewold E., Brkljacic J. Standardized method for high-throughput sterilization of Arabidopsis seeds. *J Vis Exp.* 2017;128:56587. doi 10.3791/56587

Lippert T.P., Marzec P., Idilli A.I., Sarek G., Vancevska A., Bower M., Farrell P.J., Ojala P.M., Feldhahn N., Boulton S.J. Oncogenic her-

pesvirus KSHV triggers hallmarks of alternative lengthening of telomeres. *Nat Commun.* 2021;12(1):512. doi 10.1038/s41467-020-20819-4

Liu C., Liu Q., Mou Z. Redox signaling and oxidative stress in systemic acquired resistance. *J Exp Bot*. 2024;75(15):4535-4548. doi 10.1093/jxb/erae193

Liu X., Dakic A., Chen R., Disbrow G.L., Zhang Y., Dai Y., Schlegel R. Cell-restricted immortalization by human papillomavirus correlates with telomerase activation and engagement of the hTERT promoter by Myc. *J Virol.* 2008;82(23):11568-11576. doi 10.1128/JVI.01318-08

López Sánchez A., Stassen J.H., Furci L., Smith L.M., Ton J. The role of DNA (de)methylation in immune responsiveness of *Arabidopsis*. *Plant J.* 2016;88(3):361-374. doi 10.1111/tpj.13252

Lyčka M., Peska V., Demko M., Spyroglou I., Kilar A., Fajkus J., Fojtová M. WALTER: an easy way to online evaluate telomere lengths from terminal restriction fragment analysis. *BMC Bioinformatics.* 2021;22(1):145. doi 10.1186/s12859-021-04064-0

Maekawa S., Yanagisawa S. Ribosome biogenesis factor OLI2 and its interactor BRX1-2 are associated with morphogenesis and lifespan extension in *Arabidopsis thaliana*. *Plant Biotechnol* (*Tokyo*). 2021;38(1):117-125. doi 10.5511/plantbiotechnology.20.1224a

Mayer K., Schüller C., Wambutt R., Murphy G., Volckaert G., Pohl T., Düsterhöft A., … Johnson A., Chen E., Marra M., Martienssen R., McCombie W.R. Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature.* 1999;402(6763):769-777. doi 10.1038/47134

McKnight T.D., Riha K., Shippen D.E. Telomeres, telomerase, and stability of the plant genome. *Plant Mol Biol.* 2002;48(4):331-337. doi 10.1023/a:1014091032750

Pavet V., Quintero C., Cecchini N.M., Rosa A.L., Alvarez M.E. *Arabidopsis* displays centromeric DNA hypomethylation and cytological alterations of heterochromatin upon attack by *Pseudomonas syringae*. *Mol Plant Microbe Interact.* 2006;19(6):577-587. doi 10.1094/MPMI-19-0577

Perner S., Brüderlein S., Hasel C., Waibel I., Holdenried A., Ciloglu N., Chopurian H., Nielsen K.V., Plesch A., Högel J., Möller P. Quantifying telomere lengths of human individual chromosome arms by centromere-calibrated fluorescence *in situ* hybridization and digital imaging. *Am J Pathol.* 2003;163(5):1751-1756. doi 10.1016/S0002-9440(10)63534-1

Peska V., Garcia S. Origin, diversity, and evolution of telomere sequences in plants. *Front Plant Sci.* 2020;11:117. doi 10.3389/fpls.2020.00117

Puschhof J., Sears C.L. Microbial metabolites damage DNA. *Science.* 2022;378(6618):358-359. doi 10.1126/science.ade6952

Ravindran S., Underwood S.L., Dorrens J., Seeker L.A., Watt K., Wilbourn R.V., Sparks A.M., … McNeilly T.N., Harrington L., Pemberton J.M., Nussey D.H., Froy H. No correlative evidence of costs of infection or immunity on leucocyte telomere length in a wild population of Soay sheep. *Proc Biol Sci.* 2024;291(2020):20232946. doi 10.1098/rspb.2023.2946

Riha K., Shippen D.E. Telomere structure, function and maintenance in *Arabidopsis*. *Chromosome Res.* 2003;11(3):263-275. doi 10.1023/a:1022892010878

Riha K., Watson J.M., Parkey J., Shippen D.E. Telomere length deregulation and enhanced sensitivity to genotoxic stress in *Arabidopsis* mutants deficient in Ku70. *EMBO J.* 2002;21(11):2819-2826. doi 10.1093/emboj/21.11.2819

Sahu P.K., Jayalakshmi K., Tilgam J., Gupta A., Nagaraju Y., Kumar A., Hamid S., Singh H.V., Minkina T., Rajput V.D., Rajawat M.V.S. ROS generated from biotic stress: effects on plants and alleviation by endophytic microbes. *Front Plant Sci*. 2022;13:1042936. doi 10.3389/fpls.2022.1042936

Salimi-Jeda A., Badrzadeh F., Esghaei M., Abdoli A. The role of telomerase and viruses interaction in cancer development, and telomerase-dependent therapeutic approaches. *Cancer Treat Res Commun.* 2021;27:100323. doi 10.1016/j.ctarc.2021.100323

Schosserer M., Minois N., Angerer T.B., Amring M., Dellago H., Harreither E., Calle-Perez A., … Wilson I.B., Polacek N., Grillari-Voglauer R., Breitenbach-Koller L., Grillari J. Methylation of ribosomal RNA by NSUN5 is a conserved mechanism modulating organismal lifespan. *Nat Commun.* 2015;6:6158. doi 10.1038/ncomms7158

Sewelam N., Kazan K., Schenk P.M. Global plant stress signaling: reactive oxygen species at the crossroad. *Front Plant Sci.* 2016;7:187. doi 10.3389/fpls.2016.00187

Shakirov E.V., Shippen D.E. Length regulation and dynamics of individual telomere tracts in wild-type *Arabidopsis*. *Plant Cell.* 2004;16(8):1959-1967. doi 10.1105/tpc.104.023093

Shakirov E.V., Chen J.-L., Shippen D.E. Plant telomere biology: the green solution to the end-replication problem. *Plant Cell.* 2022;34(7):2492-2504. doi 10.1093/plcell/koac122

Sholes S.L., Karimian K., Gershman A., Kelly T.J., Timp W., Greider C.W. Chromosome-specific telomere lengths and the minimal functional telomere revealed by nanopore sequencing. *Genome Res.* 2022;32(4):616-628. doi 10.1101/gr.275868.121

Siodmak A., Martinez-Seidel F., Rayapuram N., Bazin J., Alhoraibi H., Gentry-Torfer D., Tabassum N., Sheikh A.H., Kise J.K.G., Blilou I., Crespi M., Kopka J., Hirt H. Dynamics of ribosome composition and ribosomal protein phosphorylation in immune signaling in *Arabidopsis thaliana*. *Nucleic Acids Res.* 2023;51(21):11876-11892. doi 10.1093/nar/gkad827

Solomon W., Janda T., Molnár Z. Unveiling the significance of rhizosphere: implications for plant growth, stress response, and sustainable agriculture. *Plant Physiol Biochem.* 2024;206:108290. doi 10.1016/j.plaphy.2023.108290

Sood M. Reactive oxygen species (ROS): plant perspectives on oxidative signalling and biotic stress response. *Discov Plants*. 2025;2:187. doi 10.1007/s44372-025-00275-4

Sudyka J., Podmokła E., Drobniak S.M., Dubiec A., Arct A., Gustafsson L., Cichoń M. Sex-specific effects of parasites on telomere dynamics in a short-lived passerine – the blue tit. *Naturwissenschaften.* 2019;106(1-2):6. doi 10.1007/s00114-019-1601-5

Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E., Miyajima N., … Bevan M., Fransz P.; Kazusa DNA Research Institute; Cold Spring Harbor and Washington University in St Louis Sequencing Consortium; European Union Arabidopsis Genome Sequencing Consortium. Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. *Nature*. 2000;408(6814):823-826. doi 10.1038/35048507

Tieu Ngoc L.N., Jung Park S., Thi Huong T., Lee K.H., Kang H. N4-methylcytidine ribosomal RNA methylation in chloroplasts is crucial for chloroplast function, development, and abscisic acid response in *Arabidopsis*. *J Integr Plant Biol.* 2021;63(3):570-582. doi 10.1111/jipb.13009

Toller I.M., Neelsen K.J., Steger M., Hartung M.L., Hottiger M.O., Stucki M., Kalali B., Gerhard M., Sartori A.A., Lopes M., Müller A. Carcinogenic bacterial pathogen *Helicobacter pylori* triggers DNA double-strand breaks and a DNA damage response in its host cells. *Proc Natl Acad Sci USA.* 2011;108(36):14944-14949. doi 10.1073/pnas.1100959108

Tornesello M.L., Cerasuolo A., Starita N., Tornesello A.L., Bonelli P., Tuccillo F.M., Buonaguro L., Isaguliants M.G., Buonaguro F.M. The molecular interplay between human oncoviruses and telomerase in cancer development. *Cancers* (*Basel*). 2022;14(21):5257. doi 10.3390/cancers14215257

Tunnicliffe L., Muzambi R., Bartlett J.W., Howe L., Abdul Basit K., Warren-Gash C. Infection and telomere length: a systematic review protocol. *BMJ Open.* 2024;14(4):e081881. doi 10.1136/bmjopen-2023-081881

Valeeva L.R., Abdulkina L.R., Agabekian I.A., Shakirov E.V. Telomere biology and ribosome biogenesis: structural and functional interconnections. *Biochem Cell Biol.* 2023;101(5):394-409. doi 10.1139/bcb-2022-0383

Vanderauwera S., Suzuki N., Miller G., van de Cotte B., Morsa S., Ravanat J.L., Hegie A., Triantaphylidès C., Shulaev V., Van Mon-

tagu M.C., Van Breusegem F., Mittler R. Extranuclear protection of chromosomal DNA from oxidative stress. *Proc Natl Acad Sci USA.* 2011;108(4):1711-1716. doi 10.1073/pnas.1018359108

Vos S., Martens D.S., De Waele E., Dewyspelaere G., Mistiaen G., Goeminne P., Nawrot T.S. Telomere length and COVID-19 disease severity: insights from hospitalized patients. *Front Aging*. 2025;6: 1577788. doi 10.3389/fragi.2025.1577788

Wang Z., Zhang X., Liu C., Duncan S., Hang R., Sun J., Luo L., Ding Y., Cao X. AtPRMT3-RPS2B promotes ribosome biogenesis and coordinates growth and cold adaptation trade-off. *Nat Commun.* 2024;15(1):8693. doi 10.1038/s41467-024-52945-8

Watson J.M., Riha K. Comparative biology of telomeres: where plants stand. *FEBS Lett.* 2010;584(17):3752-3759. doi 10.1016/j.febslet.2010.06.017

Watson J.M., Shippen D.E. Telomere rapid deletion regulates telomere length in *Arabidopsis thaliana*. *Mol Cell Biol.* 2007;27(5):1706-1715. doi 10.1128/MCB.02059-06

Wilson M.R., Jiang Y., Villalta P.W., Stornetta A., Boudreau P.D., Carrá A., Brennan C.A., Chun E., Ngo L., Samson L.D., Engelward B.P., Garrett W.S., Balbo S., Balskus E.P. The human gut bacterial genotoxin colibactin alkylates DNA. *Science.* 2019;363(6428):eaar7785. doi 10.1126/science.aar7785

Wolf S.E., Zhang S., Clotfelter E.D. Experimental ectoparasite removal has a sex-specific effect on nestling telomere length. *Ecol Evol.* 2023;13(3):e9861. doi 10.1002/ece3.9861

Xue M., Kim C.S., Healy A.R., Wernke K.M., Wang Z., Frischling M.C., Shine E.E., Wang W., Herzon S.B., Crawford J.M. Structure elucidation of colibactin and its DNA cross-links. *Science.* 2019; 365(6457):eaax2685. doi 10.1126/science.aax2685

Zakian V.A. Telomeres: the beginnings and ends of eukaryotic chromosomes. *Exp Cell Res.* 2012;318(12):1456-1460. doi 10.1016/j.yexcr.2012.02.015

Zhang M., Zhang S. Mitogen-activated protein kinase cascades in plant signaling. *J Integr Plant Biol.* 2022;64(2):301-341. doi 10.1111/jipb.13215

Zhu Z., Tran H., Mathahs M.M., Moninger T.O., Schmidt W.N. HCV induces telomerase reverse transcriptase, increases its catalytic activity, and promotes caspase degradation in infected human hepatocytes. *PLoS One*. 2017;12(1):e0166853. doi 10.1371/journal.pone.0166853

Conflict of interest. The authors declare no conflict of interest.
Received August 15, 2025. Revised February 24, 2026. Accepted April 14, 2026.

PLANT GENETICS
Original article

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):735-746

doi 10.18699/vjgb-26-74

Molecular cytogenetic analysis of some *Picea* species from Eurasia and East Asia

O.V. Goryachkina [1] ✉, E.D. Badaeva [2, 3], A.B. Shcherban [4], E.N. Muratova [1]

[1] V.N. Sukachev Institute of Forest of the Siberian Branch of the Russian Academy of Sciences, Federal Research Center "Krasnoyarsk Science Center of the Siberian Branch of the Russian Academy of Sciences", Krasnoyarsk, Russia
[2] N.I. Vavilov Institute of General Genetics of the Russian Academy of Sciences, Moscow, Russia
[3] Engelhardt Institute of Molecular Biology of the Russian Academy of Sciences, Moscow, Russia
[4] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
✉ kvitko@ksc.krasn.ru

Abstract. Karyotypes of seven *Picea* species distributed across Eurasia and East Asia were analyzed using fluorescence *in situ* hybridization (FISH) and DAPI fluorescent staining. For *P. meyeri* and *P. purpurea*, the 5S and 45S rDNA sites were mapped here for the first time, while karyotypes of *P. abies*, *P. obovata*, *P. schrenkiana*, *P. crassifolia*, and *P. omorika* were examined from new locations. All the species studied had a similar distribution of 5S rDNA loci on both arms of the large metacentric chromosome III: a more intense signal was observed in the middle of the long arm, partially overlapping with the 45S rDNA signal, and a minor 5S rDNA site was localized terminally on the short arm. The distribution of 5S rDNA loci varied in the karyotypes of the studied species: from 7 to 9 pairs of chromosomes carried 45S rDNA signals of varying intensity. Strong (major) signals were localized intercalarily and coincided with persistent secondary constrictions. Their number per haploid genome was as follows: 8 in *P. omorika*, 6 in *P. abies* and *P. obovata*, and 5 in *P. schrenkiana*, *P. crassifolia*, *P. purpurea*, and *P. meyeri*. We also detected several minor 45S rDNA sites on the *Picea* chromosomes that had not been previously described in the literature. Supernumerary (B) chromosomes found in some seedlings of *P. obovata*, *P. meyeri*, and *P. purpurea* did not carry 45S and 5S rDNA loci. Intensely fluorescent DAPI-positive bands were observed in the intercalary regions of the spruce chromosomes, and the centromere regions of certain chromosomes were also stained. Several constant DAPI-positive blocks were identified on chromosomes I, III, and X, with similar locations across all studied species. Fluorescence *in situ* hybridization with 45S and 5S rDNA probes in combination with DAPI staining allowed us to identify virtually all homologous chromosomes and compare karyotypes of different species. The intra- and interspecific karyotypic diversity in the genus *Picea* is discussed considering the distribution of 5S and 45S rDNA loci and DAPI signals.

Key words: chromosome; karyotype; FISH; 45S rDNA; 5S rDNA; DAPI staining; *Picea*

For citation: Goryachkina O.V., Badaeva E.D., Shcherban A.B., Muratova E.N. Molecular cytogenetic analysis of some *Picea* species from Eurasia and East Asia. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2026;30(5):735-746. doi 10.18699/vjgb-26-74

Funding. This work was carried out as part of the Program of the V.N. Sukachev Institute of Forest of the Siberian Branch of the Russian Academy of Sciences, Project No. FWES-2024-0028 "Biodiversity of Siberian Forests: ecological-dynamic, genetical-breeding, physical-chemical and resource-technological aspects".

Acknowledgements. The authors thank Prof. E.A. Salina, E.M. Timonova and T.N. Kapko from Institute of Cytology and Genetics SB RAS for collaboration; N.V. Oreshkova from V.N. Sukachev Institute of Forest SB RAS for *Larix sibirica* DNA extraction; A.B. Bezdelev from Primorsky Aquarium FEB RAS, I.V. Tikhonova from V.N. Sukachev Institute of Forest SB RAS, Y.V. Tkacheva from Donetsk Botanical Garden and O.A. Mochalova from the Institute of Biological Problems of the North FEB RAS for supplying the seed samples. Isolation and cloning of the 5S rDNA probe were conducted in the Laboratory of molecular genetics and cytogenetics of plants, Institute of Cytology and Genetics SB RAS.

Молекулярно-цитогенетический анализ некоторых видов *Picea* из Евразии и Восточной Азии

О.В. Горячкина [1] ✉, Е.Д. Бадаева [2, 3], А.Б. Щербань [4], Е.Н. Муратова [1]

[1] Институт леса им. В.Н. Сукачева Сибирского отделения Российской академии наук – обособленное подразделение Федерального исследовательского центра «Красноярский научный центр Сибирского отделения Российской академии наук», Красноярск, Россия
[2] Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, Россия
[3] Институт молекулярной биологии им. В.А. Энгельгардта Российской академии наук, Москва, Россия
[4] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
✉ kvitko@ksc.krasn.ru

Аннотация. Представлены результаты молекулярно-цитогенетического исследования семи видов *Picea*, распространенных на территории Евразии и Восточной Азии, методом флуоресцентной гибридизации *in situ* (FISH) и флуоресцентного окрашивания DAPI. Кластеры рибосомных генов 45S и 5S рРНК на хромосомах *P. meyeri* и *P. purpurea* картированы впервые, *P. abies*, *P. obovata*, *P. schrenkiana*, *P. crassifolia*, *P. omorika* изучались из новых местопроизрастаний. Все виды показали сходное распределение локусов 5S рДНК на обоих плечах крупной метацентрической хромосомы III: более интенсивный сигнал наблюдался на длинном плече, частично перекрываясь с сигналом 45S рДНК, минорный сайт 5S рДНК был локализован терминально на коротком плече. Сигналы 45S рДНК разной интенсивности были выявлены на 7–9 парах хромосом в кариотипах исследуемых видов. Яркие (мажорные) сигналы локализованы интеркалярно и совпадали с постоянными вторичными перетяжками. Их количество на гаплоидный геном составило: у *P. omorika* – 8, у *P. abies* и *P. obovata* – 6, у *P. schrenkiana*, *P. crassifolia*, *P. purpurea* и *P. meyeri* – 5. Выявлено также несколько минорных сайтов 45S рДНК на хромосомах *Picea*, не описанных ранее в литературе. Добавочные (B) хромосомы, обнаруженные у некоторых проростков *P. obovata*, *P. meyeri* и *P. purpurea*, не несли локусов рибосомных генов 45S и 5S рРНК. На хромосомах ели наблюдались интенсивно флуоресцирующие DAPI-позитивные полосы в интеркалярных районах, а у некоторых хромосом окрашивалась также область центромеры. На хромосомах I, III и X пар удалось выявить несколько постоянных позитивных DAPI-блоков, имеющих сходную локализацию у всех изученных видов. FISH с зондами 45S и 5S рДНК в сочетании с окрашиванием DAPI позволяет идентифицировать практически все гомологичные пары хромосом и провести сравнение кариотипов разных видов ели. Обсуждается внутри- и межвидовое кариотипическое разнообразие в роде *Picea* на основе распределения локусов 5S и 45S рДНК и рисунка DAPI.
Ключевые слова: хромосома; кариотип; FISH; 45S рДНК; 5S рДНК; окрашивание DAPI; *Picea*

Introduction

The genus *Picea* A. Dietr. (Spruce) consists of 35–50 species; the number of species discriminated within a genus depends on the taxonomic system (Sukachev, 1928; Dallimore, Jackson, 1966; Bobrov, 1971; Liu T.S., 1982; Farjon, 1990, 2001; Lockwood et al., 2013). More than half *Picea* species grow in Siberia, Russian and foreign Far East, in the Pacific basin. They are among the main forest formers of the Northern Hemisphere and are of great practical importance. The spruce species exhibit significant morphological diversity, which might be due to their wide geographical range from temperate to subarctic regions.

Picea karyotypes consist of 12 chromosome pairs ($2n = 24$): eight pairs (I–VIII) are long metacentrics, two pairs (X, XI) are short metacentrics, and two pairs (IX, XII) are short submetacentric chromosomes (Liu Y.H., Li, 1985; Hizume, 1988; Li et al., 2001; Nkongolo, Mehes-Smith, 2012) The chromosomes of I–VIII pairs are similar in length and arm ratio, so it is not always possible to define pairs of homologs using traditional staining methods. Supernumerary B chromosomes, more often of the metacentric type, have been found in 21 *Picea* species (Muratova et al., 2023).

Fluorescence *in situ* hybridization (FISH) has great prospects in plant karyotype analysis, allowing direct localizing of various DNA sequences and physical chromosome mapping. Ribosomal DNA (rDNA) sequences belong to a family of conserved, moderately repetitive DNA, and are often used as cytogenetic markers for chromosome identification. They are organized into clusters of hundreds or thousands of gene copies separated by intergenic spacers (Roa, Guerra, 2012). Their study is used to identify patterns of chromosomal evolution, species and population differentiation, and evolutionary relationships (Rossello et al., 2022).

To date, less than half of *Picea* species have been analyzed by FISH (Ohri, 2021). 5S and 45S rDNA loci were mapped on chromosomes of 16 *Picea* species (Brown et al., 1993; Hizume, Kuzakawa, 1995; Lubaretz et al., 1996; Brown, Carlson, 1997; Hizume et al., 1999; Friesen et al., 2001; Šiljak-Yakovlev et al., 2002; Vischi et al., 2003; Sarri et al., 2008; Shibata, Hizume, 2008). The *Picea*-specific Sau3A family of satellite DNA was mapped on chromosomes of 11 *Picea* species (Shibata, Hizume, 2008). Most studies have been conducted on a small number of species from one or two localities, which is insufficient for plants with such extensive ranges as many spruce species. Furthermore, there is no unified system for classifying chromosomes in *Picea* karyotypes, making comparison of data obtained by different teams often problematic. It should be noted that correct identification of all homologous chromosomes is very important for comparing karyotypes and assessing phylogenetic relationships between taxa.

The aim of the present study is the comparative cytogenetic analysis of seven *Picea* species growing in Europe and Asia using FISH with the 5S and 45S ribosomal RNA gene probes. The intra- and interspecific karyotype diversity in the genus *Picea* is discussed considering the distribution of 5S and 45S rDNA loci and DAPI patterns.

The 5S and 45S rDNA sites on the chromosomes of *P. purpurea* Masters [syn. *P. likiangensis* var. *purpurea* (Masters) Dallim. et A.B. Jacs.] and *P. meyeri* Rehder et E.H. Wilson naturally occurring in China were mapped here for the first time. Another species from China, *P. crassifolia* Komarov [syn. *P. schrenkiana* var. *crassifolia* (Komarov) Komarov and *P. complanata* var. *crassifolia* (Komarov) Gaussen], was studied from a new locality. According to E.G. Bobrov (1971) classification, *P. crassifolia* and *P. meyeri* belong to the section *Picea*, series *Asperatae*, and *P. purpurea*, to the section *Casicta*, series *Likiangense*s.

The study included two Eurasian species with overlapping large ranges: Norway spruce *P. abies* (L.) H. Karst. (section *Picea*, series *Excelsae*) and Siberian spruce *P. obovata* Ledeb. (section *Picea*, series *Obovatae*). Within a continuous range, the European and Siberian spruce populations gradually diverge from one species to the other. At the contact zone, these two species hybridize, forming the interspecific hybrid

P. × *fennica* (Regel.) Kom. (Popov, 1999). The material for the study was collected from the area of "pure" species distribution.

The study also included *P. schrenkiana* Fisch. et C.A. Mey., section *Picea*, series *Obovatae*, which has a fairly extensive range in the mountains of Central Asia, and Serbian spruce (*P. omorika* (Pančić) Purkine, section *Omorika*) – an endemic species with an extremely limited range from Eastern Europe.

Materials and methods

Plant material and primary cytogenetic analysis. Population seed mixtures collected from natural stands and artificial plantation were used for the study (Table 1). The seeds were germinated in Petri dishes on moist filter paper at room temperature. Root tips with a length of approximately 0.5–1.0 cm were pretreated with 0.2 % colchicine for 15–18 h at 20 °C, followed by ice water for 1 h, and then fixed in a mixture of 96 % ethanol and glacial acetic acid (3 : 1) for a minimum of 3 h.

To count the number of chromosomes, identify chromosomal and genomic abnormalities, and analyze the morphology of chromosomes, we used standard methods for conifers with our own modifications (Pravdin et al., 1972). For FISH, the roots were treated with enzymes to soften the tissue. Materials were washed in distilled water for 1 h, macerated in 0.6 % Cellulysine (Calbiochem, Switzerland) water solution, pH ~ 3–4 at 24 °C overnight or in an enzyme mix consisting of 2 % cellulase Onozuka R-10 (Sigma-Aldrich, USA) and 2 % Pectolyase Y-23 (Fisher Scientific, USA) in citrate buffer (pH 4.0) at 37 °C for 30 min.

Fluorescent *in situ* hybridization. Wheat DNA probe, pTa71 (Gerlach, Bedbrook, 1979), labeled with digoxygenin by nick translation was used for 45S ribosomal RNA gene mapping on *P. purpurea* chromosomes. The probe was detected with antibodies to anti-digoxigenin-rhodamine (Fab fragments, Sigma-Aldrich, USA). The 5S rDNA probe was amplified from genomic DNA of *Larix sibirica* Ledeb. Total DNA was extracted using a modified CTAB method (Devey et al., 1996) and used as a template for PCR amplification. Two primers, P3 (5′-GAGTTCTGATGGGATCCGGTG-3′) and P4 (5′-CGCTTGGGCTAGAGCAGTAC-3′), were chosen for amplification of the entire spacer region and partial amplification of the 5S rRNA gene (Brown, Carlson, 1997; Trontin et al., 1999).

The following program was used for PCR amplification: 3 min at 96 °C, 30 cycles of (40 s at 94 °C, annealing at 53 °C for 1 min, and primer extension at 72 °C for 30 s), 10 min at 72 °C. PCR products were separated on 1 % agarose gel, stained with ethidium bromide and visualized under UV light. The fragment sizes of the 5S rDNA PCR product were approx. 250 bp. The corresponding band was excised following electrophoresis and DNA was purified using the QIAquick PCR purification kit (QIAGEN, Germany). Cloning of the 5S rDNA fragment was performed using the QIAGEN PCR Cloning Kit (QIAGEN, USA). Competent *Escherichia coli* XL-blue cells were used for transformation.

Recombinant colonies were selected on LB agar medium with ampicillin and Xgal/IPTG. The presence of inserts was checked by PCR with the above primers. A recombinant clone containing an insert of the expected length was selected for

Table 1. Geographic origins and cytogenetic characteristics of *Picea* samples

Species	Locality	Number of plants observed, pcs	Chromosome number, 2*n* (%)*	Number of nucleoli, max (x ± SE)	Number of 45S rDNA signals	
					major	minor
P. abies	Volyn Oblast, Ukraine	23	24	9 (5.63 ± 0.06)	12	4
P. crassifolia	Washington State, USA (cultivated trees)	23	24	10 (5.42 ± 0.09)	10	4
P. meyeri	Shanxi Province, China	29	24; 24 + 2B (4.8)	10 (5.48 ± 0.06)	12	4
P. obovata	(a) Krasnoyarsky Krai, Russia	64	24; 24 + 1B (18.7); 24 + 2B (3.1); 24 + 3B (1.6)	12 (6.48 ± 0.08)	12	2
	(b) Magadan Oblast, Russia	71	24	nd	12	2
	(c) Piy-Khemsky District, Tuva Republic, Russia	23	24	10 (6.20 ± 0.06)	12	2
P. omorika	(a) Serbia	54	24	nd	16	2
	(b) Washington State, USA (cultivated trees)	23	24	14 (8.20 ± 0.07)	16	2
P. purpurea	Sichuan Province, China	31	24; 24 + 1B (6.5)	10 (4.50 ± 0.06)	10	4
P. schrenkiana	Issyk-Kul Region, Kyrgyzstan	31	24	10 (6.29 ± 0.08)	10	4

* The frequency (in %) of plants with B chromosomes is indicated in parentheses.

use as a probe for FISH. For this purpose, the cloned rDNA PCR product was labelled with Biotin-16-dUTP (Roche, Germany) by nick translation, according to the manufacturer's protocol. The biotinylated probes were detected using avidin, conjugated with fluorescein (Fluorescein Avidin D, Vector Laboratories, USA). The hybridization signal was enhanced using fluorescein anti-avidin (Fluorescein Anti-Avidin D, Vector Laboratories, USA).

Two wheat DNA probes, pTa794 (5S rDNA) and pTa71 (45S rDNA), were used for *in situ* hybridization of other six *Picea* species studied (Gerlach, Bedbrook, 1979; Gerlach, Dyer, 1980). Fluorescent *in situ* hybridization was conducted as described in our previous work (Goryachkina et al., 2013).

The preparations were embedded in Vectashield mounting medium (Vector Laboratories, USA), containing 0.5 μg/mL DAPI (4′,6-diamidino-2-phenylindole dihydrochloride, Sigma-Aldrich, USA) for chromosome staining. The chromosomes were examined with an Axio Imager D1 (Carl Zeiss, Germany) microscope and recorded with an AxioCam CCD camera using the AxioVision v. 4.6 software package. Images were processed using Adobe Photoshop CS4 (Adobe Systems, USA).

Karyotype analysis. Chromosomes were measured on 7 to 27 (for different samples) well-spread metaphase plates. The following parameters were determined for each chromosome: total length, arm ratio, position of secondary constrictions (distance from the centromere to the secondary constriction / length of the arm ×100), the presence or absence of rDNA loci, the presence or absence of DAPI positive and negative bands, the relative position of the FISH signal and DAPI-bands (distance from the centromere to the middle of signal / length of the arm ×100). Pairs of homologous chromosomes were determined on the basis of chromosome morphology, the size and position of 5S and 45S rDNA loci and DAPI bands, when available. The chromosomes were numbered in order of decreasing absolute length. In the group of similar-sized metacentrics of II–VII pairs, the chromosome pairs were arranged in the same order as in the work of F. Shibata and M. Hizume (2008) to simplify the comparison of our results.

Results

Karyotypes of all seven *Picea* species consist of 24 somatic chromosomes ($2n = 24$). Supernumerary (B) chromosomes were found in several seedlings of *P. meyeri* and *P. purpurea*; in *P. obovata* they were observed only in the population from Krasnoyarsky Krai (Table 1). In all species, B chromosomes were metacentric (arm ratio 1.1–1.3 %), with the length ranging from 4.05 to 4.65 μm without rDNA sites. Morphometric parameters of *P. abies*, *P. obovata*, *P. omorika*, *P. meyeri*, and *P. schrenkiana* chromosomes are presented in Table 2. Chromosome morphologies of *P. crassifolia* and *P. purpurea* were not measured, and for these species we provide only the position of FISH signals.

All species showed similar distribution of 5S rDNA clusters; two unequal signals were detected on both arms of a single pair of large metacentric chromosomes III (Fig. 1). A larger signal of the 5S rDNA probe was always observed in the middle of the long arm; it partially overlapped with major 45S rDNA site, but was localized closer to the centromere. A smaller, minor site was located subterminally in the opposite arm. The pTa794 probe showed more intense hybridization signals than the probe amplified from the *L. sibirica* genome. This was a possible reason why we were unable to identify a minor 5S rDNA site on the short arm of chromosome III in *P. purpurea*.

45S rDNA signals of varying intensity were detected on 7–9 pairs of chromosomes in the karyotypes of the studied species. Major signals were always located intercalarily and coincided with the positions of secondary constrictions. Their number varied from eight (per haploid genome) in *P. omorika*, six in *P. abies* and *P. obovata*, to five in *P. schrenkiana*, *P. crassifolia*, *P. purpurea*, and *P. meyeri* (Table 3, Fig. 1).

Interspecific differences in the localization and intensity of 45S rDNA signals were identified in chromosome pairs II, VI, VIII, IX and X. The remaining pairs of chromosomes in the studied species did not differ: chromosome pairs I, VII, XI and XII did not carry clusters of 5S and 45S rRNA ribosomal genes, and long metacentrics of pairs III, IV and V showed a similar hybridization pattern with weak variation in signal intensity.

The 45S rDNA site on the short arm of chromosome II was observed in all species, but several differences in signal location and intensity were identified (Fig. 1). As measurements showed, in *P. schrenkiana* and *P. omorika*, the 45S rDNA locus was located more proximally than in the other studied species (Table 3); in addition, hybridization signal was weaker. On chromosome VI, the 45S rDNA site was observed in four of the seven species studied, but its localization varied: in *P. meyeri* and *P. purpurea*, the signal was detected in a medial position, while in *P. abies* and *P. omorika*, it was closer to the centromere; the signal intensity also varied (Fig. 1). Chromosome pairs VIII and X carried 45S rDNA loci of similar localization (Table 3); the studied species differed in the intensity of the hybridization signal.

A minor 45S rDNA site in pericentromeric regions of large submetacentric chromosome IX was observed in six *Picea* species (except for *P. purpurea*). The arm location of this site varied depending on the species: in *P. obovata* and *P. omorika*, it was located on the long arm of chromosome IX, while in *P. abies P. schrenkiana*, *P. crassifolia*, *P. meyeri*, it was found on the short arm. In the karyotype of *P. omorika*, chromosome IX also carried a unique major 45S rDNA signal in the medial region of the long arm, coinciding with the secondary constriction.

After FISH, several intensely fluorescent DAPI-positive bands were detected in the intercalary regions of most chromosomes; rarely, the centromere region was also stained (Fig. 2). DAPI-negative bands most often corresponded to the localization regions of the ribosomal genes 5S and 45S rRNA and the centromeres of most chromosomes.

Permanent intercalary DAPI-positive bands with bright fluorescence appeared on chromosomes I, III and X (Fig. 2); their distribution was conservative across the species and facilitated the identification of homologous chromosomes in karyotypes. One can observe proximal positive DAPI bands on both arms of chromosome I; a double proximal and distal band on the short arm of chromosome III; and a distal band on the long arm of chromosome X.

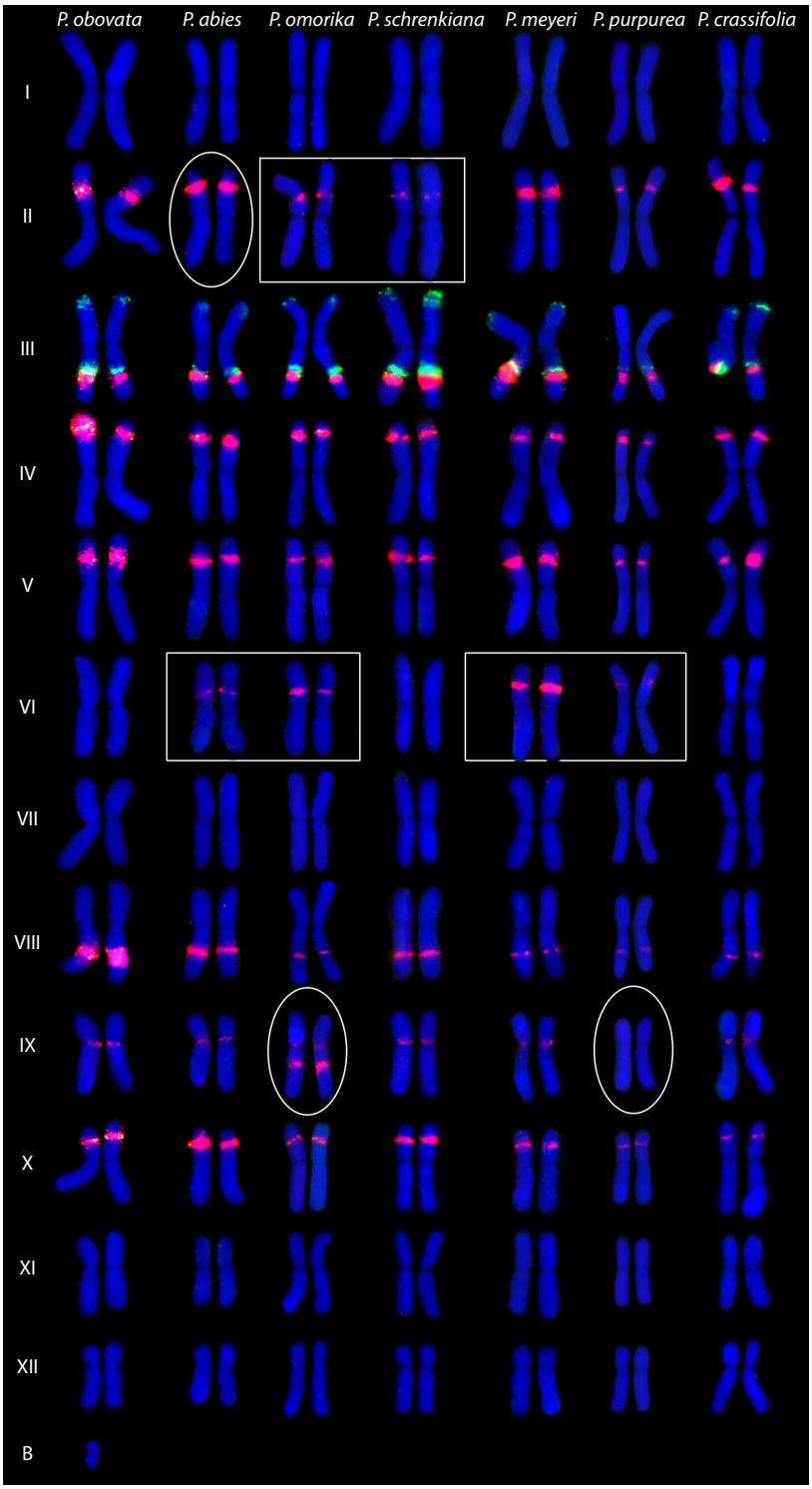


Fig. 1. FISH mapping of the 45S rDNA (red) and 5S rDNA (green) probes on metaphase chromosomes of seven *Picea* species.

Chromosomes of different species carrying polymorphic pTa71 signals in similar position are boxed. The chromosomes possessing unique signals or lacking any marker signals are outlined by an oval.

Other DAPI-positive bands were observed irregularly on chromosomes, in some cases, only in one homologue in a pair. The intercalary DAPI bands on the short arm of chromosome II, corresponding to the 45S rDNA sites in *P. obovata*, *P. abies*, and *P. meyeri*, were one of the brightest. In the remaining studied species, an intercalary DAPI band was also observed on the short arm of chromosome II, but it was weaker, located closer to the centromere and did not coincide with the 45S rDNA sites. Chromosome IV in most species showed numerous thin intercalary DAPI bands on both arms.

Table 2. Morphometric data of the chromosome pairs of different *Picea* species and localities: total length / arm ratio

Chromosome pair	*P. abies* (28#)	*P. meyeri* (30)	*P. obovata*			*P. omorika*		*P. schrenkiana* (50)
			a (15)	b (7)	c (27)	a (13)	b (18)	
I	13.60 ± 0.24 / 1.06 ± 0.01	15.76 ± 0.32 / 1.06 ± 0.01	19.30 ± 0.49 / 1.06 ± 0.02	19.09 ± 0.53 / 1.05 ± 0.02	14.37 ± 0.19 / 1.07 ± 0.01	18.00 ± 0.24 / 1.06 ± 0.01	14.82 ± 0.35 / 1.08 ± 0.01	16.33 ± 0.44 / 1.07 ± 0.01
II	12.64 ± 0.32 / 1.07 ± 0.01	14.78 ± 0.31 / 1.08 ± 0.01	17.27 ± 0.40 / 1.09 ± 0.02	17.06 ± 0.28 / 1.13 ± 0.03	13.27 ± 0.20 / 1.08 ± 0.01	16.18 ± 0.23 / 1.06 ± 0.02	12.95 ± 0.34 / 1.05 ± 0.01	14.83 ± 0.38 / 1.07 ± 0.01
III	12.51 ± 0.18 / 1.14 ± 0.02	14.62 ± 0.35 / 1.15 ± 0.02	16.78 ± 0.43 / 1.08 ± 0.01	17.52 ± 0.34 / 1.08 ± 0.02	12.44 ± 0.15 / 1.06 ± 0.01	15.50 ± 0.20 / 1.06 ± 0.01	13.12 ± 0.27 / 1.04 ± 0.01	14.42 ± 0.34 / 1.03 ± 0.01
IV	12.06 ± 0.17 / 1.11 ± 0.01	13.13 ± 0.33 / 1.12 ± 0.01	16.02 ± 0.45 / 1.11 ± 0.02	16.42 ± 0.20 / 1.08 ± 0.02	11.82 ± 0.12 / 1.10 ± 0.01	14.95 ± 0.17 / 1.09 ± 0.01	12.69 ± 0.34 / 1.11 ± 0.01	14.27 ± 0.39 / 1.03 ± 0.01
V	11.95 ± 0.26 / 1.11 ± 0.02	13.85 ± 0.36 / 1.09 ± 0.02	16.66 ± 0.41 / 1.12 ± 0.02	17.03 ± 0.41 / 1.08 ± 0.03	12.76 ± 0.19 / 1.09 ± 0.01	15.41 ± 0.21 / 1.15 ± 0.02	12.58 ± 0.29 / 1.17 ± 0.02	14.53 ± 0.39 / 1.15 ± 0.01
VI	11.74 ± 0.22 / 1.13 ± 0.01	13.55 ± 0.34 / 1.08 ± 0.01	16.13 ± 0.28 / 1.06 ± 0.02	16.37 ± 0.33 / 1.08 ± 0.02	12.26 ± 0.17 / 1.06 ± 0.01	14.82 ± 0.29 / 1.08 ± 0.04	12.16 ± 0.23 / 1.06 ± 0.02	14.13 ± 0.39 / 1.07 ± 0.01
VII	11.28 ± 0.17 / 1.09 ± 0.01	13.23 ± 0.31 / 1.10 ± 0.01	16.06 ± 0.51 / 1.06 ± 0.01	16.16 ± 0.34 / 1.08 ± 0.02	12.01 ± 0.13 / 1.05 ± 0.01	15.10 ± 0.20 / 1.03 ± 0.01	13.05 ± 0.36 / 1.03 ± 0.01	13.49 ± 0.41 / 1.06 ± 0.01
VIII	11.00 ± 0.18 / 1.33 ± 0.02	12.35 ± 0.32 / 1.20 ± 0.01	15.83 ± 0.35 / 1.34 ± 0.02	16.08 ± 0.47 / 1.34 ± 0.05	11.66 ± 0.19 / 1.27 ± 0.02	14.76 ± 0.29 / 1.28 ± 0.02	12.33 ± 0.29 / 1.29 ± 0.02	13.74 ± 0.34 / 1.28 ± 0.01
IX	9.83 ± 0.14 / 1.79 ± 0.03	11.23 ± 0.24 / 1.80 ± 0.02	13.93 ± 0.27 / 1.75 ± 0.03	14.46 ± 0.21 / 1.70 ± 0.06	10.56 ± 0.14 / 1.72 ± 0.03	13.25 ± 0.16 / 1.69 ± 0.01	11.37 ± 0.35 / 1.71 ± 0.03	12.27 ± 0.31 / 1.72 ± 0.02
X	9.75 ± 0.15 / 1.36 ± 0.03	10.59 ± 0.20 / 1.56 ± 0.02	13.46 ± 0.27 / 1.32 ± 0.02	13.72 ± 0.23 / 1.44 ± 0.03	10.34 ± 0.11 / 1.29 ± 0.01	13.07 ± 0.21 / 1.27 ± 0.02	10.21 ± 0.24 / 1.22 ± 0.01	11.21 ± 0.25 / 1.36 ± 0.01
XI	8.84 ± 0.16 / 1.26 ± 0.02	9.89 ± 0.24 / 1.23 ± 0.02	12.41 ± 0.21 / 1.23 ± 0.02	12.57 ± 0.24 / 1.26 ± 0.02	9.07 ± 0.09 / 1.16 ± 0.01	11.75 ± 0.14 / 1.22 ± 0.02	9.52 ± 0.18 / 1.20 ± 0.02	10.84 ± 0.31 / 1.27 ± 0.01
XII	8.34 ± 0.14 / 1.96 ± 0.03	8.91 ± 0.20 / 1.91 ± 0.03	10.87 ± 0.20 / 1.97 ± 0.04	11.04 ± 0.20 / 1.90 ± 0.05	8.14 ± 0.09 / 1.79 ± 0.02	10.49 ± 0.11 / 1.86 ± 0.04	8.42 ± 0.14 / 2.00 ± 0.03	9.70 ± 0.22 / 2.01 ± 0.02
B	–	4.15 ± 0.11 / 1.32 ± 0.03	4.65 ± 0.22 / 1.13 ± 0.02	–	–	–	–	–

Note. The total length is indicated as x ± SE (μm), arm ratio is indicated as x ± SE (%); x – means, SE – standard errors; # – number of metaphases measured.

Table 3. Relative position of the 45S rDNA loci (x ± SE), % on chromosomes of *Picea* species

Species and provenances	II (s)	III (l)	IV (s)	V (l)	VI (s)	VIII (l)	IX (l)	X (s)
P. abies	66.5 ± 0.9	54.9 ± 1.0	73.0 ± 0.5	53.6 ± 0.6	*42.0 ± 2.2*	48.4 ± 0.4	–	54.9 ± 0.2
P. crassifolia	53.3 ± 0.9	55.1 ± 2.2	72.4 ± 1.2	52.0 ± 1.3	–	*49.1 ± 0.5*	–	54.8 ± 1.2
P. meyeri	55.9 ± 0.6	54.1 ± 0.6	73.0 ± 0.5	50.0 ± 0.7	54.7 ± 0.6	*47.3 ± 0.4*	–	54.9 ± 0.2
P. obovata:								
a	55.0 ± 0.6	54.4 ± 0.4	74.1 ± 0.4	49.6 ± 0.6	–	46.2 ± 0.5	–	54.6 ± 0.5
b	54.9 ± 0.7	54.2 ± 0.8	72.9 ± 0.9	50.0 ± 0.5	–	47.5 ± 1.9	–	54.8 ± 0.6
c	52.8 ± 0.4	55.1 ± 0.6	73.0 ± 0.4	48.9 ± 0.5	–	48.4 ± 0.7	–	54.5 ± 0.3
P. omorika:								
a	38.9 ± 0.9	54.1 ± 0.5	73.9 ± 0.6	54.7 ± 0.7	44.3 ± 1.0	46.6 ± 0.5	36.5 ± 0.7	55.0 ± 0.2
b	41.0 ± 0.7	55.0 ± 0.6	74.3 ± 0.7	54.9 ± 0.8	43.4 ± 1.2	48.4 ± 0.9	38.7 ± 0.5	55.0 ± 0.4
P. purpurea	55.4 ± 0.4	54.2 ± 1.4	72.6 ± 0.4	51.0 ± 0.6	*54.1 ± 1.3*	*48.5 ± 0.5*	–	54.9 ± 0.4
P. schrenkiana	*40.0 ± 0.9*	55.2 ± 0.4	73.1 ± 0.2	51.1 ± 0.4	–	49.6 ± 0.2	–	54.8 ± 0.4
*P. glauca**	57.4 ± 2.4	52.5 ± 2.8	72.8 ± 3.6	51.6 ± 2.7	35.8 ± 1.3	49.8 ± 1.1	–	54.6 ± 3.1
*P. sitchensis**	56.3 ± 2.2	54.8 ± 4.2	73.6 ± 2.0	52.7 ± 4.6	35.8 ± 2.3	–	–	–

Note. Minor 45s rDNA sites are given in italics and underlined. *According to (Brown, Carlson, 1997).

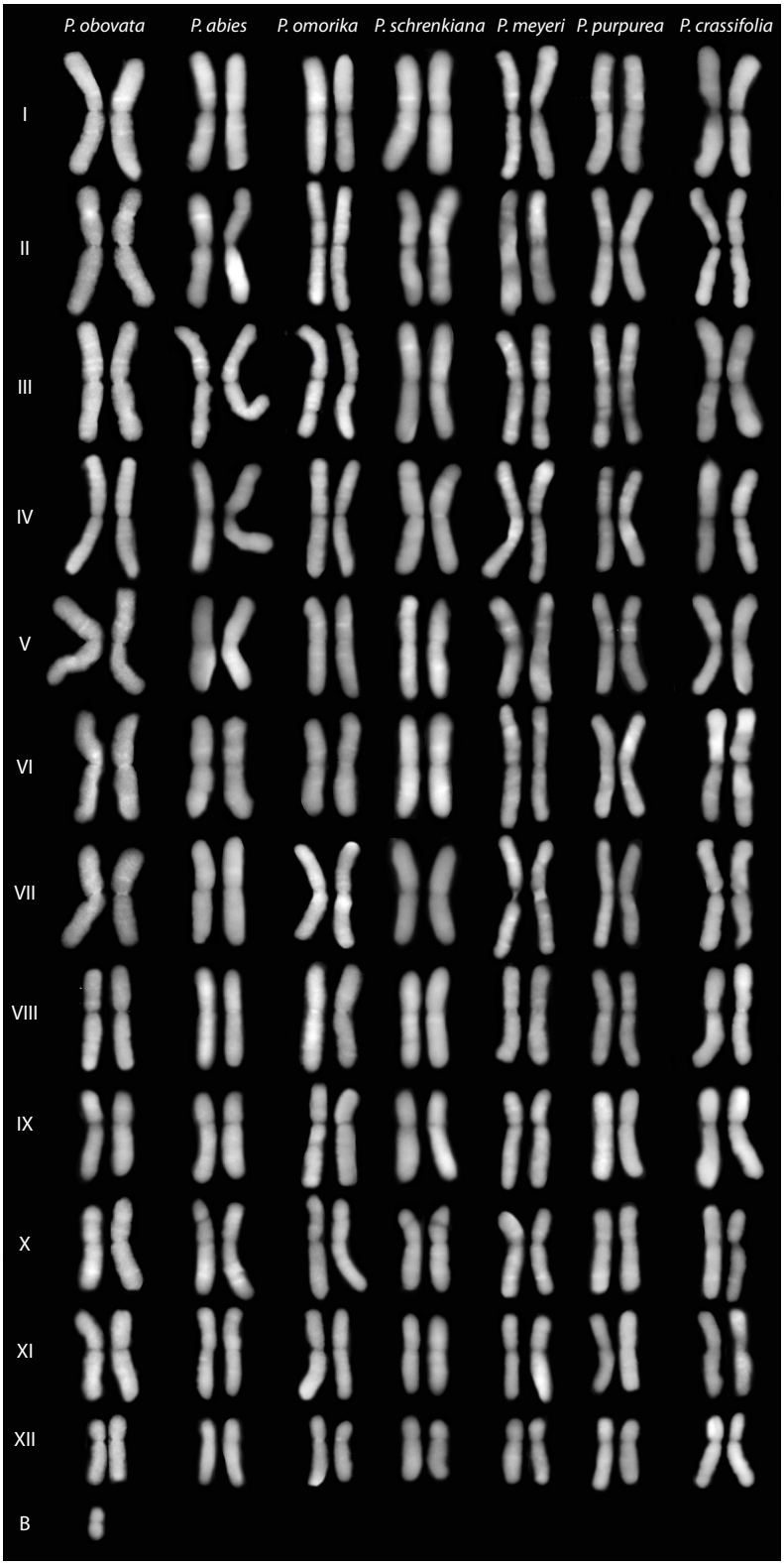


Fig. 2. DAPI staining of metaphase chromosomes of the studied *Picea* species.

In *P. obovata*, *P. omorika*, *P. meyeri* and *P. purpurea*, bright DAPI-positive blocks were observed in the proximal region of the short arm of chromosome V. In *P. omorika*, *P. meyeri* and *P. purpurea*, positive DAPI bands were observed in the middle of the long arm of chromosome XII.

Molecular cytogenetic methods coupled with traditional karyological analysis based on chromosome morphologies allowed us to define pairs of homologous chromosomes in karyotypes of seven *Picea* species. Eight chromosome pairs, namely I, III, IV, VIII, IX, X, XI, XII, were easily identified in

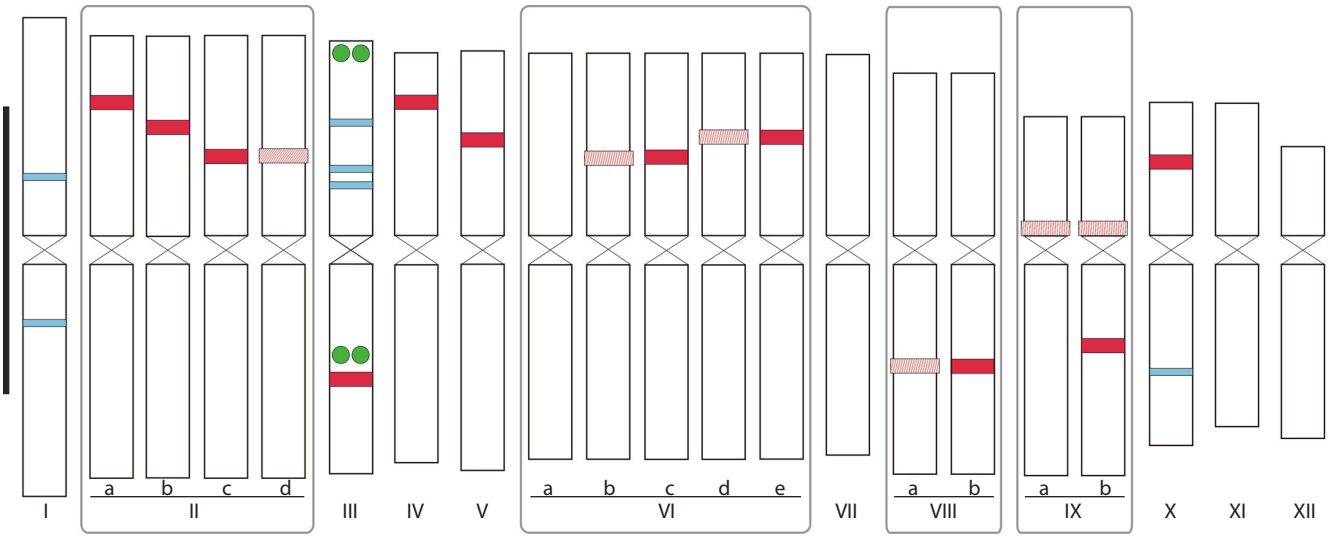


Fig. 3. Comparative idiogram of the *Picea* chromosomes showing the variation of chromosomal locations of 45S rDNA.

Major loci are shown as red bands, minor loci are shown as shaded red bands. 5S rDNA sites are shown as green circles; DAPI bands – as blue bands. Scale bar is 10 μm. I–XII – chromosome numbers. Chromosome II types (a) – *P. abies*; (b) – *P. crassifolia*, *P. meyeri*, *P. obovata*, *P. purpurea*; (c) – *P. omorika*; (d) – *P. schrenkiana*. Chromosome VI types (a) – *P. schrenkiana*, *P. crassifolia*, *P. obovata*; (b) – *P. abies*; (c) – *P. omorika*; (d) – *P. purpurea*; (e) – *P. meyeri*. Chromosome VIII types (a) – *P. crassifolia*, *P. purpurea*, *P. meyeri*; (b) – *P. omorika*, *P. schrenkiana*, *P. obovata*, *P. abies*. Chromosome IX types (a) – *P. crassifolia, P. schrenkiana, P. meyeri, P. obovata, P. abies*; (b) – *P. omorika*.

all species studied. Owing to similar morphological parameters of chromosomes II and V, VI and VII, it was more difficult to find their homologous pairs. Chromosomes II and V of *P. obovata*, *P. meyeri*, *P. purpurea*, and *P. crassifolia* all carried major intercalary 45S rDNA sites on short arms with similar localization (Table 3), but homologous chromosome pairs can be discriminated based on morphology and DAPI-banding. Chromosomes VI and VII of *P. obovata*, *P. schrenkiana*, *P. crassifolia* were difficult to distinguish because they lacked any FISH signals and had similar DAPI-patterns.

The diversity of spruce species with respect to the distribution of 5S and 45S rDNA loci and DAPI staining is reflected on a generalized idiogram of the *Picea* chromosomes shown on Figure 3. As can be seen, variability between species in the presence, intensity and localization of 45S rDNA signals exists for chromosomes II, VI, VIII, IX. These FISH patterns have been used to trace the relationships between *Picea* species, the basic premise being that more closely related species will have more similar karyotypes than distantly related ones.

Discussion

The genes encoding the 5S and 45S ribosomal RNA are widely used as molecular-cytogenetic markers in evolutionary and genomic studies of forest trees because of their highly conserved structure and tandem organization (Ribeiro et al., 2008). Within a species, the number of rDNA loci is generally stable, making them a convenient tool for chromosome identification. However, the number and location of sites varies between taxa, due to karyotype transformations during evolution.

Gymnosperms are known to possess one or four 5S rDNA loci per haploid genome, these sites can have both terminal and interstitial localization on the chromosome (Ribeiro et al., 2008; Garcia et al., 2017; Ohri, 2021). Much more diversity is observed in the number of 45S rDNA loci: from 2 (in most *Podocarpus* and *Juniperus* species) to 38 in *Pinus taeda* (Ohri, 2021). The position of 45S rDNA sites on the chromosomes of *Picea* and other gymnosperms is usually interstitial, whereas in angiosperms, the sites are mostly terminal (Roa, Guerra, 2012; Garcia et al., 2017).

In the Pinaceae, both ribosomal RNA gene families frequently appear on the same chromosome; many of these have both sites on the same chromosome arm. Partially overlapping 5S and 45S FISH signals are observed on *Picea*, *Pinus* and *Abies* chromosomes, but the Southern blot hybridization did not reveal significant cohybridization of the 26S (18S) and 5S rDNA probes in any of these. The most likely interpretation of colocalized signals is the juxtaposition of independent 5S and 45S arrays (Garcia, Kovařík, 2013). The arrangement of these signal sites was different; in some *Abies* species, the 5S rDNA signal is localized closer to the telomere (Shibata et al., 2004; Puizina et al., 2008), while in *Picea* species, the 5S rDNA signals were localized closer to the centromere. Both these 45S and 5S rDNA patterns could be recognized in different *Pinus* species (Liu Z. et al., 2003; Cai et al., 2006).

In the karyotypes of all studied species of *Larix*, and *Pseudotsuga menziesii*, the 5S and 45S rDNA loci were also located on the same chromosome, but on different arms (Hizume et al., 1996; Liu B. et al., 2006, 2007; Zhang et al., 2010; Goryachkina et al., 2013). Interspecific variability in the number and localization of 5S rDNA sites has been described in *Pinus* and *Abies* species (Liu Z. et al., 2003; Shibata et al., 2004; Cai et al., 2006; Puizina et al., 2008; Nkongolo, Mehes-Smith, 2012; Ohri, 2021). In the *Picea* and *Larix* species studied to date, the number of 5S rDNA loci is constant.

The chromosome carrying clusters of both 45S and 5S rDNA was found in all *Picea* species studied to date and

therefore can be used as cytogenetic marker of spruce karyotype (Brown et al., 1993; Hizume, Kuzukawa, 1995; Brown, Carlson, 1997; Hizume et al., 1999; Šiljak-Yakovlev et al., 2002; Shibata, Hizume, 2008). A similar pattern of clustering of 5S and 45S rDNA loci on one of the long metacentrics was observed in the karyotypes of *Larix* species: a single 5S rDNA locus is localized terminally on the short arm of chromosome III, which also carries a major 45S rDNA site on the other arm (Liu B. et al., 2006, 2007; Zhang et al., 2010; Goryachkina et al., 2013).

The number of 45S rDNA loci on chromosomes of different *Picea* species has been shown to vary from 8 to 26; American and Eurasian species possess more loci than East Asian ones (Ohri, 2021). Two chromosomes, namely I and XI, did not carry clusters of 45S rDNA ribosomal genes in any spruce species studied to date. Interspecific differences were observed in the number of 45S rDNA sites and their localization on the remaining chromosomes in the karyotype. Major 45S rDNA signals coincided with the permanent secondary constrictions, indicating the functional activity of the corresponding loci, as shown in previous studies (Šiljak-Yakovlev et al., 2002).

Two loci of 45S rDNA genes were characterized by high constancy and were detected in the karyotypes of all studied *Picea* species: there was a medial site on the long arm of chromosome III and a subterminal site on the short arm of chromosome IV. These loci facilitated the identification of homologous chromosomes in the group of long metacentrics of pairs II–VII. The distal 45S rDNA site on the short arm of chromosome V was also quite stable; it has not been identified at present only in the North American species *P. breweriana* (Shibata, Hizume, 2008). In the karyotype of *P. breweriana*, unique major 45S rDNA sites have been found on the short arm of chromosome III and in the pericentromeric region of chromosome X; these sites have not been observed on the chromosomes of other spruce species (Shibata, Hizume, 2008). Brewer spruce has been considered to occupy an isolated position in the *Omorika* section of the genus *Picea*, and its closeness to some Asian spruce species has been confirmed by studying the nucleotide sequences of the nuclear, mitochondrial and chloroplast genomes (Ran et al., 2006, 2015; Lockwood et al., 2013).

Interesting data have been obtained by comparing the localization of polymorphic sites of 45S rDNA on chromosome pairs II, VI and VIII. These results can be considered preliminary, since to date, the 5S and 45S rRNA genes have been mapped on the chromosomes of only half of the *Picea* species (Brown et al., 1993; Lubaretz et al., 1996; Brown, Carlson, 1997; Hizume et al., 1999; Šiljak-Yakovlev et al., 2002; Vischi et al., 2003; Sarri et al., 2008; Shibata, Hizume, 2008). However, certain regularities that may indicate the relationship or common origin of species should be noted.

Earlier, F. Shibata and M. Hizume (2008) described two putative positions of the major 45S rDNA signal on the short arm of chromosome II: subterminal (*P. abies*, *P. sitchensis*, *P. glauca*) and median (*P. asperata*, *P. crassifolia*, *P. jezoensis* var. *hondoensis*, *P. koraiensis*, *P. obovata*, *P. pungens* var. *glauca*). These locations are shown in the idiogram (Fig. 3). However, G.R. Brown et al. (1993) in their original work indicated the medial localization of this signal in *P. sitchensis* and *P. glauca* (Table 3), which allows us to classify these species into the second group. So, the 45S rDNA site on chromosome II of *P. abies* has a unique position in the subterminal part of the short arm (Fig. 3, Table 3). According to our results, the 45S rDNA cluster on chromosome II of *P. schrenkiana* and *P. omorika* was located more proximally compared to other species, and therefore these species can be assigned to the third group (Fig. 3). Thus, the 45S rDNA locus on the short arm of chromosome II can have three variants of location in different *Picea* species and appears to be the most polymorphic in localization on the chromosome arm.

We found that chromosome VI could exhibit several 45S rDNA patterns, but it is present in only half of the *Picea* species studied to date. Among the species studied, this site had a proximal localization in *P. omorika* (major locus) and *P. abies* (minor locus). In the North American species *P. glauca* and *P. sitchensis*, major 45S rDNA loci were noted on chromosome VI, also with a proximal localization of 35.8 % (Brown, Carlson, 1997). F. Shibata and M. Hizume (2008) showed that in the karyotype of *P. engelmannii*, chromosome VI has a 45S rDNA signal in the proximal part of the short arm, whereas in *P. koraiensis*, it has a signal in the medial part of the long arm. Morphometric data showed that in *P. purpurea* and *P. meyeri*, the 45S rDNA signal on chromosome VI was located medially on the short arm, although in metacentrics, it is sometimes difficult to identify the short and long arms (Table 3). Chromosome VI in all other studied spruce species (Eurasian *P. schrenkiana* and *P. obovata*, East Asian *P. crassifolia*, *P. asperata*, *P. jezoensis* var. *hondoensis* and *P. koyamai*, North American *P. breweriana*, *P. mariana* and *P. rubens*) did not have 45S rDNA signals.

The 45S rDNA site on chromosome VIII was observed in the middle of the long arm in all studied species, but the signal intensity varied among the species. Most East Asian species (*P. meyeri*, *P. crassifolia*, *P. purpurea*) showed weak hybridization signals, while European and Eurasian species (*P. abies*, *P. omorika*, *P. schrenkiana*, *P. obovata*) had major 45S rDNA loci. Among the species from East Asia studied by other authors, the 45S rDNA site of similar localization on chromosome VIII was observed only in *P. jezoensis* var. *hondoensis* and was absent in *P. asperata*, *P. koraiensis*, and *P. koyamai* (Hizume et al., 1999; Shibata, Hizume, 2008). However, in the last three species, fluorescent staining revealed intercalary CMA^+ bands on the long arm of chromosome VIII; among the East Asian species, they were not observed only in *P. maximowiszii* var. *senanensis* and *P. wilsonii* (Hizume, 2017). As is known, intercalary SMA^+ bands can correspond to the localization regions of 45S rRNA genes.

In the species we studied, except for *P. purpurea*, a minor 45S rDNA site was identified in the pericentromeric region of chromosome IX. This locus was previously detected in *P. abies* by hybridization with the PAF1 probe containing the 45S rDNA intergenic spacer (Vischi et al., 2003). Fluorescent staining showed similar results: spruce species from North America and Eurasia contained CMA^+ bands on chromosome IX, while a number of East Asian species (*P. wilsonii*, *P. koyamae*, *P. maximowiczii* var. *senanensis*, *P. likiangensis*)

lacked these bands (Hizume, 2017). *P. wilsonii*, *P. likiangensis* and *P. purpurea* are shown to be phylogenetically related as can be seen from the results of recent molecular genetic analyses of the nuclear, chloroplast and mitochondrial genes (Sun et al., 2014). *P. purpurea* and *P. likiangensis* also exhibit similar morphological traits: the shape and structure of seed scales, buds and needles, color, and pubescence of their shoots (Liu T.S., 1982).

Some minor 45S rDNA sites can represent low-copy-number sites of 45S rDNA that may have been active in ancestral forms of gymnosperms and then moved to distal regions of chromosomes as a result of structural rearrangements or the activity of mobile elements (Garcia, Kovařík, 2013; Garcia et al., 2017). This phenomenon known as amphiplasty frequently occurs in allopolyploid plants (Pikaard, 2000); however, the function of "latent" loci can be restored if the existing active nucleolus organizer regions (NORs) have been lost. Thus, the minor NOR on chromosome 1R of tetraploid *Triticale* become functional in genotypes lucking wheat chromosomes 1B and 6B (Badaev et al., 1992). Most minor NORs do not possess true rDNA genes, but pseudogenes (Pedersen, Linde-Laursen, 1994), and they may occur due to the action of mobile elements (Dubkovsky, Dvorak, 1995; Raskina et al., 2004).

Minor pericentromeric 45S rDNA sites were common in pines; their 2*n* number varied from 2 to 18 in different species (Liu Z. et al., 2003; Islam-Faridi et al., 2007; Bogunić et al., 2011; Shibata et al., 2016). M. Hizume et al. (2001) isolated the 18S and 26S rDNA sequences using microdissection of pericentromeric regions of *P. densiflora* chromosomes and proposed that these sequences are nonfunctional because the number of nucleoli in the interphase nuclei was lesser than the number of major NORs. The minor pericentromeric 45S rDNA sites were also observed in several *Larix* species (Goryachkina et al., 2013).

The FISH pattern has been used to suggest relationships between the species. From this point, signal locations on chromosomes II, VI, VII, VIII, and X seem to be most informative. Close location of 45S rDNA sites had been recorded for species belonging to the same sections (or subgenera in some scientist's opinion), established by different authors based on similarities in morphological and anatomical features (Sukachev, 1928; Dallimore, Jackson, 1966; Bobrov, 1971; Liu T.S., 1982; Schmidt, 1989; Ledig et al., 2004; Farjón, 1990).

It should be noted that intrasectional classification of the genus remains quite controversial. Furthermore, many *Picea* species cross readily with each other in contact ranges; cross experiments also showed hybridization between many geographically isolated spruce species, even from different continents (Wright, 1955; Mikkola, 1969). Our FISH results are largely consistent with the data of molecular phylogenetic study of the genus (Sigurgeirsson, Szmidt, 1993; Nkongolo, 1999; Ran et al., 2006, 2015; Lockwood et al., 2013).

On the other hand, similar location of 45S rDNA sites has been recorded for species from different groups of the genus *Picea*: *P. omorika vs P. schrenkiana* (the proximal 45S rDNA locus is on the short arm of chromosome II); *P. omorika vs P. abies* (the medial 45S rDNA is on the short arm of chromosome VI); *P. meyeri vs P. purpurea* (the proximal 45S rDNA is on the short arm of chromosome VI); *P. omorika vs P. pungens* var. *glauca* (the 45S rDNA is on the long arm of chromosome IX). Therefore, many aspects of taxonomic and phylogenetic relationships of *Picea* species are still awaiting their clarification by means of a broad range of molecular cytogenetic and genetic approaches.

Conclusion

The combination of molecular cytogenetic methods with traditional methods of karyological analysis based on the study of chromosome morphology allows the identification of pairs of homologous chromosomes in *Picea* karyotypes. It was established that karyotypes of *Picea* species differ in number, position, and intensity of 45S rDNA signals. Similar localization of signals is usually observed in species belonging to the same sections. In some cases, similar localization of 45S rDNA sites was noted in species from different series; it requires new comparative cytogenetics and molecular genetic studies. Therefore, the method of fluorescence *in situ* hybridization can provide valuable information on the relationship between species and gives a new possibility for analysis of genetic diversity, evolution, intra- and interspecific variability of the genus *Picea*.

References

Badaev N.S., Badaeva E.D., Dubovets N.I., Bolsheva N.L., Bormotov V.E., Zelenin A.V. Formation of a synthetic karyotype of tetraploid triticale. *Genome*. 1992;35(2):311-317. doi 10/1139/g92-047

Bobrov E.G. The history and systematics of the genus *Picea* A. Dietr. *Novitates Systematicae Plantarum Vascularium*. 1971;7:5-40 (in Russian)

Bogunić F., Šiljak-Yakovlev S., Muratović E., Ballian D. Different karyotype patterns among allopatric *Pinus nigra* (Pinaceae) populations revealed by molecular cytogenetics. *Plant Biol.* 2011;13(1):194-200. doi 10.1111/j.1438-8677.2010.00326.x

Brown G.R., Carlson J.E. Molecular cytogenetics of the genes encoding 18s-5.8s-26.s rDNA and 5s rRNA in two species of spruce (*Picea*). *Theor Appl Genet.* 1997;95(1-2):1-9. doi 10.1007/s001220050526

Brown G.R., Amarasinghe V., Kiss G., Carlson J.B. Preliminary karyotype and chromosomal localization of ribosomal DNA sites in White spruce using fluorescence *in situ* hybridization. *Genome*. 1993;36(2):261-267. doi 10.1139/g93-043

Cai Q., Zhang D.M., Liu Z.L., Wang X.R. Chromosomal localization of 5S and 18S rDNA in five species of subgenus *Strobus* and their implications for genome evolution of *Pinus*. *Ann Bot*. 2006;97(5):715-722. doi 10.1093/aob/mcl030

Dallimore W., Jackson A.B. A Handbook of Coniferae and Ginkgoaceae. London: Arnold, 1966

Devey M.E., Bell J.C., Smith D.N., Neale D.B., Moran G.F. A genetic linkage map for *Pinus radiata* based on RFLP, RAPD, and microsatellite markers. *Theor Appl Genet.* 1996;92(6):673-679. doi 10.1007/BF00226088

Dubcovsky J., Dvorak J. Ribosomal RNA multigene loci: nomads of the Triticeae genomes. *Genetics*. 1995;140(4):1367-1377. doi 10.1093/genetics/140.4.1367

Farjon A. Pinaceae: Drawings and Descriptions of the Genera *Abies*, *Cedrus*, *Pseudolarix*, *Keteleeria*, *Nothotsuga*, *Tsuga*, *Cathaya*, *Pseudotsuga*, *Larix* and *Picea*. Königstein: Koeltz Scientific Books, 1990

Farjon A. World Checklist and Bibliography of Conifers. Kew: Royal Botanic Gardens, 2001

Friesen N., Brandes A., Heslop-Harrison J.S. Diversity, origin, and distribution of retrotransposons (*gypsy* and *copia*) in conifers. *Mol Biol*

Evol. 2001;189(7):1176-1188. doi 10.1093/oxfordjournals.molbev.a003905

Garcia S., Kovařík A. Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organization. *Heredity*. 2013;111(1):23-33. doi 10.1038/hdy.2013.11

Garcia S., Kovařik A., Leitch A.R., Garnatje T. Cytogenetic features of rRNA genes across land plants: analysis of the Plant rDNA database. *Plant J.* 2017;89(5):1020-1030. doi 10.1111/tpj.13442

Gerlach W.L., Bedbrook J.R. Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucleic Acids Res*. 1979;77(7):1869-1885. doi 10.1093/nar/7.7.1869

Gerlach W.L., Dyer T.A. Sequence organization of the repeated units in the nucleus of wheat which contains 5S-rRNA genes. *Nucleic Acids Res*. 1980;8(21):4851-4865. doi 10.1093/nar/8.21.4851

Goryachkina O.V., Badaeva E.D., Muratova E.N., Zelenin A.V. Molecular cytogenetic analysis of Siberian *Larix* species by fluorescence *in situ* hybridization. *Plant Syst Evol.* 2013;299(2):471-479. doi 10.1007/s00606-012-0737-y

Hizume M. Karyomorphological studies in the family Pinaceae. *Mem Fac Educ Ehime Univ Nat Sci.* 1988;8(2):1-108

Hizume M. Chromosome banding in *Picea* IV. Comparative karyotype analysis of fluorescent band patterns in 25 taxa. *Chromosome Bot.* 2017;12(2):17-32. doi 10.3199/iscb.12.17

Hizume M., Kuzakawa Y. Fluorescent chromosome banding in *Picea*. II. Relationships between rDNA loci and chromomycin A_3-bands of *P. jezoensis* var. *hondoensis*. *La Kromosomo*. 1995;79-80:2754-2759

Hizume M., Kuzukawa Y., Kondo T. Physical mapping of 5S rDNA locus on chromosomes in *Pseudotsuga menziesii*, Pinaceae. *La Kromosomo*. 1996;83-84:2901-2908

Hizume M., Shibata F., Kondo T., Hoshi H., Kondo T., Ge S., Yang Q., Hong D. Identification of chromosomes in two Chinese spruce by multicolor fluorescence in hybridization. *Chromosome Sci*. 1999;3(1):37-41

Hizume M., Shibata F., Maruyama Y., Kondo T. Cloning of DNA sequences localized on proximal fluorescent chromosome bands by microdissection in *Pinus densiflora* Sieb. & Zucc. *Chromosoma.* 2001;110(5):345-351. doi 10.1007/s004120100149

Islam-Faridi M.N., Nelson C.D., Kubisiak T.L. Refererence karyotype and cytomolecular map for loblolly pine (*Pinus taeda* L.). *Genome*. 2007;50(2):241-251. doi 10.1139/g06-153

Ledig F.T., Hodgskiss P.D., Krutovskii K.V., Neale D.B., Eguiluz-Piedra T. Relationships among the spruces (*Picea*, Pinaceae) of Southwestern North America. *Systematic Botany*. 2004;299(2):275-295. doi 10.1600/036364404774195485

Li L.C., Wang G., Su S. Karyotype analysis of four species of *Picea* in Chengdu (Pinaceae). *Guihaia.* 2001;21(1):43-46 (in Chinese)

Liu B., Zhang S.-G., Zhang Y., Lan T.-Y., Qi L.-W., Song W.Q. Molecular cytogenetics analysis of four *Larix* species by bicolor fluorescence *in situ* hybridization and DAPI banding. *Int J Plant Sci.* 2006;167(2):367-372. doi 10.1086/499317

Liu B., Qi L.-W., Chen R., Song W.Q. Multicolor fluorescence *in situ* hybridization with combinatorial labeling probes enables a detailed karyotype analysis of *Larix principis-rupprechtii*. *Biol Res*. 2007;40(1):23-28. doi 10.4067/s0716-97602007000100003

Liu T.S. A new proposal for the classification of the genus *Picea*. *Acta Phytotax Geobot*. 1982;33(4-6):227-245. doi 10.18942/bunruichiri.KJ00001079164

Liu Y.H., Li M.X. Karyotype analysis of five species of genus *Picea*. *J Wuhan Bot Res.* 1985;3(3):203-207 (in Chinese)

Liu Z., Zhang D., Hong D., Wang X. Chromosomal localization of 5S and 18S-5.8S-25S ribosomal DNA sites in five Asian pines using fluorescence in situ hybridization. *Theor Appl Genet.* 2003;106(2):198-204. doi 10.1007/s00122-002-1024-z

Lockwood J.D., Aleksić J.M., Zou J., Wang J., Liu J., Renner S.S. A new phylogeny for the genus *Picea* from plastid, mitochondrial, and nuclear sequences. *Mol Phylogenet Evol*. 2013;69(3):717-727. doi 10.1016/j.ympev.2013.07.004

Lubaretz O., Fuchs J., Ahne R., Meister A., Schubert I. Karyotyping of three Pinaceae species via fluorescent *in situ* hybridization and computer-aided chromosome analysis. *Theor Appl Genet.* 1996;92(3-4):411-416. doi 10.1007/BF00223687

Mikkola L. Observations on interspecific sterility in *Picea*. *Ann Bot Fenn.* 1969;6(4):285-339

Muratova E.N., Sedel'nikova T.S., Goryachkina O.V., Pimenov A.V., Karpjuk T.V. Supernumerary chromosomes in coniferous plants: a case study of the *Picea* A. Dietr. genus. *Journal of Siberian Federal University. Biology.* 2023;16(4):403-419 (in Russian)

Nkongolo K.K. RAPD and cytological analysis of *Picea* spp. from different provenances: genomic relationship among taxa. *Hereditas*. 1999;130(2):137-144. doi 10.1111/j.1601-5223.1999.00137.x

Nkongolo K.K., Mehes-Smith M. Karyotype evolution in the Pinaceae: implication with molecular phylogeny. *Genome*. 2012;55(11):735-753. doi 10.1139/g2012-061

Ohri D. Karyotype evolution in conifers. *Feddes Repertorium*. 2021;132(3):232-262. doi 10.1002/fedr.202100014

Pedersen C., Lindle-Laursen I. Chromosomal locations of four minor rDNA loci and a marker microsatellite sequence in barley. *Chromosome Res*. 1994;2(1):65-71. doi 10.1007/BF01539456

Pikaard C.S. Nucleolar dominance: uniparental gene silencing on a multu-megabase scale in genetic hybrids. *Plant Mol Biol*. 2000;43(2-3):163-177. doi 10.1023/a:1006471009225

Popov P.P. Spruce in the East Europe and Western Siberia: Population-geographic variation and its silviculture value. Novosibirsk: Nauka Publ., 1999 (in Russian)

Pravdin L.F., Budaragin V.A., Kruklis M.V., Shershukova O.P. Methods of karyological investigation of conifers. *Lesovedenie*. 1972;2:67-75 (in Russian)

Puizina J., Sviben T., Krajačić-Sokol I., Zoldoš-Pećnik V., Siljak-Yakovlev S., Papeš D., Besendorfer V. Cytogenetic and molecular characterization of the *Abies alba* genome and its relationship with other members of the Pinaceae. *Plant Biol*. 2008;10(2):256-267. doi 10.1111/j.1438-8677.2007.00018.x

Ran J.-H., Wei X.-X., Wang X.-Q. Molecular phylogeny and biogeography of *Picea* (Pinaceae): implications for phylogeographical studies using cytoplasmic haplotypes. *Mol Phylogenet Evol.* 2006;41(2):405-419. doi 10.1016/j.ympev.2006.05.039

Ran J.-H., Shen T.-T., Liu W.-J., Wang P.-P., Wang X.-Q. Mitochondrial introgression and complex biogeographic history of the genus *Picea*. *Mol Phylogenet Evol*. 2015;93(1):63-76. doi 10.1016/j.ympev.2015.07.020

Raskina O., Belyayev A., Nevo E. Activity of *En*/*Spm*-like transposons in meiosis as a base for chromosome repatterning in a small. Isolated, peripheral populations of *Aegilops speltoides* Tausch. *Chromosome Res*. 2004;12(2):153-161. doi 10.1023/B:CHRO.0000013168.61359.43

Ribeiro T., Barão A., Viegas W., Morais-Cecíli L. Molecular cytogenetics of forest trees. *Cytogenet Genome Res*. 2008;120(3-4):220-227. doi 10.1159/000121070

Roa F., Guerra M. Distribution of 45S rDNA sites in chromosomes of plants: structural and evolutionary implications. *BMC Evolutionary Biology*. 2012;12:225. doi 10.1186/1471-2148-12-225

Rossello J.A., Maravilla A.J., Rosato M. The nuclear 35S rDNA world in plant systematics and evolution: a primer of cautions and common misconceptions in cytogenetic studies. *Front Plant Sci.* 2022;13:788911. doi 10.3389/fpls.2022.788911

Sarri V.V., Minelli S., Panara F.F., Morgante M., Jurman I., Zuccolo A.A., Cionini P.G. Characterization and chromosomal organization of satellite DNA sequences in *Picea abies*. *Genome*. 2008;51(9):705-713. doi 10.1139/G08-048

Schmidt P.A. Contributions to the taxonomy and evolution of the genus *Picea* A. Dietr. *Flora*. 1989;182(5-6):435-461. doi 10.1016/S0367-2530(17)30431-0

Shibata F., Hizume M. Comparative FISH karyotype analysis of 11 *Picea* species. *Cytologia.* 2008;73(2):203-211. doi 10.1508/cytologia.73.203

Shibata F., Hizume M., Garajova Z. Conserved FISH karyotype in four species of *Abies* (Pinaceae). *Chromosome Sci*. 2004;8(3):95-98

Shibata F., Matsusaki Y., Hizume M. A comparative analysis of multiprobe fluorescence *in situ* hybridization (FISH) karyotypes in 26 *Pinus* species (Pinaceae). *Cytologia*. 2016;81(4):409-421. doi 10.1508/cytologia.81.409

Sigurgeirsson A., Szmidt A.E. Phylogenetic and biogeographic implications of chloroplast DNA variation in *Picea*. *Nordic J Bot*. 1993; 13(3):233-246. doi 10.1111/j.1756-1051.1993.tb00043.x

Šiljak-Yakovlev S., Cerbah M., Coulaud J., Stoian V., Brown S.C., Zoldos V., Jelenić S., Papeš D. Nuclear DNA content, base composition, heterochromatin and rDNA in *Picea omorika* and *Picea abies*. *Theor Appl Genet.* 2002;104(2-3):505-512. doi 10.1007/s001220100755

Sukachev V.N. Forest Woody Species, Systematics, Geography and their Phytosociology. Vol. 1. Conifers. Moscow: Novaya Derevnya Publ., 1928 (in Russian)

Sun Y.-Sh., Abbott R.J., Li L.L., Li L., Zou J., Liu J.Q. Evolutionary history of Purple cone spruce (*Picea purpurea*) in the Qinghai–Tibet Plateau: homoploid hybrid origin and Pleistocene expansion. *Mol Ecol*. 2014;23(2):343-359. doi 10.1111/mec.12599

Trontin J.-F., Grandemange C., Favre J.-M. Two highly divergent 5S rDNA unit size classes occur in composite tandem array in European larch (*Larix decidua* Mill.) and Japanese larch (*Larix kaempferi* (Lamb.) Carr.). *Genome*. 1999;42(5):837-848. doi 10.1139/g99-015

Vischi M., Jurman J., Bianchi G., Morgante M. Karyotype of Norway spruce by multicolor FISH. *Theor Appl Genet*. 2003;107(4):591-597. doi 10.1007/s00122-003-1306-0

Wright J.W. Species crossability in spruce in relation to distribution and taxonomy. *For Sci.* 1955;1(4):319-340. doi 10.1093/forestscience/1.4.31

Zhang S.G., Yang W.H., Han S.Y., Han B.T., Li M.S., Qi L.W. Cytogenetic analysis of reciprocal hybrids and their parents between *Larix leptolepis* and *Larix gmelinii*: implications for identifying hybrids. *Tree Genet Genomes*. 2010;6(3):405-412. doi 10.1007/s11295-009-0258-1

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

Received November 28, 2025. Revised April 16, 2026. Accepted April 17, 2026.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):747-754

doi 10.18699/vjgb-26-75

Phylogeographic analysis of mitochondrial genome polymorphism in the Shors

B.A. Malyarchuk, G.A. Denisova, A.N. Litvinov, N.V. Pokhilyuk

Institute of Biological Problems of the North of the Far Eastern Branch of the Russian Academy of Sciences, Magadan, Russia
malbor@mail.ru

Abstract. The variability of nucleotide sequences of complete mitochondrial genomes in the Shors, a Turkic-speaking people from the Altai–Sayan region, was studied. The Shors exhibit a high level of nucleotide diversity due to the presence of two components of the East and West Eurasian origin in their gene pool (79.5 and 20.5 %, respectively). However, a relatively low level of mtDNA haplotype diversity was found among the Shors, which is associated with the high prevalence of certain identical mitochondrial haplotypes. Multidimensional scaling of interpopulation genetic distances indicates that the Shors fall in between East Asian ethnic groups and indigenous East Siberian peoples. By examining the polymorphism of whole mitochondrial genomes in individuals with the same hypervariable segment 1 haplotype belonging to haplogroup F1b1, found in ~37 % of the Shors, we identified six mtDNA haplotypes belonging to haplogroups F1b1-b1 and F1b1-b2. Phylogeographic analysis shows that the Shor mitochondrial gene pool is characterized by a high frequency (65.9 %) of mtDNA haplogroups found solely in the Altai–Sayan region: A-a1c1a, B4b1a3a1a, C4a1a4a1a, C4e, C5c*, D4b1a2a1a, D4m2a2, D5c1a2b, F1b1-b1, F1b1-b2, F2b1b1, and H8b1a. According to molecular dating results, the evolutionary age of these mtDNA haplogroups varies from 0.9 to 5.0 thousand years. All of the mtDNA haplogroups listed are of East Asian origin, except for the West Eurasian H8b1a. The isolated evolution of mitochondrial haplogroups specific to the peoples of the Altai–Sayan region is assumed to have been facilitated by their isolation from East Asian neighbors. This isolation occurred during the Bronze Age, when the Afanasyevo culture formed as a result of the mixing of indigenous populations and migrants – representatives of the Yamnaya culture, which existed in the Urals and Volga regions. In the absence of a significant influx of additional East Asian mtDNA haplotypes to the populations of the Altai–Sayan region over several thousand years, it appears that existing genetic lineages of East Asian origin diversified.

Key words: mitochondrial genome; phylogeographic analysis; human populations; Siberia; Shors

For citation: Malyarchuk B.A., Denisova G.A., Litvinov A.N., Pokhilyuk N.V. Phylogeographic analysis of mitochondrial genome polymorphism in the Shors. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(5):747-754. doi 10.18699/vjgb-26-75

Филогеографический анализ полиморфизма митохондриальных геномов шорцев

Б.А. Малярчук, Г.А. Денисова, А.Н. Литвинов, Н.В. Похилюк

Институт биологических проблем Севера Дальневосточного отделения Российской академии наук, Магадан, Россия
malbor@mail.ru

Аннотация. Получены данные об изменчивости нуклеотидных последовательностей целых митохондриальных геномов у шорцев – тюркоязычного народа Алтае-Саянского региона. У шорцев выявлен высокий уровень нуклеотидного разнообразия, обусловленный наличием в генофонде двух компонентов восточно- и западноевразийского происхождения (79.5 и 20.5 % соответственно). При этом отмечен относительно низкий (в масштабах Сибири) уровень разнообразия гаплотипов мтДНК, что связано с высокой распространенностью у шорцев некоторых одинаковых митохондриальных гаплотипов. Результаты многомерного шкалирования межпопуляционных генетических дистанций свидетельствуют о том, что шорцы занимают промежуточное положение между этническими группами Восточной Азии и коренными народами Восточной Сибири. Исследование полиморфизма целых митохондриальных геномов у индивидуумов, характеризующихся одинаковым гаплотипом гипервариабельного сегмента 1, который относится к гаплогруппе F1b1 и обнаружен у ~37 % шорцев, позволило определить 6 гаплотипов мтДНК, принадлежащих к гаплогруппам F1b1-b1 и F1b1-b2. Результаты филогеографического анализа показали, что особенностью митохондриального генофонда шорцев является высокая частота (65.9 %) гаплогрупп мтДНК, которые распространены исключительно в популяциях Алтае-Саянского региона: A-a1c1a, B4b1a3a1a, C4a1a4a1a, C4e, C5c*, D4b1a2a1a, D4m2a2, D5c1a2b, F1b1-b1,

F1b1-b2, F2b1b1, H8b1a. Эволюционный возраст этих гаплогрупп мтДНК согласно результатам молекулярного датирования варьирует от 0.9 до 5.0 тыс. лет. Все указанные выше гаплогруппы мтДНК восточноазиатские по происхождению, за исключением западноевразийской H8b1a. Обособленному развитию митохондриальных гаплогрупп, специфичных для народов Алтае-Саянского региона, возможно, способствовала изоляция от восточноазиатских соседей, вызванная формированием афанасьевского населения в эпоху бронзового века в результате смешения популяций коренных жителей региона и мигрантов – представителей ямной культурно-исторической общности, существовавшей в Приуралье и Поволжье. В условиях отсутствия интенсивного притока дополнительных восточноазиатских гаплотипов мтДНК в течение нескольких тысяч лет в популяциях Алтае-Саянского региона, по всей видимости, произошла диверсификация уже имевшихся генетических линий восточноазиатского происхождения.

Ключевые слова: митохондриальный геном; филогеографический анализ; популяции человека; Сибирь; шорцы

Introduction

Studies of mitochondrial DNA (mtDNA) polymorphisms have demonstrated that the gene pools of most ethnic groups in Western Siberia contain significant contributions from West Eurasian mtDNA lineages. This component is most prevalent (comprising 20–40 % of the gene pool) among modern populations in the Altai–Sayan region, including the Khakass, Shors, Teleuts, Altai-Kizhi, Telengits and Altai Kazakhs (Derenko et al., 2003; Derenko, Malyarchuk, 2010; Gubina et al., 2013). Paleogenomic studies have shown that the Neolithic population of Altai and neighboring regions primarily consisted of East Asian mtDNA lineages, although some West Eurasian mtDNA lineages, such as U2e1, were already present in the region approximately 7.5 thousand years ago (Wang K. et al., 2023; Gill et al., 2025). The influx of Eastern European populations into southwestern Siberia and Mongolian steppes during the Bronze Age (5.0–4.0 thousand years ago) led to a significant increase in the proportion of European genetic components in the gene pool of the ancient population (Allentoft et al., 2015). However, about 2.0 thousand years ago, the expansion of the Xiongnu caused a significant increase in the proportion of East Asian mtDNA haplotypes in the gene pools of the peoples of Central Asia and the Altai–Sayan region (Damgaard et al., 2018; Jeong et al., 2020).

The Shors are an indigenous people inhabiting Mountain Shoria in southern Kemerovo Oblast. They speak a language belonging to the Uyghur group of Turkic languages. According to ethnographic data, they are related to other ethnic groups of the Northern Altai: the Chelkans, Kumandins, and Tubalars. In anthropological terms, however, the Shors and Kumandins exhibit similarities to certain Ugric and Samoyedic peoples, such as the Khanty, Mansi, and Selkups, who are characterized by the Uralic anthropological type (Moiseev, 1999). The Shors appear to be Turkicized descendants of the Ugric-, Samoyedic-, and Ket-speaking populations of the northern taiga region of the Altai–Sayan Highlands (Peoples of Siberia, 1956). According to the 2021 Russian Federation census, there are 10,500 Shors.

An analysis of hypervariable segment 1 (HVS1) of the mitochondrial genome has shown that the Shors inhabiting the Altai and Sayan Mountains are genetically similar to other ethnic groups in the region, including the Khakass, Teleuts, Altai-Kizhi, Tuvinians, Todjins, and Tofalars (Derenko, Malyarchuk, 2010; Gubina et al., 2022). However, a phylogenetic analysis of HVS1 nucleotide sequences shows that the Shors are somewhat distant from the main cluster of Altai–Sayan populations (Derenko, Malyarchuk, 2010). This distance may be due to their distinct anthropological composition and their presumed kinship with Ugric and Samoyedic peoples. This hypothesis is supported by the presence of certain mitochondrial lineages (J1b1 and U4b1) in the mitochondrial gene pool of the Shors, which are characteristic of indigenous populations in West Siberia and the Volga-Ural region (Derenko, Malyarchuk, 2010).

An interesting feature of the mitochondrial gene pool of the Shors, as well as the related Khakass people, is the high prevalence of mtDNA haplotypes belonging to the East Asian haplogroup F (ranging from 15 to 40 %), although their frequencies among other Siberian ethnic groups are significantly lower (Derbeneva et al., 2002; Derenko et al., 2007; Gubina et al., 2013). The greatest diversity of mitochondrial lineages within haplogroup F is found in Southeast Asian populations (Hill et al., 2007).

The gene pools of the Shor and Khakas peoples are dominated by F1b mtDNA lineages, which originated in southern China (Hill et al., 2007). One of these F1b haplotypes, with the 16189-16232CA-16249-16304-16311 HVS1 motif, is widespread among the Shors (up to 37 %) and the Khakass (up to 26 %) (Derenko et al., 2007; Gubina et al., 2013). It is believed that this is due to the "founder effect" that occurred when these populations formed (Derenko, Malyarchuk, 2010). Due to the low resolution of mtDNA HVS1 polymorphism analysis, it is reasonable to employ a more informative method, such as analysis of nucleotide sequence variability in whole mitochondrial genomes. This will clarify the extent to which previously identified F1b mtDNA HVS1 sequences in the Shors are similar. Additionally, analyzing the polymorphism of the entire mitochondrial genome of the Shors at the population level would provide the most comprehensive understanding of their gene pool diversity and enable us to trace the phylogenetic relationships among mtDNA haplotypes using phylogeographic analysis. This work aims to study these issues.

Materials and methods

Whole blood samples were taken from 33 Shors chosen from residents of the Tashtagol District of the Kemerovo Region.

The Regional Ethics Committee for Medical and Biological Research under the Presidium of the Far Eastern Scientific Center of the Russian Academy of Sciences, Magadan, approved the study (Protocol No. 001/011, January 21, 2011).

We determined the nucleotide sequences of whole mitochondrial genomes by Sanger DNA sequencing on an ABI 3500xL genetic analyzer (Applied Biosystems), as in our previous work (Derenko, Malyarchuk, 2010). Amplification products were sequenced using the BrilliantDye™ Terminator Cycle Sequencing kit v3.1 (Netherlands). The SeqScape 2.7 software package (Applied Biosystems) was used for alignment and assembly of the mitochondrial genomes. The nucleotide sequences of the Shors' mitochondrial genomes have been deposited in GenBank (www.ncbi.nlm.nih.gov) under accession numbers PX641521–PX641539. The analysis also utilized 11 previously sequenced mitochondrial genomes from Shors in the Tashtagol District of the Kemerovo Region (Table S1)[1].

We performed phylogenetic analysis of mitochondrial genome nucleotide sequences and estimated the evolutionary age of mtDNA monophyletic clusters using the mtPhyl v4.015 software package (https://isogg.org/wiki/MtPhyl), as previously described (Malyarchuk et al., 2025). We used the phylogenetic classifications proposed by the PhyloTree (www.phylotree.org) and YFull MTree (https://www.yfull.com/mtree/) resources to determine mtDNA haplogroups.

Our phylogeographic analysis of mtDNA involved data on the variability of complete mitochondrial genomes from various human populations as reported in GenBank, the Logan DNA Project (http://www.ianlogan.co.uk), and YFull MTree. As of November 2025, GenBank contains over 62,500 mitochondrial genomes from individuals of different ethnic backgrounds worldwide (https://www.mitomap.org/foswiki/bin/view/MITOMAP/GBFreqInfo). The ethnic affiliation of the studied samples was determined based on information provided in the databases.

A comparative analysis of mtDNA variation at the interethnic level is based on 2,923 mitochondrial genomes from various East Asian populations. These include Han Chinese and Japanese (Zheng et al., 2011), Evenks, Yakuts, and Evens from Yakutia, Nivkhs, Udegeys, and Yukaghirs (Duggan et al., 2013), Koryaks and Evens from the Magadan region, Chukchi (Derenko et al., 2023), Buryats (Derenko et al., 2018), Mongols, Barghuts, and Khamnigans (Derenko et al., 2022), Uyghurs (Zheng et al., 2017), Tajiks and Kyrgyz (Peng et al., 2018), Daurs (Wang C.-Z. et al., 2022). We used the DnaSP 5.10.01 software package (Librado, Rozas, 2009) to calculate genetic diversity parameters in the populations. We performed an analysis of molecular variance (AMOVA, *F*st analysis) based on pairwise nucleotide differences between mitogenomes using the Arlequin 3.5.1.2 software package (Excoffier, Lischer, 2010). We used the multidimensional scaling method of interpopulation *F*st differences (STATISTICA 10 software package, StatSoft Inc.) to investigate the distribution of populations in two-dimensional space. We calculated interpopulation genetic distances (D_A и D_{XY}) using the equation

$$D_A = D_{XY} - (D_X + D_Y)/2,$$

where D_{XY} is the average number of nucleotide differences between populations X and Y; D_X and D_Y are the average numbers of nucleotide differences within populations X and Y, respectively (Nei, 1987).

[1] Supplementary Tables S1, S2 and Figure S1 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Mal_Engl_30_5.xlsx

Results and discussion

Diversity of the mitochondrial gene pool among the Shor people and genetic differentiation of East Asian populations

Analysis of mitochondrial genome nucleotide sequence polymorphisms encompassing the Shors (33 new samples and 11 previously published ones) and other ethnic groups in Siberia, Central Asia, and East Asia showed that the mitochondrial haplotypes belong to 23 haplogroups (Table 1 and Table S1).

Compared to other East Asian populations, the Shor sample studied is characterized by relatively low haplotype diversity (H = 0.979), comparable to data from populations in Northeast Siberia, such as the Chukchi, Koryaks, Yukaghirs, and Nivkhs (Derenko et al., 2023). This is apparently due to the high prevalence of identical haplotypes among the Shors, which mainly belong to haplogroup F1b1 (Table 1). At the same time, the nucleotide diversity among the Shors is at its maximum level, with $\pi = 0.00236$ and an average number of pairwise nucleotide differences of $k = 39.01$. Previous studies have shown that similar nucleotide diversity values are found among the Daurs, Kyrgyz, and Khamnigans (Derenko et al., 2023).

An analysis of genetic differences based on the number of nucleotide substitutions per site between populations (D_{XY} and D_A) showed the least genetic differences of Shors from the Han, Uyghurs, Mongols, and Kyrgyz (Table 2). The results of *F*st-analysis also demonstrated minimal differences between the Shors and the Mongols, Han, Kyrgyz, and Uyghurs (*F*st values of 5.5–6.0 %), while *F*st differences between the Shors and Paleo-Asian populations (Chukchi, Koryaks, and Nivkhs) exceeded 20 % (Table S2).

Our analysis of interpopulation *F*st differences conducted using multidimensional scaling showed that the Shors are situated between the compactly clustered East Asian ethnic groups and the scattered Eastern Siberian populations, as illustrated in Figure 1 in a two-dimensional genetic space. In the latter group, only the Evens, Evenki, Koryaks, and Yukaghirs show a tendency toward clustering. Unfortunately, the Samoyedic and Ugric groups are absent from the populations studied. This is due to the lack of population-level data on whole mitochondrial genome variability for many ethnic groups in Siberia and Eastern Europe. Therefore, a more detailed study of the origins of the Shors – for instance, based on the mixing of Samoyedic and Ugric tribes with Turkic-speaking groups –

Table 1. mtDNA haplogroups of the Shors, their evolutionary age, and geographic distribution

mtDNA haplogroup	*n*	Age, thousands of years	Distribution in Altai	Distribution in other regions
A-a1c1a	1	1.29	Shors, Kumandins, Tubalars	None
B4b1a3a1a	1	–	Shors, Kumandins, Chelkans	
C4a1a4a1a	1	1.72	Shors, Teleuts	
C4a2a1	4	–	Shors	Yakuts, Evenki
C4e	2	1.72	Shors, Teleuts	None
C5c*	1	4.98	Shors, Teleuts, Kumandins, Chelkans, Tubalars	
D4b1a2a1a	1	3.24	Shors, Kumandins, Chelkans, Tubalars	
D4j7	1	4.16	Shors, Teleuts, Kumandins, Tubalars	Buryats, Barghuts
D4m2a2	1	0.86	Shors, Chelkans, Tubalars	None
D5c1a2b	2	2.06	Shors, Kumandins, Chelkans, Tubalars	
F1b1-b1	6	0.86	Shors, Chelkans	
F1b1-b2	9	0.86	Shors	
F2b1b1	1	0.86	Shors, Chelkans, Tubalars, Altaians	
G3a1'2	1	19.01	Shors	East Asia
M10a1-a1	2	4.6		Mongols, Daurs
Z4a	1	–		East Asia
H1b2c	1	–		Europe (Russians, Poles)
H6a1b	1	–		Europe, West Asia
H8b1a	3	4.22	Shors, Tubalars, Teleuts, Altaians	None
J1b1a1	1	–	Shors	West Eurasia
K1-a4	1	–		
U4b1b1-a	1	3.54	Shors, Kumandins, Teleuts, Altaians	West Eurasia (Tatars, Mansi, Persians)
U5b2b4c	1	1.29	Shors	Tuvinians

Note. *n* – the number of samples belonging to the haplogroup. Dashes (–) indicate cases where the evolutionary age cannot be determined because of data shortage.

Table 2. Interpopulation genetic distances (D_{XY} and D_A are numbers of substitutions per site between populations) between the Shors and other East Asian ethnic groups

Ethnic group	D_{XY}	D_A	Ethnic group	D_{XY}	D_A
Chinese (Han)	0.0022	0.00012	Tajiks	0.00247	0.00024
Uighurs	0.00228	0.00013	Udege	0.00223	0.00028
Mongols	0.00236	0.00013	Evenki	0.00239	0.00035
Kyrgyz	0.00243	0.00013	Evens	0.0024	0.00039
Daurs	0.00245	0.00016	Koryaks	0.00241	0.00047
Japanese	0.00218	0.00017	Nivkhs	0.00225	0.00048
Khamnigans	0.00244	0.00018	Yukaghirs	0.00244	0.00051
Buryats	0.00238	0.00019	Chukchi	0.00245	0.00074
Barghuts	0.00244	0.00021			

Note. The distance values are ranked by D_A.

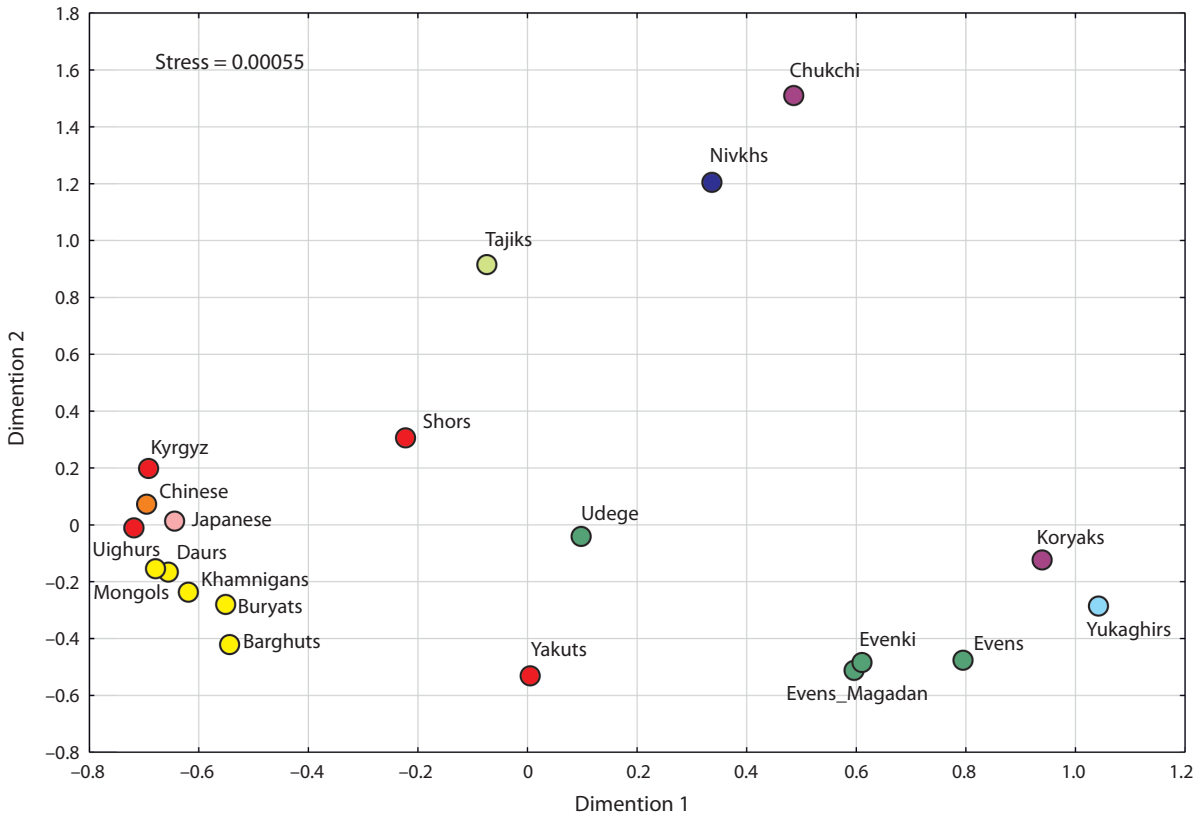


Fig. 1. Results of multidimensional scaling of interpopulation Fst values based on pairwise nucleotide differences between whole mitochondrial genomes in populations from Siberia, East Asia, and Central Asia.

The colors indicate the linguistic affiliation of the populations: red for the Turkic group of the Altaic family; yellow for the Mongolic group of the Altaic family; green for the Tungusic group of the Altaic family; purple for the Chukotko-Kamchatkan family; orange for the Sino-Tibetan family; pale pink for the Japanese language; light yellow for the Indo-Iranian group of the Indo-European family (Tajiks), and other colors for linguistic isolates (Nivkhs and Yukaghirs).

remains to be conducted after databases on mtDNA variation among indigenous Siberian and neighboring populations are expanded.

Diversity of F1b1 haplotypes in the Shors

Previous studies have shown that 30 of the 82 Shor individuals examined (36.6 % of the sample) exhibit the HVS1 haplotype 16189-16232CA-16249-16304-16311, belonging to haplogroup F1b1 (Derenko et al., 2007). In this study, we sequenced the entire mitochondrial genomes of 15 Shor individuals carrying this haplotype. Our study demonstrated that the HVS1 haplotype belongs to two haplogroups: F1b1-b1 and F1b1-b2. The former has already been recorded among the Chelkans and Khakass, according to the YFull MTree database. The latter is a new haplogroup, which we have identified only among the Shors (Fig. 2). The evolutionary age of both haplogroups is 0.86 thousand years. However, if the age is estimated without considering identical samples, i. e., based solely on haplotypes, it increases to 1.72 (0–4.13) thousand years for both.

At the same time, haplogroup F1b1-b is approximately 7.7 thousand years old, and it includes mitochondrial lineages found among Japanese and Chinese peoples (Fig. 2). This suggests that F1b1-b mtDNA lineages were introduced into the Altai peoples' gene pools from East Asia quite some time ago. Conversely, the high prevalence of a few F1b1-b mtDNA haplotypes in the Shor population could be due to a founder effect in their demographic history.

The high population specificity of Altai mtDNA variants

The results of the phylogeographic analysis show that the mitochondrial gene pool of the Shor sample is predominantly composed of East Asian haplotypes (79.5 %) (Table 1). The West Eurasian mitochondrial lineages identified in the Shors are represented by single haplotypes (H1b2c, H6a1b, J1b1a1, and K1-a4) or subclades (U4b1b1-a and U5b2b4c) of relatively recent evolutionary origin (within the last 3.5 thousand years). However, some subclades of West Eurasian mtDNA haplotypes evolved over a relatively long period within the gene pools of Siberian peoples (Table 1, Table S1). One example is haplogroup U4b1b1-a, which is primarily found in the Altai region but has also been observed in populations in the Volga region and Iran. All other U4b1b1 subgroups have been identified in European populations. Haplogroup U5b2b4 is primarily found in European populations, but its subgroup, U5b2b4c, has been identified among the Shors and Tuvinians. Haplogroup H8b

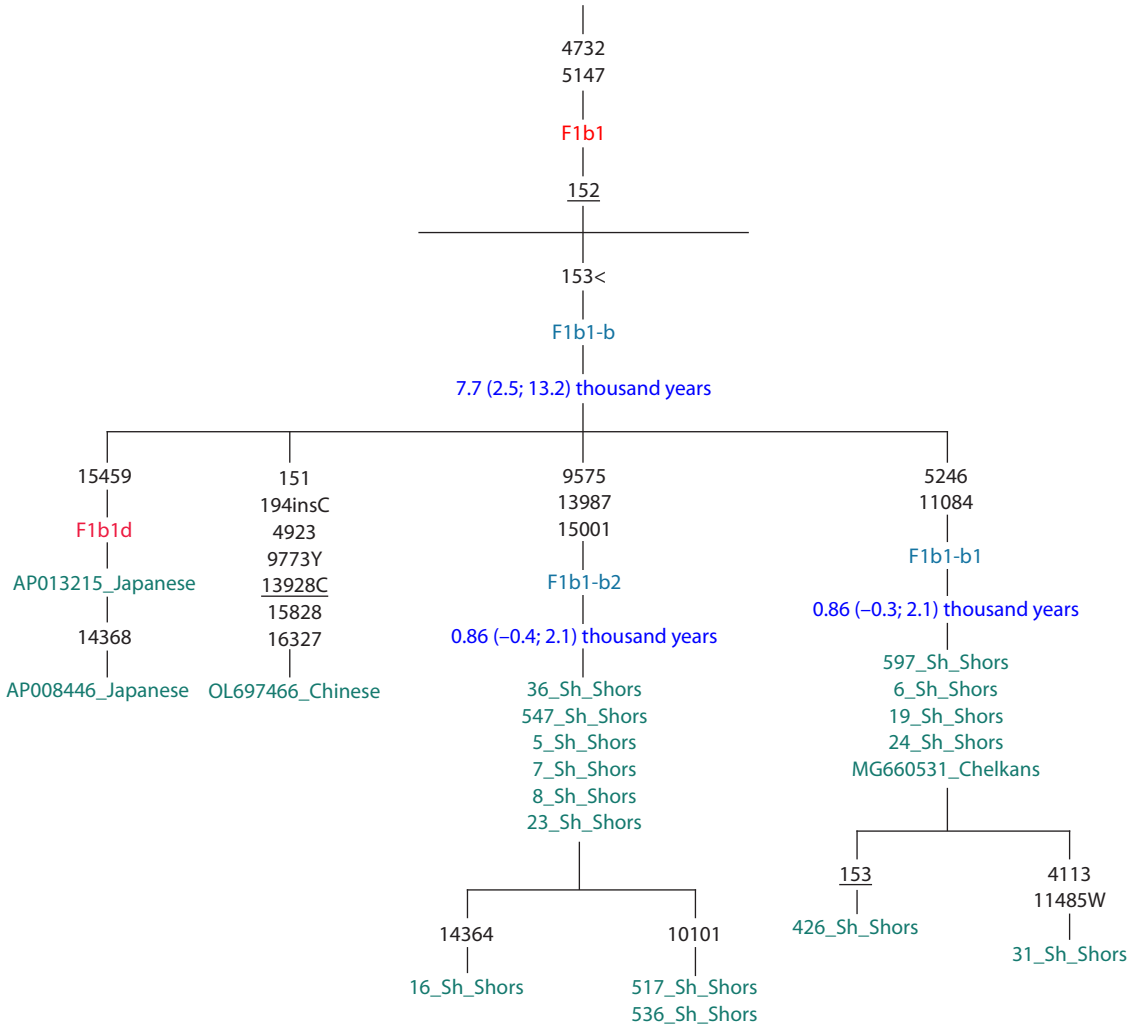


Fig. 2. Phylogenetic tree of mtDNA haplogroup F1b1-b.

The branches indicate the nucleotide positions where transitions occurred. For transversions and cases of heteroplasmy, the nucleotides are labeled. "Ins" denotes a nucleotide insertion. Nucleotide positions where reverse mutations were recorded are underlined, and recurrent mutations are marked with the symbol "<". GenBank accession numbers and ethnic affiliation are shown.

is of particular interest because its distribution is associated with Western Eurasia (subgroups H8b2 and H8b3). However, the most diverse subgroup is H8b1, which has been recorded in various Siberian and Central Asian populations. Its evolutionary age is estimated to be between 8.0 and 9.5 thousand years old (Fig. S1) (Derenko et al., 2014), which points to a long-standing presence of the H8b1 haplogroup in eastern North Eurasia. The H8b1a subgroup is primarily found in the Altai region.

A distinctive feature of the mitochondrial gene pool of the Shors is the high frequency (65.9 %) of mtDNA haplogroups that are found exclusively in the Altai region: A-a1c1a, B4b1a3a1a, C4a1a4a1a, C4e, C5c*, D4b1a2a1a, D4m2a2, D5c1a2b, F1b1-b1, F1b1-b2, F2b1b1, and H8b1a. According to molecular dating results, their evolutionary ages range from 0.86 to 4.98 thousand years (Table 1, Fig. S1). All of the aforementioned mtDNA haplogroups are of East Asian origin, except for H8b1a.

Conclusion

Archaeological, anthropological, and paleogenomic data show that the emergence of the Afanasyevo archaeological culture in the Altai and the steppes of the Minusinsk Basin during the Bronze Age (approximately about 5.0 thousand years ago) was associated with the migration of the Yamnaya culture population from the Volga-Ural region (Allentoft et al., 2015; Jeong et al., 2020). This population was characterized by the Proto-European anthropological type and genetic lineages of West Eurasian origin. This explains the presence of a wide variety of West Eurasian mtDNA haplotypes among the populations of the Altai–Sayan region. It is of particular interest, though, that the component of the Shor mitochondrial gene pool specific to Altai peoples is mainly represented by East Asian mitochondrial lineages. Their evolutionary age, approximately up to 5.0 thousand years ago, falls within the period of isolation from East Asian neighbors' gene pools caused by the formation of the mixed Afanasyevo population.

In the absence of a significant influx of additional East Asian haplotypes over the course of several thousand years, it appears that the existing genetic lineages of East Asian origin began to diversify within the populations of the Altai–Sayan region. The isolated development of mitochondrial haplogroups may have also been facilitated by the local natural features, as the Altai–Sayan mountain range could have served as a refuge for preserving "relict" gene pools and as a regulator of human migration patterns (Balanovska et al., 2014).

The study of the variability in full-length mtDNA nucleotide sequences demonstrates that the mitochondrial gene pool of the Shors is characterized by a high frequency (65.9 %) of a genetic component specific to ethnic groups inhabiting Altai, such as the Kumandins, Chelkans, Tubalars, and Teleuts. This component is primarily represented by mitochondrial lineages of East Asian origin. These lineages diversified during the formation of a mixed population based on the indigenous Altai populations and migrants from the Yamnaya cultural-historical community, which existed in the Ural and Volga regions.

References

Allentoft M.E., Sikora M., Sjögren K.G., Rasmussen S., Rasmussen M., Stenderup J., Damgaard P.B., … Sicheritz-Pontén T., Brunak S., Nielsen R., Kristiansen K., Willerslev E. Population genomics of Bronze Age Eurasia. *Nature.* 2015;522(7555):167-172. doi 10.1038/nature14507

Balanovska E.V., Balaganskaya O.A., Damba L.D., Dibirova Kh.D., Agdzhoyan A.T., Bogunov Yu.V., Zhabagin M.K., Isakova J.T., Lavryashina M.B., Balanovsky O.P. Impact of the environment on the gene pool of Turkic-speaking populations of mountains and steppes in Altay, Sayan, Tien Shan and Pamir. *Lomonosov J Anthropol.* 2014;2:46-55 (in Russian)

Damgaard P.B., Marchi N., Rasmussen S., Peyrot M., Renaud G., Korneliussen T., Moreno-Mayar J.V., … Nielsen R., Sikora M., Heyer E., Kristiansen K., Willerslev E. 137 ancient human genomes from across the Eurasian steppes. *Nature.* 2018;557(7705):369-374. doi 10.1038/s41586-018-0094-2

Derbeneva O.A., Starikovskaya E.B., Volodko N.V., Wallace D.C., Sukernik R.I. Mitochondrial DNA variation in the Kets and Nganasans and its implications for the initial peopling of Northern Eurasia. *Russ J Genet.* 2002;38:1316-1321. doi 10.1023/A:1021111530654

Derenko M.V., Malyarchuk B.A. Molecular phylogeography of populations of Northern Eurasia based on mitochondrial DNA variability data. Magadan, 2010 (in Russian)

Derenko M.V., Grzybowski T., Malyarchuk B.A., Dambueva I.K., Denisova G.A., Czarny J., Dorzhu Ch.M., Kakpakov V.T., Miścicka-Sliwka D., Woźniak M., Zakharov I.A. Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet.* 2003;67(Pt. 5):391-411. doi 10.1046/j.1469-1809.2003.00035.x

Derenko M., Malyarchuk B., Grzybowski T., Denisova G.A., Dambueva I., Perkova M., Dorzhu C., Luzina F., Lee H.K., Vanecek T., Villems R., Zakharov I. Phylogeographic analysis of mitochondrial DNA in North Asian populations. *Am J Hum Genet.* 2007;81:1025-1041. doi 10.1086/522933

Derenko M., Malyarchuk B., Denisova G., Perkova M., Litvinov A., Grzybowski T., Dambueva I., Skonieczna K., Rogalla U., Tsybovsky I., Zakharov I. Western Eurasian ancestry in modern Siberians based on mitogenomic data. *BMC Evol Biol.* 2014;14:217. doi 10.1186/s12862-014-0217-9

Derenko M., Denisova G., Malyarchuk B., Dambueva I., Bazarov B. Mitogenomic diversity and differentiation of the Buryats. *J Hum Genet.* 2018;63(1):71-81. doi 10.1038/s10038-017-0370-2

Derenko M., Denisova G., Dambueva I., Malyarchuk B., Bazarov B. Mitogenomics of modern Mongolic-speaking populations. *Mol Genet Genomics.* 2022;297(1):47-62. doi 10.1007/s00438-021-01830-w

Derenko M., Denisova G., Litvinov A., Dambueva I., Malyarchuk B. Mitogenomics of the Koryaks and Evens of the northern coast of the Sea of Okhotsk. *J Hum Genet.* 2023;68(10):705-712. doi 10.1038/s10038-023-01173-x

Duggan A.T., Whitten M., Wiebe V., Crawford M., Butthof A., Spitsyn V., Makarov S., Novgorodov I., Osakovsky V., Pakendorf B. Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers. *PLoS One*. 2013;8(12):e83570. doi 10.1371/journal.pone.0083570

Excoffier L., Lischer H.E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10(3):564-567. doi 10.1111/j.1755-0998.2010.02847.x

Gill H., Seidualy M., Lee J., Lee J., Moon H., Walter A., Bianco R.A., … Hermes T., Zhabagin M., Doumani Dupuy P., Warinner C., Jeong C. Ancient genomes from eastern Kazakhstan reveal dynamic genetic legacy of Inner Eurasian hunter-gatherers. *Sci Adv.* 2025;11(42):eadw8219. doi 10.1126/sciadv.adw8219

Gubina M.A., Damba L.D., Babenko V.N., Romaschenko A.G., Voevoda M.I. Haplotype diversity in mtDNA and Y-chromosome in populations of Altai-Sayan region. *Russ J Genet.* 2013;49:329-343. doi 10.1134/S1022795412120034

Gubina M.A., Babenko V.N., Batsevich V.A., Leibova N.A., Zabiyako A.P. Polymorphism of mitochondrial DNA and six nuclear genes in the Amur Evenk population. *Russ J Genet.* 2022;58:42-56. doi 10.1134/S1022795422010033

Hill C., Soares P., Mormina M., Macaulay V., Clarke D., Blumbach P.B., Vizuete-Forster M., Forster P., Bulbeck D., Oppenheimer S., Richards M. A mitochondrial stratigraphy for island southeast Asia. *Am J Hum Genet.* 2007;80(1):29-43. doi 10.1086/510412

Jeong C., Wang K., Wilkin S., Taylor W.T.T., Miller B.K., Bemmann J.H., Stahl R., … Krause J., Boivin N., Erdene M., Hendy J., Warinner C. A Dynamic 6,000-year genetic history of Eurasia's Eastern steppe. *Cell.* 2020;183(4):890-904.e29. doi 10.1016/j.cell.2020.10.015

Librado P., Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009;25(11):1451-1452. doi 10.1093/bioinformatics/btp187

Malyarchuk B.A., Denisova G.A., Litvinov A.N. Mitochondrial genome polymorphism in the East Slavic population of Northeastern Siberia. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2025;29(5):704-710. doi 10.18699/vjgb-25-77

Moiseyev V.G. Origin of Uralic-speaking peoples based on craniological data. St. Petersburg: Nauka Publ., 1999 (in Russian)

Nei M. Molecular evolutionary genetics. NY: Columbia University Press, 1987

Peng M.S., Xu W., Song J.J., Chen X., Sulaiman X., Cai L., Liu H.Q., … Yang X.Y., Rakha A., Yao Y.G., Upur H., Zhang Y.P. Mitochondrial genomes uncover the maternal history of the Pamir populations. *Eur J Hum Genet.* 2018;26(1):124-136. doi 10.1038/s41431-017-0028-8

Peoples of Siberia. Eds. M.G. Levin, L.P. Potapov. Moscow; Leningrad: AN SSSR Publ., 1956 (in Russian)

Wang C.-Z., Yu X.-E., Shi M.-S., Li H., Ma S.-H. Whole mitochondrial genome analysis of the Daur ethnic minority from Hulunbuir in

the Inner Mongolia Autonomous Region of China. *BMC Ecol Evol.* 2022;22:66. doi 10.1186/s12862-022-02019-4

Wang K., Yu H., Radzevičiūtė R., Kiryushin Y.F., Tishkin A.A., Frolov Y.V., Stepanova N.F., … Buzhilova A., Slon V., Jeong C., Krause J., Posth C. Middle Holocene Siberian genomes reveal highly connected gene pools throughout North Asia. *Curr Biol.* 2023;33(3): 423-433.e5. doi 10.1016/j.cub.2022.11.062

Zheng H.X., Yan S., Qin Z.D., Wang Y., Tan J.Z., Li H., Jin L. Major population expansion of East Asians began before Neolithic time: evidence of mtDNA genomes. *PLoS One.* 2011;6(10):e25835. doi 10.1371/journal.pone.0025835

Zheng H.X., Li L., Jiang X.Y., Yan S., Qin Z., Wang X., Jin L. MtDNA genomes reveal a relaxation of selective constraints in low-BMI individuals in a Uyghur population. *Hum Genet*. 2017;136(10):1353-1362. doi 10.1007/s00439-017-1829-0

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

Received November 28, 2025. Revised March 21, 2026. Accepted April 6, 2026.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):755-760

doi 10.18699/vjgb-26-76

Evaluation of genetic parameters affecting the reliability and effectiveness of kinship reconstruction in Forest and Tundra Nenets populations using X-STR markers

K.V. Vagaytseva, N.A. Kolesnikov, O.M. Burenkova, I.A. Volkova, O.D. Ruzavina, V.N. Kharkov, V.A. Stepanov

Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia
kseniya.simonova@medgenetics.ru

Abstract. X-chromosomal STR markers (X-STR) are informative tools for kinship reconstruction, particularly in cases where standard autosomal panels have limited discriminatory power: when reconstructing kinship between full sisters, paternal half-sisters, paternal grandmother and granddaughter, maternal aunt and niece. However, the use of X-STR markers in forensic genetics requires consideration of linkage disequilibrium structure and the use of haplotype frequencies in calculating the likelihood ratio (LR). In this work, the genetic structure of Forest and Tundra Nenets populations was assessed using 16 X-STR markers in order to obtain reference values. A sample of 504 Nenets males was analyzed, divided into subethnic groups (Forest and Tundra Nenets), phratries of Tundra Nenets (Haryuchi and Vanuito), as well as geographical local groups (individuals living in the Antipayutinskaya, Gydan and Nakhodka tundras). No statistically significant differences were found between the samples, and the pairwise Fst values (exact test) did not exceed 0.01. The Nenets population demonstrated an intermediate level of genetic diversity typical of Indigenous Siberian populations. The loci DXS8377 (He = 0.910) and DXS10079 (He = 0.839) were found to be the most polymorphic in the Nenets population. Of the 16 markers, 13 were highly informative (PIC > 0.5). The combined power of exclusion (CPE) was 0.99996, which indicates that the panel is highly informative for reconstructing kinship relationships. Pairwise analysis of linkage disequilibrium (LD) revealed statistically significant deviations for 46 pairs of markers out of 120; however, the r^2 coefficient for all pairs of alleles did not exceed 0.09, which indicates a weak correlation between alleles. This study provides the reference data of allele and genotype frequencies recommended for use in assessing the likelihood ratio in the reconstruction of kinship relationships. The obtained reference data will contribute to improving the reliability and efficiency of forensic genetic analyses in the Russian Federation.

Key words: Nenets; X-STR; forensic genetic; kinship testing; linkage disequilibrium

For citation: Vagaytseva K.V., Kolesnikov N.A., Burenkova O.M., Volkova I.A., Ruzavina O.D., Kharkov V.N., Stepanov V.A. Evaluation of genetic parameters affecting the reliability and effectiveness of kinship reconstruction in Forest and Tundra Nenets populations using X-STR markers. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(5):755-760. doi 10.18699/vjgb-26-76

Funding. This work was funded by the Russian Science Foundation (project No. 23-74-10058, https://rscf.ru/project/23-74-10058/).

Оценка генетических показателей, влияющих на надежность и эффективность реконструкции родословной в популяциях лесных и тундровых ненцев, при использовании X-STR маркеров

К.В. Вагайцева, Н.А. Колесников, О.М. Буренкова, И.А. Волкова, О.Д. Рузавина, В.Н. Харьков, В.А. Степанов

Научно-исследовательский институт медицинской генетики Томского национального исследовательского медицинского центра Российской академии наук, Томск, Россия
kseniya.simonova@medgenetics.ru

Аннотация. X-хромосомные STR-маркеры (X-STR) являются информативным инструментом для реконструкции родственных связей, особенно в случаях, когда стандартные аутосомные панели недостаточно эффективны: при реконструкции родственных отношений между полнородными сестрами, сводными по отцу сестрами, бабушкой по отцовской линии и внучкой, тетей по материнской линии и племянницей. Однако использование

X-STR маркеров в ДНК-идентификации требует учета структуры неравновесия по сцеплению и использования частот гаплотипов при расчете отношения правдоподобия (likelihood ratio, LR). В настоящей работе проведена оценка генетической структуры популяций лесных и тундровых ненцев по 16 X-STR маркерам, необходимая для получения референсных значений. Исследована выборка из 504 мужчин ненцев, подразделенная на субэтнические группы (лесные и тундровые ненцы), фратрии тундровых ненцев (Харючи и Вануйто), а также географические локальные группы (индивиды, проживающие в Антипаютинской, Гыданской и Находкинской тундрах). Статистически значимые различия между выборками не обнаружены, попарные значения Fst (exact test) не превысили 0.01. Показан средний уровень генетического разнообразия, характерный для популяций коренных народов Сибири. Наиболее полиморфными в популяции ненцев оказались локусы DXS8377 (He = 0.910) и DXS10079 (He = 0.839). Тринадцать из 16 маркеров были высокоинформативными (PIC > 0.5). Суммарная исключающая способность (CPE) составила 0.99996, что указывает на высокую информативность набора при реконструкции родственных отношений. При попарном анализе структуры неравновесия по сцеплению (LD) статистически значимые отклонения выявлены для 46 из 120 пар маркеров, однако коэффициент r^2 для всех пар аллелей не превысил 0.09, что свидетельствует о слабой силе корреляции между аллелями. В статье приведены референсные значения частот аллелей и генотипов, рекомендуемые для использования в оценке отношения правдоподобия при реконструкции родственных связей. Полученные параметры позволят повысить надежность и эффективность генетической экспертизы на территории Российской Федерации.

Ключевые слова: ненцы; X-STR; ДНК-идентификация; реконструкция родственных связей; структура неравновесия по сцеплению

Introduction

Kinship reconstruction based on the analysis of genetic markers is a key task in forensic genetics. A standard set of autosomal STR markers is a highly effective tool for reconstructing first-degree relationships when biological samples from at least three individuals who are presumably related are available for analysis (for example, a mother, a child, and a presumed father, the so-called "trio"). At the same time, the reliability of this approach is lower in more complex cases due to overlapping ranges of likelihood ratio (LR) values for true relatives and unrelated individuals (Ge, Budowle, 2020).

To increase the informativeness in such cases, it is common to use additional genetic markers: STR, SNP, and InDEL. Markers localized on the X chromosome are particularly informative for some variants of pedigree reconstruction. Due to the specific inheritance pattern of the X chromosome and hemizygosity in males, X-chromosomal markers can be used to establish a relationship between an alleged father and daughter even when the biological sample of the man himself is unavailable; in such cases, a sample from the man's mother may be included in the forensic genetic examination. It is shown that the inclusion of 12 X-STR markers in the analysis leads to an increase in the effectiveness of forensic genetic examination for the following types of kinship pairs: full sisters, paternal half-sisters, paternal grandmother–granddaughter, as well as maternal aunt–niece, maternal aunt–nephew. For other kinship categories, the effect of including X-STR is comparable to the inclusion of additional autosomal STR markers, and in some cases (for example, maternal uncle–nephew) it is inferior to it (Zhang et al., 2021a). At the same time, the use of X-chromosome markers in the practice of forensic genetics leads to a complication of statistical calculations due to the linkage between loci.

According to the recommendations of the International Society for Forensic Genetics (ISFG), when marker sets containing linked loci are used, haplotype frequencies should be applied in probabilistic assessments, and the structure of linkage disequilibrium should be taken into account (Tillmar et al., 2017). However, neither of these parameters is universal: they vary depending on the genetic and demographic history of the population. Therefore, before introducing any set of X-STR markers into forensic practice for a specific population, it is necessary to conduct population genetic studies.

The aim of the present study is to analyze the genetic structure of the Forest and Tundra Nenets populations using 16 X-STR markers to form a reference base necessary to increase the effectiveness of kinship reconstruction in the practice of forensic genetics.

This study was approved by the Biomedical Ethics Committee of the Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences (Protocol No. 19 of October 13, 2023).

Materials and methods

The study was based on a sample of Nenets males (760 individuals) from the biobank "Biobank of the Population of Northern Eurasia" of the Research Institute of Medical Genetics. Individuals identified as close relatives (father and son, brothers, uncle and nephew) were excluded from this sample. Kinship, population affiliation, place of birth, and place of residence were established based on a questionnaire. After filtering, the sample size was reduced to 504 individuals. The final sample comprises several subsamples: Forest Nenets (46 samples) and Tundra Nenets (458 samples); Tundra Nenets of the Haryuchi phratry (288 samples) and Tundra Nenets of the Vanuito phratry (158 samples); Nenets living in the Antipayutinskaya (107 samples), Gydan (56 samples), and Nakhodka (59 samples) tundras. All samples originate from the Yamalo-Nenets Autonomous Okrug. Venous blood samples were collected after obtaining written informed consent from the donors.

Genotyping of samples using X-STR markers was performed by multiplex polymerase chain reaction followed by analysis of the products using a Nanophore 05 genetic analyzer (Syntol, Russia). The primer sequences and marker information are presented in Y. Zhang et al. (2021b). The PCR reaction

was carried out using a reaction mixture and an activator from Gordiz (Russia). The PCR mixture per sample included 2 μL of activator, 2 μL of reaction mixture, 1 μL of DNA sample with a concentration of more than 10 ng/μL, and 5 μL of deionized water. The polymerase chain reaction was performed under the following conditions: 95 °C for 5 min; 36 cycles of 95 °C for 30 s, 60 °C for 45 s, and 72 °C for 45 s; final elongation: 72 °C for 4 min. Fluorescent peaks were evaluated using the GeneMarker software. Allele assignment was performed by comparing fragment sizes with those of the 9947DNA control sample (Applied Biosystems, USA) and the MK01 control sample from Gordiz (Russia). The experimental studies were conducted at the Center for the Collective Use of Scientific Research Equipment "Medical Genomics" (Research Institute of Medical Genetics, Tomsk National Research Medical Center).

Polymorphism information content (PIC), expected heterozygosity (He), and power of exclusion (PE) were calculated using the ChrX-STR.org online calculator (ChrX-STR.org database, accessed January 10, 2026). The combined power of exclusion (CPE) for the entire set of markers (SoftGenetics LLC, 2012), the effective number of alleles (Ne), and allele frequencies were calculated in a standard manner using the MS Excel software package. Analysis of molecular variance (AMOVA), evaluation of pairwise Fst and Rst values, assessment of linkage disequilibrium (LD) structure, and estimation of haplotype frequencies were performed using the Arlequin v.3.5 software tool (Excoffier, Lischer, 2010; Gusmão et al., 2025). Principal component analysis (PCA) was performed in the R environment using the FactoMineR (PCA function) and factoextra visualization packages (Lê et al., 2008).

Results

Genetic differentiation of populations

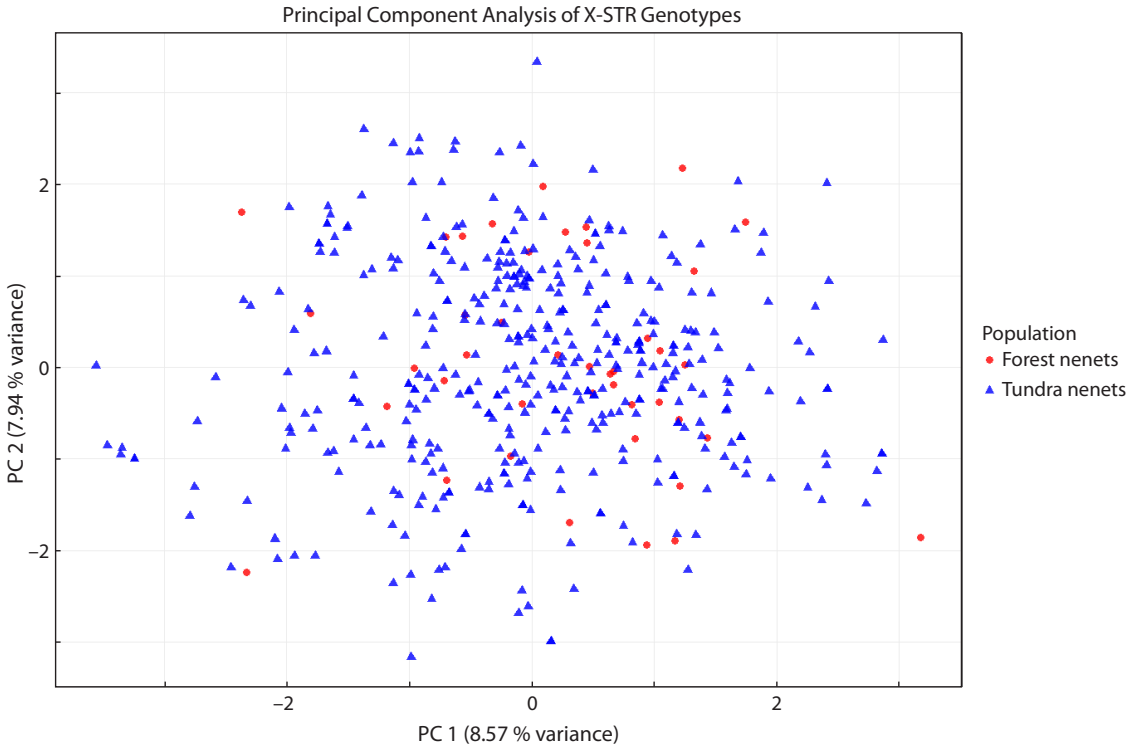
The first stage of the study was to evaluate the possibility of pooling the Nenets samples into a metapopulation. To this end, a principal component analysis (PCA) plot was constructed to visualize the general genetic relationships between individuals (see the Figure).

The distribution of individuals in the space of the first two principal components (which together account for 16.51 % of the total variability) does not demonstrate the formation of isolated clusters corresponding to the sample of Forest or Tundra Nenets. We used three indicators to evaluate genetic differentiation using F-statistics. Two of them characterize the strength of differentiation (Fst, $\sigma\mu^2$), and one (exact test) characterizes the statistical significance of differentiation (Table 1).

The pairwise Fst values in all comparisons are very low (from 0.001 to 0.010), which indicates weak genetic subdivision of the populations, explained by intensive gene flow between populations and a relatively recent divergence time. The exact test showed no statistically significant differences between the populations.

Genetic diversity and identification potential

Since pairwise Fst values among the studied samples did not exceed 0.01, the level of genetic differentiation was considered negligible, following common practice in population genetic studies (Gusmão et al., 2025), it was decided to treat the samples as a single metapopulation. All samples were



Distribution of forest and tundra samples in the space of the first two main components.

Table 1. Indicators of genetic differentiation of Nenets populations

Comparison groups	Fst (*p*)	$\sigma\mu^2$	*p*-value (exact test)
Forest/Tundra	0.007 (0.009)	0.062	1
Haryuchi/Vanuito	0.001 (0.233)	0.022	1
Antipayutinskaya/Gydan tundra	0.006 (0.041)	0.103	1
Antipayutinskaya/Nakhodka tundra	0.010 (0.002)	0.139	
Gydan/Nakhodka tundra	0.003 (0.231)	0.028	

Note. Fst (*p*) are the pairwise values of Fst (population comparisons), the significance level values are shown in parentheses; $\sigma\mu^2$ is the genetic distance between microsatellite loci; *p*-value is the significance level value for an accurate population differentiation test.

combined into a single metapopulation (504 samples) to assess the genetic diversity and identification potential of X-STR markers (Table 2).

The calculation results showed that the Nenets population demonstrates average genetic diversity, which is typical for the Indigenous peoples of Siberia who are in relative isolation (Vagaitseva et al., 2015, 2025). As can be seen from Table 2, the effective number of alleles varies from 1.882 (DXS7423) to 11.169 (DXS8377), and the expected heterozygosity (He), from 0.469 (DXS7423) to 0.910 (DXS8377), which indicates a pronounced variability in the level of polymorphism between loci. According to the classification of D. Botstein et al. (1980), 13 out of the 16 markers (81.3 %) are highly informative (PIC > 0.5), including the most polymorphic DXS10079 and DXS8377 (PIC > 0.8). The combined power of exclusion (CPE) of the entire set of markers was 0.99996. The values obtained are comparable to those typical for other Indigenous peoples of Siberia, but they are inferior to those in European populations (Vagaitseva et al., 2015; Mršić et al., 2017).

The linkage disequilibrium

Localization of the analyzed markers on the X chromosome both gives them advantages for the reconstruction of kinship and complicates the analysis. The alleles of the linked loci are inherited together, forming haplotypes. Ignoring the linkage disequilibrium structure in probabilistic calculations, in which the observed frequency of a haplotype in a population significantly deviates from the product of the frequencies of its constituent alleles, can lead to systematic errors.

Differences in the genetic and demographic history of a population form a unique pattern of linkage disequilibrium, which requires an empirical assessment for each specific population and does not allow extrapolating the data obtained from the analysis of some samples to other populations. To empirically assess the linkage disequilibrium structure between the considered X-STR loci in the Nenets metapopulation, a pairwise analysis was performed with the calculation of two generally accepted indicators: the significance level for each pair of loci (*p*-value) and the correlation square r^2. The matrix of *p*-values (Table 3) shows many statistically significant associations. With a threshold significance level of $p < 0.05$, 46 pairs out of 120 loci showed statistically significant LD. At the same time, the r^2 values for all pairs of alleles did not exceed 0.09. This means that despite the statistical significance, the correlation strength is low.

Reference data

The reference data necessary for probabilistic calculations in the reconstruction of kinship in the Nenets population are presented in the Supplementary Materials[1]. The Supplementary Tables show the frequencies of alleles of 16 X-STR loci and the frequencies of haplotypes of loci located in the same linkage group: II linkage group (DXS7132–DXS10079), III linkage group (DXS10103–HPRTB), IV linkage group (DXS8377–DXS7423) (Database ChrX-STR.org, accessed January 10, 2026).

Discussion

The Nenets are an Indigenous people of the Russian North, belonging to the Uralic language family. They are divided into two main sub-ethnic groups, the Forest and Tundra Nenets, which differ in their geographical distribution, cultural characteristics, and dialects. The Tundra Nenets include two main phratries: Haryuchi (Haryutsa) and Vanuito (Vanueta) (Khomic, 1996). The analysis of the samples under consideration by Y-chromosomal markers conducted by V.N. Kharkov et al. demonstrated a high degree of genetic differentiation, fully corresponding to the clan and phratry organization (Zvénigorosky et al., 2020; Kharkov et al., 2025). The X-STR analysis results obtained in our study, in contrast, demonstrated the absence of significant differences between the samples (exact test, $p = 1$). Such differences in the level of genetic differentiation indicate an asymmetric gene flow due to social organization.

The key characteristics of the Nenets marriage system are patrilineality, exogamy, and patrilocality (Khomic, 1996). Unlike men, women actively moved between clans, phratries, and even sub-ethnic groups. Our results are consistent with the demographic history of the Nenets population. The results of the X-STR marker study demonstrated an intermediate level of genetic diversity in the analyzed Nenets population. This is due to the founder effect during human settlement of

[1] Supplementary Materials are available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Vag_Engl_30_5.xlsx

Table 2. Parameters of genetic diversity and power of exclusion of X-STR markers in the Nenets metapopulation

	GATA31E08	DXS10079	DXS10103	DXS7132	DXS9895	DXS7133	DXS7424	DXS7423	DXS6789	DXS9902	DXS6810	DXS8377	HPRTB	DXS6797	DXS6804	GATA165B12
Na	7	9	7	6	5	4	8	4	9	5	4	18	8	8	6	4
Ne	4.282	6.201	3.960	3.676	2.855	2.308	4.083	1.882	3.913	3.012	2.416	11.169	3.636	3.413	3.878	2.184
PIC	0.729	0.819	0.714	0.682	0.584	0.500	0.724	0.424	0.708	0.602	0.522	0.903	0.682	0.660	0.703	0.472
He	0.766	0.839	0.747	0.729	0.650	0.567	0.755	0.469	0.744	0.668	0.586	0.910	0.725	0.707	0.743	0.542
PE	0.538	0.673	0.505	0.474	0.355	0.253	0.519	0.162	0.500	0.380	0.274	0.817	0.468	0.439	0.497	0.227

Note. Na – observed number of alleles per locus; Ne – effective number of alleles; PIC – polymorphism information content; He – expected heterozygosity; PE – power of exclusion.

Table 3. Matrix of *p*-values for estimating the linkage disequilibrium (LD) structure

Loci	DXS9902	DXS9895	DXS6810	DXS7132	DXS10079	DXS6789	DXS7424	DXS6797	DXS7133	DXS6804	GATA165B12	DXS10103	HPRTB	GATA31E08	DXS8377
DXS9902	*														
DXS9895	**0.0000**	*													
DXS6810	0.0299	0.1000	*												
DXS7132	0.0136	0.1778	0.1298	*											
DXS10079	0.2127	0.4524	**0.0006**	**0.0000**	*										
DXS6789	0.1210	0.0022	0.0020	0.3529	**0.0000**	*									
DXS7424	0.3006	0.2589	0.5006	0.2559	0.3805	**0.0000**	*								
DXS6797	0.5259	**0.0000**	0.0277	0.2368	0.2640	0.0285	**0.0000**	*							
DXS7133	0.3248	0.3343	0.2124	0.2066	0.5788	0.0815	0.0015	0.0047	*						
DXS6804	0.1663	0.0081	0.3119	0.0438	0.2698	**0.0000**	**0.0000**	**0.0000**	**0.0000**	*					
GATA165B12	0.5644	0.3142	0.0650	0.7254	0.0202	0.0016	0.0355	0.0008	0.4632	0.3574	*				
DXS10103	0.4066	0.0924	0.4195	0.5312	0.1546	**0.0000**	0.4138	0.1351	0.2672	0.8459	0.0888	*			
HPRTB	0.4551	0.0358	0.2916	0.2064	0.0150	0.3056	0.1193	0.3816	0.7270	0.0472	0.0514	**0.0000**	*		
GATA31E08	**0.0000**	**0.0000**	0.0178	0.1458	0.0393	0.4363	0.2963	0.1192	0.7810	0.0983	0.2668	**0.0000**	**0.0000**	*	
DXS8377	0.2403	0.4584	0.8199	0.6650	0.0009	0.3204	0.0033	0.3578	0.0414	0.0726	**0.0000**	0.0060	0.2131	0.0148	*
DXS7423	0.1433	0.0382	0.3474	0.0152	0.6663	0.0932	0.0498	0.0873	0.3627	0.2202	0.5829	0.1132	0.9637	0.1002	**0.0000**

Northern Eurasia, and the subsequent genetic drift in isolated groups.

Pairwise Fst analysis showed a low level of genetic differentiation, values not exceeding 1 %. Of the five pairwise comparisons, statistical significance ($p < 0.05$) was observed in only two. At the same time, the differences between the samples of the Antipayutinskaya and Nakhodka tundras (Fst = 0.010, $p = 0.002$) exceed those between the Forest and Tundra Nenets (Fst = 0.007, $p = 0.009$). For comparison, in our other study using 18 X-STR markers, weak statistically significant differences (Rst = 0.007, $p = 0.002$) were also revealed between two other populations of Siberia (Telengits and Altai-Kizhi), and the level of genetic diversity (He) in Altai populations was higher (mean He in the Nenets population = 0.697, in the Altai-Kizhi population = 0.726, and in the Telengits population = 0.723) (Vagaytseva et al., 2025). The obtained value of the combined power of exclusion (CPE = 0.99996) confirms that the presented panel of 16 X-STR markers is quite informative and can be used for pedigree reconstruction in the Nenets population. It is shown that when reconstructing the kinship of individuals from the Nenets population, it is possible not to take into account the territory of residence of the analyzed individuals, due to the lack of statistically significant differentiation between the analyzed Nenets groups (exact test $p = 1$; Fst < 1 %). In this paper, the analyzed sample of Forest Nenets is small (46 samples), which may affect the robustness of the results obtained. Nevertheless, the presented data are the most comprehensive characterization of the genetic structure of the Nenets population by X-STR markers to date.

Conclusion

The present study provides reference data on 16 X-STR markers for the Nenets population, which are necessary to increase the efficiency and reliability of forensic genetic examination. This is especially important for populations of Indigenous peoples of Siberia, since the low level of genetic diversity of neutral markers used in forensic genetics makes standard autosomal STRs (aSTRs) less informative (Zvénigorosky et al., 2020). The obtained parameters can be used in the practice of forensic genetics to increase the reliability of forensic genetic examination in the Russian Federation.

References

Botstein D., White R.L., Skalnick M.H., Davies R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am J Hum Genet*. 1980;32(3):314-331

Excoffier L., Lischer H.E.L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10(3):564-567. doi 10.1111/j.1755-0998.2010.02847.x

Ge J., Budowle B. How many familial relationship testing results could be wrong? *PLoS Genet*. 2020;16(8):e1008929. doi 10.1371/journal.pgen.1008929

Gusmão L., Antão-Sousa S., Faustino M., Abovich M.A., Aguirre D., Alghafri R., Alves C., … Turchi C., Vullo C., Yurrebaso I., Pereira V., Pinto N. X-chromosomal STRs: metapopulations and mutation rates. *Forensic Sci Int Genet.* 2025;76:103232. doi 8.1016/j.fsigen.2025.103232

Kharkov V.N., Valikhova L.V., Adamov D.S., Zarubin A.A., Khitrinskaya I.Yu., Stepanov V.A. The Forest and Tundra Nenets: differences in Y-chromosome haplogroups. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(5):711-721. doi 10.18699/vjgb-25-78

Khomic L.V. The Nenets: Historical and Ethnographic Essays. St. Petersburg: Nauka Publ., 1966 (in Russian)

Lê S., Josse J., Husson F. FactoMineR: an R package for multivariate analysis. *J Stat Softw*. 2008;25(1):1-18. doi 10.18637/jss.v025.i01

Mršić G., Ozretić P., Crnjac J., Merkaš S., Račić I., Rožić S., Sukser V., Popović M., Korolija M. Analysis of 12 X-STR loci in the population of south Croatia. *Mol Biol Rep*. 2017;44(1):183-189. doi 10.1007/s11033-017-4096-1

SoftGenetics LLC. GeneMarker® HID Software User Manual. Version 2.0. State College, PA: SoftGenetics LLC, 2012

Tillmar A.O., Kling D., Butler J.M., Parson W., Prinz M., Schneider P.M., Egeland T., Gusmão L. DNA Commission of the International Society for Forensic Genetics (ISFG): guidelines on the use of X-STRs in kinship analysis. *Forensic Sci Int Genet*. 2017;29:269-275. doi 10.1016/j.fsigen.2017.05.005

Vagaitseva K.V., Kharkov V.N., Cherpinskaya K.V., Khitrinskaya I.Yu., Stepanov V.A. Genetic variability of X-linked STR markers in Siberian populations. *Mol Biol*. 2015;49(2):267-274. doi 10.1134/S0026893315020132

Vagaitseva K.V., Kolesnikov N.A., Skalin M.D., Valikhova L.V., Ruzavina O.D., Pestretsova D.E., Vysochina A.N., Kharkov V.N., Stepanov V.A. Genetic diversity of Telengite and Altai-kizhi populations by markers used in forensic genetics. *Vestnik Tomskogo Gosudarstvennogo Universiteta. Biologiya = Tomsk State University Journal of Biology*. 2025;69:29-37. doi 10.17223/19988591/69/4 (in Russian)

Zhang Y., Yu Z., Mo X., Zhao X., Li W., Liu H., Liu C., Wu R., Sun H. Comparative evaluation of autosomal STRs and X-chromosome STRs as a complement of autosomal STRs in kinship testing in Southern Han Chinese. *Ann Hum Biol.* 2021a;48(1):66-69. doi 10.1080/03014460.2020.1856926

Zhang Y., Yu Z., Mo X., Zhao X., Li W., Liu H., Liu C., Wu R., Sun H. Development and validation of a new 18 X-STR typing assay for forensic applications. *Electrophoresis.* 2021b;42(6):766-773. doi 10.1002/elps.202000168

Zvénigorosky V., Sabbagh A., Gonzalez A., Fausser J.L., Palstra F., Romanov G., Solovyev A., Barashkov N., Fedorova S., Crubézy É., Ludes B., Keyser C. The limitations of kinship determinations using STR data in ill-defined populations. *Int J Legal Med*. 2020;134(6):1981-1990. doi 10.1007/s00414-020-02298-w

ChrX-STR.org 2.0 [site]. Forensic ChrX Research; [n.d.; updated Jun 25, 2009; cited Jan 10, 2026]. Available from: https://www.chrx-str.org/xdb/index.jsf

Conflict of interest. The authors declare no conflict of interest.
Received April 13, 2026. Revised April 28, 2026. Accepted May 12, 2026.

HUMAN GENETICS
Original article

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):761-774

doi 10.18699/vjgb-26-77

Genomic profiles of the Khazar-Period inhabitants of Phanagoria

V.S. Okovantsev [1] ✉, A.D. Manakhov [1, 2], E.V. Rozhdestvenskih [1], S.E. Khrapov [1], M.V. Dobrovolskaya [3], N.G. Svirkina [1, 3], S.N. Ostapenko [1, 3, 4], V.D. Kuznetsov [1, 3, 4], T.V. Andreeva [1, 2, 5], E.I. Rogaev [1, 6] ✉

[1] Sirius University of Science and Technology, Sirius Federal Territory, Krasnodar region, Russia
[2] Vavilov Institute of General Genetics of the Russian Academy of Sciences, Moscow, Russia
[3] Institute of Archaeology of the Russian Academy of Sciences, Moscow, Russia
[4] The State Museum-Preserve "Phanagoria", Sennoy village, Krasnodar region, Russia
[5] Centre of Genetics and Genetic Technologies, Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia
[6] UMass Chan Medical School, Department of Psychiatry, Shrewsbury, USA
✉ okovantsev.vs@talantiuspeh.ru; evivrecc@gmail.com

Abstract. The Phanagoria settlement, located on the Taman Peninsula, is a unique archaeological site. The history of the ancient city spans approximately 15 centuries, making Phanagoria a witness to many historical events, including those related to the early history of the Old Rus' state. The early medieval period of Phanagoria's history is of particular interest due to its insufficient study. The Phanagoria Expedition of the Institute of Archaeology of the Russian Academy of Sciences, in collaboration with the Phanagoria Museum-Reserve, studied the city's cultural strata from the 7th–10th centuries. This period of the urban center's existence is closely linked to the history of the Khazar Khaganate and culminates in the city's destruction as a result of an enemy attack. During excavations, skeletal remains bearing signs of violent death were discovered. Anthropological analysis revealed Mongoloid features of the individuals discovered. This paper presents the results of a bioinformatics-based analysis of whole-genome sequences from the remains of three individuals from a Khazar cultural layer. Radiocarbon dating was performed for two individuals. High mitochondrial DNA coverage (43- to 96-fold) confirmed its low level of contamination, and determined the mitochondrial chromosome haplogroup. High genome coverage confirmed the authenticity of the analyzed ancient DNA and allowed the determination of the individuals' sex and Y-chromosome haplogroup. Population analysis was performed using PCA, ancestral component analysis, using ADMIXTURE, and allele frequency correlation analysis, using the *f*3 statistic. Whole-genome data analysis of the three samples demonstrated the genetic proximity of the studied individuals to the modern Turkic-speaking population of Central Asia, as well as to early medieval nomadic groups of the Eurasian steppe, Hungarian Avars, and Mongols. Principal component analysis revealed the presence of Asian and European genetic components in the genomes of the studied samples. Our study suggests that the genetic profile of the individuals studied was formed by the mixing of the gene pools of people from the eastern regions of the Eurasian Steppe and populations of Western Eurasia. The Asian components are related to those of medieval groups of Mongols and Avars.

Key words: ancient DNA; genome; paleoanthropology; Phanagoria; massive parallel sequencing; Khazar Khaganate

For citation: Okovantsev V.S., Manakhov A.D., Rozhdestvenskih E.V., Khrapov S.E., Dobrovolskaya M.V., Svirkina N.G., Ostapenko S.N., Kuznetsov V.D., Andreeva T.V., Rogaev E.I. Genomic profiles of the Khazar-Period inhabitants of Phanagoria. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(5):761-774. doi 10.18699/vjgb-26-77

Funding. This work was supported by the grant of the state program of the Sirius Federal Territory "Paleogenetics of Krasnodar Krai", Agreement No. 18-03 dated September 10, 2024, project code GEN-BFT-2407 (VSO, ADM, EVR, SEK).

Геномные профили жителей Фанагории хазарского времени

В.С. Окованцев [1] ✉, А.Д. Манахов [1, 2], Е.В. Рождественских [1], С.Е. Храпов [1], М.В. Добровольская [3], Н.Г. Свиркина [1, 3], С.Н. Остапенко [1, 3, 4], В.Д. Кузнецов [1, 3, 4], Т.В. Андреева [1, 2, 5], Е.И. Рогаев [1, 6] ✉

[1] Научно-технологический университет «Сириус», федеральная территория «Сириус», Краснодарский край, Россия
[2] Институт общей генетики им. Н.И. Вавилова Российской академии наук, Москва, Россия
[3] Институт археологии Российской академии наук, Москва, Россия
[4] Государственный историко-археологический музей-заповедник «Фанагория», пос. Сенной, Темрюкский район, Краснодарский край, Россия
[5] Центр генетики и генетических технологий, биологический факультет МГУ им. М.В. Ломоносова, Москва, Россия
[6] Медицинская школа Чан Массачусетского университета, департамент психиатрии, Шрусбери, США
✉ okovantsev.vs@talantiuspeh.ru; evivrecc@gmail.com

Аннотация. Городище Фанагория, расположенное на территории Таманского полуострова, представляет собой уникальный объект археологического наследия. История древнего города охватывает около 15 веков, что делает Фанагорию свидетелем многих исторических событий, в том числе имеющих отношение к ранней истории

древнерусского государства. Раннесредневековый период истории Фанагории представляет особый интерес в связи с ее недостаточной изученностью. Работами Фанагорийской экспедиции Института археологии РАН в содружестве с Музеем-заповедником «Фанагория» были изучены культурные напластования города VII–X вв. Этот период существования урбанистического центра тесно связан с историей Хазарского каганата и завершается гибелью города в результате вражеского нападения. В ходе раскопок обнаружены скелетные останки людей со следами насильственной смерти. Антропологический анализ выявил монголоидные особенности этих индивидов. В данной работе представлены результаты анализа полногеномных последовательностей из останков трех индивидов из культурного слоя хазарского времени с использованием биоинформатических методов. Для двух индивидов было проведено радиоуглеродное датирование. Благодаря высокому покрытию митохондриальной ДНК (от 43- до 96-кратного), подтвержден низкий уровень контаминации и определена гаплогруппа митохондриальной хромосомы. Установлена аутентичность древней ДНК и определены пол индивидов и гаплогруппа Y-хромосомы. Популяционный анализ проведен методом PCA, анализ предковых компонентов – методом ADMIXTURE и анализ корреляции частот аллелей – методом *f*3-статистики. Анализ полногеномных данных трех образцов показал генетическую близость исследованных индивидов с современным тюркоязычным населением Центральной Азии, а также с раннесредневековыми группами кочевников Евразийской степи, венгерских авар и монголов. Анализ главных компонент показал присутствие в геномах изученных образцов азиатских и европейских генетических компонентов. Наше исследование позволяет сделать вывод, что генетический профиль изученных индивидов сформировался в результате смешения генофондов выходцев из восточных районов Евразийской степи и населения Западной Евразии. Азиатские компоненты родственны таковым у средневековых групп монголов и авар.

Ключевые слова: древняя ДНК; геном; палеоантропология; Фанагория; параллельное секвенирование; Хазарский каганат

Introduction

Phanagoria is an ancient city located on the eastern shore of the Kerch Strait (Fig. 1), founded by Greek colonists presumably in the 6th century BCE (Chkhaidze, 2012). In the 5th century CE, along with other cities of the Bosporan Kingdom, Phanagoria became part of the Byzantine Empire, which maintained full control over the city until the mid-7th century CE (Chkhaidze, 2012). The city's population in the 5th–6th centuries CE is reported to have been predominantly Greek (Chkhaidze, 2012).

With the collapse of the Western Turkic Khaganate, the Khazar Khaganate emerged in the Caspian steppes; in the 7th century CE, its influence expanded westward into the Northern Black Sea region and Eastern Azov region (Artamonov, 1962; Golden, 1980). By the end of the 7th century CE, Phanagoria had become an important site of diplomatic contacts between Byzantium and Khazaria. Constantinople retained authority over the city, while two Khazar officials were also active there. This arrangement formed a so-called condominium between the two states. In the 8th century CE, there was a mass migration of the Khazar population to Crimea and the Azov region, caused by clashes between the Khazars and the Arab Caliphate in the Caucasus (Sazanov, Mogarichev, 2006). In the second half of the 9th century CE, Byzantium once again regained control over Phanagoria due to the weakening of Khazaria's positions under the pressure of the Magyars and Pechenegs. By the beginning of the 10th century CE, the city was devastated by nomads: Magyars, Pechenegs, or Oghuz (Dunlop, 1954; Pletneva, 1982; Novoseltsev, 1990).

The population of the Khaganate was highly diverse and included Bulgars, Alans, Jews, Slavs, and other ethnic and cultural groups. Our knowledge about the Khazars themselves is limited and contradictory (Artamonov, 1962; Balabanova, 2013; Ginzburg, 1963). According to the available linguistic data and information from medieval historians, the Khazars are classified as Turkic-speaking peoples. A relationship between the Khazar language and Bulgar has also been proposed (Novoseltsev, 1990). Among Soviet and Russian authors, there is a widespread hypothesis that the Khazar ethnic group was formed with the participation of the Iranian population of the North Caucasus and incoming Turks, as well as migrants from Siberia (Artamonov, 1962; Novoseltsev, 1990; Balabanova, 2013).

At present, the beginning of the Khazar period in Phanagoria is dated to the last quarter of the 7th century CE (Kuznetsov, Ostapenko, 2024). The Khazar presence is attested by archaeological finds, including ceramics and architectural remains. In the layers of the 8th–9th centuries CE, ceramics of the Saltovo-Mayaki culture are found (Chkhaidze, 2012), which are associated with the Alanian and Bulgarian components in the population of the Khazar Khaganate (Pletneva, 1967).

The late 8th century CE is particularly important in Khazar history, as this was the period when the Khazar elite adopted Judaism, likely under the influence of a significant Jewish diaspora that existed in the cities of the khaganate (Artamonov, 1962; Kalinina, 2010). It is not possible to determine the proportion of the Jewish population in the Khazar Khaganate. Many researchers believe that only the elite close to the khan were Jewish, while the majority of the population remained pagan (Artamonov, 1962; Novoseltsev, 1990; Kalinina, 2010). Historical sources indicate that there was a Jewish diaspora in Phanagoria during the 1st to 5th centuries CE, as evidenced by numerous Jewish burials (Chkhaidze, 2012). A synagogue was discovered, which existed until its destruction by fire in the 6th century CE (Kuznetsov, 2024). The community also persisted in the Bosporus region during the Khazar period (Chichurov, 1980); however, to date, there is only one archaeo-

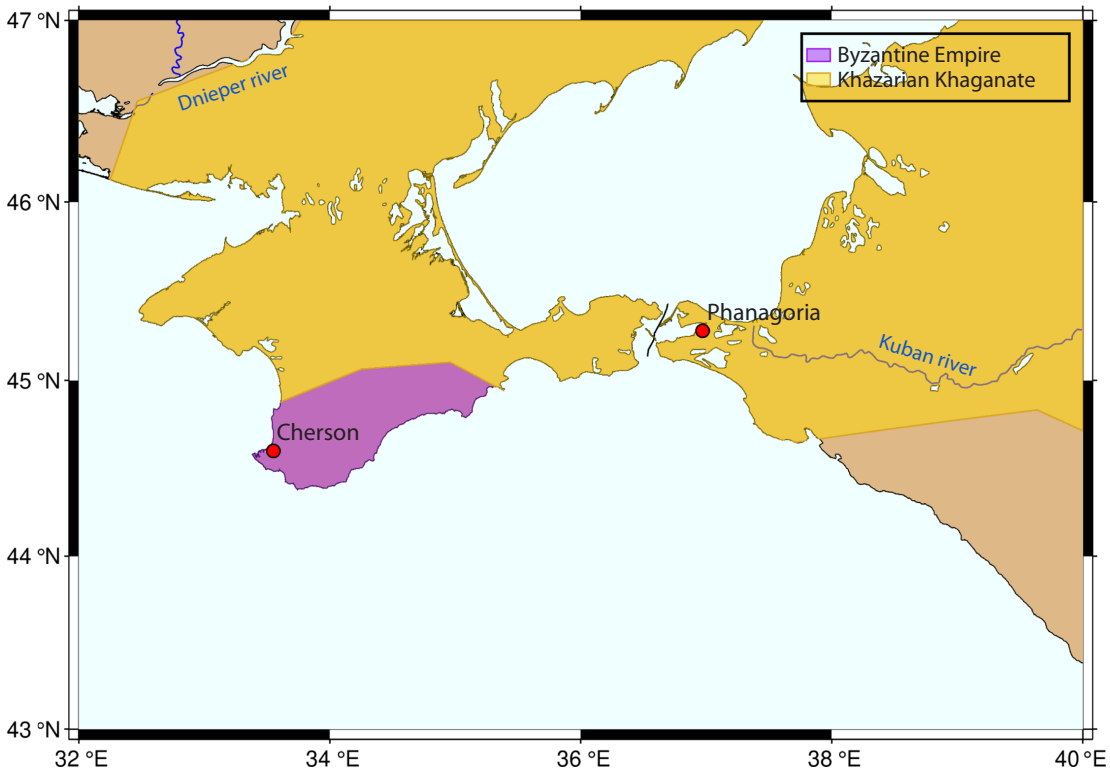


Fig. 1. Location of the archaeological site of Phanagoria in the Northern Black Sea region (second half of the 7th century CE).

The approximate boundaries of Byzantine and Khazar influence in the region are indicated (according to: Novoseltsev, 1990; Aibabin, 2013).

logical confirmation of this in the form of an amphora with an ancient Hebrew inscription (Golofast, 2020).

According to anthropological data, the population of the Lower Don and Lower Volga regions during the Khazar Khaganate period is characterized by a mixture of Mongoloid and Europid traits (Ginzburg, 1963; Batieva, 2002; Balabanova, 2006, 2013). It is noted that skulls with Mongoloid features are more frequently found in the southernmost sites of the Saltovo-Mayaki culture (Balabanova, 2005, 2006). In burial mounds with ditches, which are associated with the Khazar ethnic group (Semenov, 1983; Novoseltsev, 1990; Afanasyev, 2015), predominantly Mongoloid-type skulls are found (Ginzburg, 1946; Batieva, 2002).

Current studies of the burials in ancient Phanagoria (Manakhov et al., 2025; Rozhdestvenskikh et al., 2026) note the similarity of the population of the Hellenistic and Roman periods with modern European populations. Individuals from ancient Phanagoria have been found to possess mitochondrial DNA and Y-chromosome haplogroups that are common in modern European populations. As a result of the analysis of autosomal markers using the principal component method, individuals from the ancient period are projected to be close to modern European and Caucasian populations.

We do not have anthropological materials from the burials of Phanagoria from the Khazar period. All available remains for study come from the cultural layers of the city. The damage identified on the bones indicates an episode of brutal aggression, manifested in the killing of people of different genders and ages. Since the completeness of the materials is far from full, the racial diversity of the sample could only be assessed for a few individuals. Many are characterized by Mongoloid features: a flattened face, slight protrusion of the nasal bones (Dobrovolskaya, Svirkina, 2018).

At present, paleogenetic data on the population of the Khazar Khaganate remain extremely limited. The limited available whole-genome data for Bulgar and Alan samples of the Saltovo-Mayaki culture (14 individuals) have been presented within the framework of studies on the population of the Pontic steppe (Damgaard et al., 2018; Saag et al., 2025).

In this study, we apply genomic and population genetic methods for the first time to remains from Phanagoria dating to the period of Khazar rule in the region.

Materials and methods

The study included the skeletal remains of four individuals discovered in the “Lower City” excavation of the archaeological site of Phanagoria in 2016, dating back to the period of Khazar Khaganate rule in the Northern Black Sea region during the 7th–10th centuries CE. Anthropological descriptions of the samples and associated archaeological finds are provided in

(Dobrovolskaya, Svirkina, 2018). The KhT02 individual is described as an 8–9-year-old girl. Cranial injuries and signs of chronic disease were noted. Cranially, facial flattening is noted. In the KhT03 individual, a chronic inflammatory process has been described, which could have caused developmental delays. The individual was estimated to be 6–7 years old, and sex could not be determined anthropologically. The KhT04 individual is identified as a 45-year-old male, with a healed parietal bone injury and an inflammatory process. Anthropologically, brachycephaly and facial flattening are described. The KhT05 individual was excluded from the analysis due to low endogenous DNA content.

Radiocarbon dating of two studied samples was performed at the Isotope Research Laboratory of the Center for Collective Use "Cenozoic Geochronology" of the Institute of Archaeology and Ethnography, Siberian Branch of the Russian Academy of Sciences, and at the Center for Collective Use "Accelerator Mass Spectrometry" of Novosibirsk State University.

Ancient DNA was extracted from bone fragments at Sirius University in dedicated clean rooms for ancient DNA work, following a previously developed protocol (Andreeva et al., 2022a). Each DNA extraction was accompanied by a negative control. Fragmented genomic libraries were prepared according to a protocol based on the use of single-stranded DNA (Gansauge, Meyer, 2013). DNA extracts and prepared libraries were assessed qualitatively and quantitatively using capillary gel electrophoresis on the Agilent TapeStation 4150 device with the High Sensitivity D1000 reagent kit (Agilent). For the KhT03 sample, a fragment library was additionally prepared from the dDNA preparation, which was subjected to repair using the PreCR Repair Mix (NEB) reagent kit (Mouttham et al., 2015). Sequencing was conducted at the Sirius University on the Illumina MiSeq platform in paired-end read mode of 76+76 nucleotides (test sequencing) and on the Illumina NovaSeq 6000 platform in single-end read mode of 56 nucleotides (deep sequencing).

Adapter sequences and low-quality reads were removed from the sequencing data using the AdapterRemoval v2 program (Schubert et al., 2016). The resulting reads were then mapped to the human reference genome GRCh37 and the mitochondrial reference genome NC_012920.1 using the BWA program (Li, Durbin, 2009) with parameters adapted for ddDNA sequencing results (Schubert et al., 2012). Using the MarkDuplicates function from the Picard software package (Picard Tools – By Broad Institute, 2019), duplicates were marked to exclude the corresponding reads from further analysis.

Ancient DNA authenticity was assessed using MapDamage2 (Jónsson et al., 2013). Genetic sex was determined using the karyo_RxRy program (Anastasiadou et al., 2024).

The assessment of sample contamination levels was conducted based on mtDNA heterozygosity using the Schmutzi (Renaud et al., 2015) and ContamMix (Fu et al., 2013) programs. For the male sample, an additional analysis of contamination levels based on X chromosome heterozygosity (Rasmussen et al., 2011) was conducted.

The reconstruction of the mitochondrial DNA sequence was carried out by identifying polymorphisms in it using the BCFtools program (Danecek et al., 2021). The identified polymorphisms were filtered by quality (QUAL > 30) and visually checked using the IGV (Integrative Genomics Viewer) program (Thorvaldsdóttir et al., 2013). The determination of mtDNA haplogroups was conducted using the Haplogrep3 program (PhyloTree version 17 – Forensic Update 1.2) (Schönherr et al., 2023), with additional verification of the identified variants using the Yfull MTree 1.02 database (Van Oven, Kayser, 2009).

Phylogenetic analysis of mtDNA sequences was conducted using the mtPhyl software package (Eltsov, Volodko, 2016). For phylogenetic analysis, a sample of mtDNA sequences belonging to the haplogroup identified in the studied individual was formed, using databases containing both ancient and modern samples: HaploTree (HaploTree, 2025), AADR (Mallick et al., 2024), NCBI BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi), Yfull MTree 1.02, and AmtDB (Ehler et al., 2019). In addition, databases of ancient samples previously compiled at NTU Sirius (Andreeva et al., 2022b, 2023) were used. Regions of poly-C tracts, tandem repeats, and positions with high mutation frequency were excluded from the analysis.

Using the pileupCaller program from the SequenceTools package (Schiffels, 2025), we reconstructed pseudohaploid genomic profiles of the studied individuals based on 1.24 million genetic markers corresponding to the AADR panel (parameter --randomHaploid). For the analysis, reads with a mapping quality (MQ) and read quality (BQ) of at least 30 were used.

Genetic relatedness was assessed using READv2 (Alaçamlı et al., 2024), which can infer relationships up to the third degree.

Population analysis was conducted using the principal component method (PCA) with the smartpca function from the EIGENSOFT software package (Patterson et al., 2006). The genomic profiles of the individuals studied were projected into the principal component space formed based on the genetic profiles of 2,598 representatives from 168 modern groups from the Eurasian region in the Human Origins array, which includes 600 thousand genetic markers. The analysis was conducted using genomic profiles from which transitions (C ↔ T and G ↔ A), which may be a result of postmortem modifications characteristic of ancient DNA, were excluded.

To compare the studied samples with ancient populations, 1,435 samples from 90 ancient groups of the early Iron Age to medieval periods from Eastern Europe, the Near East, and Central Asia were included in the analysis, as represented in the AADR v.62 database and publications (Nikitin et al., 2025; Saag et al., 2025).

The ADMIXTURE analysis was conducted using a set of 560 thousand autosomal genetic markers from the AADR v.62 database and materials from publications (Lazaridis et al., 2014; Flegontov et al., 2019; Nikitin et al., 2025; Saag et al., 2025). The ADMIXTURE v1.3 program (Alexander et al., 2009) was used with default parameters. We conducted

10 independent calculations, varying the number of assumed ancestral components (K) from 1 to 12. The optimal number of components was determined based on the comparison of CV-error values calculated by the program; as a result, the optimal value was K = 5. Out of 10 options, the calculation with the minimum CV-error value was selected (Supplementary Materials)[1].

Before the analysis, sample selection and genetic variant filtering were conducted. In cases where the samples originated from the same family group, only one individual with the highest number of identified autosomal pseudodiploid genotypes was included in the analysis. Using the PLINK v2.00 program (Chang et al., 2015), genetic variants were excluded from the analysis if their genotypes were determined in only one individual (--geno 0.999); variants with a minor allele frequency of less than 0.002 (--maf 0.002); variants that are transitions, which may be a result of post-mortem modifications characteristic of ancient DNA; as well as linked genetic variants (--indep-pairwise 200 5 0.5). Samples for which fewer than 10 thousand pseudodiploid markers remained after filtering were excluded from the analysis.

A total of 2,521 samples were included in the analysis, including three representatives of the population of Phanagoria from the Khazar period; 2,248 ancient samples from the AADR v62 database and published materials (Flegontov et al., 2019; Lazaridis et al., 2025; Nikitin et al., 2025; Saag et al., 2025) from the period ranging from the Early Bronze Age to the Middle Ages in Eurasia; 188 ancient and 85 modern samples from the AADR v62 database, representing groups in which commonly accepted ancestral components in paleogenetic studies are maximized (WHG (Western European Hunters Gatherers) – component associated with Western European Mesolithic hunter-gatherers; EHG (Eastern European Hunters Gatherers) – component associated with Eastern European Mesolithic hunter-gatherers; ANF (Anatolian Neolithic Farmers) – component associated with Neolithic farmers of Anatolia; Han – East Asian component maximized in the modern Han Chinese population; Nganasan – Siberian component maximized in the modern Nganasan population; Mbuti – African component represented by the African Mbuti population) (see further).

To assess genetic similarity with previously published ancient populations, *f*3-statistics were calculated. Pseudohaploid genotypes from a panel of 1.24 million genetic markers were included in the analysis, with transitions being excluded. The *f*3-statistic calculations were performed using the AdmixTools v.7.0.1 (Patterson et al., 2012) and admixr v.0.9.1 (Petr et al., 2019) software in the form *f*3 (Mbuti, Test; KhT02/KhT03/KhT04), where Test is one of the ancient groups relative to which the sample under study is analyzed; Mbuti are representatives of the African Mbuti population used as an external group. Ancient groups from Europe, Central, and East Asia dating on average from 1,700 to 700 years ago were taken as the test. Only groups represented by more than three non-related individuals and having at least 20 thousand genetic markers overlapping with the studied samples KhT02/KhT03/KhT04 were included in the analysis.

[1] Supplementary Materials are available at: https://vavilovj-icg.ru/download/pict-2026-30/appx42.xlsx

Results

Radiocarbon dating of two samples confirmed that they date to the Khazar period (Table 1). The two obtained radiocarbon dates have very broad calibrated ranges. Both KhT02 and KhT03 are associated with the same episode of violence; however, the radiocarbon age and calibrated dates show a significant temporal discrepancy between these samples. This discrepancy calls for a cautious interpretation of the dates and for expanding both the number of dated samples and the range of dated materials (Table 1). Nevertheless, the attack on the city is more likely to date to the 8th century CE than to the 9th century CE; therefore, linking these deaths to the final destruction of the city is currently unlikely.

Genomic DNA was successfully extracted from bone samples of all four individuals and used to prepare genomic libraries for sequencing. In three samples, a high proportion of reads mapped to the human reference genome and aDNA authenticity were confirmed, allowing the use of the DNA from these samples for whole-genome analysis (Table 2, Fig. 2). Sample KhT05 was excluded from further analysis due to low endogenous DNA content.

For all three samples, the genetic sex was determined: samples KhT02 (child aged 8–9 years) and KhT04 (individual over 45 years old) were female, while sample KhT03 (child aged 6–7 years) was male. The discrepancy between the anthropological and genetic sex estimates for KhT04 is noteworthy. Based on the preserved morphological features of the skull (development of the supraorbital ridge, outer edge of the eye socket, size of the mastoid process, overall robustness of the skull), as well as the size and prominence of the relief on the bones of the upper limb, the skeleton was determined to belong to a male. It should be noted that anthropological traits provide a probabilistic determination of biological sex, so discrepancies in the results of sex diagnosis by two methods are possible. In this case, the robust physical habitus of an individual genetically identified as female is noteworthy.

The assessment of contamination levels for the male sample based on mtDNA and the X chromosome showed that the contamination does not exceed the permissible 5 % level for paleogenomic studies (Table 2).

For all samples, complete mtDNA sequences were reconstructed and their mitochondrial haplogroups were determined (KhT02 – C4a1a-a, KhT03 – B4b1a3a, KhT04 – T2d2a) (Fig. 3). Since the haplogroups differ, the studied individuals are unlikely to have been direct maternal relatives.

The kinship analysis based on whole-genome sequencing data, conducted using the READv2 program, also showed that the studied individuals are not first, second, or third-degree relatives.

For the male individual KhT03, the Y-chromosome haplogroup was determined as C-Y10420 (C2a1a1b1b).

Table 1. Archaeological and anthropological data on the examined samples from the 2016 excavation of the "Lower City" of Phanagoria

Sample	Location	Sex (according to anthropological data)	Age, years	Material	Radiocarbon dating			
					Sample code	Calibrated age, years BP	Calibrated date with 95.4 % probability	Probability distribution of time intervals
KhT02	Square Ж4, stake 12	F	8–9	Temporal bone	GV-6027	1,207 ± 34	687–947 CE	687–743 CE 10.6 % 771–894 CE 81.6 % 927–947 CE 2–3 %
KhT03	Object 10, stake 15	Infant	6–7	Temporal bone	GV-6028	1,288 ± 35	658–821 CE	658–777 CE 90.2 % 791–821 CE 5.2 %
KhT04	Square И1, stake 13	M	Over 45	Temporal bone	–	–	–	–
KhT05	Square Ж4, stake 14	F	20–25	Tooth	–	–	–	–

Table 2. Statistics of DNA sequencing, mtDNA and Y-chromosome haplogroups

Sample ID	Library type	% of reads mapped to the genome	Average mtDNA coverage	Average genome coverage	Genetic sex	mtDNA contamination (Shmutzi)	mtDNA contamination (ContaMix [95 % CI])	X chromosome contamination	mtDNA haplogroup	Y chromosome haplogroup	Number of SNPs on the 1240k panel
KhT02	Unrepaired	82.34	x43.5	x0.3	F	1 %	1.23 % [0.32 – 2.68 %]	–	C4a1a-a	–	223 379
KhT03	Unrepaired	61.55	x96.7	x2.6	M	2 %	0.74 % [0.25 – 1.48 %]	1.75 ± 0.13 %	B4b1a3a	C2a1a1b1b (C-Y10420)	1 082 901
	Repaired	64.79				–					
KhT04	Unrepaired	70.23	x48.4	x0.4	F	2 %	3.72 % [2.63 – 5.35 %]	–	T2d2a	–	310 822
KhT05	Unrepaired	6.13	x0.9	x0.002	F	–		–	–	–	–

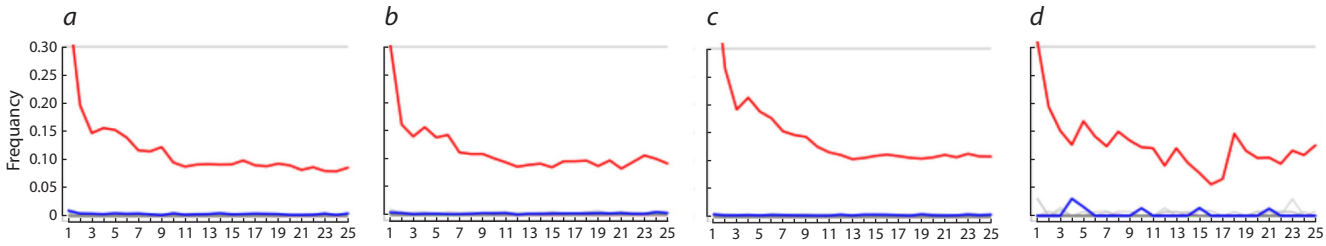


Fig. 2. Profiles of nucleotide substitutions for reads mapped to the reference mtDNA sequence, calculated using MapDamage2.

C > T substitutions specific to mtDNA are shown with a red line. The X-axis marks the nucleotide position from the 5′ end of the DNA fragments. Samples: *a* – KhT02; *b* – KhT03; *c* – KhT04; *d* – KhT05.

Population analysis conducted using the principal component method showed that the studied samples project near the modern Turkic-speaking populations of Central Asia – Turkmen, Uzbeks, Karakalpaks, and Kazakhs – as well as the Nogais of the southern European part of Russia (Fig. 4*a*). At the same time, all three samples are separated from each other and are located along the Caucasus–Mongolia axis.

Sample KhT04 is located closer than the others to the cluster of European groups, in the array of Turkmen, Uzbeks, and Stavropol Nogais. Sample KhT02 clusters with Kazakhs,

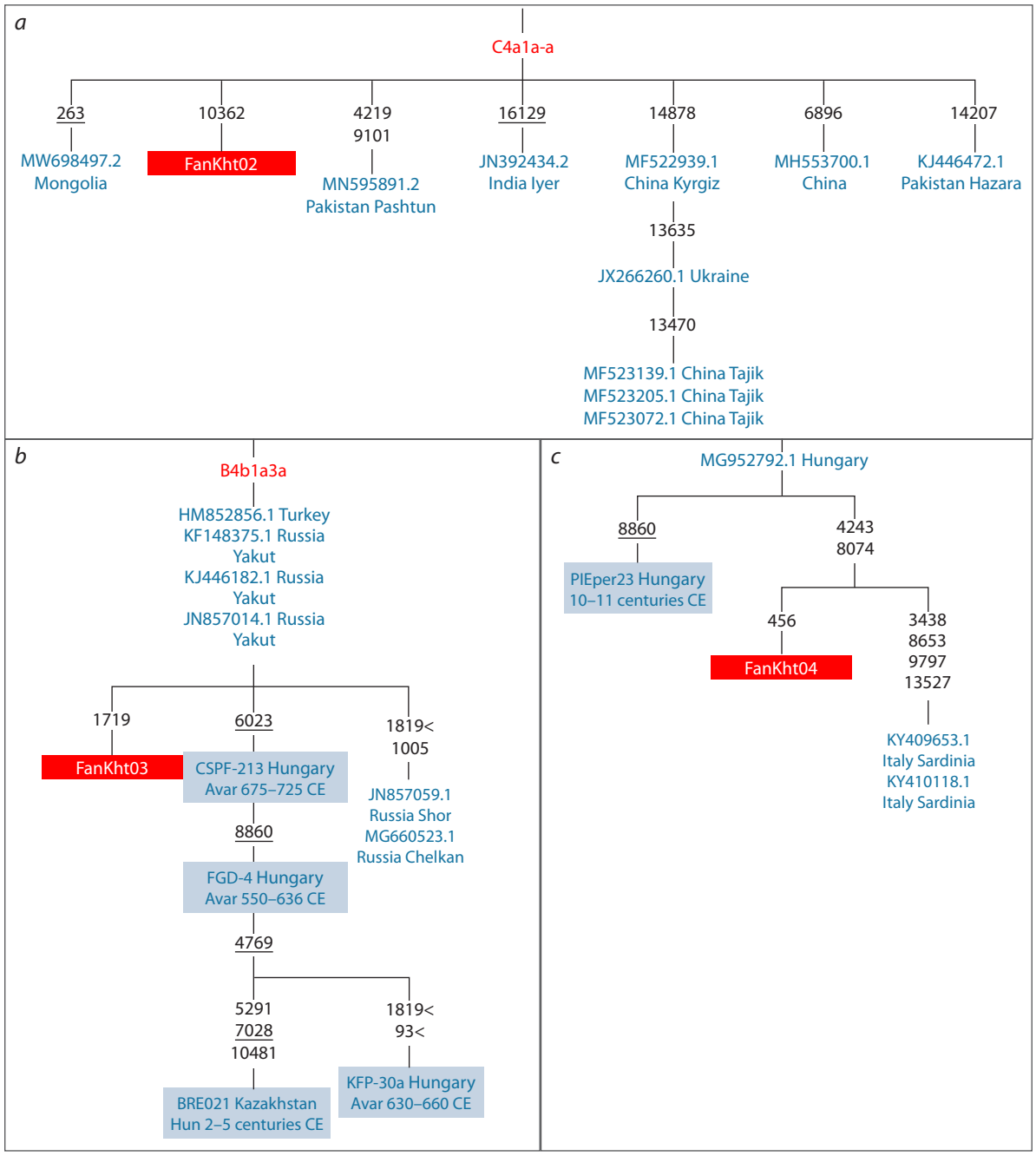


Fig. 3. Phylogenetic tree of haplogroups: *a* – C4a1a-a; *b* – B4b1a3a; *c* – T2d2a, constructed using the mtPhyl software package.

Mutations on the branches are indicated relative to the reference mtDNA sequence (rCRS). For transitions, only the nucleotide position number is indicated, while for transversions, the nucleotide replacement is provided; positions of recurrent mutations are marked with the symbol "<". Uncertain positions in the mtDNA sequences were not considered when constructing the tree. Ancient samples are highlighted with a blue background. Geographic origin is indicated for each sample, and chronological age is provided for ancient samples (Table 2).

Kyrgyz, and Karakalpaks. KhT03 is projected between them alongside Astrakhan and Stavropol Nogais. It can be noted that all three studied samples do not intersect with the cluster of modern Jewish populations, which group with other populations of the Middle East and Levant.

When compared to ancient populations of the early Iron Age, early and high Middle Ages (Fig. 4*b*), the studied samples cluster with early Iron Age and medieval steppe nomads from the territories of present-day Kazakhstan, Mongolia, China, and Hungary. All individuals project onto the Avar cluster from the territory of Hungary.

To determine the ancestral components constituting the genomes of the studied samples from Phanagoria, an ADMIXTURE analysis was conducted in the context of

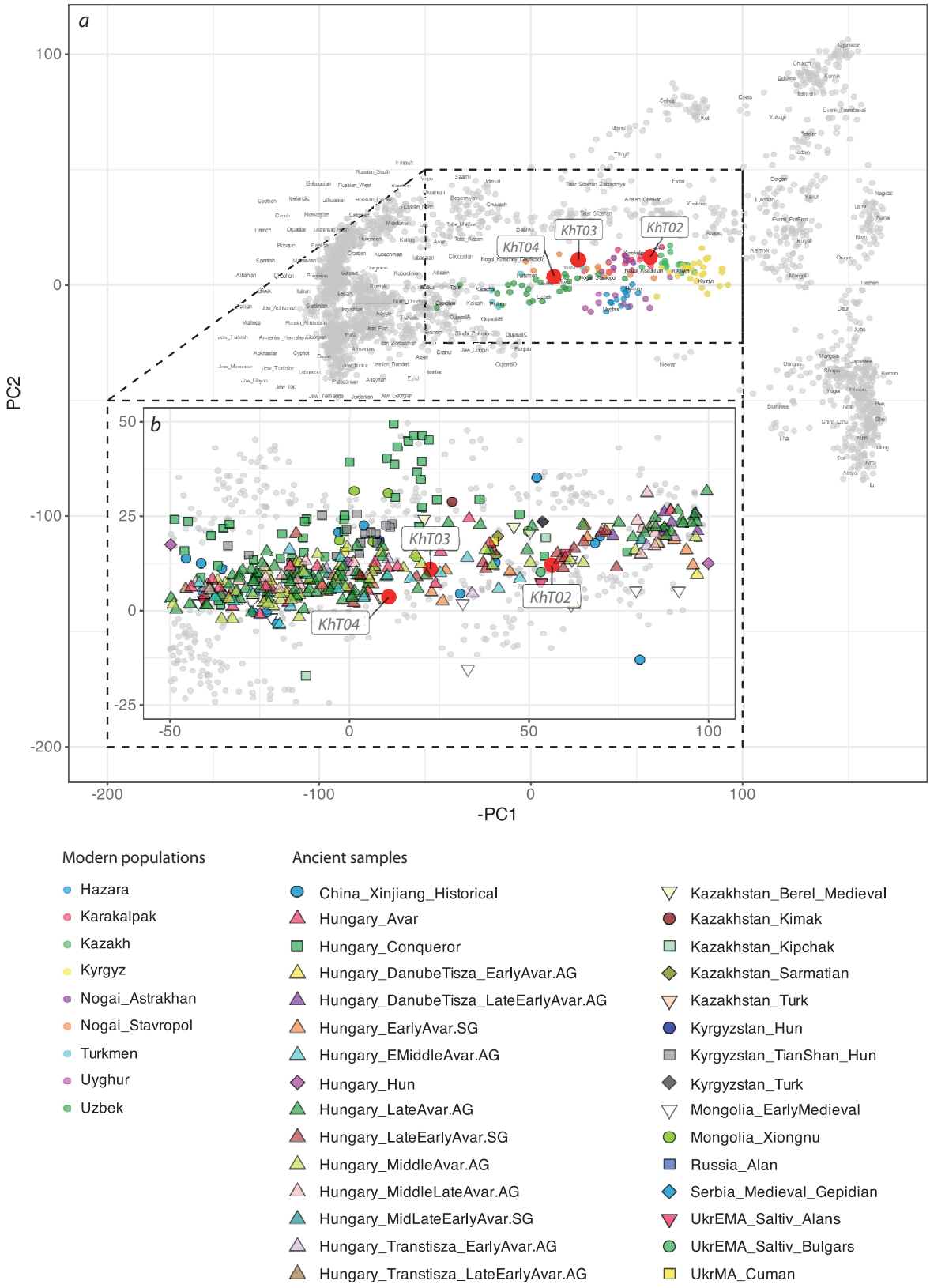


Fig. 4. Projection of genomic profiles of the studied Khazar-period individuals from Phanagoria (KhT02–KhT04) (marked in red) in the principal component space (PC1 and PC2), formed based on the genetic profiles of representatives of modern Eurasian populations (gray points), with colored points marking samples of modern Hazaras, Karakalpaks, Kazakhs, Nogais, Turkmen, Uyghurs, and Uzbeks (*a*); projection of genomic profiles of the studied individuals with ancient groups from open sources with an average dating from 1,700 years ago to 700 years ago (*b*).

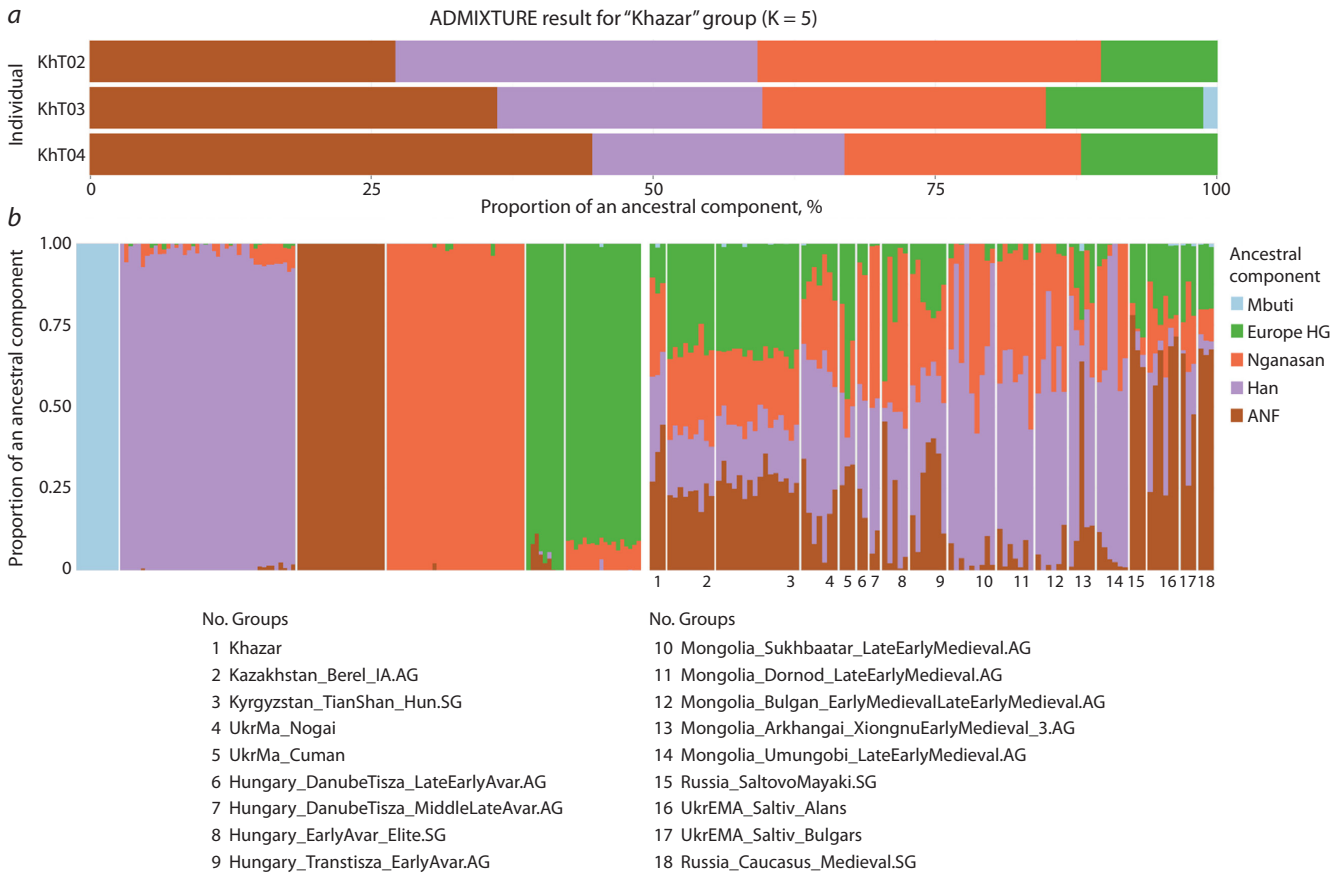


Fig. 5. ADMIXTURE results (K = 5).

a – for the studied samples KhT02–04; *b* – for six groups, where one of the ancestral components is maximized, and for some groups of Iron Age and medieval nomads (Supplementary Material 1).

ancient Neolithic, Bronze Age, Iron Age, and medieval populations (Fig. 5). In the KhT samples, four components are identified, named according to the groups in which each predominates: Anatolian Neolithic Farmers (ANF) (Turkey_Marmara_Barcin_N.AG), European Hunter-Gatherers (Europe HG) (France_Mesolithic.AG, Russia_YuzhniyOleniy Ostrov_Mesolithic.AG), East Asian (Han), and Ancient Siberian component (Nganasan).

Analysis using the *f*3-statistics method showed that samples KhT02 and KhT03 have genetic profiles similar to those of Avars from Hungary (Table 3, Fig. 6). For the KhT03 individual, a similarity is observed with the group Avar_Asia_Core2, representing early Avars from the 6th–8th centuries from the cemeteries of Apátfalva-Nagy-út-dűlő, Budapest-Csepel-Kavicsbánya, Csólyospálos-Felsőpálos, Fajsz-Garadomb, Felgyő-Ürmöstanya, Kiskőrös-Vágóhídidűlő, Kunpeszér-Felsőpeszériút-Homokbánya, Makó-Mikócsa-halom in Hungary with a predominant East Asian component (Maróti et al., 2022). For the KhT04 individual, a high value of *f*3-statistics, in addition to Avars, is observed for the group of the medieval population of South Mongolia from the cemeteries of Banzart-Khairkhan, Erdene-Mountain, Ganzagad, Gun-Tarmagtai, Ikh-Uvgun, Shar-Tolgoi in Umungobi Province, Mongolia (Mongolia_Umungobi_EarlyMedieval), as well as for the Cumans from the cemeteries of Mamai-Gora and Velika Znamyanka in Zaporizhzhia and Kumy in Kharkiv Oblast (UkrMA_Cuman) (Saag et al., 2025).

Discussion

Studies of skeletal remains from small burial mounds near the Khazar city of Sarkel, in the area of the modern Tsimlyansk Reservoir, reported southern Siberian traits characteristic of Turkic and Cuman groups (Ginzburg, 1946). In the burial mounds of the Lower Don region, associated with the ethnic Khazars (Semenov, 1983; Novoseltsev, 1990), the buried individuals were predominantly described in the anthropological literature as having Mongoloid cranial features, and "male Mongoloid skulls and skulls from female samples show the greatest similarity to the Huns of Transbaikalia and the nomadic Turks of Siberia, Altai, and Kazakhstan" (Batieva, 2002, p. 99). Anthropological studies of remains from Phanagoria during the Khazar period note the presence of a Mongoloid component (Dobrovolskaya, Svirkina, 2018).

PCA results for the studied Khazar-period samples from Phanagoria, in the context of modern and ancient populations, confirm the conclusions of anthropology and support the presence of an Asian-related component in their genomes. The genetic profile of samples from the Khazar period in the

Table 3. Values of the *f*3-statistic *f*3(KhT, Test; Mbuti.DG) for Khazar period samples and comparison groups

Studied sample	Comparison group	Out group	*f*3	stderr	Zscore	nsnps
KhT02	Hungary_DanubeTisza_EarlyAvar.AG	Mbuti	0.291885	0.004902	59.538	33322
	Hungary_DanubeTisza_MiddleLateAvar.AG	Mbuti	0.290993	0.005677	51.258	29211
	Avar_Asia_Core2	Mbuti	0.286955	0.004871	58.914	33559
	Mongolia_Umungobi_EarlyMedieval	Mbuti	0.28378	0.005348	53.059	30283
	Mongolia_EarlyMedieval	Mbuti	0.283269	0.00449	63.086	35545
KhT03	Avar_Asia_Core2	Mbuti	0.279129	0.00303	92.117	161020
	Hungary_DanubeTisza_EarlyAvar.AG	Mbuti	0.278325	0.003062	90.908	159674
	Hungary_DanubeTisza_LateEarlyAvar.AG	Mbuti	0.275194	0.00298	92.349	157505
	Mongolia_Selenge_EarlyMedieval	Mbuti	0.274352	0.003178	86.334	145533
	Hungary_DanubeTisza_MiddleLateAvar.AG	Mbuti	0.273691	0.003343	81.872	139562
KhT04	Mongolia_Umungobi_EarlyMedieval	Mbuti	0.278274	0.004533	61.386	41593
	Hungary_DanubeTisza_EarlyAvar.AG	Mbuti	0.277181	0.004095	67.69	46232
	Hungary_DanubeTisza_LateEarlyAvar.AG	Mbuti	0.276548	0.004191	65.987	45588
	UkrMA_Cuman	Mbuti	0.275737	0.004979	55.385	35218
	Avar_Asia_Core2	Mbuti	0.275528	0.004085	67.446	46719

Note. stderr (standard error) is the standard error of the *f*3-statistic estimate, characterizing the precision of its calculation; Z-score is the ratio of the *f*3 value to its standard error (*f*3/stderr), reflecting the statistical reliability of the estimate — the higher the value, the more reliable the result; nsnps is the number of single-nucleotide polymorphisms (SNPs) used to calculate the *f*3-statistic for a given pair of samples.

Lower City of Phanagoria is similar to that of modern Turkic-speaking peoples of Central Asia and the historical nomadic population of the Eurasian steppe, but this similarity does not necessarily indicate direct common origin.

The steppe population was formed as a result of complex processes involving carriers of Asian and European genetic components. The position of the KhT samples in PCA space likely reflects admixture processes similar to those that shaped the genomes of steppe populations. At the same time, it should be noted that the three studied samples are located at a certain distance from each other in the principal component space. Sample KhT04 is closer than the others to the cluster of groups with a known presence of West Eurasian roots, KhT02, to the steppe Turkic groups (Kazakhs, Kyrgyz), and KhT03 is in between them. Comparison with ancient samples showed that all three samples fall into a large cluster of Avars from the territory of Hungary (550–900 CE).

Analysis of mtDNA uniparental markers (Fig. 3) showed that the KhT02 sample belongs to haplogroup C4a1a-a (which diverged about 3.4 thousand years ago in Southern Siberia (Derenko et al., 2010; YFull, 2025)), which is widespread among the modern populations of China, Siberia, Afghanistan, and Pakistan (Van Oven, Kayser, 2009; YFull, 2025), and was also found in the population of the Khovsgol Mongolian aimag of the Bronze Age, among the Huns and Xiongnu of the early Iron Age from the territory of Kazakhstan, early medieval Visigoths from the territory of Spain, and the population of the Irkutsk region of the early Bronze Age. Phylogenetic analysis showed that the sequences most closely related to the mtDNA of this individual are found in modern samples.

Sample KhT03 belongs to haplogroup B4b1a3a (which diverged about 6.8 thousand years ago (YFull, 2025)), which is widespread among the modern populations of Southern and Eastern Siberia, while among ancient groups it was found in the Avars of the 6th–9th centuries from Hungary, in the early medieval Huns from the Berel burial ground in Kazakhstan, and in the early Iron Age population of Mongolia. The mtDNA sequences most closely related to the mtDNA of the KhT03 individual were found in modern populations of Siberia, Turkey, as well as among the Avars.

Sample KhT04 belongs to haplogroup T2d2a (which diverged about 5.4 thousand years ago (YFull, 2025)), which is widespread in modern Turkey, the Mediterranean, and the Caucasus. Among ancient groups, this lineage was found in individuals from late antique Rome, from the Iranian burial sites of Dinkha-Tepe and Shahr-i Sokhta of the Bronze Age and the Iron Age burial site of Hasanlu, an individual from the Early Bronze Age Lake culture in Bulgaria, an Etruscan from the Early Iron Age in Italy, and a Magyar from the 10th–11th centuries in Hungary. Phylogeographic analysis showed that the mtDNA sequence of the KhT04 sample is most closely related to modern samples from Hungary and Italy, as well as to a medieval sample (10th–11th centuries) from Hungary.

Thus, both based on the analysis of mtDNA uniparental markers and the analysis of autosomal profiles of the studied

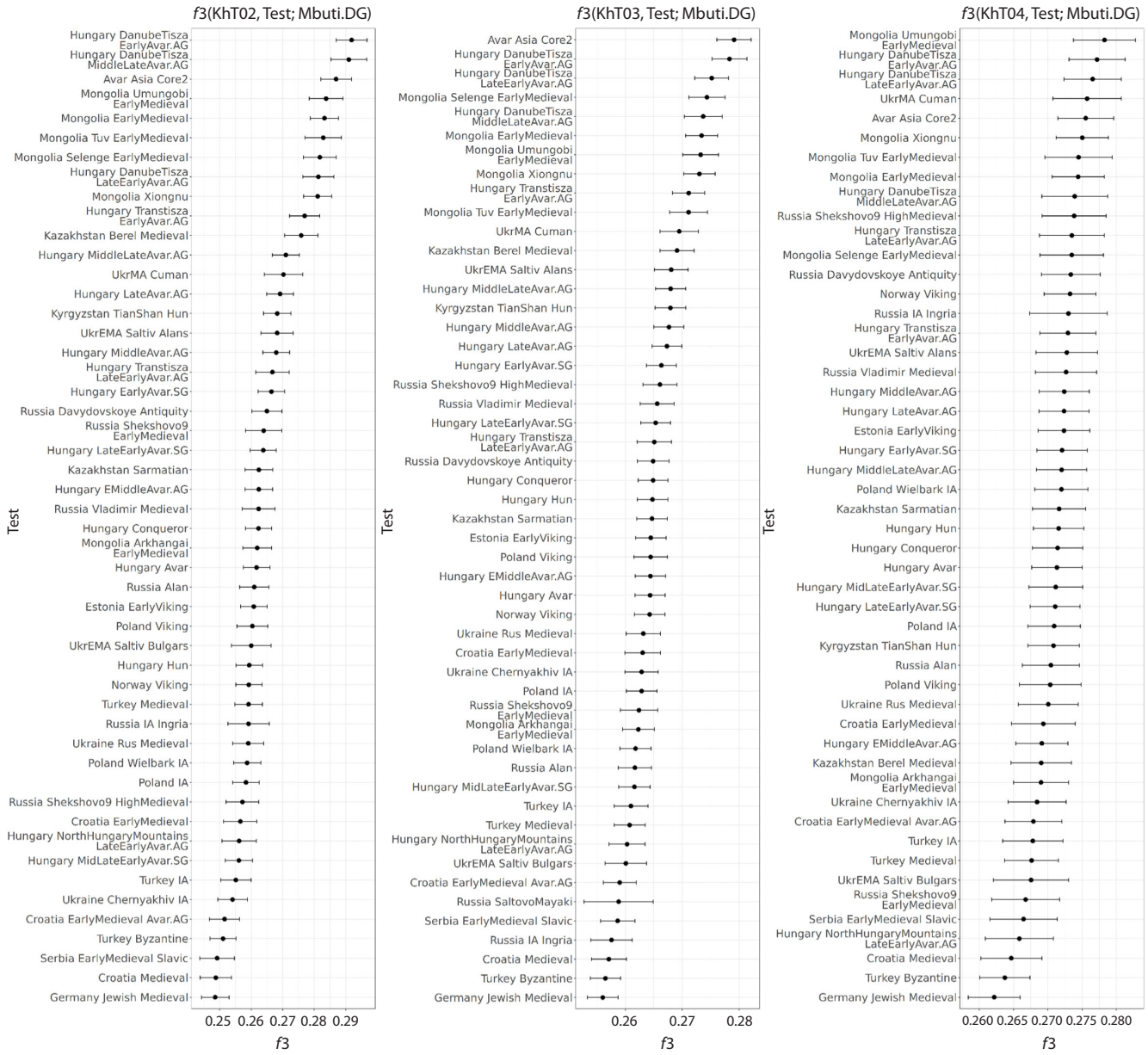


Fig. 6. *f*3-statistics for samples KhT02, KhT03, and KhT04 compared with ancient groups from the Early Iron Age to the Middle Ages.
The higher the *f*3 values, the more genetically similar the comparison group and the studied sample are.

samples, KhT02 and KhT03 show affinities with Central Asian and Siberian populations, whereas KhT04 shows stronger affinity with West Eurasian populations, whose descendants migrated to Europe as a result of migration processes in the early Iron Age.

ADMIXTURE analysis indicates that more than half of the ancestry of the studied Phanagoria samples is represented by Asian-related components: the East Asian Han component and the ancient Siberian Nganasan, which are represented in equal proportions in the genomes of medieval samples from Mongolia (Mongolia_Sukhbaatar_LateEarlyMedieval.AG, Mongolia_Dornod_LateEarlyMedieval.AG). Moreover, significant proportions of the components Europe HG and ANF, which are widespread in European and Near Eastern groups, as well as found among the nomadic groups of Central Asia, are also present.

The proportions of these components make the Phanagoria samples most similar to Avar groups of the 6th–9th centuries from the cemeteries in northeastern Hungary, Hungary_DanubeTisza_LateEarlyAvar.AG and Hungary_Transtisza_EarlyAvar.AG, which may indicate similar admixture processes. The difference in the proportions of the ANF component between the studied individuals corresponds to their positions on the PCA projection: the individual KhT04, clustering closer to the European clusters, has a higher proportion of the Anatolian ANF component, whereas the individual KhT02 has

a lower proportion of this component, which aligns with its position on the PCA graph. The increase in the proportion of ANF may be the result of mixing with groups from the Black Sea and Transcaucasia, where this component predominates, as seen in the studied groups of Alans from the Saltovo-Maiaki culture (Russia_SaltovoMayaki.SG) and the medieval population of Anapa (Russia_Caucasus_Medieval.SG).

The *f*3 results show that all three samples have genetic affinities with early and elite Avar groups, which exhibit stronger Siberian and East Asian components than others. This may suggest a shared East Asian-related ancestry component for the Avars and the ancestors of the studied samples. The presence of similarity with the UkrMA_Cuman group, which has a component profile characteristic of nomadic groups in the western part of the Eurasian steppe (Saag et al., 2025), in individual KhT04 indicates this individual's mixed ancestry involving Asian and European-related components.

The small number of synchronous samples does not allow us to definitively assert whether the genetic profile of individuals KhT02–04 is typical for representatives of the Khazar community itself or another Turkic group of the early Middle Ages living in the territory of the khaganate. These individuals differ sharply from previously studied individuals from ancient Phanagoria (Manakhov et al., 2025; Rozhdestvenskikh et al., 2026), who show similarities with European populations. These genetic differences are consistent with the cultural and demographic changes documented in the history of the city.

In the Northern Black Sea region, Khazaria emerged after the collapse of the Western part of the Turkic Khaganate, whose population included genetically diverse components. Therefore, the study of the genetic structure of Khazaria is an extremely complex task. Previous cranial studies also emphasized this complexity, identifying several distinct anthropological components (Ginzburg, 1963). Moreover, cranial features demonstrate a connection with burial rites, indicating a relatively isolated existence of certain groups within Khazar society. Our materials do not allow us to determine whether the individuals included in the genetic analysis belonged to different social, ethnic, or cultural groups. Therefore, at the current stage of research, we find it difficult to link the genetic heterogeneity of the studied individuals with their original heterogeneity or with the existing differences between groups within Phanagoria during the Khazar period.

Conclusion

Our study supports previous anthropological observations indicating the presence of an Asian-related component among the inhabitants of Phanagoria during the Khazar period. The admixture processes underlying this genetic profile may have been similar to those observed in the Avars, who carried an East Asian-related substrate combined with West Eurasian components. However, in the absence of available whole-genome data from securely identified ethnic Khazars, it is not yet possible to determine whether the studied individuals belonged to the Khazars themselves or to other early medieval nomadic groups.

References

Afanasiev G.E. On the self-identification of the Khazar Khaganate in the 9th century (based on data of the defence system). *Kratkie Soobshcheniya Instituta Arkheologii = Brief Commun Inst Archaeol*. 2015;238:98-114 (in Russian)

Aibabin A.I. Written sources on Byzantine ports in the Crimea from the fourth to seventh century. *Uchenye Zapiski Tavricheskogo Natsional'nogo Universiteta imeni V.I. Vernadskogo. Seriya: Istoricheskiye Nauki.* 2013;26(2):156-166

Alaçamlı E., Naidoo T., Güler M.N., Sağlıcan E., Aktürk Ş., Mapelli I., Vural K.B., Somel M., Malmström H., Günther T. READv2: advanced and user-friendly detection of biological relatedness in archaeogenomics. *Genome Biol.* 2024;25:216. https://doi.org/10.1186/s13059-024-03350-3

Alexander D.H., Novembre J., Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655-1664. doi 10.1101/gr.094052.109

Anastasiadou K., Silva M., Booth T., Speidel L., Audsley T., Barrington C., Buckberry J., … Holst M., Loe L., Armit I., Schulting R., Skoglund P. Detection of chromosomal aneuploidy in ancient genomes. *Commun Biol.* 2024;7(1):14. doi 10.1038/s42003-023-05642-z

Andreeva T.V., Manakhov A.D., Gusev F.E., Patrikeev A.D., Golovanova L.V., Doronichev V.B., Shirobokov I.G., Rogaev E.I. Genomic analysis of a novel Neanderthal from Mezmaiskaya Cave provides insights into the genetic relationships of Middle Palaeolithic populations. *Sci Rep.* 2022a;12(1):13016. doi 10.1038/s41598-022-16164-9

Andreeva T.V., Patrikeev A.D., Manakhov A.D., Rogaev E.I. Genetic profiles of representatives of ancient populations of the Russian plain in the Iron and Middle ages. Database. № 2022623562. 2022b (in Russian)

Andreeva T.V., Manakhov A.D., Rogaev E.I. Genetic profiles of the population of Medieval Rus. Database. № 2023624133. 2023 (in Russian)

Artamonov M.I. History of the Khazars. Leningrad, 1962 (in Russian)

Balabanova M.A. Anthropology of the population of the Lower Volga region (late 5th – first half of the 9th centuries). In: Evglevsky A.V. (Ed.) Steppes of Europe in the Middle Ages. Donetsk: Donetsk National University Publ., 2005;55-72 (in Russian)

Balabanova M.A. Features of the anthropological composition of burial complexes of the Khazar period. In: Gokhman I.I., Gromov A.V. (Eds) Some Current Issues of Modern Anthropology. St. Petersburg, 2006;59-61 (in Russian)

Balabanova M.A. Anthropological structure of the population of the Eastern European steppes in the Khazar period. In: Matishov D.G. (Ed.) International Scientific Conference "The Population of Southern Russia Across the Ages" (The Don Readings in Physical Anthropology). Rostov-on-Don, 2013;76-78 (in Russian)

Batieva E.F. Antropology of Bottom Don's population in Khazar Studies. *Donskaya Arkheologiya*. 2002;(3-4):71-101 (in Russian)

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., Lee J.J. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4:7. doi 10.1186/s13742-015-0047-8

Chichurov I.S. Byzantine Historical Works: "Chronographia" by Theophanes, "Breviarium" by Nikephoros: Texts, Translation, Commentary. Moscow: Nauka Publ., 1980 (in Russian)

Chkhaidze V.N. Phanagoria in the 6th–10th Centuries. Moscow, 2012 (in Russian)

Damgaard P. de B., Marchi N., Rasmussen S., Peyrot M., Renaud G., Korneliussen T., Moreno-Mayar J.V., … Nielsen R., Sikora M., Heyer E., Kristiansen K., Willerslev E. 137 ancient human genomes from across the Eurasian steppes. *Nature*. 2018;557(7705):369-374. doi 10.1038/s41586-018-0094-2

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

Derenko M., Malyarchuk B., Grzybowski T., Denisova G., Rogalla U., Perkova M., Dambueva I., Zakharov I. Origin and Post-Glacial dispersal of mitochondrial DNA haplogroups C and D in Northern Asia. *PLoS One*. 2010;5(12):e15214. doi 10.1371/journal.pone.0015214

Dobrovolskaya M.V., Svirkina N.G. Anthropological materials of the Khazar period from Phanagoria. In: Phanagoria. Results of Archaeological Research. Vol. 6. Moscow, 2018;73-87 (in Russian)

Dunlop D.M. The History of the Jewish Khazars. Princeton University Press, 1954

Ehler E., Novotný J., Juras A., Chyleński M., Moravčík O., Pačes J. AmtDB: a database of ancient human mitochondrial genomes. *Nucleic Acids Res.* 2019;47(D1):D29-D32. doi 10.1093/nar/gky843

Eltsov N., Volodko N. MtPhyl: Software tool for human mtDNA analysis and phylogeny reconstruction [WWW Document]. URL: http://eltsov.org

Flegontov P., Altınışık N.E., Changmai P., Rohland N., Mallick S., Adamski N., Bolnick D.A., … O'Rourke D.H., Krause J., Pinhasi R., Reich D., Schiffels S. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature*. 2019;570(7760):236-240. doi 10.1038/s41586-019-1251-y

Fu Q., Mittnik A., Johnson P.L.F., Bos K., Lari M., Bollongino R., Sun C., … Caramelli D., Castellano S., Reich D., Pääbo S., Krause J. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol*. 2013;23(7):553-559. doi 10.1016/j.cub.2013.02.044

Gansauge M.-T., Meyer M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat Protoc.* 2013;8(4):737-748. doi 10.1038/nprot.2013.038

Ginzburg V.V. Anthropological data on the ethnogenesis of the Khazars. *Sovetskaya Etnografiya = Soviet Ethnogr.* 1946;2:81-86 (in Russian)

Ginzburg V.V. Anthropological composition of the population of Sarkel – Belaya Vezha and its origin. In: Artamonov M.I. (Ed.) Proceedings of the Volga-Don Archaeological Expedition. Materials and Research on the Archaeology of the USSR. Moscow; Leningrad, 1963;260-281 (in Russian)

Golden P.B. Khazar Studies. Budapest: Akadémiai kiadó, 1980

Golofast L.A. Early medieval amphora with a Hebrew inscription on the lead seal from Phanagoria. *Rossijskaya Archaeologiya = Russian Archaeology*. 2020;3:159-172. doi 10.31857/S086960630010958-2 (in Russian)]

HaploTree. 2025. URL: https://haplotree.info/home/ (accessed 12.1.25)

Jónsson H., Ginolhac A., Schubert M., Johnson P.L.F., Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics.* 2013;29(13):1682-1684. doi 10.1093/bioinformatics/btt193

Kalinina T.M. Three stages of existence and the fall of the Khazar Khaganate. In: The Khazars: Myth and History. Moscow: Mosty Kultury Publ., 2010;25-41 (in Russian)

Kuznetsov V.D. Phanagorian synagogue (preliminary report). *Rossijskaya Archaeologiya = Russian Archaeology*. 2024;3:117-127. doi 10.31857/S0869606324030085 (in Russian)

Kuznetsov V.D., Ostapenko S.N. Medieval Phanagoria: new discoveries. *Vestnik Drevney Istorii = J Ancient Hist*. 2024;84(2):386-399. doi 10.31857/S0321039124020077 (in Russian)

Lazaridis I., Patterson N., Mittnik A., Renaud G., Mallick S., Kirsanow K., Sudmant P.H., … Slatkin M., Pääbo S., Kelso J., Reich D., Krause J. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. 2014;513(7518):409-413. doi 10.1038/nature13673

Lazaridis I., Patterson N., Anthony D., Vyazov L., Fournier R., Ringbauer H., Olalde I., … Palamara P.F., Mallick S., Rohland N., Pinhasi R., Reich D. The genetic origin of the Indo-Europeans. *Nature*. 2025;639(8053):132-142. doi 10.1038/s41586-024-08531-5

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

Mallick S., Micco A., Mah M., Ringbauer H., Lazaridis I., Olalde I., Patterson N., Reich D. The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Sci Data*. 2024;11(1):182. doi 10.1038/s41597-024-03031-7

Manakhov A.D., Rozhdestvenskikh E.V., Aituganova E.D., Kunizheva S.S., Voroshilov A.N., Svirkina N.G., Ostapenko S.N., Kuznetsov V.D., Rogaev E.I. Genetic study of individuals buried in two tombs of the Classical period from the eastern necropolis of Phanagoria. In: Proceedings of the 7th (23rd) All-Russian Archaeological Congress. Vol. 3. Krasnoyarsk, 2025;370-372 (in Russian)

Maróti Z., Neparáczki E., Schütz O., Maár K., Varga G.I.B., Kovács B., Kalmár T., … Gulyás B., Kovacsóczy B.N., Gál S.S., Tomka P., Török T. The genetic origin of Huns, Avars, and conquering Hungarians. *Curr Biol.* 2022;32(13):2858-2870.e7. doi 10.1016/j.cub.2022.04.093

Mouttham N., Klunk J., Kuch M., Fourney R., Poinar H. Surveying the repair of ancient DNA from bones via high-throughput sequencing. *BioTechniques*. 2015;59(1):19-25. doi 10.2144/000114307

Nikitin A.G., Lazaridis I., Patterson N., Ivanova S., Videiko M., Dergachev V., Kotova N., … Workman J.N., Zalzala F., Mallick S., Rohland N., Reich D. A genomic history of the North Pontic Region from the Neolithic to the Bronze Age. *Nature*. 2025;639(8053):124-131. doi 10.1038/s41586-024-08372-2

Novoseltsev A.P. The Khazar State and its Role in the History of Eastern Europe and the Caucasus. Moscow: Nauka Publ., 1990 (in Russian)

Patterson N., Price A.L., Reich D. Population structure and eigenanalysis. *PLoS Genet.* 2006;2(12):e190. doi 10.1371/journal.pgen.0020190

Patterson N., Moorjani P., Luo Y., Mallick S., Rohland N., Zhan Y., Genschoreck T., Webster T., Reich D. Ancient admixture in human history. *Genetics*. 2012;192(3):1065-1093. doi 10.1534/genetics.112.145037

Petr M., Vernot B., Kelso J. *admixr* – R package for reproducible analyses using ADMIXTOOLS. *Bioinformatics*. 2019;35(17):3194-3195. doi 10.1093/bioinformatics/btz030

Picard Tools – By Broad Institute [WWW Document] 2019. URL: https://broadinstitute.github.io/picard/ (accessed 12.1.25)

Pletneva S.A. From Nomadic Camps to Cities. The Saltovo-Mayaki Culture. Moscow: Nauka Publ., 1967 (in Russian)

Pletneva S.A. Nomads of the Middle Ages. In Search of Historical Patterns. Moscow: Nauka Publ., 1982 (in Russian)

Rasmussen M., Guo X., Wang Y., Lohmueller K.E., Rasmussen S., Albrechtsen A., Skotte L., … Sicheritz-Pontén T., Villems R., Nielsen R., Wang J., Willerslev E. An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science*. 2011;334(6052):94-98. doi 10.1126/science.1211177

Renaud G., Slon V., Duggan A.T., Kelso J. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* 2015;16:224. doi 10.1186/s13059-015-0776-0

Rozhdestvenskikh E.V., Manakhov A.D., Voroshilov A.N., Svirkina N.G., Aituganova E.D., Kunizheva S.S., Andreeva T.V., Ostapenko S.N., Kuznetsov V.D., Rogaev E.I. Historical genetics of Phanagoria: genomic analysis of individuals from a Hellenistic tomb. *Kratkie Soobshcheniya Instituta Arkheologii = Brief Commun Inst Archaeol*. 2026;282:309-324. doi 10.25681/IARAS.0130-2620.282.309-324 (in Russian)

Saag L., Utevska O., Zadnikov S., Shramko I., Gorbenko K., Bandrivskyi M., Pavliv D., … Barrington C., Gilardet A., Macleod R., Skoglund P., Thomas M.G. North Pontic crossroads: mobility in Ukraine from the Bronze Age to the early modern period. *Sci Adv.* 2025;11(2):eadr0695. doi 10.1126/sciadv.adr0695

Sazanov A.V., Mogarichev Yu.M. The Bosporus and the Khazar Khaganate in the late 7th–early 8th centuries. In: Zhuravlev D.V., Lomtadze G.A. (Eds) Northern Black Sea Region in Antiquity and the Middle Ages. Moscow, 2006;117-129 (in Russian)

Schiffels S. stschiff/sequenceTools. 2025. URL: https://github.com/stschiff/sequencetools

Schönherr S., Weissensteiner H., Kronenberg F., Forer L. Haplogrep 3 – an interactive haplogroup classification and analysis platform. *Nucleic Acids Res.* 2023;51(W1):W263-W268. doi 10.1093/nar/gkad284

Schubert M., Ginolhac A., Lindgreen S., Thompson J.F., Al-Rasheid K.A., Willerslev E., Krogh A., Orlando L. Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics*. 2012;13:178. doi 10.1186/1471-2164-13-178

Schubert M., Lindgreen S., Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*. 2016;9:88. doi 10.1186/s13104-016-1900-2

Semenov I.G. The Romanovka burial and Don region sites of the pre-Saltovo culture of the Khazar period. In: Problems of Chronology of Archaeological Sites of the Steppe Zone of the North Caucasus. Rostov-on-Don, 1983;98-102 (in Russian)

Thorvaldsdóttir H., Robinson J.T., Mesirov J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 2013;14(2):178-192. doi 10.1093/bib/bbs017

Van Oven M., Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 2009;30(2):E386-E394. doi 10.1002/humu.20921

YFull [WWW Document] 2025. URL: https://www.yfull.com/ (accessed 12.01.25)

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

Received December 23, 2025. Revised December 29, 2025. Accepted December 29, 2025.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):775-781

doi 10.18699/vjgb-26-78

Analysis of *CFTR* mutation spectrum in Yugra region (Russian Federation)

M.Yu. Donnikov [1], A.V. Morozkina [1], L.V. Kovalenko [1], N.A. Donnikova [2], V.V. Mescheryakov [1]

[1] Medical Institute of Surgut State University, Surgut, KHMAO-Yugra, Russia
[2] Surgut Regional Clinical Center for Protection of Maternity and Childhood, Surgut, KHMAO-Yugra, Russia
donnikov@gmail.com

Abstract. Rapid progress in both molecular diagnostics tools and targeted therapy for cystic fibrosis (CF) requires clear understanding of *CFTR* mutation spectrum in population constantly living in any large region of any country. There are still no published results of comprehensive genetic study of CF patients from the Yugra region (Western Siberian part of Russia). Three-step molecular genetic approach was used for gDNA samples derived from a group of 54 CF patients living in the Yugra region: (1) AFLP/RFLP was used for the panel of 35 major *CFTR* mutations; (2) NGS, for full sequencing of exons and exon-intron boundaries of *CFTR* gene; and (3) the MLPA technique, to search for large gene rearrangements. The search of major *CFTR* mutations in multiethnic Yugra population accounted for only 76.6 % of all CF-causing mutations. Most common were variants F508del (rs113993960), 1677delTA (rs121908776), CFTRdele2,3, E92K (rs121908751), R1066C (rs78194216). Genomic DNA samples from 19 CF patients from the tested group in which only one mutant allele was identified were analyzed by the NGS approach followed by MLPA. This allowed to reveal 19 rare pathogenic *CFTR* variants, including five new variants absent in CF mutation databases. Generally, three-tier molecular genetics approach applied for the first time for a Yugra-originated cohort of CF patients allowed to find 12 missense mutations, 9 nonsense ones, 6 in/dels, 4 splice-site variants, 1 CNV. Three alleles remained unidentified. Deciphering the genetic background for CF patients living in Yugra allowed to support rational administration of CFTR modulators with profound clinical effect. The need to develop a region-specific testing panel for common *CFTR* mutations is demonstrated.

Key words: cystic fibrosis; *CFTR* gene; common and rare mutations; KHMAO-Yugra

For citation: Donnikov M.Yu., Morozkina A.V., Kovalenko L.V., Donnikova N.A., Mescheryakov V.V. Analysis of *CFTR* mutation spectrum in Yugra region (Russian Federation). *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(5):775-781. doi 10.18699/vjgb-26-78

Funding. The study was conducted with the financial support of the Fund of Scientific and Technological Development of Yugra (scientific project # 2023-574-05).

Acknowledgements. We deeply appreciate the aid of pulmonologist Natalya Satsuk from Nizhnevartovsk Regional Children Hospital in data retrieval from the Regional CF Patients Registry.

Анализ спектра мутаций гена *CFTR* в ХМАО-Югре

М.Ю. Донников [1], А.В. Морозкина [1], Л.В. Коваленко [1], Н.А. Донникова [2], В.В. Мещеряков [1]

[1] Медицинский институт Сургутского государственного университета, Сургут, ХМАО-Югра, Россия
[2] Сургутский окружной клинический центр охраны материнства и детства, Сургут, ХМАО-Югра, Россия
donnikov@gmail.com

Аннотация. Быстрый прогресс как в молекулярной диагностике, так и в таргетной терапии муковисцидоза (МВ) требует четкого понимания спектра мутаций в гене *CFTR* у населения, постоянно проживающего в крупном регионе любой страны. До сих пор нет опубликованных результатов ни одного комплексного генетического исследования пациентов с МВ из ХМАО-Югры (Западно-Сибирская часть России). Образцы геномной ДНК, полученные от группы из 54 пациентов с МВ, проживающих в Югре, были изучены с использованием трехэтапного молекулярно-генетического подхода: 1) ПДРФ-анализ для панели из 35 мажорных мутаций в гене *CFTR*; 2) NGS для полного секвенирования экзонов и экзон-интронных границ гена *CFTR*; 3) метод MLPA для поиска крупных генных перестроек. Поиск мажорных мутаций в гене *CFTR* в многоэтнической популяции Югры позволяет выявить лишь 76.6 % от всех мутаций, вызывающих МВ. Наиболее распространенными являются варианты F508del (rs113993960), 1677delTA (rs121908776), CFTRdele2,3, E92K (rs121908751), R1066C (rs78194216). Геномная ДНК 19 пациентов с МВ из исследуемой группы, у которых идентифицирован только один мутантный аллель, была проанализирована с помощью метода NGS с последующим исследованием методом MLPA. В результате выявлены 19 редких патогенных вариантов гена *CFTR*, включая пять новых вариантов, информация о которых отсутствует в базах данных мутаций гена *CFTR*. В целом трехступенчатый молекулярно-генетический

подход, впервые примененный для когорты пациентов с МВ из ХМАО-Югры, позволил обнаружить 12 миссенс-мутаций, 9 нонсенс-мутаций, 6 инсерций/делеций, 4 варианта сайта сплайсинга и 1 CNV. Три аллеля остались неидентифицированными. Расшифровка генетического груза пациентов с МВ, проживающих в ХМАО-Югре, позволяет обосновать рациональное назначение модуляторов белка CFTR с выраженным клиническим эффектом. Показана необходимость разработки регион-специфической панели тестирования на распространенные мутации в гене *CFTR*.

Ключевые слова: муковисцидоз; ген *CFTR*; частые и редкие мутации; ХМАО-Югра

Introduction

Yugra is officially known as Khanty-Mansi Autonomous Okrug – Yugra and is located in the Western Siberian part of the Russian Federation, with a predominantly urban population of 1.7 million people scattered over an area of 534,000 km^2 in several large towns and a string of small settlements. The indigenous population (Khanty and Mansi) currently comprises only 27,000 people. Rapid development of oil and gas production over the last 40 years led to radical changes in population structure. The incoming migrant population is characterized by ethnic heterogeneity and a high proportion of Turk-originated and Finno-Ugric cohorts with a predominance of people of Slavic origin. The formation of an ethnically and genetically heterogenous population creates a need of genetic studies aimed to reveal regional features of hereditary disorders.

Among these, cystic fibrosis (CF, OMIM #219700) is one of the most frequent genetic conditions with an autosomal recessive type of inheritance; it is common among Europeans and residents of North America (Ong, Ramsey, 2023; Amaral, Pankonien, 2025). CF is a multiple organ pathology caused by mutations in the *CFTR* gene encoding the chloride channel (Callebaut et al., 2018). Until recently, the diagnosis of CF was determined by characteristic clinical manifestations of the respiratory system and exocrine glands, indicating the onset of irreversible pathological changes (Sosnay et al., 2017b). With the spread of neonatal (newborn) screening (NBS), the diagnosis of CF becomes presymptomatic (Farrell et al., 2017), which largely determines the early initiation of therapy, prevents the development of severe complications, and improves prognosis.

Currently, more than 2,300 mutations of the *CFTR* gene are known (Stepanova et al., 2025), and their spectrum is strongly population-specific, which makes it difficult to correctly diagnose CF in heterogeneous populations (Petrova et al., 2020; Ayupova et al., 2024). A complete genetic diagnosis is of particular relevance in connection with optimized genetic counseling and facilitation of mutation-specific targeted CF therapy in clinical practice (Parisi et al., 2025; Terlizzi, Lopes-Pacheco, 2025). The goal of CF genetic diagnostics is to detect at least one clinically significant mutation in the *CFTR* gene in the majority of patients with clinical symptoms of CF (Sosnay et al., 2017a), which is ensured by using a panel of frequent (major) mutations. Federal genetic centers in the Russian Federation use panels of 33 mutations characteristic of the Russian population (Petrova et al., 2020), which identify up to 85 % of mutant CF alleles as described previously (Petrova et al., 2016). The search for the second mutation for patients from remote Yugra locations was until recently hampered by a number of resource and logistics issues and thus was performed mostly retrospectively.

According to annually updated data from CF regional registry, there are 54 patients with confirmed CF diagnosis (as of June 2025), and each year this number is gradually increased by 3–4 children. Since 2006, CF has been one of the five diseases monitored through newborn screening (NBS) program in Russia, and CF diagnostics begins immediately after birth by continuous screening examination (Therrell et al., 2024). In Yugra, as throughout Russia, mass screening of newborns for CF has been carried out as part of a national project, an important component of which is the genetic diagnostics of CF. The results of many years of continuous research make it possible to calculate and predict the incidence of CF, assess the effectiveness of the genetic diagnostics of CF NBS. In this regard, there is a strong need to study the structure of *CFTR* mutations in a cohort of patients in a given territory, which determines the necessity for a complete examination of patients using a method of gene diagnostics available for regional conditions. Thus, the aim of the study was to describe the range of *CFTR* mutations revealed in a local Russian region characterized by a small general population scattered over a large territory and having a healthcare system capable of applying modern genetic technologies to study hereditary diseases like CF.

Materials and methods

The study was approved by the Local Ethics Committee of the Medical Institute of Surgut State University (protocol No. 6, February 12, 2025). All participants (or their official representatives) gave their informed consent. Clinical and demographic characteristics of the studied CF patients living in Yugra are listed in Table 1. All EDTA blood samples were retrieved as frozen aliquots from the regional biobank located in the Medical Institute of Surgut State University.

AFLP/RFLP. In-house amplified fragment length (AFLP) and restriction fragment length polymorphism (RFLP) techniques were used to detect insertion/deletion variants and nucleotide substitutions, respectively (Ivaschenko et al., 1995). The panel included 35 CF-causing mutations (legacy names given) relevant for the North-Western region of Russia: F508del (rs113993960), I507del (rs121908745), CFTRdele2,3, E92K (rs121908751), 3849+10kb C>T (rs75039782), 1677delTA (rs121908776), N1303K (rs80034486), W1282X (rs77010898), L138ins (rs397508686), G542X (rs113993959), 394delTT (rs121908769), R334W (rs121909011), W1282R (rs766509028), S1196X (rs886061212), 2143delT (rs121908812), 1367del5 (rs397508184), 2183AA>G (rs121908799), R1066C (rs78194216), W1310X (rs397508645), 712-1G>T (rs121908793), 621+1G>T (rs78756941), R553X (rs74597325), L1335P (rs397508658), 4015delA (rs1553429051), R1162X (rs74767530),

3120+1G>A (rs75096551), 2184insA (rs121908769), 2184delA (rs121908798), 2113delA, 2118del4, 2176insC, 2183delAA (rs121908769), R347P (rs77932196), G551D (rs75527207), W1282G. In cases where only one *CFTR* mutation was found, we performed NGS of the entire gene. Positive findings were confirmed in parents to determine mutant allele segregation.

Next generation sequencing (NGS) was performed on the MiSeq platform (Illumina, USA) with library preparation based on NimbleGen SeqCap EZ Choice: 151012_HG38_CysFib_EZ_HX3, fully covering the *CFTR* gene. Reference genome NCBI37/hg19 was used for NGS data analysis. Genetic variants were identified by BWA (http://bio-bwa.sourceforge.net/), GATK (https://www.broadinstitute.org/gatk/), SNPeff (http://snpeff.sourceforge.net/), Basespace Illumina (https://basespace.illumina.com/apps/). Variant pathogenicity was assessed using data from 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), dbSNP (http://www.ncbi.nlm.nih.gov/SNP/), ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/), OMIM (http://omim.org/), HGMD (http://www.hgmd.cf.ac.uk/ac/index.php). Final results were verified by direct Sanger sequencing using an ABI 3130 genetic analyzer (Applied Biosystems, USA).

MLPA. A search for large rearrangements (deletions/duplications – CNV) involving the *CFTR* locus in chromosome region 7q31.2 was performed using multiplex ligation-dependent probe amplification (MLPA) in cases where no pathogenic variant was found. MLPA was performed using SALSA MLPA probemix P091-D2 CFTR (MRC-Holland, the Netherlands) according to the manufacturer's recommendations. The MLPA results were analyzed using Coffalyser.Net (MRC-Holland).

To assess the disease course in patients living in Yugra and included in regional CF registry, the following key clinical parameters were taken into account: the patient age at follow-up (first half of 2025), sweat chloride concentration (mmol/L), body mass index (BMI) (kg/m^2), lung function estimation (FEV_1, % predicted), microbiological status of the lower respiratory tract.

Statistical analysis was performed using the STATISTICA 8.0 program. To compare the observed categorical variables, the Fisher test was used, while for quantitative tests, the Mann–Whitney test was utilized. Results were considered significant at $p \leq 0.05$.

Results

As of June 2025, 54 CF patients were officially registered in Yugra (Table 1). The long-term mean annual prevalence (2016–2024) was 48.6 cases, and the incidence was one case

Table 1. Demographic, clinical and brief genetics characteristic of the CF patients studied

Parameter	Group 1 (without NBS), *N* = 11	Group 2 (with NBS), *N* = 43	Statistics
Age at follow-up, years			
M ± SD	22.0 ± 5.8	6.9 ± 4.2	<0.01*
Me (Q25; Q75)	19.0 (18.5; 25.6)	7.2 (3.1; 10.5)	
Sex distribution, *n* (%)			
Males	6 (54.5)	21 (47.7)	>0.05†
Females	5 (45.5)	23 (52.3)	
Age of diagnosis, years			
Me (Q25; Q75)	5.90 (0.79–7.56)	0.28 (0.14–1.25)	0.001*
Sweat test			
Sweat conductivity, mmol/L			
Me (Q25; Q75)	101.5 (98.9–113.4)	102.0 (77.5–108.5)	0.82*
BMI, kg/m^2			
Me (Q25; Q75)	20.1 (18.4; 21.4)	19.4 (17.7–19.9)	0.045*
Lung function			
FEV_1, % of predicted			
Me (Q25; Q75)	93.0 (86.5–95.0)	100.0 (95.0–111.3)	0.011*
Microbiology			
Intermittent *Pseudomonas aeruginosa*, *n* (%)	10 (45.5)	4 (12.5)	<0.05†
Chronic *Staphylococcus aureus*, *n* (%)	16 (72.7)	22 (68.8)	<0.05†
CFTR mutation testing			
Three mutations known (complex allele), *n* (%)	0	2 (4.7)	0†
Two mutations known, *n* (%)	9 (81.8)	40 (93.0)	>0.05†
One variant known, *n* (%)	11 (100.0)	43 (100.0)	>0.05†
F508del/F508del, *n* (%)	1 (9.1)	9 (20.9)	<0.05†
F508del/non-F508del, *n* (%)	8 (72.7)	25 (58.1)	>0.05†
non-F508del/non-F508del, *n* (%)	2 (18.2)	9 (20.9)	>0.05†

* Mann–Whitney U test (when comparing median values).
† Fischer's exact test (when comparing %).

per 9,025 newborns per year, which corresponds to nationwide data (1:9,000) (Registry…, 2025).

The diagnosis of CF was confirmed according to the guidelines of the European Cystic Fibrosis Society as well as the Russian National Consensus on Cystic Fibrosis: for adult patients, the presence of typical pulmonary or gastrointestinal symptoms was established, for patients with NBS, at least one of the following was found: two positive sweat chloride tests, or the identification of two *CFTR* mutations *in trans* (Kondratyeva et al., 2019; Castellani et al., 2023).

The studied cohort consisted of 54 patients, mostly children (45 individuals) and 9 adults (the oldest was born in 1988). To demonstrate NBS efficiency, we subdivided the cohort in two subgroups depending on NBS availability (Table 1). All of the known CF patients residing in Yugra and included in regional CF registry were genetically studied in consecutive steps. As the first step, we used the testing panel of our colleagues from St. Petersburg (Glotov O., personal communication) developed for routine testing of CF-suspected individuals living in Northwestern Russia. Samples with unidentified mutant alleles underwent NGS of the *CFTR* gene. As the last step, MLPA was performed to confirm or identify large gene rearrangements.

Analysis of the patient cohort using a panel of 35 *CFTR* mutations common in Northwestern Russia revealed a particularly high frequency of two mutations in the Yugra population: c.1521_1523delCTT (p.Phe508del; F508del) – 53/111 alleles (47.7 %), and c.1545_1546delTA (p.Tyr515X; 1677delTA) – 8/111 alleles (7.2 %) (Table 2). Overall, only 15 out of the 35 mutations included in the panel were detected (Table 2), whereas 20 mutations were not found at all. Thus, the overall efficiency of the panel was detection of 85/111 alleles (76.6 %), which is less than that for a similar panel developed by our colleagues from Moscow (83 %); it obviously needs further customization for regional-specific *CFTR* variants (Petrova et al., 2020).

Sequencing of all the samples confirmed the results of panel testing and also additionally identified 19 rare mutations, 15 of which were unique (reported in individual families). Using the MLPA method, only one type of large rearrangement of the *CFTR* gene (c.54-5940_273+10250del21kb, CFTRdele2,3) was found in 5 unrelated patients, which constitutes 5/111 (4.5 %) of all tested mutant alleles. In three patients (two adults and one kid) with a confirmed CF diagnosis, only one mutation was revealed despite the use of NGS and MLPA. Thus, three mutant alleles remain unidentified and require further in-depth study.

As a result, 34 pathogenic genetic variants in the *CFTR* gene were detected (Table 2). Among these, there are 12 missense mutations, 9 nonsense mutations, 5 frameshift deletions, 1 frameshift insertion, 4 splice-site mutations, 1 in-frame insertion, and 1 CNV. Frequent mutation E92K (prevalent for Chuvash nationality) was attributed to both missense and splice-site mutation classes.

Mutation F508del (c.1521_1523delCTT) is among the most common *CFTR* variants worldwide (Bobadilla et al., 2002), and it is the most frequent in Russian CF patients (52.61 % in the 2023 registry (Registry…, 2025)). Its frequency in Yugra (47.7 %) precedes the lowest number observed in the North Caucasian region of Russia (27.1 %). F508del homozygotes were observed in 10 patients, F508del in a heterozygous compound state was observed in 33 patients (61.1 %).

The second most frequent mutation identified (1677delTA) was described previously as common for Georgian patients (Petrova et al., 2021); but in Yugra, these eight alleles are distributed among patients of Chechen origin (two siblings and one unrelated proband are homozygous and two unrelated patients are heterozygous for this variant). These cases changed the expected distribution of *CFTR* mutations within the Yugra population, in spite of the fact that the Chechens comprise only 0.41 % of the Yugra population structure (Rosstat data, available online). A possible explanation is health-related internal migration due to a high level of medical care for CF patients in Yugra compared to North Caucasus. Before these families relocated to Yugra, the second most frequent variant had been the Slavic-associated CFTRdele2,3 variant.

The reported predominant distribution of the CFTRdele2,3 (c.54-5940_273+10250del21kb) variant for Eastern European populations (Dörk et al., 2000) is confirmed in our study, but compared to the All-Russian CF registry, this mutation is the third by frequency in Yugra: 5/111 alleles (4.5 % vs 6.15 %, respectively), preceded by the 1677delTA mutation.

The fourth most frequent mutation is E92K (c.274G>A) – 4/111 alleles (3.6 %). It has been described in the Chuvash population (Stepanova et al., 2012). In our study all four alleles were detected in the heterozygous state in unrelated patients of Russian origin.

The fifth most frequent mutation is R1066C (c.3196C>T) – 3/111 alleles (2.7 %) – revealed in a homozygous state in one CF patient and in one unrelated heterozygous patient, both of non-Slavonic nationalities (Azerbaijani and Dagestan).

Five other relatively frequent mutations (L138ins, S466X(TAA), R1070Q, L467F, R668C) were observed in 2/111 alleles (1.8 %) each in a compound heterozygous state. Moreover, S466X(TAA) mutation formed a complex allele with R1070Q, which was revealed in two siblings. L138ins was also revealed in two siblings. N1303K was found in a homozygous state in one proband. The remaining 23 mutant alleles were private and were observed only once in a compound heterozygous state.

Additionally, five novel variants absent from the main CF mutation databases (CFTR1, CFTR2, ExAC) were identified for the first time: three of them are nonsense mutations (c.252T>A (p.Tyr84X), c.831G>A (p.Trp277X), c.1580dupA (p.Glu528Argfs*40)), one mutation is missense (c.3983T>A (p.Ile1328Lys), and one mutation is a frameshift deletion (c.1845_1846delAA). Clinical significance of these newly found mutations is estimated as “likely pathogenic” (Zhang et al., 2020) and requires additional functional study. All five children carrying these variants *in trans* with F508del had high sweat chloride values and a pancreatic-insufficient phenotype with pulmonary manifestations typical of CF.

Over half of the studied alleles (62) corresponding to a relatively low variety of mutations *per se* (only four – F508del, E92K, R1066C, N1303K) are attributed to the II class of *CFTR* mutation. Among these, ten patients were homozygous and 33 patients were heterozygous for F508del. That makes a

Table 2. The *CFTR* gene variants identified in Yugra-living patients

	No.	cDNA variant	Protein change	Legacy name	Exon/intron	Mutation type	Mutation class	Allele number	%	% in Russian CF registry (2023)	% in ECFS registry (2023) §	% in CFTR2 DB ‡
	1	c.1521_1523delCTT	p.Phe508del	F508del	11e	i	II	53	**47.7**	52.61	60.41	69.70
	2	c.1545_1546delTA	p.Tyr515X	1677delTA	11e	sd	I	8	7.2	2.12		0.09
	3	c.54-5940_273+10250 del21kb	p.Ser18Argfs*16	CFTRdele2,3	2-3e	CNV	I	5	4.5	6.15	0.96	0.29
	4	c.274G>A	p.Glu92Lys	E92K	4e	m, s	II	4	3.6	3.25		0.03
	5	c.3196C>T	p.Arg1066Cys	R1066C	20e	m	II	3	2.7	0.38		0.15
	6	c.412_413insACT	p.Leu137_Leu138insThr	L138ins	4e	si	IV	2	1.8	1.53		0.01
	7	c.1397C>G#	p.Ser466X	S466X(TAA)	11e	n	I	2	1.8	0.59		0.03
	8	c.1399C>T	p.Leu467Phe	L467F	11e	m	?	2	1.8	–		
	9	c.2002C>T	p.Arg668Cys	R668C	14e	m	?	2	1.8	0.01		0.05
	10	c.3209G>A#	p.Arg1070Gln	R1070Q	20e	m	IV	2	1.8	0.59		0.01
	11	c.3909C>G	p.Asn1303Lys	N1303K	24e	m	II	2	1.8	1.53	2.18	1.58
	12	c.43delC	p.Leu15PhefsX1	175delC	1e	sd	I	1	0.9	0.07		
	13	c.53+1G>T (p.(=))	–	185+1G>T	1i	s	I	1	0.9	0.04		
†	14	c.252T>A	p.Tyr84X	–	3e	n	I	1	0.9	0.07		
	15	c.262_263delTT	p.Leu88Ilefs*22	394delTT	3e	sd	I	1	0.9	0.85		
	16	c.489+1G>T	–	621+1g>t	4i	s	I	1	0.9	0.20		
	17	c.653T>A	p.Leu218X	L218X	6e	n	I	1	0.9	0.01		
†	18	c.831G>A	p.Trp277X	–	7e	n	I	1	0.9	0.03		
	19	c.1000C>T	p.Arg334Trp	R334W	8e	m	IV	1	0.9	0.74		
	20	c.1040G>A	p.Arg347His	R347H	8e	m	IV	1	0.9	0.13		
†	21	c.1580dupA	p.Glu528Argfs*40	E527fs	11e	n	I	1	0.9	0.01		
	22	c.1624G>T	p.Gly542X	G542X	12e	n	I	1	0.9	1.46	2.75	2.54
	23	c.1704G>T	p.Leu568Phe	L568F	13e	m	?	1	0.9	0.01		
†	24	c.1845_1846delAA	p.Lys615fs	–	13e	sd	?	1	0.9	–		
	25	c.2353C>T	p.Arg785X	R785X	14e	n	I	1	0.9	0.17		
	26	c.2657+5G>A	–	2789+5G>A	16i	s	V	1	0.9	0.38	1.07	
	27	c.3208C>T	p.Arg1070Trp	R1070W	20e	m	IV	1	0.9	0.01		
	28	c.3454G>C	p.Asp1152His	D1152H	21e	m	IV	1	0.9	0.09		
	29	c.3718–2477C>T	–	3849+10kbC>T	22i	s	V	1	0.9	0.01	1.00	
	30	c.3485G>T	p.Arg1162Leu	3617G/T	22e	m	?	1	0.9	–		
	31	c.3816_3817delGT	p.Ser1273LeufsX28	3944delGT	23e	sd	I	1	0.9	0.29		
	32	c.3846G>A	p.Trp1282X	W1282X	23e	n	I	1	0.9	1.73	1.07	1.21
	33	c.3929G>A	p.Trp1310X	W1310X	24e	n	I	1	0.9	0.32		
†	34	c.3983T>A	p.Ile1328Lys	–	25e	m	?	1	0.9	0.01		
		Identified						108	97.3			
		Not identified						3	2.7	–		
		Total						111				

Note. Mutation types used: (m) – missense mutation, (n) – nonsense mutation, (sd) – frameshift deletion, (si) – frameshift insertion, (s) – splice-site, (i) – in-frame insertion, (CNV) – copy number variation (large rearrangement).
Complex allele S466X-R1070Q.
† Newly identified *CFTR* variants.
§ Out of 7 most frequent (Zolin et al., 2025).
‡ Out of 6 most frequent.

significant number of patients (at least 44) carrying these alleles eligible for targeted drug therapy. Currently, 36 children with CF living in Yugra are receiving targeted therapy and this number is rapidly growing. Unfortunately, there are 28 mutant alleles belonging to the I class (and carried by 10 patients) with potential targeted drugs still in development.

Discussion

According to the Russian Registry of CF patients in 2023 (RF CF Registry), among at least 275 pathogenic *CFTR* variants, only eleven mutations are the most frequent ones in Russia with frequencies exceeding 1 % in the tested sample: F508del shares 51.4 %, CFTRdele2,3 – 6.1 %, E92K – 3.7 %, 1677delTA – 2.5 %, 3849+10kbC>T – 2.2 %, 2143delT – 2.0 %, 2184insA – 1.9 %, W1282X – 1.7 %, L138ins and N1303K each sharing 1.7 %, G542X – 1.5 % (Registry…, 2025). It is conceivable that the spectrum of *CFTR* mutations varies in different regions depending on specific ethnic background of the population. The East Slavic population comprises 77.7 % of the population living in Russia, according to the 2010 census. Moreover, in the European part, Russians comprise 85–90 % of the population (Rosstat data). In Yugra, top three nationalities according to the latest statistical Rosstat data are: Slavonic (Russians, Ukrainians and Belarusians) – 74.1 %, Tatar – 6.3 %, Bashkir – 2.3 %. Indigenous people (Khanty and Mansi) comprise a small part: 1.5 and 0.9 %, respectively. To our knowledge, no CF patients of indigenous origin have been identified to date, in spite of equal involvement of all newborn children in neonatal screening.

A comparison of frequency profiles for 34 *CFTR* mutations revealed in Yugra with that of reported officially by Russian CF registry (data for 2023 (Registry…, 2025)) reveals a similarity of frequency distributions for Yugra patients and for patients from the All-Russian sample (Table 2), which is not surprising as ethnic Russians make up the majority of CF patients in the Russian Federation. The differences between values are determined by the much larger sample size in Russian CF registry than in Yugra regional CF registry (4,058 and 54 patients, respectively). Any small change in patient number reported in Yugra (for example, during updating genetics data for newly found newborns with CF) will inevitably cause the range of frequent regional mutations to be restructured. Comparison of mutation profiles in Yugra with those from European CF registry (Zolin et al., 2025) and the CFTR2 database (Sosnay et al., 2017a) appears to be inappropriate (see Table 2), except for the F508del mutation. This may reflect greater differences in population structures between Western Europe and Eastern territories of Russia.

Conclusion

The relatively small (1.78 million) size of the population constantly living in the Yugra region and its highly developed healthcare system allowed us to perform comprehensive genetic testing for all 54 CF patients currently included in regional CF registry. The shaping of CF patients' cohort in Yugra actually is the result of efficient NBS efforts which allowed to sieve up to 396,000 kids born since 2006 in terms of biochemical analysis of immunoreactive trypsinogen levels and sweat testing. Also, a comprehensive genetic analysis of CF patients became possible with the advent of NGS in routine laboratory work. Targeted CF therapy with CFTR modulators that started in 2022 in Yugra greatly improved both quality of life and prognosis for patients. The three-step model of genetic testing proposed by leading Russian geneticists (Petrova et al., 2020) and the results of the current study, which demonstrated a relatively low efficiency of using a panel of *CFTR* mutations not adapted to a regional population structure, clearly indicate the need for development of a region-focused panel for the preliminary search for most frequent mutations in Yugra-born patients as the 1st-tier genetic test. Comprehensive genetic study of *CFTR* mutations shows that out of 34 identified pathogenic variants, only 11 mutant alleles had the frequencies exceeding 1 %, whereas the remaining 23 variants were unique and were observed only in individual cases. This fact strongly supports the usefulness of consistent use of NGS and MLPA for identifying rare *CFTR* mutations. Functional analysis for novel genetic variants is also recommended.

The obtained genetic information is useful for further improvement of medical care in CF patients and related families with regard to effective genetic counseling and resource allocation in cases where CF-targeted drug therapy with CFTR modulators is available. Identification of unknown pathogenic variants contributes to the general knowledge about *CFTR* mutations heterogeneity at the national and worldwide scale.

References

Amaral M.D., Pankonien I. Cystic fibrosis as a paradigmatic disease in bringing science to the bedside. *FEBS Lett.* 2025;599(22):3196-3207. doi 10.1002/1873-3468.70101

Ayupova G., Litvinov S., Akhmetova V., Minniakhmetov I., Mokrysheva N., Khusainova R. Population characteristics of the spectrum and frequencies of *CFTR* gene mutations in patients with cystic fibrosis from the Republic of Bashkortostan (Russia). *Genes* (*Basel*). 2024;15(10):1335. doi 10.3390/genes15101335

Bobadilla J.L., Macek M., Jr., Fine J.P., Farrell P.M. Cystic fibrosis: a worldwide analysis of *CFTR* mutations – correlation with incidence data and application to screening. *Hum Mutat.* 2002;19(6):575-606. doi 10.1002/humu.10041

Callebaut I., Chong P.A., Forman-Kay J.D. CFTR structure. *J Cyst Fibros*. 2018;17(2S):S5-S8. doi 10.1016/j.jcf.2017.08.008

Castellani C., Simmonds N.J., Barben J., Addy C., Bevan A., Burgel P.R., Drevinek P., … Nährlich L., Raraigh K., Sermet-Gaudelus I., Sommerburg O., Southern K.W. Standards for the care of people with cystic fibrosis (CF): a timely and accurate diagnosis. *J Cyst Fibros*. 2023;22(6):963-968. doi 10.1016/j.jcf.2023.09.008

Dörk T., Macek M., Mekus F., Tümmler B., Tzountzouris J., Casals T., Krebsová A., … Stuhrmann M., Bauer I., Seydewitz H.H., Neumann T., Jakubiczka S. Characterization of a novel 21-kb deletion, CFTRdele2,3(21 kb), in the *CFTR* gene: a cystic fibrosis mutation of Slavic origin common in Central and East Europe. *Hum Genet*. 2000;106(3):259-268. doi 10.1007/s004390000246

Farrell P.M., White T.B., Howenstine M.S., Munck A., Parad R.B., Rosenfeld M., Sommerburg O., … Wilschanski M., Sermet-Gaudelus I., Blau H., Gartner S., McColley S.A. Diagnosis of cystic fibrosis in screened populations. *J Pediatr.* 2017;181S:S33-S44.e2. doi 10.1016/j.jpeds.2016.09.065

Ivaschenko T., Bakay M., Baranov V. Novel frameshift mutation in exon 4 of CFTR gene. *Hum Mutat.* 1995;5(2):184-185. doi 10.1002/humu.1380050215

Kondratyeva E.I., Kashirskaya N.Y., Kapranov N.I. (Eds) National Consensus Project. Cystic Fibrosis: Definition, Diagnostic Criteria, Treatment. Moscow, 2019 (in Russian)

Ong T., Ramsey B.W. Cystic fibrosis: a review. *JAMA*. 2023;329(21): 1859-1871. doi 10.1001/jama.2023.8120

Parisi G.F., Terlizzi V., Manti S., Papale M., Pecora G., Presti S., Tosto M., Leonardi S. Cutting-edge advances in cystic fibrosis: from gene therapy to personalized medicine and holistic management. *Genes* (*Basel*). 2025;16(4):402. doi 10.3390/genes16040402

Petrova N.V., Kondratyeva E.I., Krasovsky S.A., Polyakov A.V., Ivachshenko T.E., Pavlov A.E., Zinchenko R.A., … Semykin S.Y., Sergienko D.F., Simonova O.I., Shabalova L.A., Kashirskaya N.Y. National Consensus Project "Cystic fibrosis: definition, diagnostic criteria, treatment" Section "Genetics of Cystic Fibrosis. Molecular genetic diagnosis of cystic fibrosis". *Med Genet*. 2016;15(11):29-45 (in Russian)

Petrova N.V., Kashirskaya N.Y., Vasilyeva T.A., Kondratyeva E.I., Zhekaite E.K., Voronkova A.Y., Sherman V.D., Galkina V.A., Ginter E.K., Kutsev S.I., Marakhonov A.V., Zinchenko R.A. Analysis of *CFTR* mutation spectrum in ethnic Russian cystic fibrosis patients. *Genes* (*Basel*). 2020;11(5):554. doi 10.3390/genes11050554

Petrova N., Balinova N., Marakhonov A., Vasilyeva T., Kashirskaya N., Galkina V., Ginter E., Kutsev S., Zinchenko R. Ethnic differences in the frequency of *CFTR* gene mutations in populations of the European and North Caucasian part of the Russian Federation. *Front Genet.* 2021;12:678374. doi 10.3389/fgene.2021.678374

Registry of Patients with Cystic Fibrosis in the Russian Federation. 2023. Moscow, 2025. doi 10.61726/1981.2025.26.16.001 (in Russian)

Sosnay P.R., Salinas D.B., White T.B., Ren C.L., Farrell P.M., Raraigh K.S., Girodon E., Castellani C. Applying cystic fibrosis transmembrane conductance regulator genetics and CFTR2 data to facilitate diagnoses. *J Pediatr*. 2017a;181S:S27-S32.e1. doi 10.1016/j.jpeds.2016.09.063

Sosnay P.R., White T.B., Farrell P.M., Ren C.L., Derichs N., Howenstine M.S., Nick J.A., De Boeck K. Diagnosis of cystic fibrosis in nonscreened populations. *J Pediatr.* 2017b;181S:S52-S57.e2. doi 10.1016/j.jpeds.2016.09.068

Stepanova A.A., Abrukova A.V., Savaskina E.N., Poliakov A.V. Mutation p.E92K is the primary cause of cystic fibrosis in Chuvashes. *Russ J Genet*. 2012;48:731-737. doi 10.1134/S1022795412060166

Stepanova A.A., Galeeva N.M., Schagina O.A., Ryzhkova O.P., Krasovsky S.A., Voronkova A.Y., Zhekaite E.K., Melyanovskaya Y.L., Kondratyeva E.I., Polyakov A.V. Results of the extended genotyping program for patients with suspected cystic fibrosis. *Med Genet*. 2025;24(7):100-103. doi 10.25557/2073-7998.2025.07.100-103 (in Russian)

Terlizzi V., Lopes-Pacheco M. Cystic fibrosis: new challenges and perspectives beyond elexacaftor/tezacaftor/ivacaftor. *Ther Adv Respir Dis*. 2025;19:17534666251323194. doi 10.1177/17534666251323194

Therrell B.L., Padilla C.D., Borrajo G.J.C., Khneisser I., Schielen P.C.J.I., Knight-Madden J., Malherbe H.L., Kase M. Current status of newborn bloodspot screening worldwide 2024: a comprehensive review of recent activities (2020–2023). *Int J Neonatal Screen*. 2024;10(2):38. doi 10.3390/ijns10020038

Zhang J., Yao Y., He H., Shen J. Clinical interpretation of sequence variants. *Curr Protoc Hum Genet.* 2020;106(1):e98. doi 10.1002/cphg.98

Zolin A., Adamoli A., Bakkeheim E. ECFSPR Annual Report 2023. European Cystic Fibrosis Society, 2025. Available: https://www.ecfs.eu/sites/default/files/Annual%20Report_2023_vs1.2_ECFSPR_20250721.pdf

Conflict of interest. The authors declare no conflict of interest.
Received September 9, 2025. Revised April 17, 2026. Accepted April 20, 2026.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):782-791

doi 10.18699/vjgb-26-79

Analysis of distribution of *IL13* and *ADAM33* gene haplotypes in Russians, Khakass, Tuvans in relation to the risk of developing bronchial asthma

M.V. Smolnikova, K.V. Afonicheva, A.G. Mileyko

Scientific Research Institute of Medical Problems of the North – a separate division of the Federal Research Center "Krasnoyarsk Science Center" of the Siberian Branch of the Russian Academy of Sciences, Krasnoyarsk, Russia
smarinv@yandex.ru

Abstract. Bronchial asthma (BA) is a heterogeneous disease, the development of which is determined by a complex interaction of genetic factors and environmental factors. Despite significant progress in understanding the molecular mechanisms of the disease, population-specific genetic characteristics, especially in multi-ethnic regions of Russia, remain understudied. The relevance of this study is determined by the need to identify ethnicity-specific genetic markers to improve approaches for personalized prediction of the risk of developing BA and its severity. The aim of the work was to conduct a comparative analysis of haplotypic combinations in the *IL13* (rs20541, rs1800925, rs1295686) and *ADAM33* (rs2280090, rs2787094, rs44707) genes between ethnic groups to identify potential population-specific genetic markers at BA-associated loci. The study involved 524 adolescents aged 12 to 18 years from Krasnoyarsk, Kyzyl and Abakan with verified ethnicity (Slavs – 293, Tuvans – 158, Khakass – 73). DNA was isolated from saliva samples followed by genotyping of polymorphic loci of the *IL13* (rs20541, rs1800925, rs1295686) and *ADAM33* (rs2280090, rs2787094, rs44707) genes using real-time PCR. It was found that rare variants of *IL13* polymorphisms were more common in Russians than in Tuvans and the Khakass. The ATC haplotype of *IL13* is 2–3 times more common among the Khakass and Tuvans than in Russians. In the Russian population, the GAC haplotype of the *ADAM33* gene, containing rare alleles of all the polymorphisms studied, is detected significantly more often. The data obtained emphasize the relevance of haplotype analysis and the importance of taking into account ethno-geographical features when assessing the genetic risks of BA.

Key words: bronchial asthma; gene polymorphism; haplotypes; populations; *IL13*; *ADAM33*

For citation: Smolnikova M.V., Afonicheva K.V., Mileyko A.G. Analysis of distribution of *IL13* and *ADAM33* gene haplotypes in Russians, Khakass, Tuvans in relation to the risk of developing bronchial asthma. *Vavilovskii Zhurnal Genetiki i Selektsii* = *Vavilov J Genet Breed*. 2026;30(5):782-791. doi 10.18699/vjgb-26-79

Funding. The research was carried out within the subject of the state assignment of Scientific Research Institute of Medical Problems of the North – a separate subdivision of FRC KSC SB RAS No. 124020100065-3 "The study of the molecular and cellular regulating mechanisms of the immune response and the interaction of the immune system with other body systems in residents of Eastern Siberia".

Анализ распределения гаплотипов генов *IL13* и *ADAM33* у русских, хакасов и тувинцев в связи с риском развития бронхиальной астмы

М.В. Смольникова, К.В. Афоничева, А.Г. Милейко

Научно-исследовательский институт медицинских проблем Севера – обособленное подразделение Федерального исследовательского центра «Красноярский научный центр Сибирского отделения Российской академии наук», Красноярск, Россия
smarinv@yandex.ru

Аннотация. Несмотря на значительные успехи в исследованиях молекулярных механизмов бронхиальной астмы (БА), остаются малоизученными популяционно-генетические особенности, особенно в мультиэтнических регионах России. Актуальность изучения патогенеза астмы определяется необходимостью выявления этноспецифичных генетических маркеров для совершенствования подходов к персонифицированному прогнозированию риска развития и тяжести течения БА. Целью работы был сравнительный анализ гаплотипических комбинаций в генах *IL13* (rs20541, rs1800925, rs1295686) и *ADAM33* (rs2280090, rs2787094, rs44707) между этническими группами для выявления потенциальных популяционно-специфичных генетических маркеров в локусах, ассоциированных с БА. В исследовании приняли участие 524 подростка в возрасте от 12 до 18 лет из Красноярска, Кызыла и Абакана, с верифицированной национальной

принадлежностью (русские – 293, тувинцы – 158, хакасы – 73). ДНК выделяли из образцов слюны с последующим генотипированием полиморфных локусов генов *IL13* (rs20541, rs1800925, rs1295686) и *ADAM33* (rs2280090, rs2787094, rs44707) методом ПЦР в реальном времени. Установлено, что у русских редкие варианты полиморфизмов *IL13* встречаются чаще по сравнению с тувинцами и хакасами. У хакасов и тувинцев в два-три раза чаще относительно русских встречается гаплотип ATC *IL13*. В русской популяции значимо чаще выявлялся гаплотип GAC гена *ADAM33*, содержащий редкие аллели всех исследованных полиморфизмов. Полученные данные подчеркивают актуальность анализа гаплотипов и важность учета этногеографических особенностей при оценке генетических рисков развития БА.

Ключевые слова: бронхиальная астма; генетический полиморфизм; гаплотипы; популяции; *IL13*; *ADAM33*

Introduction

Genetic predisposition, influenced by environmental factors, underlies the development of bronchial asthma (BA). The heterogeneity of BA is manifested in various endotypes (atopic, non-atopic, aspirin-induced, obesity-associated, severe steroid-dependent, steroid-insensitive eosinophilic, neutrophilic, eosinophilic, etc.), variability in response to drug therapy, and multiple pathogenetic mechanisms.

According to the Russian Ministry of Health, a high prevalence of asthma was revealed by epidemiological data. In 2019, there were more than 350 million patients worldwide, with children comprising 10 %. In the Russian Federation, 1.591 million patients were diagnosed with asthma in 2022, including 84,000 adolescents aged 15–17 years and 229,000 children aged 0–14 years. In 2019, the overall incidence of asthma in the Siberian Federal District exceeded the Russian average (1,507.1 per 100,000 total population) (Bystritskaya, Bilichenko, 2022). The overall incidence of asthma in the Republics of Khakassia and Tyva proved to be consistent with national trends; however, precise data were not provided by the available official sources. According to media reports, approximately 6,500 cases of asthma have been registered in the Republic of Khakassia as of 2022. The Khakassia Ministry of Health has noted an increase in the prevalence of severe asthma. In the Republic of Tyva, as of 2023, there has been about 1,000 cases of asthma.

The high heritability coefficient for asthma development indicated the role of genetic factors, i. e. up to 70 % of asthma cases are caused by genetic factors, with the heritability coefficient being estimated to be up to 74 % in adults and up to 90 % in children (Hernandez-Pacheco et al., 2019). The proportion of phenotypic dispersion of a trait attributable to common single nucleotide polymorphisms, estimated within the framework of the inferred predisposition model (h^2_g), is higher for childhood asthma (aged 0 to 19 years; (h^2_g) = 25.6 %) than for asthma with the onset recorded between the ages of 20 and 60; (h^2_g) = 10.6 % (Ferreira et al., 2019). Furthermore, genome-wide association studies (GWAS) have identified more than 150 asthma risk loci, including *IL13*, *IL33*, *IL1RL1*, *ADAM33*, *ORMDL3*, *ADRB2*, *FLG* (Pividori et al., 2019). Ethnic specificity for the identified genetic associations is also worth noting. For example, when analyzing subgroups for ethnicity, a significant association between the rs2243250 *IL4* polymorphism and asthma was found in Asians and Caucasians, but not in African Americans (Nie et al., 2013). In addition to genes encoding proinflammatory cytokines, polymorphisms in genes regulating the metabolic processes make a significant contribution to the development and control of asthma. Thus, the G allele of rs37973 *GLCCI1* was associated with both uncontrolled asthma and a decrease in respiratory function indicators in Tatar population of Russia, and the T allele of rs2305089 *TBXT*, in the Bashkirs (Fedorova et al., 2019).

A total of 167 genes have been identified for the Russian population, with 11 polymorphic variants demonstrating a significant association with asthma. The data obtained on a large sample of the Federal Medical and Biological Agency of Russia have indicated specific genetic variants that can both increase and decrease the chance of developing the disease, which will certainly be important for the development of preventive diagnostic methods (Akhmerova et al., 2023). The distribution frequency of significant polymorphisms in European and Asian populations has not been sufficiently studied, which puts emphasis on the currently important ethnic factor for the indigenous peoples of Russia, with significant differences in genetic profile associated with liability to disease and behavior being revealed (Smolnikova, Tereshchenko, 2020; Smolnikova et al., 2022; Malarchuk, 2024; Malarchuk et al., 2024; Gusareva et al., 2025; Malarchuk, Litvinov, 2025).

Searching the PubMed database for the keywords “bronchial asthma AND gene AND ethnic group AND nationality” yielded 12,950 results over the past five years (2020–2025). Searching the eLibrary for a similar query “bronchial asthma, gene, ethnic group, nationality” for the same period yielded 125 publications, demonstrating not only research interest in this problem but also an imbalance in global populations studied, as most publications concentrate on large ethnic groups, while indigenous peoples of Russia, including those of the Angara-Yenisei region, are understudied, limiting the possibility for personalized approaches to diagnosis and treatment to be developed.

In contrast to calculating the prevalence of genotypes and alleles of isolated polymorphisms, calculating haplotypes of polymorphisms of asthma-predisposing genes is important for identifying synergistic effects, since protein function is often affected by the allele combinations in a haplotype in a non-additive manner (Belyaeva, 2020), as well as for personalizing therapy, since taking into account the haplotype allows one to optimize the prediction of treatment response (Ouyang, Krontiris, 2006). In addition, defining haplotypes as ethno-specific risk profiles allows for characterization of regional features of morbidity and disease prevention (Tereshchenko, Smolnikova, 2020).

IL13 and *ADAM33* are some of the key asthma-predisposing genes.

Interleukin-13 (IL-13) is a central mediator of immunoglobulin E (IgE) synthesis, regulates allergic inflammation, and promotes excessive mucus production, including for asthma (Gao et al., 2017). IL-13-induced periostin (Krasilnikova et al., 2020) modulates airway inflammation and remodeling (Izuhara et al., 2016), thereby promoting epithelial-mesenchymal transition and fibrosis, respectively (Sidhu et al., 2010). The risk of developing allergic asthma in Asian and Caucasian populations has been shown to raise with IL-13 overexpression (Resende et al., 2017; Smolnikova et al., 2023), but in the Saudi population, for example, no significant association was found (Halwani et al., 2018).

The minor A allele of the rs20541 polymorphism (G>A), located in exon 4 of the *IL13* gene (5q31), is associated with increased total IgE levels in the serum, high serum IL-13 and periostin concentrations (Hafez et al., 2022). The significance of this polymorphism for the *IL13* gene has been verified by population studies in Russia. So, in the Republic of Bashkortostan, polymorphic variants of *IL13*, *IL4* and *CCL11* have been shown to be significant risk factors associated with the development of asthma (Karunas et al., 2012). The G allele of rs20541 *IL13* is more common in asthma children of Kazakh and Han (China) ethnic groups (Bai et al., 2023). Carriers of the G allele of this polymorphism respond better to combination therapy with budesonide/fluticasone propionate in the treatment of moderate to severe asthma (Liu et al., 2020). Korean children carrying the A allele of rs20541 *IL13* showed increased total IgE levels and a greater maximum percent fall in FEV_1 during exercise-induced bronchoconstriction (EIB) compared to GG homozygotes.

The rare T allele of the rs1295686 *IL13* polymorphism is associated with a high risk of developing bronchial asthma, but the exact molecular mechanism is poorly understood (Halwani et al., 2018). In the Saudi population, the CT and TT genotypes raise the risk of asthma in a dominant inheritance pattern (Halwani et al., 2018).

The single nucleotide polymorphism (SNP) rs1800925 (C>T) is located in the promoter region of the *IL13* gene. Overexpression of interleukin-13 caused by the presence of the T allele at rs1800925 causes is likely to be due to increased binding of nuclear proteins at this site (van der Pouw Kraan et al., 1999; Gaceja et al., 2023). An association between the T allele of rs1800925 *IL13* and a better response to Montelukast treatment has been revealed (Kang et al., 2008).

The CTA haplotype having risk alleles of *IL13* polymorphisms (rs1881457-C, rs1800925-T, rs20541-A) was shown to be more frequently found in individuals who did not respond to Montelukast (Kang et al., 2008). Carriers of the CGTG haplotype for *IL13* (rs762534, rs20541, rs1295686, rs1800925) have a low risk of developing asthma in the Saudi population, whereas CATG and CATA have a higher risk, although this result is not statistically significant (Halwani et al., 2018).

The ADAM33 protein belongs to members of the metalloprotease family containing disintegrin and metalloprotease domains. It plays a pivotal role in airway remodeling due to its protease activity (breakdown of extracellular matrix proteins with an effect on bronchial structure), adhesive function (cell-matrix interactions via the disintegrin domain), and specific location in lung fibroblasts and bronchial smooth muscle cells (Camoretti-Mercado, Lockey, 2021). Changes in the activity of the ADAM33 protein encoded by the polymorphic *ADAM33* gene lead to airway remodeling and a high risk of developing asthma in children (Lambrecht, Hammad, 2012; Sun et al., 2017). Furthermore, ADAM33 is involved in the activation of cytokines produced by Th2 cells (Zheng et al., 2015). Increased ADAM33 expression plays a potential role in the eosinophilic/type 2 inflammatory endotype of asthma, considered to be the most severe and treatment-resistant disease variant (Sunadome et al., 2017). ADAM expression has been shown to increase in human fibroblasts *in vitro* in asthmatic patients exposed to interleukin 4 (IL-4) and IL-13 in a time- and concentration-dependent manner. In asthmatic patients, IL-4 and IL-13 can significantly increase ADAM33 mRNA expression in human fibroblasts in a concentration- and time-dependent manner (Jie et al., 2009).

The A allelic variant of the rs44707 (G>A) polymorphism of *ADAM33* (20p13) alters the conformation of the disintegrin domain, breaking down the adhesive properties of the ADAM33 protein (Van Eerdewegh et al., 2002). For the Li population (China), the A allele of rs44707 *ADAM33* was shown to be significantly more common among patients with asthma compared to the control group (Shen et al., 2017). The meta-analysis (4,096 asthma patients and 3,767 controls) did not reveal a significant association between the rs44707 *ADAM33* polymorphism and the risk of developing asthma in the overall group and in subgroup analysis by ethnic groups. However, age-adjusted analysis demonstrated that the A allele and AA+GA genotypes are associated with asthma in children (Li et al., 2019).

The rare A allele of the rs2280090 polymorphism of *ADAM33* (G>A) is associated with an increased risk of developing asthma and accelerated decline in lung function in children in China (Li et al., 2019), with this being verified in the meta-analysis (Sun et al., 2017). In the Bengali population (India), carriers of the GG genotype of rs2280090 *ADAM33* and the GGA haplotype (rs2280090, rs44707, rs3918396) have an increased risk of developing asthma (Sultana et al., 2023); however, in the Punjabi population of Pakistan, no significant associations were found, whereas the strong allelic co-genetic linkage was shown in the study population (Ghani et al., 2019).

The G allele of the rs2787094 (G>C) polymorphism of *ADAM33* enhances transcription factor binding and ADAM33 expression in the bronchial wall (Holgate et al., 2006; Li et al., 2019). In Chinese patients with asthma, the CG and GG genotypes of rs2787094 *ADAM33* were shown to be associated with a nearly 2-fold increased risk of developing asthma in children (Qu et al., 2011) and in the Li population (Hainan Island, China) (Shen et al., 2017), with the same association being found in an Iranian cohort (Shahhosseini, Zare, 2023). However, in the Indonesian population, no association of rs2787094 with asthma was shown (Viyati et al., 2023). The meta-analysis (7,109 asthma patients and 7,794 controls) did not give a significant association between the rs2787094 *ADAM33* polymorphism and asthma; however, when analyzing subgroups by ethnicity, the risk of developing asthma was shown for Caucasians, and by age, for adults (Li et al., 2019).

When analyzing haplotypes, it was shown that haplotypes encoded H1 (rs511898-T, rs2280089-G, rs2280090-G, rs2280091-A, rs2787094-G, rs612709-G), H3 (rs511898-C, rs2280089-G, rs2280090-G, rs2280091-A, rs2787094-G, rs612709-G), H5 (rs511898-T, rs2280089-A, rs2280090-G, rs2280091-A, rs2787094-C, rs612709-G) and H8 (rs511898-C, rs2280089-G, rs2280090-G, rs2280091-A, rs2787094-C, rs612709-G), containing risk alleles of the *ADAM33* polymorphisms, double the risk of developing childhood asthma (China) (Qu et al., 2011).

Given the limited data on haplotypes and their role in liability to asthma in global and Russian populations, the analysis of allele combinations of genes regulating the immune response (*IL13*) and airway remodeling (*ADAM33*) is a promising area of research. In this regard, and taking into account the pronounced ethnic specificity of associations of gene polymorphisms with diseases, the aim of this study is to make a comparative analysis of haplotype combinations in the *IL13* (rs20541, rs1800925, rs1295686) and *ADAM33* (rs2280090, rs2787094, rs44707) genes in different ethnic groups to identify candidate population-specific genetic markers at asthma-associated loci.

Materials and methods

In the present work, a genetic study of adolescents aged 12 to 18 years in Eastern Siberia was carried out. The nationality of the studied adolescents was verified by the nationality of both parents and was defined as Russians (n = 293), Tuvans (n = 158), and Khakassians (n = 73). Demographic data (i. e. age, gender, and nationality of parents) were taken from questionnaires collected within the framework of the State Assignment of the Scientific Research Institute of Medical Problems of the North (2018–2020).

The study object was DNA isolated from saliva. Saliva samples were collected using "Saliva DNA Collection and Preservation Devices" (Norgen Biotek, Canada). DNA extraction was performed using the DIAtomTM DNA Prep 100 reagent kit (Isogen Laboratory LLC, Russia). DNA concentration was measured using a Qubit 4.0 fluorimeter (Thermo Fisher Scientific, USA). Genotyping of *IL13* (rs20541, rs1800925, rs1295686) and *ADAM33* (rs2280090, rs2787094, rs44707) polymorphisms was performed by real-time polymerase chain reaction (RT-PCR) using oligonucleotide primers, fluorescently labeled TagMan probes (DNA Synthesis LLC, Russia) and RotorGene Q 6plex (QIAGEN, Germany).

The study was approved by the Ethics Committee of the Federal Research Center Krasnoyarsk Science Center of the Siberian Branch of the Russian Academy of Sciences (Protocol No. 6 dated December 6, 2012). The parents gave written informed consent for the study.

The frequency of occurrence of the analyzed traits was expressed as absolute and relative values. Differences at a significance level of $p < 0.05$ were considered to be statistically significant. Statistical data processing was performed via the RStudio 4.3.2 software environment (Posit, USA). Genotype frequencies were analyzed using the specialized SNPassoc package, haplotype analysis and assessment of linkage disequilibrium patterns between the studied gene polymorphisms were provided using the "haplo.stats" package for R. EM analysis (Expectation-maximization) was applied with the LR calculation (Likehood Ratio Test), df (Degrees of Freedom), D′ (D prime, the standardized coefficient of linkage disequilibrium) and r^2 (the square of the Pearson correlation coefficient). The genotype distribution across polymorphic loci was tested for compliance with Hardy–Weinberg equilibrium using the χ^2 test. Comparative Population genetic analysis of the distribution of the haplotypes studied in this work with global distributions was performed with data from the 1000 Genomes Project. Reference haplotype frequencies for the *IL13* and *ADAM33* loci were downloaded using the LDlink (LDhap) resource (https://ldlink.nih.gov/ldhap?ref=82878) for the CEU (Utah residents from North and West Europe) and CHB (Han Chinese in Beijing, China) cohorts.

Results

The analysis revealed the haplotype frequency distribution of the *IL13* and *ADAM33* polymorphic genes in Russian, Tuvan, and Khakass populations. Haplotype combination of the *IL13* gene includes the rs20541, rs1295686, and rs1800925 polymorphisms, while that of the *ADAM33* gene consists of the rs44707, rs2280090, and rs2787094 polymorphisms.

Data on the frequency distribution of haplotypes of the *IL13* gene polymorphisms (rs20541, rs1295686, and rs1800925) are presented in Table 1. The most common haplotype in all examined groups turned out to be GCC (Russians – 37.5 %, Tuvans – 49.4 %, Khakassians – 31.5 %). This haplotype consists of wild allelic variants of the three studied polymorphisms. The ATT haplotype was the second most common in Russians (29.4 %); in Tuvans and Khakassians, it was ATC (21.5 and 31.5 %, respectively). The ATT haplotype consisting of rare allelic variants rs20541 + rs1295686 + rs1800925 of *IL13* was significantly more common among Russians compared to Tuvans and Khakassians (29.4 % versus 17.1 and 23.3 %, respectively).

To assess the presence of linkage disequilibrium between the studied loci, EM analysis was used, and the results revealed a statistically significant presence of linkage disequilibrium between the rs1295686, rs1800925 and rs20541 *IL13* polymorphisms (Table 2).

A significant deviation from linkage equilibrium was shown in all three populations ($p < 0.0001$), with the most pronounced general disequilibrium observed in Russians (LR = 345.81), whereas this finding was almost 3.5 times lower in Khakassians, indicating population differences in haplotype structure. The maximum interethnic differences were revealed for the rs20541 and rs1800925 polymorphisms. So, in Tuvans, a strong linkage was observed (D′ = 0.90), developing a stable haplotype block; this indicator was lower (D′ = 0.36) in Khakassians, indicating a partial loss of association between the alleles. On the other hand, the maximum D′ value was shown for the rs1295686 and rs1800925 pair in Tuvans (D′ = 0.86). The correlation coefficient (r^2 = 0.82) was shown to exceed the threshold ($r^2 > 0.8$) only in the Khakass people in the rs20541 and rs1295686 pair, which allows one to consider these polymorphisms as interchangeable research markers. In Russians and Tuvans, this indicator was at the

Table 1. Haplotype frequency distribution for the *IL13* gene (rs20541, rs1295686, and rs1800925), % (*n*)

Order of SNPs in the combination and nucleotide substitutions	Haplotype	Comparison groups			*p*
		Russians (*n* = 293) (1)	Tuvans (*n* = 158) (2)	Khakassians (*n* = 73) (3)	
rs20541 (G>A) + rs1295686 (C>T) + rs1800925 (C>T)	**ATT**	29.4 (86)	17.1 (27)	23.3 (17)	1.2 < 0.001 1.3 < 0.001 2.3 < 0.001
	ATC	11.9 (35)	21.5 (34)	31.5 (23)	
	ACT	2.4 (7)	0.6 (1)	0.0 (0)	
	ACC	3.4 (10)	7.6 (12)	1.4 (1)	
	GTT	4.1 (12)	0.0 (0)	1.4 (1)	
	GTC	2.0 (6)	3.2 (5)	2.7 (2)	
	GCT	9.2 (27)	0.6 (1)	8.2 (6)	
	GCC	37.5 (110)	49.4 (78)	31.5 (23)	

Note. Zero haplotype frequency values are not included in the table.

level of moderate correlation (0.63 and 0.69, respectively), which does not allow one to completely replace one polymorphism with another. In the Khakass people, a low correlation coefficient for pairs of polymorphisms involving rs1800925 (0.06 and 0.07) was revealed, limiting the use of this locus as an interchangeable risk marker for other polymorphisms in this population. At the same time, in Tuvans, the rs1800925 polymorphism demonstrated a strong linkage with other polymorphisms, suggesting the development of a haplotype block including all the studied *IL13* polymorphisms.

The haplotype frequencies of *ADAM33* (rs44707, rs2280090 and rs2787094) were presented in Table 3. In Russians, the most frequent haplotypes were GGG (24.9 %) and GGC (22.2 %); however, in Tuvans and Khakassians, the most frequent was the GGC haplotype (42.4 and 35.6 %, respectively). The second most common haplotype among Tuvans and Khakassians was TGG (27.8 and 24.7 %, respectively). The TGC haplotype prevalence was lower in Russians compared to Tuvans and Khakassians (8.5 % versus 14.6 and 13.7 %, respectively). And, conversely, the TAG haplotype was more common in the Russian population compared to Tuvans and Khakassians (11.3 % versus 3.8 and 6.8 %, respectively). The GAC haplotype was significantly more common among Russians (4.8 %) compared to Khakassians and Tuvans (1.4 and 0.0 %).

To assess the presence of linkage disequilibrium between the studied loci, EM analysis was used, with a statistically significant presence of linkage disequilibrium between the rs44707, rs2280090 and rs2787094 *ADAM33* polymorphisms found (Table 4).

A significant deviation from linkage equilibrium was revealed in all the studied populations ($p < 0.0001$): the most pronounced general disequilibrium was observed in Russians (LR = 59.43), and in Khakassians, this indicator was almost 2.7 times lower. For the pair of rs2280090 and rs2787094 polymorphisms, a complete linkage disequilibrium (D′ = 1.0) was recorded for all populations; accordingly, the alleles of these loci are always inherited together, developing a stable haplotype block. Similar results were shown for the pair of rs44707 and rs2280090 polymorphisms (in Russians D′ = 0.92, in Tuvans and Khakassians D′ = 1.0). Different results were obtained for the rs44707 and rs2787094 polymorphisms, i. e. the D′ values, having moderate linkage, ranged from 0.43 to 0.53, respectively, pointing to partial recombination between these loci. The ratio of the D′ and r^2 values is worth noting. Despite the complete linkage of rs2280090 and rs2787094, the r^2 values were below the moderate correlation threshold ($r^2 > 0.33$), which does not allow these risk markers to

Table 2. Linkage disequilibrium analysis of the *IL13* gene polymorphisms in Russians, Tuvans and Khakassians

Parameter	Russians	Tuvans	Khakassians
LR	345.81	204.21	94.76
df	4.0	3.0	4.0
p	<0.0001	<0.0001	<0.0001
rs20541 *vs* rs1295686			
D′	0.80	0.89	0.93
r^2	0.63	0.69	0.82
p	<0.0001	<0.0001	<0.0001
rs20541 *vs* rs1800925			
D′	0.52	0.90	0.36
r^2	0.23	0.25	0.06
p	<0.0001	<0.0001	0.0046
rs1295686 *vs* rs1800925			
D′	0.58	0.86	0.41
r^2	0.29	0.26	0.07
p	<0.0001	<0.0001	0.0015

Table 3. Haplotype frequency distribution for the *ADAM33* gene (rs44707, rs2280090, and rs2787094), % (*n*)

Order of SNPs in the combination and nucleotide substitutions	Haplotype	Comparison groups			*p*
		Russians (*n* = 293) (1)	Tuvans (*n* = 158) (2)	Khakassians (*n* = 73) (3)	
rs44707 (T>G) + rs2280090 (G>A) + rs2787094 (G>C)	**GAC**	4.8 (14)	0.0 (0)	1.4 (1)	1.2 < 0.001 1.3 < 0.001 2.3 < 0.001
	GAG	5.5 (16)	1.9 (3)	1.4 (1)	
	GGC	22.2 (65)	42.4 (67)	35.6 (26)	
	GGG	24.9 (73)	8.9 (14)	12.3 (9)	
	TAC	2.0 (6)	0.6 (1)	4.1 (3)	
	TAG	11.3 (33)	3.8 (6)	6.8 (5)	
	TGC	8.5 (25)	14.6 (23)	13.7 (10)	
	TGG	20.8 (61)	27.8 (44)	24.7 (18)	

Note. Zero haplotype frequency values are not included in the table.

Table 4. Linkage disequilibrium analysis of the *ADAM33* gene polymorphisms in Russians, Tuvans and Khakassians

Parameter	Russians	Tuvans	Khakassians
LR	59.43	50.80	22.21
df	3.0	2.0	3.0
p	<0.0001	<0.0001	<0.0001
rs44707 *vs* rs2280090			
D′	0.92	1.0	1.0
r^2	0.07	0.02	0.03
p	<0.0001	0.02	0.04
rs44707 *vs* rs2787094			
D′	0.43	0.53	0.51
r^2	0.09	0.23	0.26
p	<0.0001	<0.0001	<0.0001
rs2280090 *vs* rs2787094			
D′	1.0	1.0	1.0
r^2	0.04	0.02	0.03
p	<0.0001	0.0111	0.0349

be used interchangeably in studies. The highest r^2 values for the rs44707 and rs2787094 pair were found in Khakassians (0.26) and Tuvans (0.23), this value being much lower (0.09) in Russians, implying the population differences in the balance of allele frequencies of the *ADAM33* gene polymorphisms.

The prevalence of haplotypes of *IL13* and *ADAM33* in samples of Russians, Tuvans, and Khakassians differed significantly from those presented for the CEU and CHB reference populations. The populations studied did not completely correspond to the European (CEU) or East Asian (CHB) profile (Table 5).

Discussion

In the current study, a comparative analysis of the prevalence of haplotypes of the *IL13* (rs20541, rs1295686, and rs1800925) and *ADAM33* (rs44707, rs2280090, and rs2787094) polymorphisms in Russian, Tuvan, and Khakass populations was carried out.

The "risk" haplotype ATT combines rare alleles of the rs20541-A, rs1295686-T, and rs1800925-T polymorphisms of the *IL13* gene. This haplotype contains functionally significant variants that mediate chronic inflammation and lead to mucus hyperproduction (Gaceja et al., 2023). In particular, the presence of the T allele of rs20541 causes an increase in the affinity of the IL-13 ligand for the receptor, which, in turn, results in hyperproduction of IgE and periostin (Halwani et al., 2018; Hafez et al., 2022). At the same time, the T allele of rs1295686 is associated with an increased synthesis of IL-13 due to an enhancement in the transcriptional activity of the *IL13* gene (Halwani et al., 2018). The presence of the T allele of rs1800925 is associated with a decreased affinity of IL-13 for the receptor and an increased IL-13 concentration in the blood (Arima et al., 2002). So, the risk haplotype ATT is characterized by a combination of increased IL13 expression and high affinity for the receptor, leading to bronchial hyperreactivity and eosinophilic inflammation.

Our data revealed statistically significant differences in the prevalence of the risk haplotype ATT among Russians (29.4 %), Tuvans (17.1 %), and Khakassians (23.3 %), which is consistent with the known population patterns in the distribution of individual alleles for *IL13*. So, rare alleles of the rs20541 and rs1295686 polymorphisms also predominate in the Saudi population of patients with asthma; however, their low prevalence in the general population limits the statistical power of the study (Halwani et al., 2018). In addition, the

Table 5. The prevalence of haplotypes of *IL13* (rs20541, rs1295686 and rs1800925) and *ADAM33* (rs44707, rs2280090 and rs2787094) in the European (CEU) and East Asian (CHB) populations, % (*n*)

IL13				*ADAM33*			
CEU		CHB		CEU		CHB	
CCG	72.3 (143)	CCG	66.5 (137)	GGT	38.9 (77)	CGG	35.4 (73)
TTA	12.6 (25)	CTA	16.5 (34)	GGG	25.8 (51)	GGT	30.1 (62)
CTA	10.1 (20)	TTA	15 (31)	CGG	14.1 (28)	GGG	20.8 (43)
TCG	5.0 (10)	TCG	1.5 (3)	GAT	12.1 (24)	GAT	7.7 (16)
		CCA	0.5 (1)	CGT	8.6 (17)	CGT	3.0 (6)
				CAT	0.5 (1)	GAG	1.5 (3)
						CAT	1.0 (2)
						CAG	0.5 (1)

CATG and CATA haplotypes (rs762534, rs20541, rs1295686, rs1800925) of the *IL13* gene, which contain the ATC and ATT haplotypes of the polymorphisms studied, demonstrate a tendency toward an increased risk of developing asthma in the Saudi population (Halwani et al., 2018). The CGTG haplotype (rs762534, rs20541, rs1295686, rs1800925) demonstrates a protective effect on the asthma development in the Saudi population (Halwani et al., 2018).

It is important to note that even though the CGTG haplotype contains a minor risk allele T of the rs1295686 polymorphism associated with airway inflammation, its combination with the G allele of rs20541 neutralizes the risk, emphasizing the epistatic effect of loci, when the effects of the T allele of rs1295686 are modulated by the G allele of rs20541. Indeed, the CTA haplotype (rs1881457, rs1800925, rs20541) of the *IL13* gene in Korean children with asthma is associated with non-response to therapy with Montelukast, a leukotriene receptor antagonist (Kang et al., 2008). The key factor in this combination is the risk allele T of the rs1800925 polymorphism, which causes overexpression of *IL-13.* In the Chinese population, carriers of haplotypes that included both the T allele of rs20541 and the T allele of rs1800925 had very high IgE level in cord blood, predicting asthma development in childhood (Hua et al., 2021).

The Russian population has a higher frequency of haplotypes including rare alleles of the studied *IL13* gene polymorphisms, which is likely to point to an increased genetic predisposition to asthma development in this group compared to Tuvans and Khakassians. Interestingly, in Khakass and Tuvans, the frequency of the ATC haplotype of the *IL13* gene, which contains two of the three studied alleles associated, according to the literature, with asthma, is 2–3 times higher than in Russians. The absence of the T allele of rs1800925 in this haplotype, which is considered to be a key marker of asthma risk in some studies, suggests that the functional significance of this haplotype may be different. The problem of its contribution to disease predisposition warrants further investigation in case-control studies involving asthma patients from these ethnic groups.

The GAC haplotype (rs44707-G, rs2280090-A, rs2787094-C) of the *ADAM33* gene combines rare risk alleles of the studied polymorphisms associated with changes in the structural and functional properties of the ADAM33 protein. In particular, the presence of the G allele of rs44707 leads to increased transcriptional activity and, in turn, to hyperproduction of ADAM33 in airway tissue cells (Xue et al., 2013). Therefore, the phosphorylation of the ADAM33 protein is affected by the A allele of rs2280090, changing its proteolytic activity and the ability to mediate airway remodeling. The C allele of rs2787094 is associated with increased *ADAM33* expression, which facilitates bronchial smooth muscle hypertrophy (Li et al., 2019). According to our study, the risk haplotype GAC of *ADAM33* is statistically significantly more common among Russians (4.8 %) compared to Khakassians (1.4 %) and Tuvans (0.0 %). These findings are consistent with the published data. Thus, in the study (Zhou et al., 2011), another haplotype H5 of the *ADAM33* gene, also including rare alleles at the rs2280090 and rs44707 loci in a different combination, was associated with a 2-fold increased risk of asthma in children of the Han population of China. Given the higher frequency of the GAC haplotype identified in the Russian population, the distribution patterns of *ADAM33* gene variants can be assumed to be able to contribute to the inter-ethnic differences in asthma prevalence previously noted for the population of Eastern Siberia (Konopleva et al., 2017; Afonicheva et al., 2025a).

According to our data, the GGG haplotype of the *ADAM33* gene is more common in Russians (24.9 %) than in Tuvans and Khakassians (8.9 and 12.3 %). The haplotype encoded as H1 (rs2280090-G + rs2787094-G + rs44707-G) in the Northern Chinese population is associated with severe asthma in children (Qu et al., 2011). Considering that carriers of rare alleles G of the rs44707 polymorphism and A of the rs2787094 polymorphism of *ADAM33* have a greater allergen sensitivity and pronounced lung function decrease (Sleziak et al., 2024; Afonicheva et al., 2025b), the risk of developing asthma exacerbations among Russians is assumed to be definitely higher than among other populations studied in the work. It should be taken into account that the low prevalence of the GGG

haplotype of *ADAM33* in the Tuvan and Khakass populations limits the statistical analysis power of their association with the disease.

The GGC haplotype containing two of the three studied risk variants of the *ADAM33* gene polymorphisms is significantly more common in Tuvans and Khakassians than in Russians. Due to the absence of the A allele of the *ADAM33* gene for rs2280090 in this haplotype, considered in the literature to be a marker of underlying risk for asthma, we hypothesize that the functional effects of the GGC haplotype may differ from those with a complete set of risk alleles. However, weather the clinical significance of this haplotype and its contribution to genetic predisposition to asthma development remains unclear.

The identified strong linkage disequilibrium between the studied *IL13* and *ADAM33* gene loci means that the allelic variants of the studied gene polymorphisms are inherited as haplotype blocks. Therefore, the consideration of polymorphism combinations as haplotypes has been justified. Population differences in the distribution of haplotype frequencies, indicating both the different levels of LD in Russians, Tuvans and Khakassians and the presence of genetic drift, have been demonstrated in our study.

The differences identified in the haplotype distribution of the *IL13* and *ADAM33* loci between our sample data and the CEU and CHB populations reveals a unique genetic structure of the studied ethnic groups.

Conclusion

The distribution patterns of *IL13* and *ADAM33* gene haplotypes in Russian, Tuvan, and Khakass populations were shown in the study. We identified statistically significant differences in the frequencies of haplotypes, including polymorphism alleles, which, according to the literature, are associated with a predisposition to asthma. *IL13* gene haplotypes containing rare "risk" alleles were established to occur at a higher incidence in the Russian population compared to Tuvans and Khakassians. Our results also demonstrated that the risk GAC haplotype of the *ADAM33* gene is significantly more common in Russians than in Khakassians and Tuvans.

The differences identified in haplotype distribution may reflect genetic characteristics of the studied populations, potentially influencing interpopulation differences in asthma prevalence. However, for this hypothesis to be validated, case-control studies in each of the studied ethnic groups are needed. The obtained data highlight the multifactorial nature of asthma and the need to take into account ethno-geographical characteristics when studying genetic susceptibility to the disease. The results and the role of the identified haplotypes in asthma development have to be verified in functional studies with larger sample sizes.

References

Afonicheva K.V., Mileyko A.G., Tereshchenko S.Yu., Smolnikova M.V. Variants of *IL4* (rs2243250, rs2070874) and *IL13* (rs1800925) genes in ethnic groups of Eastern Siberia. *Sibirskij Nauchnyj Medicinskij Zhurnal = Siberian Scientific Medical Journal*. 2025a;45(3):53-60. doi 10.18699/SSMJ20250305 (in Russian)

Afonicheva K.V., Tereshchenko S.Yu., Smolnikova M.V. Prevalence of allele variants of *ADAM33* (rs2280090) and *ADRB2* (rs1042714) in Siberian populations. *Russ J Genet.* 2025b;61(7):885-890. doi 10.1134/S1022795425700401

Akhmerova Y.N., Shpakova T.A., Grammatikati K.S., Mitrofanov S.I., Kazakova P.G., Mkrtchian A.A., Zemsky P.Y., … Yudin V.S., Keskinov A.A., Kraevoy S.A., Yudin S.M., Skvortsova V.I. Genetic variants associated with bronchial asthma specific to the population of the Russian Federation. *Acta Naturae*. 2023;15(1):31-41. doi 10.32607/actanaturae.11853

Arima K., Umeshita-Suyama R., Sakata Y., Akaiwa M., Mao X.-Q., Enomoto T., Dake Y., … Adra C.N., Hamasaki N., Hopkin J.M., Shirakawa T., Izuhara K. Upregulation of IL-13 concentration in vivo by the *IL13* variant associated with bronchial asthma. *J Allergy Clin Immunol*. 2002;109(6):980-987. doi 10.1067/mai.2002.124656

Bai S., Lu J., Hua L., Liu Q., Chen M., Gu Y., Zhang J., Ma D., Bao Y. Prediction of asthma using a four-locus gene model including *IL13*, *IL4*, *FCER1B*, and *ADRB2* in children of Kazak nationality. *Ital J Pediatr*. 2023;49(1):162. doi 10.1186/s13052-023-01564-y

Belyaeva T.M. Study of associations of haplotypes of *FLG* gene polymorphism with the development of chronic true eczema in men. *Res Results Biomed*. 2020;6(2):160-171. doi 10.18413/2658-6533-2020-6-2-0-2 (in Russian)

Bystritskaya E.V., Bilichenko T.N. The review of the bronchial asthma morbidity in the population of the Russian Federation. *Pulmonologiya*. 2022;32(5):651-660. doi 10.18093/0869-0189-2022-32-5-651-660 (in Russian)

Camoretti-Mercado B., Lockey R.F. Airway smooth muscle pathophysiology in asthma. *J Allergy Clin Immunol*. 2021;147(6):1983-1995. doi 10.1016/j.jaci.2021.03.035

Fedorova Y.Y., Karunas A.S., Murzina R.R., Savelieva O.N., Gimalova G.F., Gatiyatullin R.F., Etkina E.I., Khusnutdinova E.K. Association between allelic variants of the genes involved in glucocorticoids metabolism and asthma. *Russ J Genet*. 2019;55(12):1520-1527. doi 10.1134/S1022795419120044

Ferreira M.A.R., Mathur R., Vonk J.M., Szwajda A., Brumpton B., Granell R., Brew B.K., … Karlsson R., Hinds D.A., Paternoster L., Koppelman G.H., Almqvist C. Genetic architectures of childhood- and adult-onset asthma are partly distinct. *Am J Hum Genet*. 2019;104(4):665-684. doi 10.1016/j.ajhg.2019.02.022

Gaceja K.V., Ancheta Z.F.R., Buna A.C.A., Clarencio S.M.S., Garrido M.A.R., Ramos J.D.A. Association of interleukin-13 gene single nucleotide polymorphism rs1800925 with allergic asthma in Asian population: a meta-analysis. *Asia Pac Allergy.* 2023;13(4):148-157. doi 10.5415/apallergy.0000000000000119

Gao M., Liu L.-X., Wu F.-L., Zhang X., Li Y.-Y., Shi T., Li D.-Z., Han T.-T. The changes of Th17/Treg and related cytokines: IL-17, IL-23, IL-10, and TGF-β in respiratory syncytial virus bronchiolitis rat model. *Iran J Allergy Asthma Immunol*. 2017;16(5):386-395

Ghani M.U., Sabar M.F., Bano I., Shahid M., Akram M., Khalid I., Maryam A., Khan M.U. Inheritance pattern of rs2280089, rs2280090, rs2280091 SNP variants in Punjabi population and association with asthma disease. *CHEST*. 2019;155(4):168A. doi 10.1016/j.chest.2019.02.162

Gusareva E.S., Ghosh A.G., Kharkov V.N., Khor S.-S., Zarubin A., Moshkov N., Kalsi N., … Nakagome S., Tatarinova T.V., Stepanov V.A., Schuster S.C., Kim H.L. From North Asia to South America: tracing the longest human migration through genomic sequencing. *Science*. 2025;388(6748):eadk5081. doi 10.1126/science.adk5081

Hafez R.A., Hassan M.E., Haggag M.G., Atef N., Abdallah A.L., Gerges M.A. Association of interleukin 13 rs20541 gene polymorphism and serum periostin with asthma and allergic conjunctivitis among Egyptian patients. *J Asthma Allergy*. 2022;15:971-982. doi 10.2147/JAA.S373098

Halwani R., Vazquez-Tello A., Kenana R., Al-Otaibi M., Alhasan K.A., Shakoor Z., Al-Muhsen S. Association of IL-13 rs20541

and rs1295686 variants with symptomatic asthma in a Saudi Arabian population. *J Asthma*. 2018;55(10):1157-1165. doi 10.1080/02770903.2017.1400047

Hernandez-Pacheco N., Pino-Yanes M., Flores C. Genomic predictors of asthma phenotypes and treatment response. *Front Pediatr*. 2019;7:6. doi 10.3389/fped.2019.00006

Holgate S.T., Yang Y., Haitchi H.-M., Powell R.M., Holloway J.W., Yoshisue H., Pang Y.Y., Cakebread J., Davies D.E. The genetics of asthma: *ADAM33* as an example of a susceptibility gene. *Proc Am Thorac Soc*. 2006;3(5):440-443. doi 10.1513/pats.200603-026AW

Hua L., Liu Q., Li J., Zuo X., Chen Q., Li J., Wang Y., … Chen Y., Zhong W., Zhang J., Zhang J., Bao Y. Gene-gene and gene-environment interactions on cord blood total IgE in Chinese Han children. *Allergy Asthma Clin Immunol*. 2021;17(1):69. doi 10.1186/s13223-021-00571-0

Izuhara K., Conway S.J., Moore B.B., Matsumoto H., Holweg C.T.J., Matthews J.G., Arron J.R. Roles of periostin in respiratory disorders. *Am J Respir Crit Care Med.* 2016;193(9):949-956. doi 10.1164/rccm.201510-2032PP

Jie Z., Jin M., Cai Y., Bai C., Shen Y., Yuan Z., Hu Y., Holgate S. The effects of Th2 cytokines on the expression of ADAM33 in allergen-induced chronic airway inflammation. *Respir Physiol Neurobiol*. 2009;168(3):289-294. doi 10.1016/j.resp.2009.07.019

Kang M.-J., Lee S.-Y., Kim H.-B., Yu J., Kim B.-J., Choi W.-A., Jang S.-O., Hong S.-J. Association of IL-13 polymorphisms with leukotriene receptor antagonist drug responsiveness in Korean children with exercise-induced bronchoconstriction. *Pharmacogenet Genomics*. 2008;18(7):551-558. doi 10.1097/FPC.0b013e3282fe94c5

Karunas A.S., Fedorova Yu.Yu., Ramazanova N.N., Galimova E.S., Gimalova G.F., Guryeva L.L., Levasheva S.V., Biktasheva A.R., Mukhtarova L.A., Zagidullin Sh.Z., Etkina E.I., Khusnutdinova E.K. Evaluation of a role of cytokine gene polymorphisms in development of bronchial asthma in the Republic of Bashkortostan. *Pulmonologiya*. 2012;5:37-40. doi 10.18093/0869-0189-2012-0-5-37-40 (in Russian)

Konopleva O.S., Smolnikova M.V., Smirnova S.V., Gorbacheva N.N., Tereshchenko S.Yu. Genes of the immune response in the pathogenesis of atopic bronchial asthma in children of the East Siberian region. *Cytokines and Inflammation*. 2017;16(3):39-41 (in Russian)

Krasilnikova S.V., Tush E.V., Frolov P.A., Ovsyannikov D.Y., Terentyeva A.B., Kubysheva N.I., Eliseeva T.I. Periostin as a biomarker of allergic inflammation in atopic bronchial asthma and allergic rhinitis (a pilot study). *Mod Technol Med.* 2020;12(5):37-44. doi 10.17691/stm2020.12.5.04

Lambrecht B.N., Hammad H. The airway epithelium in asthma. *Nat Med.* 2012;18(5):684-692. doi 10.1038/nm.2737

Li H., Yan L., Wang K., Li X., Liu H., Tan W. Association between ADAM33 polymorphisms and asthma risk: a systematic review and meta-analysis. *Respir Res*. 2019;20(1):38. doi 10.1186/s12931-019-1006-1

Liu L., Yue D., Hu L., Wang F., Huang Y., Liao Y. Relationship between interleukin-13 rs20541 single nucleotide polymorphisms and therapeutic efficacy in children with asthma. *J Int Med Res*. 2020;48(6):030006052092917. doi 10.1177/0300060520929179

Malyarchuk B.A. Genetic features of lipid and carbohydrate metabolism in Arctic populations. *Biochemistry* (*Moscow*). 2024;89(7):1192-1201. doi 10.1134/S0006297924070034

Malyarchuk B.A., Litvinov A.N. Polymorphism of the rs1421085 locus of the *FTO* gene in human populations: genogeographic aspect. *Russ J Genet*. 2025;61(4):453-459. doi 10.1134/S1022795424701850

Malyarchuk B.A., Derenko M.V., Denisova G.A., Litvinov A.N., Dambueva I.K. *FADS* gene polymorphism and the history of the formation of the indigenous populations of Siberia. *Russ J Genet*. 2024;60(2):199-209. doi 10.1134/S1022795424020091

Nie W., Zhu Z., Pan X., Xiu Q. The interleukin-4-589C/T polymorphism and the risk of asthma: a meta-analysis including 7,345 cases and 7,819 controls. *Gene*. 2013;520(2):22-29. doi 10.1016/j.gene.2013.02.027

Ouyang C., Krontiris T.G. Identification and functional significance of SNPs underlying conserved haplotype frameworks across ethnic populations. *Pharmacogenet Genomics*. 2006;16(9):667-682. doi 10.1097/01.fpc.0000220569.82842.9b

Pividori M., Schoettler N., Nicolae D.L., Ober C., Im H.K. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. *Lancet Respir Med*. 2019;7(6):509-522. doi 10.1016/S2213-2600(19)30055-4

Qu S., Sun D., Wang Y., Zhang C., Lv Y., Yao L. Association of *ADAM33* polymorphisms with childhood asthma in a northern Chinese population. *Exp Mol Pathol*. 2011;91(3):775-779. doi 10.1016/j.yexmp.2011.09.001

Resende E.P., Todo-Bom A., Loureiro C., Mota Pinto A., Oliveiros B., Mesquita L., Silva H.C. Asthma and rhinitis have different genetic profiles for *IL13*, *IL17A* and *GSTP1* polymorphisms. *Rev Port Pneumol*. 2017;23(1):10-16. doi 10.1016/j.rppnen.2016.06.009

Shahhosseini S., Zare M. Association of rs2787094 in ADAM33 and rs11650354 in TBX21 gene polymorphisms with asthma in Iranian patients. *Iran South Med J*. 2023;25(6):489-504. doi 10.61186/ismj.25.6.489

Shen B., Lin R., Wang C.-C., Rei J., Sun Y., Yang Y.-L., Lin Y.-Y. *ADAM33* gene polymorphisms identified to be associated with asthma in a Chinese Li population. *Biomed Rep*. 2017;6(3):323-328. doi 10.3892/br.2017.854

Sidhu S.S., Yuan S., Innes A.L., Kerr S., Woodruff P.G., Hou L., Muller S.J., Fahy J.V. Roles of epithelial cell-derived periostin in TGF-β activation, collagen production, and collagen gel elasticity in asthma. *Proc Natl Acad Sci USA*. 2010;107(32):14170-14175. doi 10.1073/pnas.1009426107

Sleziak J., Gawor A., Błażejewska M., Antosz K., Gomułka K. ADAM33′s role in asthma pathogenesis: an overview. *Int J Mol Sci*. 2024;25(4):2318. doi 10.3390/ijms25042318

Smolnikova M.V., Tereshchenko S.Yu. Internet addiction in adolescents: the search for genetic associations. *Med Genet*. 2020;19(4):46-48. doi 10.25557/2073-7998.2020.04.46-48 (in Russian)

Smolnikova M.V., Malinchik M.A., Tereschenko S.Y. Gene polymorphism of mannose-binding lectin-associated serine protease (MASP2) in indigenous populations of the Russian Arctic Territories. *Russ J Infect Immun*. 2022;12(3):543-550. doi 10.15789/2220-7619-GPO-1712 (in Russian)

Smolnikova M.V., Kasparov Ed.W., Malinchik M.A., Kopylova K.V. Genetic markers of children asthma: predisposition to disease course variants. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2023;27(4):393-400. doi 10.18699/VJGB-23-47

Sultana S., Banerjee P., Ganai I., Laha A., Sultana N., Biswas H., Saha N.C., Moitra S., Podder S. Polymorphism in ADAM33 gene associated with asthmatics in West Bengal, India – an investigation by in-silico analysis. *World Allergy Organ J*. 2023;16(11):100834. doi 10.1016/j.waojou.2023.100834

Sun F.J., Zou L.Y., Tong D.M., Lu X.Y., Li J., Deng C.B. Association between ADAM metallopeptidase domain 33 gene polymorphism and risk of childhood asthma: a meta-analysis. *Braz J Med Biol Res.* 2017;50(10):e6148. doi 10.1590/1414-431X20176148

Sunadome H., Matsumoto H., Petrova G., Kanemitsu Y., Tohda Y., Horiguchi T., Kita H., … Tamari M., Yokoyama T., Niimi A., Izuhara K., Mishima M. *IL4Rα* and *ADAM33* as genetic markers in asthma exacerbations and type-2 inflammatory endotype. *Clin Exp Allergy*. 2017;47(8):998-1006. doi 10.1111/cea.12927

Tereshchenko S.Yu., Smolnikova M.V. Polymorphism of the mannose-binding lectin gene in the Arctic indigenous populations of the Rus-

sian Federation. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed.* 2020;24(8):868-875. doi 10.18699/VJ20.685 (in Russian)

van der Pouw Kraan T., van Veen A., Boeije L.C.M., van Tuyl S.A.P., de Groot E.R., Stapel S.O., Bakker A., Verweij C.L., Aarden L.A., van der Zee J.S. An IL-13 promoter polymorphism associated with increased risk of allergic asthma. *Genes Immun.* 1999;1(1):61-65. doi 10.1038/sj.gene.6363630

Van Eerdewegh P., Little R.D., Dupuis J., Del Mastro R.G., Falls K., Simon J., Torrey D., … Rorke S., Clough J.B., Holloway J.W., Holgate S.T., Keith T.P. Association of the *ADAM33* gene with asthma and bronchial hyperresponsiveness. *Nature*. 2002;418(6896):426-430. doi 10.1038/nature00878

Viyati K., Prayuni K., Zulhamidah Y., Razari I., Yuliwulandari R. Association between rs2787094 genetic variants in ADAM33 gene and asthma in Indonesian population: preliminary study. *Makara J Health Res.* 2023;27(3):149-153. doi 10.7454/msk.v27i2.1431

Xue W., Han W., Zhou Z.-S. ADAM33 polymorphisms are associated with asthma and a distinctive palm dermatoglyphic pattern. *Mol Med Rep*. 2013;8(6):1795-1800. doi 10.3892/mmr.2013.1733

Zheng W., Wang L., Su X., Hu X.F. Association between V4 polymorphism in the ADAM33 gene and asthma risk: a meta-analysis. *Genet Mol Res*. 2015;14(1):989-999. doi 10.4238/2015.February.6.2

Zhou J., Sun D., Xu L., Sun L., Fu S., Li Y. ADAM33 as a psoriasis susceptibility gene in the Han population of northeastern China. *Dermatology*. 2012;223(4):356-362. doi 10.1159/000335417

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

Received December 1, 2025. Revised March 20, 2026. Accepted April 21, 2026.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):792-802

doi 10.18699/vjgb-26-80

Inherited and somatic components in the pathogenetics of common diseases

M.S. Nazarenko [1, 2], A.A. Sleptcov [1, 2], V.P. Puzyrev [1]

[1] Research Institute of Medical Genetics, Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russia
[2] Tyumen Cardiology Research Center – Branch of the Tomsk National Research Medical Center of the Russian Academy of Sciences, Tyumen, Russia
maria.nazarenko@medgenetics.ru

Abstract. It is known that the pathogenesis of common diseases (CDs) is determined by a complex and nonlinear interaction of genetic predisposition, epigenetic modifications, and environmental factors. However, in addition to inherited genetic variants, somatic mutations play a key role in the pathological phenotype, shaping cell-type-specific genetic landscapes and influencing regional predisposition of target organs, age of onset, and variability in clinical manifestations. Thus, according to current concepts, CDs develop as a result of the combined influence of inherited germline variants and somatic mutations acquired during life, the interaction of which is realized through cell-specific molecular networks. Therefore, this review sequentially examines the main factors, classical and modern models of complex pathological phenotypes, as well as the achievements and prospects of analyzing inherited genetic variants in relation to CDs. The role of somatic mutations in the development of pathology is discussed using the example of atherosclerosis ontogenesis and the concept of athero-oncology through the assessment of somatic genomic variability in smooth muscle cells of atherosclerotic plaques, where the sequential accumulation of "driver" mutations confers a selective advantage and triggers clonal evolution in the tissue microenvironment. Special attention is given to the concept of "paradominant" inheritance – a special form of non-Mendelian transmission of complex traits that combines inherited predisposition and somatic mutations that arise early in ontogenesis. Based on an analysis of current data, we propose an integrative hypothesis postulating that inherited genetic variants form a body-wide predisposition to CDs, while somatic mutations, arising and selectively increasing in specific cellular compartments of target organs, are critical events explaining the regional specificity, temporal dynamics (age of onset), and variability in the severity of the clinical phenotype. A new paradigm for predictive medicine for CDs is based on a comprehensive characterization of the continuum of genetic variants (inherited and somatic), an analysis of their interactions with cell-specific molecular networks, and a transition from population-based risk assessments to causal models of individual pathogenesis.

Key words: pathogenetics; common diseases; omnigenic model; atherosclerosis; cancer; germinal variants; somatic mutations

For citation: Nazarenko M.S., Sleptcov A.A., Puzyrev V.P. Inherited and somatic components in the pathogenetics of common diseases. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(5):792-802. doi 10.18699/vjgb-26-80

Funding. This work was partially supported by an internal grant of Tomsk National Research Medical Center "Clonal hematopoiesis of uncertain potential as a risk factor for chronic heart failure: the role of somatic mutations and systemic inflammation" and Addendum to Federal Subsidy Agreement No. 075-03-2026-698/7.

Унаследованная и соматическая компоненты в патогенетике многофакторных заболеваний

М.С. Назаренко [1, 2], А.А. Слепцов [1, 2], В.П. Пузырев [1]

[1] Научно-исследовательский институт медицинской генетики Томского национального исследовательского медицинского центра Российской академии наук, Томск, Россия
[2] Тюменский кардиологический научный центр – филиал Томского национального исследовательского медицинского центра Российской академии наук, Тюмень, Россия
maria.nazarenko@medgenetics.ru

Аннотация. Хорошо известно, что патогенез многофакторных заболеваний определяется сложным и нелинейным взаимодействием генетической предрасположенности, эпигенетических модификаций и факторов внешней среды. Однако помимо унаследованных генетических вариантов, важную роль в развитии патологического фенотипа играют соматические мутации, образуя клеточно-специфичные генетические ландшафты и влияя на регионарную предрасположенность органов-мишеней, возраст манифестации и вариабельность клинических проявлений. Таким образом, согласно современным представлениям, многофакторные заболевания формируются в результате

совместного влияния унаследованных герминативных вариантов и соматических мутаций, приобретенных в течение жизни, взаимодействие которых реализуется через клеточно-специфичные молекулярные сети. В связи с этим в обзоре последовательно рассматриваются основные факторы, классические и современные модели сложных патологических фенотипов, а также достижения и перспективы анализа наследуемых генетических вариантов в отношении многофакторных заболеваний. Роль соматических мутаций в формировании патологии обсуждается на примере онтогенеза атеросклероза и концепции атеро-онкологии через оценку соматической вариабельности генома в гладкомышечных клетках атеросклеротической бляшки, когда последовательное накопление «драйверных» мутаций дает селективное преимущество и запускает процесс клональной эволюции в тканевом микроокружении. Отдельное внимание уделено описанию концепции «парадоминантного» наследования – особой форме неменделевской передачи сложных признаков, сочетающей унаследованную предрасположенность и соматические мутации, возникающие на ранних стадиях онтогенеза. На основе анализа современных данных нами предлагается интегративная гипотеза, постулирующая, что унаследованные генетические варианты формируют общеорганизменный фон предрасположенности к многофакторным заболеваниям, в то время как соматические мутации, возникающие и селективно увеличивающиеся в специфических клеточных компартментах органов-мишеней, являются критичным событием, объясняющим топическую специфичность, временну́ю динамику (возраст манифестации) и вариабельность тяжести клинического фенотипа. Новая парадигма предиктивной медицины в отношении многофакторных заболеваний основывается на комплексной характеристике континуума генетических вариантов (унаследованных и соматических), анализе их взаимодействия с клеточно-специфичными молекулярными сетями и переходе от популяционных оценок риска к каузальным моделям индивидуального патогенеза.

Ключевые слова: патогенетика; многофакторные заболевания; омнигенная модель; атеросклероз; злокачественные новообразования; герминальные варианты; соматические мутации

Introduction

Pathogenetics is a discipline that studies the etiological role of hereditary factors and the molecular genetic mechanisms underlying the initiation, development and progression of pathological processes. It focuses on the causal pathways through which a genetic predisposition is transformed into a clinical disease phenotype (Puzyrev, Kucher, 2011). This transformation is considered through the prism of the interaction of genetic variants with epigenetic regulatory systems and networks of intergenic interactions. Moreover, pathogenetics includes the description of processes occurring at all levels of biological organization, from the molecular to the population level, within the framework of interaction with the environment.

The fundamental goal of pathogenetics in medicine is to establish causal relationships between genetic variants and pathological phenotypes, thereby providing the basis for the development of pathogenetically grounded diagnostic strategies, targeted therapies, and preventive interventions (Puzyrev, Kucher, 2011). The greatest progress in this field has been achieved in the study of monogenic and chromosomal diseases. However, the pathogenetics of common diseases (CDs) remains, to a large extent, an insufficiently studied area.

Since the beginning of the era of genome-wide studies, the formal genetic approach to the investigation of CDs, which was primarily based on heritability estimates derived from twin and family studies, has been complemented by the paradigm of medical genomics. This new paradigm was aimed at identifying specific genetic loci and structural variants associated with the predisposition to CDs from evolutionary and ontogenetic perspectives (Puzyrev, Kucher, 2011).

The subsequent methodological shift, which marked the transition from the cataloging of genetic associations to the construction of causal models, required the convergence of genome-wide sequencing, functional genomics, and systems biology data. This integration enabled the interpretation of identified genetic associations through the prism of cell type-specific regulatory networks, epigenetic mechanisms, and a hierarchy of pathophysiological pathways that determine disease trajectories.

The convergence of these methodological approaches has shaped the modern paradigm of CD pathogenetics, in which critical importance is attributed not only to the spectrum of genetic variants but also to their ontogenetic origin. In this framework, genetic variants are classified as either inherited germline variants or *de novo* somatic mutations, each making fundamentally distinct contributions to disease penetrance, age at onset, and clinical polymorphism.

In this regard, the aim of this review is to examine the inherited and somatic components involved in the pathogenetics of CDs.

Common diseases: factors and models

According to current concepts, CDs are the result of complex nonlinear interactions in the “genome–epigenome–environment” system. The intensity with which each of these components has been investigated varies for historical reasons and is largely determined by the level of development of biomolecular and bioinformatic technologies. Consequently, the dominant research paradigms in this field have consistently shifted from classical genetics toward integrative analyses of polygenic networks and their interactions with environmental factors.

Approaches to identifying predisposition loci for CDs have evolved from paradigms based on candidate gene analysis and microarray-based genome-wide association studies (GWAS) to population-specific analysis of whole genome

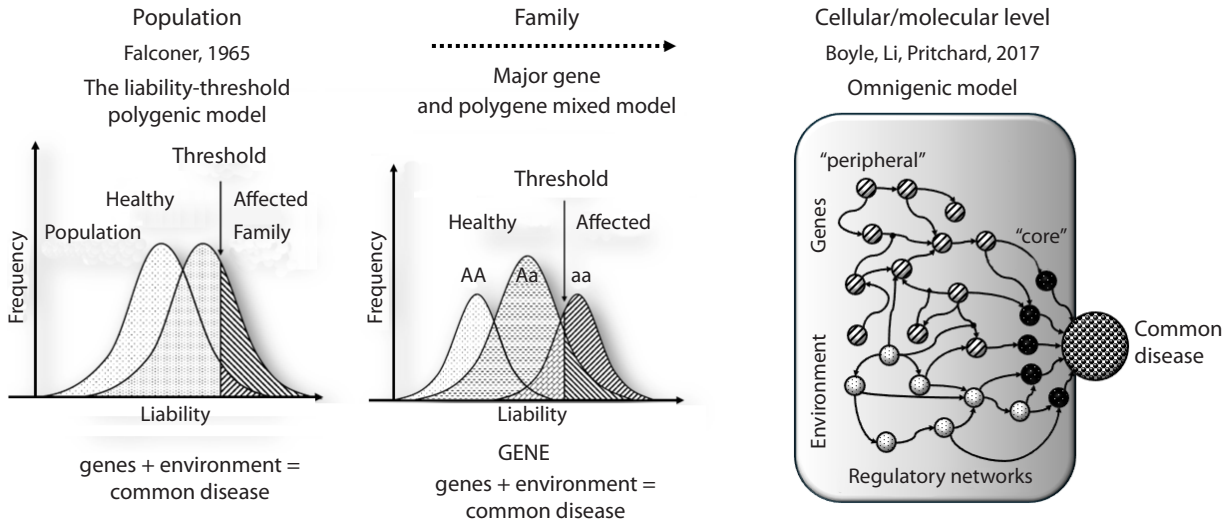


Fig. 1. Evolution of inheritance models for common diseases.

sequencing data with the detection of the full spectrum of genetic variation, from common to unique variants, in combination with *in vivo* and *in vitro* functional genomics approaches. These paradigms are further complemented by epigenomic annotation maps for prediction of cell type-specific regulatory effects and identification of causal genes.

An important aspect of understanding genotype-phenotype relationships in CDs is determining how genetic variants influence gene regulation and expression across different tissues and cell types. Such analyses may be performed using *in silico* bioinformatic approaches or through the integration of GWAS data with epigenomic and transcriptomic datasets, including single-cell datasets.

The central concept in the study of CDs is genetic predisposition, defined as an integral probabilistic measure that quantifies the contribution of polygenic burden, inherited rare high-penetrance variants, and the modifying effects of environmental factors to the individual profile of the disease risk. This profile is described using probability distribution functions for different clinical phenotypes while accounting for the temporal dynamics of phenotype manifestation.

According to established models, the formation of predisposition to CDs is described within the framework of either polygenic or oligogenic architectures. The polygenic architecture involves the additive effects of numerous genetic variants (polygenes), each exerting a small effect, whereas the oligogenic architecture is characterized by a limited number of driver genes with substantial effects on penetrance. In both cases, additional genetic and epigenetic modifiers influence disease onset and the severity of the clinical phenotype (Fig. 1).

A key implication of the liability-threshold polygenic model is that genetic predisposition alone does not inevitably result in disease development but only shifts the distribution of disease risk within a population. Meanwhile, clinical manifestation occurs when the individual multifactorial burden exceeds a certain systemic threshold of tolerance, determined by the reserve capacity of compensatory physiological mechanisms.

One of the influential hypotheses proposed to explain the complex genetic architecture of CDs is the omnigenic model (Boyle et al., 2017) (Fig. 1). According to this concept, a large repertoire of genetic variants, predominantly exerting *cis*-regulatory effects, modulates the expression of genes functioning on the periphery of pathogenetic pathways. These dysregulated transcription programs are integrated through complex *trans*-regulatory networks, converging in the altered activity of a limited number of highly conserved core genes that directly determine key pathophysiological processes. It should be emphasized that the dichotomy of "core" and "peripheral" genes itself is functional and context-dependent rather than strictly structural, since the same gene may function as a "core" gene in one tissue-specific network and as a "peripheral" gene in another. Statistically significant GWAS associations are detected predominantly for "peripheral" genes because such variants have higher population frequencies and because current statistical approaches possess limited statistical power to detect the effects of rare variants located within "core" gene loci.

The omnigenic model provides a unified explanation for the key features of the genetic architecture of CDs, including high polygenicity, the predominance of low-effect variants, widespread pleiotropy, and the critical importance of regulatory interactions mediated by transcriptional networks. This model claims to be a conceptual shift in the interpretation of the genetic architecture of CDs, offering a mechanistic explanation of polygenicity through the prism of regulatory convergence. However, its full verification requires a transition from correlational expression maps to direct experimental demonstration of causal relationships within regulatory networks, for example through Mendelian randomization approaches or gene-editing technologies, as well

as taking into account the contribution of non-transcriptional mechanisms such as post-translational modifications and metabolic control.

Despite its heuristic value, the omnigenic model remains a simplified representation and leaves several fundamental questions to be solved. What objective biomolecular and functional criteria should be used to stratify genes into "core" and "peripheral" categories and to identify them reliably across different tissues and ontogenetic stages? Are pathogenetically relevant regulatory networks exclusively intracellular, or do they fundamentally incorporate intercellular signaling cascades that form the tissue-specific disease contexts? How can the contribution of converging regulatory influences be quantified relative to direct causal effects in specific cell types? Furthermore, in its current formulation, the model does not fully account for the role of chromatin spatial organization and intercellular communications in the formation of an integral pathogenetic context.

Inheritable genome and common diseases

Modern studies of the genetic architecture of CDs, initially based on GWAS, have cumulatively identified a vast repertoire of statistically significant associations between genetic variants, primarily single nucleotide polymorphisms (SNPs), and diseases catalogued in repositories such as NHGRI-EBI GWAS Catalog (https://www.ebi.ac.uk/gwas/). Analysis of these data has revealed the key features of such associations.

The vast majority of identified genetic associations are characterized by small effect sizes and increase the odds ratio (OR) of disease development to approximately 1.1–1.5, corresponding to a relative risk increase of 10–50 %. As a rule, a single common genetic variant (minor allele frequency [MAF] > 1–5 %) explains only a small proportion of heritability for most CDs, which limits its clinical utility as an isolated prognostic biomarker.

Cohort and prospective studies are currently being conducted to refine these risk estimates and to elaborate methods for genetic risk stratification. To maximize the statistical power of GWAS and minimize systematic errors resulting from population stratification, it is critically important to form large, phenotypically and genotypically well-characterized cohorts that adequately reflect the ethnic background of study participants. Such cohorts are also essential for the elaboration and calibration of polygenic risk scores (PRSs) relevant to specific populations, since differences in allele frequencies and haplotype structure between populations substantially limit the applicability of scores created on a different genetic background.

In the Russian Federation, the elaboration of population-specific PRS is a topic of active research, with research groups led by E.K. Khusnutdinova (Academician of the Bashkortostan Academy of Sciences), O.M. Drapkina (Academician of the Russian Academy of Sciences, RAS), M.I. Voevoda (Academician of RAS), A.V. Polonikov (Doctor of Research Medicine), A.N. Meshkov (Doctor of Research Medicine), M.I. Churnosov (Doctor of Research Medicine), V.E. Golimbet (Doctor of Biological Sciences), A.O. Kibitov (Doctor of Research Medicine), and researchers from the Centre for Strategic Planning of the Federal Medical and Biological Agency of Russia.

It has been established that combinations (ensembles) of allelic variants of various genes may substantially increase both disease susceptibility and the probability of disease complications. The combined effect of such polygenic combinations, aggregated into PRSs, may explain a considerable proportion of heritability for certain CDs; however, their contribution varies significantly among diseases and depends on the ethnic composition of the studied population. Nevertheless, a substantial fraction of CD heritability remains unexplained within the framework of the "common disease-common variant" model, a phenomenon referred to as "missing heritability". Current concepts suggest that this missing component reflects the combined contribution of rare genetic variants, structural genomic rearrangements, epistatic interactions. The limitations of statistical models used to estimate heritability also play a role, as they do not fully account for the contribution of somatic mutagenesis and inaccuracies in phenotyping.

Whole genome sequencing (WGS) and whole exome sequencing (WES) are critical for identifying a range of rare and unique variants, including single nucleotide variants (SNVs) and copy number variations (CNVs). To evaluate their contribution to the "missing heritability" and establish associations with disease phenotypes, specialized statistical approaches are used, such as the transmission disequilibrium test (TDT) for *de novo* mutation analysis and variant aggregation methods (burden test, SKAT), which assess the cumulative effects of rare alleles within genes or biological pathways.

The variants associated with CDs are located in the nuclear and mitochondrial genomes, highlighting the complexity of the functional interactions underlying the pathogenesis of the respective diseases. Genetic studies have revealed widespread pleiotropy, in which individual genetic loci and specific variants are associated with the risk of developing several CDs, which can vary in their clinical manifestations. This phenomenon provides a molecular basis for genetic correlations observed clinically among diseases, and for phenomena such as syntropy (comorbidity) (Bragina, Puzyrev, 2023). These findings indicate the presence of shared molecular mechanisms in the pathogenesis of various CDs and open new perspectives for understanding their genetic architecture.

The GWAS results demonstrate that the vast majority of significant SNPs are located within non-coding genomic regions enriched with epigenetic marks of active regulatory elements such as enhancers and promoters. This observation suggests that SNPs predominantly influence CD risk through indirect regulatory mechanisms, including allele-specific effects on transcription factor binding and subsequent dysregulation of target gene expression, rather than through direct alterations in protein structure. However, the interpretation

of functional significance of SNPs requires caution because of potential linkage disequilibrium with truly causal variants, including rare or structural variants.

Moreover, despite the large number of identified associations, the "causal" or "driver" genetic variants in relation to CDs often remain unknown. Consequently, modern studies of the CD pathogenesis rely on integrative systems biology approaches aimed at constructing causal models through the combined analysis of multi-omics datasets. However, the key methodological challenge remains not simply the integration of heterogeneous datasets (genome, epigenome, translatome, transcriptome, proteome, metabolome, and exposome), but also the reconstruction of the temporal sequence and hierarchical organization of molecular events within pathogenetic cascades, as well as taking into account the spatial organization of the genome and intracellular compartments that determine the specificity of molecular interactions.

Furthermore, the relationship between genetic variants and CDs is now being analyzed not only for individual diseases but also for patterns of disease co-occurrence (syntropy) and mutual exclusion (dystropy), as well as within the broader framework of the diseasome, which conceptualizes diseases according to their shared molecular networks and genetic relatedness (Bragina, Puzyrev, 2023). However, accurate quantification of the diseasome requires more complex metrics than the simple enumeration of shared loci.

Generalizations within the framework of the study of the phenomenon of disease combinations provide an opportunity to approach the systematization and natural classification of diseases based not on organ systems or clinical symptoms, but rather on the commonality of their underlying molecular networks and pathogenetic pathways. This approach opens up prospects for the development of fundamentally new therapeutic strategies targeting key nodes shared among multiple diseases rather than individual nosological entities.

The strategy of deconvolution of the molecular genetic mechanisms of CDs requires integrative profiling of a broad spectrum of genetic variants (SNV, CNV, mobile elements), epigenetic landscapes (DNA methylation, histone modifications, chromatin accessibility patterns) and gene expression profiles obtained with single-cell resolution and spatial-temporal organization in target organ tissues using single-cell multi-omics and spatial transcriptomics technologies.

The statistical analysis of such sparse data and identification of biologically meaningful patterns remains methodologically challenging. A key step is the integration of GWAS data with multi-omics profiles of relevant cell types to infer causal genes and cell-specific regulatory pathways. However, the success of such approaches fundamentally depends on the completeness and accuracy of cellular atlases and annotations of functional genomic elements for specific human tissues, since incomplete or inaccurate reference datasets may lead to distorted conclusions regarding causality.

Historically, most studies of CD genetics have focused on inherited germline variants, primarily SNPs, analyzed in easily accessible surrogate tissues such as peripheral blood. This approach imposes substantial limitations because it underestimates the tissue and cell specificity of molecular processes, as well as the dynamics of transcriptional and epigenetic states during ontogenesis. Such studies are based on the assumption that most cells in the human body have an identical genome and are genetically stable throughout ontogenesis. However, this traditional perspective ignores the growing body of evidence indicating a significant contribution of somatic mutations to CD pathogenesis, which manifests through the formation of mosaic genetic landscapes in target tissues (dysfunction of tissue-specific cell pools) and can significantly modify the phenotype determined by germinal variants. The role of somatic mutations in the development of CDs may be illustrated using the example of atherosclerosis.

Somatic genome variability in atherosclerotic plaque smooth muscle cells

The pathogenesis of atherosclerosis may be viewed through the prism of clonal expansion of somatic cells carrying pathogenic genetic variants within the arterial wall. In this context, atherogenesis demonstrates conceptual parallels with the model of carcinogenesis, in which the sequential accumulation of "driver" mutations gives a selective advantage and promotes clonal evolution within the tissue microenvironment. It is well known that atherosclerosis and malignancy share numerous common features, from a prolonged latent period and overlapping risk factors to similar molecular and cellular processes and mechanisms (Baylis et al., 2023).

At the molecular level, these similarities are reflected in the common characteristics of the microenvironment, including chronic unresolved inflammation, hypoxia, angiogenesis and extracellular matrix remodeling. They also involve fundamental cellular phenotypes such as the metabolic shift towards aerobic glycolysis, resistance to apoptotic signals, dysregulation of cellular senescence mechanisms, escaping immune surveillance in a manner similar to cancer stem cell (CD47 expression, impairment in efferocytosis), epithelial-mesenchymal transition, increased invasiveness and, most importantly, clonal expansion driven by somatic mutations (DiRenzo et al., 2017).

The discovery of monoclonality of smooth muscle cells (SMCs) within atherosclerotic plaques formed the basis for the hypothesis that atherogenesis represents a neoplastic process in which a series of somatic driver mutations, for example in cell cycle regulatory genes (*TP53*, *NOTCH1*), determines the clonal expansion of cells with a selective advantage (Benditt E., Benditt J., 1973). According to this hypothesis, the initiation and progression of atherosclerotic lesions are associated with the accumulation of somatic mutations induced by both endogenous processes (oxidative stress, replication errors) and exogenous factors (chemical mutagens, oncogenic viruses), which creates the molecular basis for clonal evolution within the vessel wall. Although

the monoclonal theory of atherogenesis was proposed a long time ago, experimental studies aimed at its verification began to emerge only after 2015.

The validation of the monoclonal hypothesis of atherosclerosis became possible through the development of high-resolution approaches, such as lineage tracing in animal models, as well as single-cell multi-omics analysis (scRNA-seq, scATAC-seq) combined with spatial transcriptomics technologies. These approaches make it possible not only to visualize the clonal architecture of the plaque, but also to establish causal relationships among specific somatic mutations, clonal dynamics, and pathogenic transcriptomic programs, thereby revealing the sequence of cellular and molecular events underlying disease development.

Application of an integrated methodological approach has demonstrated that phenotypically altered SMCs within atherosclerotic plaques accumulate oncogenic "driver" mutations and exhibit properties resembling those of tumor cells, including disturbances in cell cycle regulation and apoptosis mechanisms. These findings provide direct experimental support for the hypothesis of a deep commonality of the fundamental molecular and cellular mechanisms underlying atherogenesis and carcinogenesis.

SMCs and their derivatives within atherosclerotic plaques exhibit accumulation of oxidative DNA damage, which serves as a source of somatic mutations and leads to an increase in genomic instability that correlates with the extent of SMC transdifferentiation toward macrophage-like and mesenchymal cells. Phenotypically altered SMCs also harbor somatic copy number variants (CNVs) affecting oncogenes (*EGFR*, *KRAS*) and tumor suppressor genes (*FAT1*, *UBR5*), as well as genes associated with coronary artery disease (*CCM2*, *PALLD*). Notably, the spectrum of these genomic alterations overlaps with that observed in solid tumors. The oncogenic *KrasG12D* mutation accelerates SMC phenotypic switching and promotes atherosclerosis progression (Pan et al., 2024).

According to the current paradigm, mature SMCs undergo dedifferentiation in response to pathogenic stimuli and give rise to oligoclonal populations. These clones migrate from media to intima and form a fibrous plaque cap, but their subsequent fate is heterogeneous. Some clones transdifferentiate into macrophage-like phenotypes that promote inflammation and destabilization, whereas others acquire osteogenic properties leading to vascular calcification. Together, these processes influence plaque structural integrity and susceptibility to rupture. Thus, clonal expansion of SMCs represents a dynamic pathological process that may initially contribute to formation of a stabilizing fibrous cap during the early stages of atherogenesis, but later promotes plaque destabilization through the accumulation of additional somatic mutations and activation of pro-inflammatory transdifferentiation programs, thereby increasing the risk of plaque rupture and acute vascular events.

The molecular basis of SMC clonal expansion may involve the existence of discrete subpopulations of progenitor cells within the vascular wall characterized by unique transcriptomic and epigenetic landscapes that determine enhanced proliferative capacity and resistance to apoptosis under conditions of atherogenic stress. In particular, experimental studies have demonstrated that the C3high Sca1+ SMC subpopulation functions as a key regulatory node. These cells not only promote autonomous SMC proliferation through autocrine signaling but also mediate macrophage recruitment and functional reprogramming toward a pro-inflammatory phenotype characterized by impaired efferocytosis, thereby enhancing the inflammatory response and exacerbating tissue damage (Dobnikar et al., 2018).

An alternative model suggests that, although all SMCs of media initially possess proliferative potential, clonal dynamics is determined by the selection of subpopulations that acquire a selective advantage through somatic mutations, enabling hyperproliferation and evasion of efferocytosis. In parallel, selective advantage may also result from increased tolerance to a pathological environment characterized by metabolic stress and inflammation, or through mechanisms of lateral inhibition in which dominant clones actively suppress neighboring cells via short-range (juxtacrine) signaling (Luo et al., 2023).

Accumulating evidence suggests that the contribution of structural mutations themselves to SMC phenotypic alterations may be limited. Instead, epigenetic reprogramming, including alterations in DNA methylation and histone modifications, may function as an alternative or synergistic mechanism underlying SMC clonal expansion. In this case, SMC clonality may be partially explained by dynamic changes in the chromatin architecture that contribute to maladaptive cellular responses to inflammatory stress. Notably, the transcription factor ATF3 has been identified as a key suppressor of SMC transition toward a pro-inflammatory phenotype. Its protective effect is realized not through the activation of anti-inflammatory programs, but rather through the direct suppression of pro-inflammatory signaling pathways, particularly through the inhibition of the complement activation cascade (Wang et al., 2022).

The molecular mechanisms of SMC clonal expansion may be mediated by intercellular interactions in the macrophage–SMC axis. Moreover, myeloid cells carrying functional defects, such as decreased expression of the epigenetic regulator TET2 and integrin β3, generate a pro-inflammatory environment through hyperactivation of the TNF signaling pathway. This process contributes not to oligoclonal but to polyclonal proliferation of SMC progenitor cells and the formation of more extensive atherosclerotic lesions (Kabir et al., 2023).

The heterogeneity of SMC transcriptional profiles, determined by their embryonic origin (Dobnikar et al., 2018), may underlie the variable potential for clonal expansion among various segments of the arterial bed, which, in turn, may explain the regional predisposition to pathological lesion formation. It has been hypothesized that clonal expansion of vascular cells, including SMCs, endothelial cells, and macro-

phages, in various vasculopathies (atherosclerosis, aneurysm, pulmonary hypertension, and cavernous malformation of the brain) may be regulated by conserved molecular pathways shared across distinct diseases (Lin et al., 2023). These pathways are likely integrated with fundamental ontogenetic programs that determine cellular plasticity and proliferative potential of the cells.

Moreover, it has been demonstrated that the epigenomic landscape, particularly chromatin accessibility, in individual cells (SMCs, fibroblasts, endothelial cells) within different segments of the ascending aorta is determined by their embryonic origin. These ontogenetically inherited regulatory domains spatially co-localize with cardiovascular disease risk loci identified in genome-wide association studies (GWAS). In particular, studies have shown that in normal mice the transcriptomic and epigenomic landscapes of individual arterial cells (SMCs, fibroblasts, and endothelial cells) depend on cell type and vascular bed, and the "open chromatin" regions are enriched with human cardiovascular disease risk genes (Weldy et al., 2025).

It has been assumed that enhancers of genes of developmental transcription factors (*Tbx20*, *Hand2*, *Gata4*, and *Hoxb*), which are specific to particular cell types and vascular regions, serve as a key element in the interpretation of complex genetic risk factors for arterial pathology by transforming inherited polymorphisms into cell-specific pathological programs (Weldy et al., 2025). At the same time, aneurysm/dissection and atherosclerosis of the ascending aorta in humans are associated with multidirectional alterations in DNA methylation patterns within regulatory regions linked to the transcription factor genes *NR2F1* and *TBX20*, which play a key role in the vascular development in embryogenesis. These findings indicate reactivation of ontogenetic programs during disease pathogenesis (Koroleva et al., 2024).

Thus, the cumulative and dynamic interaction of multiple genetic and epigenetic factors affecting cellular and molecular systems throughout ontogenesis forms a pathological context that promotes disease comorbidity. Within this context, somatic variants, including both structural genomic alterations and persistent epigenetic reprogramming, function as universal mechanisms modifying cellular phenotypes and integrating pathogenic influences.

Somatic mutations are detected not only in parenchymal cells of target organs in CDs, but also in hematopoietic cells, where they manifest as clonal hematopoiesis (CH). Of particular clinical and biological significance is clonal hematopoiesis of indeterminate potential (CHIP), characterized by the presence of a "driver" mutation in the absence of diagnostic criteria for hematologic neoplasm (Jaiswal et al., 2014). Intensive study of CHIP has revealed its role as a systemic pro-inflammatory factor mediating the effect of somatic mosaicism on the pathogenesis of a broad range of CDs through the secretion of pro-inflammatory cytokines by mutant macrophage clones and other myeloid-derived cells (Schuermans, Honigberg, 2025).

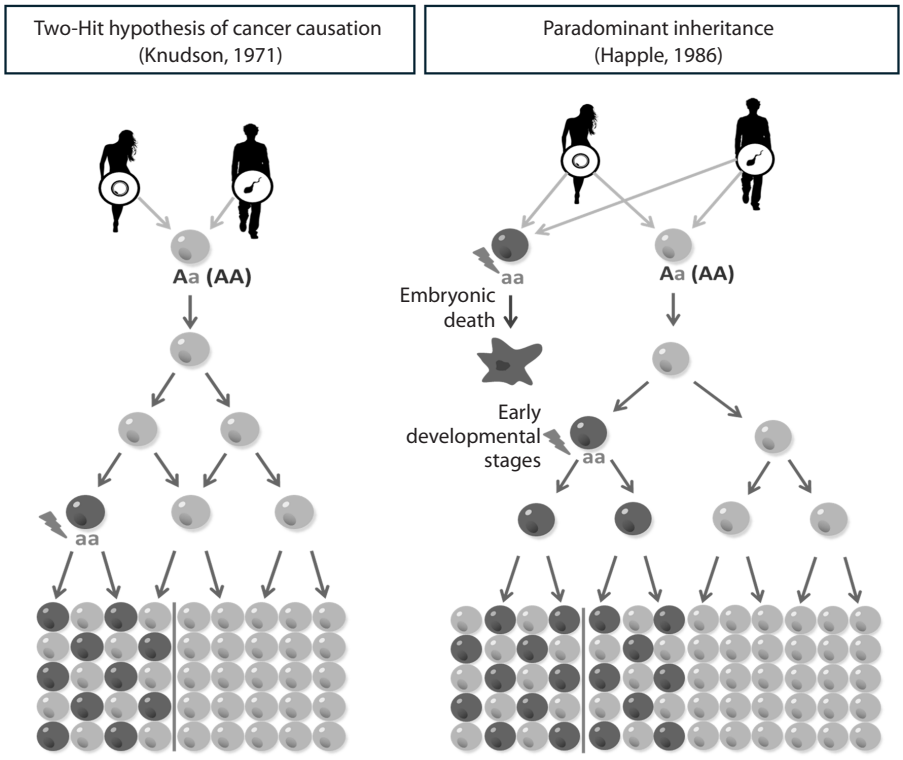
However, the detailed molecular and cellular mechanisms linking CHIP to such a wide range of CDs remain incompletely understood. The concept of "paradominance", in which a somatic mutation occurring on the background of an inherited heterozygous mutation in the same gene (for example, *TET2*, *DNMT3A*) leads to a complete loss of function and subsequent manifestation of a pathological phenotype, represents an important conceptual bridge between classical genetics and the genetics of somatic cells. This concept offers a mechanistic explanation for cases of incomplete penetrance and mosaic expressivity that cannot be adequately explained by traditional inheritance models and may be particularly relevant for genes in which constitutional homozygous inactivation is embryonically lethal.

The concept of "paradominant" inheritance in relation to common diseases

"Paradominant" inheritance represents a distinct form of non-Mendelian inheritance that combines inherited predisposition with somatic mutations acquired during ontogenesis (Happle, 1991, 1993). Unlike classical Mendelian models, this mechanism assumes that the pathological phenotype manifests only when an inherited heterozygous germline mutation is followed by a somatic mutation resulting in loss of heterozygosity (LOH) of the remaining wild-type allele in progenitor cells. This process leads to clonal expansion of cells carrying biallelic gene inactivation. Importantly, constitutional homozygosity for such mutations is usually embryonically lethal (Fig. 2), which explains why these alleles persist in the population primarily in the heterozygous state and manifest phenotypically only after somatic second-hit events.

A key feature of "paradominant" inheritance is the formation of a mosaic genotype resulting from LOH events occurring in a subset of somatic cells at an early stage of embryo development. Interestingly, similar mosaic alterations may arise independently in multiple members of the same family, thereby mimicking autosomal dominant inheritance despite a fundamentally different underlying mechanism. This phenomenon formed the basis of the term "paradominance".

Unlike the Knudson model, according to which constitutional heterozygosity for non-lethal mutations predisposes individuals to tumor development, paradominance is characterized by constitutional heterozygosity specifically for embryonically lethal mutations, which excludes Mendelian inheritance and manifests exclusively through somatic mosaicism. In this context, the somatic event – most commonly post-replicative mitotic recombination – occurs at an early stage of embryogenesis. The specific time point of mutation occurrence determines both the size and tissue prevalence of the mutant clone, which in turn determine not only the mosaic pattern of abnormal cell distribution but also the extent of tissue involvement, which directly influences clinical severity (Happle, 2009).



Nazarenko M.S. et al. Genome variability of somatic cells in human complex diseases. *Medical Genetics*. 2017;16(12):4-8 (in Russian)

Fig. 2. Molecular mechanisms of somatic mutations.

Initially, this concept was proposed to explain the pathogenesis of several dermatological syndromes, such as McCune–Albright syndrome and Becker's pigmented nevus (Happle, 1991, 1993). A classic example is provided by multifocal venous malformations, in which biallelic somatic mutations in the *TEK* (receptor tyrosine kinase) gene arise in the background of an inherited germline p.R849W mutation. This combination results in autonomous receptor activation and local clonal expansion of endothelial cells, highlighting the importance of targeted analysis of somatic variants directly in pathological tissues (Limaye et al., 2009). However, subsequent studies have revealed the role of "paradominant" inheritance in other pathologies, including atherosclerosis (Fig. 2).

Modern studies employing high-resolution sequencing technologies and mosaicism analyses are likely to clarify the contribution of "paradominant" mechanisms to the pathogenesis of CDs, particularly in diseases characterized by sporadic occurrence, familial clustering, and mosaic patterns of organ involvement that cannot be adequately explained by classical inheritance models.

Translational potential in the practice of common diseases research

At the current stage of development, genetic testing for hereditary predisposition to CDs is characterized by its gradual integration into clinical practice. The elaboration of population-specific PRSs based on genome-wide data enables stratification of individuals in cohorts for predictive assessment of the risk of developing CDs, their complications, and potential response to pharmacotherapy. However, the widespread clinical implementation of these approaches remains limited by the low reproducibility of predictive power when used in genetically distinct populations.

When whole-exome sequencing (WES) or whole-genome sequencing (WGS) is performed in CD cohorts, analysis of rare genetic variants in individuals may facilitate the identification of novel molecular targets for pharmacological intervention. Moreover, it has been demonstrated that combining common and rare genetic variants within a single model improves the accuracy of CD risk prediction (Fiziev et al., 2023).

The integration of the so-called "non-genetic" determinants of CDs – including epigenetic markers, exposome parameters, as well as their complex interactions, analyzed within the framework of multi-omics approaches and comorbid pattern analysis, – into unified predictive models has considerable, although methodologically more complex, translational potential compared with purely genetic scales.

Understanding the prevalence, spectrum, and mechanisms underlying CHIP in CDs is important for the development of novel screening strategies, individual risk assessment and the implementation of diagnostic and treatment programs. For example, an individual prognosis may be established through integration of traditional risk factors and clinical status with the molecular characteristics of CHIP. The type of somatic

mutation and VAF are of key importance. An individual with large *TET2* CHIP (>10 % VAF) and borderline 10-year risk of cardiovascular complications would be counseled regarding aggressive risk reduction with intensification of lifestyle modifications and aggressive lipid-lowering and anti-hypertensive therapies. In contrast, such interventions are not indicated in individuals with *DNMT3A* CHIP and a low 10-year CVD risk. While there are currently no randomized clinical trials specifically supporting these interventions on the basis of CHIP status, therapeutic tactics relies on shared decision-making, when the doctor and the patient carefully balance potential benefits and risks (Oren, Libby, 2025). Translation of findings from fundamental studies investigating the role of somatic genetic variants in target organ cells in CDs into clinical practice is not so obvious.

Targeting the pathogenetic mechanisms underlying clonal expansion represents a promising but methodologically complex direction for the development of novel therapeutic strategies for CDs. Such approaches include the selective elimination of mutant hematopoietic stem cell clones, for example, through senolytic therapies, modulation of the pro-inflammatory phenotype of smooth muscle cell clones, or inhibition of specific inflammatory mediators produced by these clones.

The proof of the fundamental possibility of such targeted therapeutic approaches is the atheroprotective effect observed for some antitumor agents. These include the blockade of the "don't eat me" signaling pathway using anti-CD47 antibodies to enhance the efferocytosis of apoptotic plaque cells (Kojima et al., 2016; Tao et al., 2020); inhibition of SMC clone proliferation using tretinoin (Pan et al., 2020); and the use of selective vulnerability of SMCs with impaired DNA repair through the use of PARP inhibitors such as niraparib (Pan et al., 2024). These findings open new prospects for the repositioning of oncological drugs in cardiovascular medicine.

Conclusion

The evolution of CD pathogenetics demonstrates a shift from the identification of disease-associated loci to the systemic functional characterization of cellular phenotypes driven by these genetic variants within cell-specific molecular networks, which marks the transition from genetic association to a causal understanding of pathogenesis.

The pathogenesis of CDs is characterized by multifactorial nature and is realized through nonlinear dynamic interactions between genetic predisposition, epigenetic modifications and environmental factors, forming complex causal networks. In this regard, progress in the prediction, diagnosis, treatment and prevention of CDs in the context of personalized medicine requires not merely the accumulation of data, but the elaboration of fundamentally new computational and conceptual models capable of integrating heterogeneous causal factors and their nonlinear interactions at various levels of biological organization, from the molecular to the organismal level, in order to construct predictive dynamic models of individual pathogenesis.

Somatic mutations that form mosaic genetic landscapes (somatic mosaicism) at the level of intra- and interstitial heterogeneity, together with tissue- and cell-specific epigenetic programs, act as key modifiers of penetrance, expressivity, and clinical polymorphism in CDs. Since the burden and spectrum of somatic mutations, including "driver" events, accumulate in a tissue-specific manner throughout life during ontogenesis, this process represents a universal mechanism explaining such fundamental characteristics of CDs as the regional predisposition of target organs and variable expressivity in different individuals. At the same time, the spatial distribution and clinical manifestation of these mutations critically depend on local selective processes and the rate of cellular turnover within each specific tissue.

The greatest success in characterizing the frequency and spectrum of somatic mutations at the molecular level in target organ cells and tissues in "non-cancerous" CDs has been achieved for SMCs in atherosclerosis. The key tasks remain unresolved: the systematic identification and functional verification of the full repertoire of "driver" somatic mutations, the determination of the contribution of structural variants and mobile elements to pathogenesis, as well as the full-scale characterization of the frequency and spectrum of somatic mosaicism across different cell types (populations of macrophages, endotheliocytes, smooth muscle cells, neurons) in a broad range of CDs.

Genotype-specific patterns of clonal hematopoiesis (CHIP) determined by mutations in distinct "driver" genes (*TET2*, *DNMT3A*, *JAK2*, *ASXL1*) are associated with divergent cardiovascular risk profiles, indicating the need for the development of targeted therapeutic interventions. Studies in cellular and mouse models, in particular using bone marrow transplantation from mutant animals, confirm the causal relationship between CHIP and cardiovascular diseases. This relationship is realized through a complex of alterations, including a loss of balance between self-renewal, proliferation and differentiation of hematopoietic stem cells, activation of pro-inflammatory and profibrotic mechanisms in clones, as well as increased resistance of clones to external stressors. However, translation of these findings into human medicine requires careful consideration of the fundamental differences in lifespan, metabolism, and immune system organization between model organisms and humans.

Based on the analysis of modern data, we propose an integrative hypothesis according to which inherited genetic variants form an organizational background of predisposition to CDs, while somatic mutations arising and selectively expanding within specific cellular compartments of target organs represent critical events explaining topical specificity, temporal dynamics (age at onset) and variability in the severity of clinical phenotype.

The 9p21 locus containing the tumor suppressor genes *CDKN2A* and *CDKN2B* provides a pathogenetic illustration

of this relationship. Inherited variants at this locus, associated with coronary heart disease risk, create a state of haploinsufficiency, which manifests in weaker control over cell cycle regulation and DNA repair in vascular smooth muscle cells. However, realization of the full pathogenic potential of this predisposition requires an additional event, such as somatic inactivation of the remaining allele. This initial vulnerability constitutes the essence of a "permissive environment", since it increases the probability that a somatic event (such as loss of heterozygosity for the remaining wild-type allele) will lead not to cell death, but to its selective advantage and subsequent clonal expansion, finally inactivating the tumor suppressor locus and fixing the proliferative phenotype.

Moreover, in atherosclerotic vessels, activation of the pro-inflammatory SMC phenotype is associated with epigenome remodeling, in particular, with a decrease in the availability of the *ATF3* gene promoter, which is a key transcriptional repressor of inflammation. Notably, germline variants at the *ATF3* locus are also associated with coronary artery disease risk according to GWAS (Wang et al., 2022). The combined presence of inherited predisposition and acquired epigenetic dysregulation within the same molecular node highlights the principle of convergence of heterogeneous pathogenic effects on key regulatory elements, which can serve as a general model for explaining the pathogenesis of multiple diseases. This principle of convergence demonstrates that the pathogenesis of CDs is determined not by a simple sum of independent factors, but by their nonlinear dynamic interaction, in which inherited variants create a context that determines the specifics of the response to acquired epigenetic alterations, forming a unique pathological profile for each individual.

Thus, a significant proportion of the population heritability of CDs is explained by the spectrum of frequent inherited genetic variants, and rare, especially somatic mutations make a critical contribution to individual risk by modulating the penetrance of hereditary predisposition and determining the specifics of the pathological phenotype at the level of tissues and organs. Consequently, comprehensive characterization of the continuum of genetic variants, encompassing both inherited polymorphisms and somatic mutations acquired during life, in their dynamic interaction with cell-specific molecular networks in normal and pathological conditions forms a new paradigm of predictive medicine. This paradigm marks the transition from population-based probabilistic risk assessments to causal models of individual pathogenesis based on the principles of clonal somatic evolution.

References

Baylis R.A., Gao H., Wang F., Bell C.F., Luo L., Björkegren J.L.M., Leeper N.J. Identifying shared transcriptional risk patterns between atherosclerosis and cancer. *iScience*. 2023;26(9):107513. doi 10.1016/j.isci.2023.107513

Benditt E.P., Benditt J.M. Evidence for a monoclonal origin of human atherosclerotic plaques. *Proc Natl Acad Sci USA*. 1973;70(6):1753-1756. doi 10.1073/pnas.70.6.1753

Boyle E.A., Li Y.I., Pritchard J.K. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169(7):1177-1186. doi 10.1016/j.cell.2017.05.038

Bragina E.Yu., Puzyrev V.P. Genetic outline of the hermeneutics of the diseases connection phenomenon in human. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2023;27(1):7-17. doi 10.18699/VJGB-23-03

DiRenzo D., Owens G.K., Leeper N.J. "Attack of the Clones": commonalities between cancer and atherosclerosis. *Circ Res*. 2017;120(4):624-626. doi 10.1161/CIRCRESAHA.116.310091

Dobnikar L., Taylor A.L., Chappell J., Oldach P., Harman J.L., Oerton E., Dzierzak E., Bennett M.R., Spivakov M., Jørgensen H.F. Disease-relevant transcriptional signatures identified in individual smooth muscle cells from healthy mouse vessels. *Nat Commun*. 2018;9(1):4567. doi 10.1038/s41467-018-06891-x

Fiziev P.P., McRae J., Ulirsch J.C., Dron J.S., Hamp T., Yang Y., Wainschtein P., … Marques-Bonet T., Rehm H.L., O'Donnell-Luria A., Khera A.V., Farh K.K. Rare penetrant mutations confer severe risk of common diseases. *Science*. 2023;380(6648):eabo1131. doi 10.1126/science.abo1131

Happle R. The McCune-Albright syndrome: a lethal gene surviving by mosaicism. *Clin Genet*. 1986;29(4):321-324. doi 10.1111/j.1399-0004.1986.tb01261.x

Happle R. Allelic somatic mutations may explain vascular twin nevi. *Hum Genet*. 1991;86(3):321-322. doi 10.1007/BF00202421

Happle R. Mosaicism in human skin. Understanding the patterns and mechanisms. *Arch Dermatol*. 1993;129(11):1460-1470

Happle R. Monoallelic expression on autosomes may explain an unusual heritable form of pigmentary mosaicism: a historical case revisited. *Clin Exp Dermatol*. 2009;34(7):834-837. doi 10.1111/j.1365-2230.2009.03543.x

Jaiswal S., Fontanillas P., Flannick J., Manning A., Grauman P.V., Mar B.G., Lindsley R.C., … Atzmon G., Wilson J.G., Neuberg D., Altshuler D., Ebert B.L. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014;371(26):2488-2498. doi 10.1056/NEJMoa1408617

Kabir I., Zhang X., Dave J.M., Chakraborty R., Qu R., Chandran R.R., Ntokou A., … Hwa J., Kluger Y., Martin K.A., Fernández-Hernando C., Greif D.M. The age of bone marrow dictates the clonality of smooth muscle-derived cells in atherosclerotic plaques. *Nat Aging*. 2023;3(1):64-81. doi 10.1038/s43587-022-00342-5

Knudson A.G. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci USA*. 1971;68(4):820-823. doi 10.1073/pnas.68.4.820

Kojima Y., Volkmer J.P., McKenna K., Civelek M., Lusis A.J., Miller C.L., Direnzo D., … Maegdefessel L., Matic L.P., Hedin U., Weissman I.L., Leeper N.J. CD47-blocking antibodies restore phagocytosis and prevent atherosclerosis. *Nature*. 2016;536(7614):86-90. doi 10.1038/nature18935

Koroleva Iu.A., Goncharova I.A., Zarubin A.A., Shipulina S.A., Sleptsov A.A., Panfilov D.S., Kozlov B.N., Nazarenko M.S. Methylation levels in the 5′ region of the *TBX20* gene in the ascending aorta change in opposite direction in atherosclerosis and aneurysm. *Russ J Genet*. 2024;60(7):977-981. doi 10.1134/S102279542470042X

Limaye N., Wouters V., Uebelhoer M., Tuominen M., Wirkkala R., Mulliken J.B., Eklund L., Boon L.M., Vikkula M. Somatic mutations in angiopoietin receptor gene *TEK* cause solitary and multiple sporadic venous malformations. *Nat Genet*. 2009;41(1):118-124. doi 10.1038/ng.272

Lin A., Brittan M., Baker A.H., Dimmeler S., Fisher E.A., Sluimer J.C., Misra A. Clonal expansion in cardiovascular pathology. *JACC Basic Transl Sci*. 2023;9(1):120-144. doi 10.1016/j.jacbts.2023.04.008

Luo L., Fu C., Bell C.F., Wang Y., Leeper N.J. Role of vascular smooth muscle cell clonality in atherosclerosis. *Front Cardiovasc Med*. 2023;10:1273596. doi 10.3389/fcvm.2023.1273596

Oren O., Libby P. Incorporating management of clonal hematopoiesis of indeterminate potential into cardiovascular practice. *Circulation*. 2025;152(14):975-977. doi 10.1161/CIRCULATIONAHA.125.075769

Pan H., Xue C., Auerbach B.J., Fan J., Bashore A.C., Cui J., Yang D.Y., … Califano A., Sims P.A., Zhang H., Li M., Reilly M.P. Single-cell genomics reveals a novel cell state during smooth muscle cell phenotypic switching and potential therapeutic targets for atherosclerosis in mouse and human. *Circulation*. 2020;142(21):2060-2075. doi 10.1161/CIRCULATIONAHA.120.048378

Pan H., Ho S.E., Xue C., Cui J., Johanson Q.S., Sachs N., Ross L.S., Li F., Solomon R.A., Connolly E.S., Jr., Patel V.I., Maegdefessel L., Zhang H., Reilly M.P. Atherosclerosis is a smooth muscle cell-driven tumor-like disease. *Circulation*. 2024;149(24):1885-1898. doi 10.1161/CIRCULATIONAHA.123.067587

Puzyrev V.P., Kucher A.N. Evolutionary ontogenetic aspects of pathogenetics of chronic human diseases. *Russ J Genet*. 2011;47(12):1395-1405. doi 10.1134/S102279541112012X

Schuermans A., Honigberg M.C. Clonal haematopoiesis in cardiovascular disease: prognostic role and novel therapeutic target. *Nat Rev Cardiol*. 2025;22(11):845-856. doi 10.1038/s41569-025-01148-9

Tao W., Yurdagul A., Jr., Kong N., Li W., Wang X., Doran A.C., Feng C., Wang J., Islam M.A., Farokhzad O.C., Tabas I., Shi J. siRNA nanoparticles targeting CaMKIIγ in lesional macrophages improve atherosclerotic plaque stability in mice. *Sci Transl Med*. 2020;12(553):eaay1063. doi 10.1126/scitranslmed.aay1063

Wang Y., Gao H., Wang F., Ye Z., Mokry M., Turner A.W., Ye J., … Björkegren J.L.M., Pasterkamp G., Miller C.L., Ross E.G., Leeper N.J. Dynamic changes in chromatin accessibility are associated with the atherogenic transitioning of vascular smooth muscle cells. *Cardiovasc Res*. 2022;118(13):2792-2804. doi 10.1093/cvr/cvab347

Weldy C.S., Kundu S., Monteiro J., Gu W., Pedroza A.J., Dalal A.R., Worssam M.D., … Fischbein M.P., Engreitz J., Kundaje A.B., Cheng P.P., Quertermous T. Epigenomic landscape of single vascular cells reflects developmental origin and disease risk loci. *Mol Syst Biol*. 2025;21(11):1522-1546. doi 10.1038/s44320-025-00140-2

Conflict of interest. The authors declare no conflict of interest.
Received December 26, 2025. Revised May 5, 2026. Accepted May 8, 2026.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):803-813

doi 10.18699/vjgb-26-81

In vitro analysis of the effects of intronic variants c.940+3_940+6del, c.941-3C>G, and c.2389+5G>A in the *LDLR* gene on pre-mRNA splicing using a minigene assay

V.Yu. Danilchenko [1#], D.E. Ivanoshchuk [1#], O.V. Timoshchenko [2], E.A. Panina [1], P.S. Orlov [1], E.V. Shakhtshneider [1]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Institute of Internal and Preventive Medicine – Branch of the Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

danilchenko_valeri@mail.ru

Abstract. Familial hypercholesterolemia (FH) is an autosomal codominant disorder characterized by impaired clearance of low-density lipoproteins from the bloodstream, markedly elevated plasma total cholesterol and low density lipoprotein cholesterol levels, and early onset of atherosclerosis and cardiovascular disease. FH is one of the most common monogenic disorders in humans. The majority of FH cases are caused by pathogenic variants in three genes: *LDLR* (OMIM: 143890), *APOB* (OMIM: 107730), and *PCSK9* (OMIM: 607786). More than 80 % of FH cases are associated with mutations in the *LDLR* gene, located on chromosome 19. In 25–75 % of patients with a clinical FH phenotype, no pathogenic variant is identified in these genes. Whole-exome sequencing and targeted gene panel sequencing of the *LDLR*, *APOB*, *PCSK9*, and *LDLRAP1* genes were performed in patients with an FH phenotype, followed by confirmation of the identified variants by Sanger sequencing. Three unrelated probands were found to carry intronic *LDLR* variants for which functional evidence was previously unavailable. The aim of this study was to functionally assess *in vitro* the effects of three intronic *LDLR* variants (NM_000527.5: c.940+3_940+6del, c.941-3C>G, and c.2389+5G>A) on pre-mRNA splicing using a minigene assay. Minigene constructs encompassing target exons with flanking intronic sequences were generated and transfected into the HEK293 and HeLa cell lines. The deleterious effect of all three variants on pre-mRNA splicing was confirmed. The four-nucleotide deletion c.940+3_940+6del resulted in two aberrant transcripts: inclusion of six nucleotides from intron 6 and complete retention of the minigene intronic sequence. The c.941-3C>G variant caused loss of the canonical acceptor splice site and activation of a cryptic site, with inclusion of intronic nucleotides from intron 6. The c.2389+5G>A variant resulted in exon 16 skipping. Functional *in vitro* analysis is a key tool for the molecular verification of intronic variants of uncertain clinical significance and, in conjunction with clinical data, supports the establishment of a definitive molecular diagnosis.

Key words: familial hypercholesterolemia; *LDLR* gene; intronic variants; minigenes; functional analysis; NGS

For citation: Danilchenko V.Yu., Ivanoshchuk D.E., Timoshchenko O.V., Panina E.A., Orlov P.S., Shakhtshneider E.V. *In vitro* analysis of the effects of intronic variants c.940+3_940+6del, c.941-3C>G, and c.2389+5G>A in the *LDLR* gene on pre-mRNA splicing using a minigene assay. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(5):803-813. doi 10.18699/vjgb-26-81

Funding. This study was supported by the budget project No. FWNR-2026-0027 of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences. Genetic engineering work and cell line experiments were performed at the Center for Collective Use of Cell Technologies of the Institute of Cytology and Genetics SB RAS and were funded within the framework of project No. FWNR-2026-0024.

Анализ влияния интронных вариантов c.940+3_940+6del, c.941-3C>G и c.2389+5G>A гена *LDLR* на сплайсинг пре-мРНК с использованием системы минигенов

В.Ю. Данильченко [1#], Д.Е. Иванощук [1#], О.В. Тимощенко [2], Е.А. Панина [1], П.С. Орлов [1], Е.В. Шахтшнейдер [1]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Научно-исследовательский институт терапии и профилактической медицины – филиал Федерального исследовательского центра Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия

danilchenko_valeri@mail.ru

Аннотация. Семейная гиперхолестеринемия (СГХС) – аутосомно-кодоминантное заболевание, характеризующееся нарушением выведения липопротеинов низкой плотности из кровотока, выраженным повышением уровня общего холестерина и холестерина липопротеинов низкой плотности в крови и ранним развитием атеросклероза

# These authors contributed equally to this work.

и сердечно-сосудистых заболеваний. СГХС является наиболее распространенным моногенным заболеванием у человека. Самые частые причины развития заболевания – мутации в трех генах: *LDLR* (OMIM: 143890), *APOB* (OMIM: 107730) и *PCSK9* (OMIM: 607786). Более 80 % случаев СГХС обусловлено патогенными вариантами гена *LDLR*, локализованного на хромосоме 19. У 25–75 % пациентов с фенотипом СГХС патогенный вариант в указанных генах не обнаруживается. Полноэкзомное и таргетное секвенирование генов *LDLR*, *APOB*, *PCSK9*, *LDLRAP1* было проведено у пациентов с фенотипом СГХС с последующим подтверждением выявленных вариантов секвенированием по методу Сэнгера. У трех неродственных пробандов обнаружены интронные варианты гена *LDLR*, для которых ранее не проводилась функциональная верификация. Целью работы стала функциональная оценка *in vitro* влияния трех интронных вариантов гена *LDLR* (NM_000527.5: c.940+3_940+6del, c.941-3C>G, c.2389+5G>A) на процесс сплайсинга с использованием генетических конструкций – минигенов. Для функциональной верификации *in vitro* были созданы минигены, включающие исследуемые экзоны с фланкирующими интронными областями. Анализ сплайсинга проводили в клеточных линиях HEK293 и HeLa. В результате анализа подтвержден патогенный эффект всех трех выявленных вариантов на сплайсинг пре-мРНК. Делеция четырех нуклеотидов c.940+3_940+6del приводила к образованию двух аберрантных транскриптов – включению шести нуклеотидов интрона 6, а также к полному удержанию интронной последовательности минигена. Вариант c.941-3C>G обусловливал потерю акцепторного сайта сплайсинга и активацию криптического, с включением нуклеотидов интрона 6. Замена c.2389+5G>A приводила к пропуску экзона 16. Функциональный анализ *in vitro* – ключевой инструмент молекулярной верификации интронных вариантов с неопределенной клинической значимостью и в совокупности с клиническими данными помогает в установлении окончательного диагноза.

Ключевые слова: семейная гиперхолестеринемия; ген *LDLR*; интронные варианты; минигены; функциональный анализ; NGS

Introduction

Familial hypercholesterolemia (FH) is an inherited disorder characterized by impaired clearance of low-density lipoprotein (LDL) particles from the circulation, markedly elevated plasma total cholesterol and low density lipoprotein cholesterol levels, and premature atherosclerosis and cardiovascular disease (Austin et al., 2004; Nordestgaard et al., 2013). FH is a monogenic disorder with a predominantly autosomal codominant pattern of inheritance, whereby the number of pathogenic alleles determines the severity of clinical manifestations (Freiberger, 2026).

The prevalence of heterozygous FH in the general population is approximately one in 311 individuals, making it one of the most common inherited disorders in humans (Hu et al., 2020). Sustained hypercholesterolemia leads to premature atherosclerosis and early-onset cardiovascular complications. Timely molecular genetic diagnosis confirms the clinical diagnosis and guides the selection of optimal lipid-lowering therapy, thereby contributing to reduced mortality from atherosclerotic disease (Ahmad et al., 2026).

FH is most commonly caused by pathogenic variants in one of three genes: *LDLR* (low-density lipoprotein receptor, OMIM: 143890), *APOB* (apolipoprotein B, OMIM: 107730), and *PCSK9* (proprotein convertase subtilisin/kexin type 9, OMIM: 607786). Variants in *LDLR* account for 85–90 % of molecularly confirmed FH cases, and more than 3,000 pathogenic variants in this gene have been reported to date in patients with lipid metabolism disorders (Ahmad et al., 2026). The *LDLR* gene (19p13.2) includes 18 exons and encodes the LDL receptor protein comprised of 860 amino acids (Hobbs et al., 1990). The primary function of the LDLR protein is to mediate the clearance of LDL particles from the circulation via binding to apolipoprotein B-100, thereby promoting their lysosomal degradation and maintaining cellular and plasma cholesterol homeostasis (Brown, Goldstein, 1979; Südhof et al., 1985; Segrest et al., 2001).

In 25–75 % of patients with a clinical FH phenotype, no pathogenic variant is identified in the genes mentioned above (Freiberger, 2026). One contributing factor is the difficulty in clinical interpretation of variants located outside the coding sequences and canonical splice sites. Substitutions at donor site positions +1/+2 and acceptor site positions −1/−2 disrupt pre-mRNA splicing and are generally straightforward to interpret. By contrast, splice region variants and synonymous substitutions can impair spliceosomal recognition of donor and acceptor sites to varying degrees, activate cryptic splice sites, cause exon skipping or intron retention, and ultimately give rise to aberrant transcripts and loss of protein function (Abramowicz, Gos, 2018). *In silico* predictions for such variants are often contradictory, and in the absence of functional evidence, most are classified as variants of uncertain significance (VUS) (Walker et al., 2023). Variants of this type have been reported in *LDLR* in patients with FH (Kulseth et al., 2010; Reeskamp et al., 2018, 2021), however, their functional verification remains uncommon.

A widely used experimental approach for assessing the impact of intronic variants on splicing in multi-exon genes is the minigene assay. This method can be used when patient-derived RNA is unavailable, as it enables the evaluation of splicing outcomes in a cell-based *in vitro* system (Abramowicz, Gos, 2018).

The aim of this study was to functionally evaluate the effects of three intronic *LDLR* variants (c.940+3_940+6del, c.941-3C>G, and c.2389+5G>A) identified in patients with a clinical FH phenotype on pre-mRNA splicing, using an *in vitro* minigene assay.

Materials and methods

Patients with familial hypercholesterolemia enrolled in this study were examined at the Clinical Diagnostic Department of the Research Institute of Internal and Preventive Medicine – Branch of the Institute of Cytology and Genetics, Siberian

Branch of the Russian Academy of Sciences (IIPM – Branch of ICG SB RAS). The study protocol was approved by the Local Ethics Committee of IIPM – Branch of ICG SB RAS (Protocol No. 68, June 4, 2019). Written informed consent was obtained from each participant, or from a parent or legal guardian on behalf of minor participants, prior to enrollment.

The clinical diagnosis of FH was established using the Dutch Lipid Clinic Network (DLCN) diagnostic criteria (Table 1) and the Simon Broome Register criteria (Table 2). The DLCN scoring system classifies patients into four probability categories for FH diagnosis based on cumulative point scores (Tables 1 and 2) (Ezhov et al., 2019).

The diagnosis is established based on the sum of points obtained across all five categories. Within each category, scores are not cumulative – only a single criterion yielding the highest score is counted.

All study participants underwent clinical examination, vascular ultrasound imaging, and venous blood sampling for biochemical analysis, including total cholesterol (TC), LDL cholesterol (LDL-C), HDL cholesterol (HDL-C), triglycerides (TG), atherogenic index (AI), and fasting plasma glucose. Molecular genetic analysis was performed in all participants.

Genomic DNA for subsequent analysis was extracted from peripheral venous blood using the phenol–chloroform extraction method (Sambrook, Russell, 2006).

Whole-exome sequencing was performed on the FASTA Seq 300 platform (GeneMind Biosciences, China). Library preparation was carried out using the SG GM Ultra-96 kit (cat. No. #SG_GM_plus-96, Raissol Bio, Russia). Exome capture was performed using the VAHTS Target Capture Core Exome Panel (cat. No. NC001-01, Vazyme, China) according to the manufacturer's protocol. The capture panel covered exonic sequences (~33.9 Mb) with limited coverage of adjacent intronic regions. Library quality was assessed on the Qsep1 platform (BiOptic, Taiwan), and library concentration was determined using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, USA).

Table 1. DLCN diagnostic criteria for the diagnosis of heterozygous FH in adults (18 years of age and older)

Criterion	Score
1. Family history	
First-degree relative with premature atherosclerotic cardiovascular disease (men <55 years; women <60 years), including coronary artery disease, atherothrombotic ischaemic stroke, transient ischaemic attack, or peripheral atherosclerosis with plaques causing ≥50 % luminal stenosis or First-degree relative with LDL-cholesterol above the 95th percentile	1
First-degree relative with tendon xanthomas and/or corneal arcus or Children under 18 years of age with LDL-cholesterol above the 95th percentile	2
2. Personal history	
Premature coronary artery disease (men <55 years; women <60 years)	2
Premature atherothrombotic ischaemic stroke, transient ischaemic attack, or peripheral atherosclerosis with plaques causing ≥50 % luminal stenosis (men <55 years; women <60 years)	1
3. Physical examination	
Tendon xanthomas	6
Corneal arcus before the age of 45 years	4
4. LDL cholesterol level, mmol/L	
≥8.5	8
6.5–8.4	5
5.0–6.4	3
4.0–4.9	1
5. DNA analysis	
Presence of a pathogenic or likely pathogenic variant in *LDLR*, *APOB*, or *PCSK9*	8
Diagnosis is based on the cumulative score: Definite FH – >8 points Probable FH – 6–8 points Possible FH – 3–5 points Unlikely FH – ≤2 points	

Table 2. Simon Broome Register diagnostic criteria for heterozygous familial hypercholesterolemia

Definite HeFH* is diagnosed when: Total cholesterol (TC) >6.7 mmol/L or LDL-cholesterol >4.0 mmol/L in a patient aged <16 years; or TC >7.5 mmol/L or LDL-cholesterol >4.9 mmol/L in a patient aged ≥16 years. Plus at least one of the following: • tendon xanthomas in the patient, or in a first-degree relative (parent, child, or sibling), or in a second-degree relative (grandparent, aunt, or uncle); and/or • a positive DNA test confirming the presence of a pathogenic or likely pathogenic variant in *LDLR*, *APOB*, or *PCSK9*	**Probable HeFH** is diagnosed when: TC >6.7 mmol/L or LDL-cholesterol >4.0 mmol/L in a patient aged <16 years; or TC >7.5 mmol/L or LDL-cholesterol >4.9 mmol/L in a patient aged ≥16 years. Plus at least one of the following: • a family history of myocardial infarction in a second-degree relative before the age of 50 years, or in a first-degree relative before the age of 60 years and/or • TC >7.5 mmol/L in a first- or second-degree relative aged ≥16 years, or TC >6.7 mmol/L in the patient or a first-degree relative aged <16 years

* HeFH – heterozygous familial hypercholesterolemia.

Raw sequencing data were processed using NGS Wizard software (Genomenal, version 2.11.0; https://genomenal.ru/). Alignment to the GRCh38 reference genome achieved a mapping rate of 99.99 %, with a properly paired alignment rate of 99.98 %; the duplication rate was 4.82 %. Mean per-nucleotide coverage was 129×. Nucleotide and exon numbering was based on the GenBank reference sequence NM_000527.4.

Variant filtering criteria were as follows: genotype quality ≥30, read depth ≥10, and minor allele frequency (MAF) in the gnomAD database (versions 3.2.1 and 4.1.1; https://gnomad.broadinstitute.org) <0.01 %. Candidate variants were confirmed by Sanger sequencing using a GTZ G08 genetic analyzer (Gene Technologies of Health, Russia/China). For patient P03, variant detection was performed by targeted next-generation sequencing with hybridization capture-based library enrichment using a custom probe panel (Roche, Basel, Switzerland) according to the manufacturer's recommended protocol. The panel was designed to cover all exons, introns, putative promoter regions, and 5′-untranslated regions (5′UTR) of four genes: *LDLR*, *APOB*, *PCSK9*, and *LDLRAP1*. Sequencing was performed on the GS Junior platform (Roche, Basel, Switzerland).

The potential impact of the identified *LDLR* variants on pre-mRNA splicing was assessed by *in silico* analysis using the following prediction tools: Human Splicing Finder (https://genomnis.com/hsf), SpliceAI (https://spliceailookup.broadinstitute.org/), and MaxEntScan (Shamsani et al., 2019).

The pathogenicity of the identified variants was evaluated in accordance with the ACMG/AMP variant classification guidelines (Richards et al., 2015) and the gene-specific rules of the ClinGen Familial Hypercholesterolemia Variant Curation Expert Panel (FH VCEP) (Chora et al., 2022).

Genomic fragments of the *LDLR* gene (NG_009060.1) containing exons and flanking intronic sequences were selected for minigene construction. The exon-trap vector pET01 (MoBiTec GmbH, Germany) was used in the study. Oligonucleotide primers for amplification of the relevant *LDLR* genomic fragments were designed using Primer3Plus (https://www.primer3plus.com/index.html) (Table 3). Genomic DNA from patients carrying variants c.940+3_940+6del, c.941-3C>G, and c.2389+5G>A served as PCR templates.

Initial amplification was performed using the designed primers and Q5 High-Fidelity DNA Polymerase (New England Biolabs, USA), followed by re-amplification (re-PCR) with primers including recognition sites for the restriction endonucleases *BamHI* and *XbaI* at their 5′ ends (Table 3). PCR products and the pET01 vector were digested with *BamHI* and *XbaI* restriction endonucleases (New England Biolabs). The resulting fragments were cloned into the pET01 vector at the corresponding restriction sites using T4 DNA Ligase (SibEnzyme, Russia), and the constructs were transformed into *Escherichia coli*. Colony screening for minigene constructs carrying the studied variants and the wild-type sequence was performed by PCR with vector-specific primers (Table 3), agarose gel electrophoresis, and Sanger sequencing. Plasmid DNA was isolated using the Plasmid Midiprep kit (Evrogen, Russia). All minigenes were sequenced to confirm the absence of unintended sequence alterations.

Splicing analysis was initially performed using the HEK293 cell line and subsequently repeated using the HeLa cell line. Both cell lines were cultured in DMEM/F12 medium (Servicebio, China) supplemented with 10 % fetal bovine serum (Capricorn, Germany) and 2 mM penicillin/streptomycin (Capricorn) at 37 °C in a humidified incubator with 5 % CO_2. Cell lines were transfected with either mutant or wild-type minigenes in 6-well plates (Eppendorf, Germany) using Lipofectamine 3000 (Thermo Fisher Scientific, USA) according to the manufacturer's instructions.

Forty-eight hours post-transfection, total RNA was extracted from the cells using the RNA Solo kit (Evrogen). The isolated RNA was treated with RQ1 RNase-Free DNase (Promega, USA) to eliminate residual genomic DNA contamination. Complementary DNA (cDNA) was synthesized using the MMLV RT kit (Evrogen) with a primer specific to the plasmid region of the pET01 vector (Table 3). Minigene-derived transcripts were then amplified by PCR using the synthesized cDNA as template and primers specific to the exons of the pET01 vector (Table 3). PCR products were resolved by elec-

Table 3. Primers used for amplification of *LDLR* gene fragments

Fragments	Primer sequences	Product size, bp	Restriction site
Primers used for amplification of *LDLR* gene fragments			
Exon 6	F: 5′-TAATGAATCCATTTGCATGCGTTCT-3′ R: 5′-GGGCAGAGTGGAGTTCCCAAAA-3′	335	–
Exon 7	F: 5′-GATGGGTAGGGGCCCGAGA-3′ R: 5′-AAACTGAGGCATGAGGGGTTTGG-3′	250	–
Exon 16	F: 5′-CCCGTGTGGCCTCTCACAGA-3′ R: 5′-ATTCCACAGCACGGGACCCA-3′	455	–
Primers used for molecular cloning			
Exon 6	F: 5′-TAATGGATCCATTTGCATGCGTTCT-3′ R: 5′-GGGCTCTAGAGAGTTCCCAAAA-3′	335	*BamHI* *XbaI*
Exon 7	F: 5′-GATGGGATCCGGCCCGAGA-3′ R: 5′-AAACTCTAGAATGAGGGGTTTGG-3′	250	*BamHI* *XbaI*
Exon 16	F: 5′-CCCGGGATCCCTCTCACAGA-3′ R: 5′-ATTCTCTAGAACGGGACCCA-3′	455	*BamHI* *XbaI*
Primers specific to the plasmid region of the pET01 vector			
cDNA	cDNA primer 1: 5′-GATCCACGATGC-3′	–	–
pET01 exons	F: 5′-GATGGATCCGCTTCCTGCCCC-3′ R: 5′-CTCCCGGGCCACCTCCAGTGCC-3′	246	*BamHI* *SmaI*

Note. The 5′ ends containing recognition sites for restriction endonucleases are underlined.

trophoresis on a 3 % agarose gel. The nucleotide sequence of each amplified fragment was confirmed by Sanger sequencing.

Construction, editing, and annotation of nucleotide and protein sequences were performed using Unipro UGENE software, version 53.1 (Okonechnikov et al., 2012).

Results

Molecular genetic analysis identified three *LDLR* variants potentially affecting pre-mRNA splicing in three unrelated probands presenting with a clinical FH phenotype (Table 4): c.940+3_940+6del, affecting the donor splice site of intron 6; c.941-3C>G, affecting the acceptor splice site of intron 6; and c.2389+5G>A, located within the donor splice site of intron 16. None of these variants had previously been functionally characterized in the published literature.

Proband P01 was a male born in 1989 (36 years of age at the time of examination) with marked hypercholesterolaemia: total cholesterol (TC) was 9.0 mmol/L and LDL cholesterol (LDL-C) was 7.2 mmol/L. Triglycerides (TG) were 1.1 mmol/L and HDL cholesterol (HDL-C) was 1.3 mmol/L, consistent with the typical lipid profile of familial hypercholesterolaemia, characterized by pronounced elevation of TC and LDL-C (optimal values: TC <5.0 mmol/L, LDL-C <3.0 mmol/L in patients without additional risk factors) and normal TG levels (<1.7 mmol/L).

Family history was notable: his biological sister (38 years at the time of examination) had a markedly elevated TC level of 10.0 mmol/L. The father was also reported to have hypercholesterolaemia, but laboratory data were not available. The DLCN score for the proband was 7 points, corresponding to "probable FH" (6–8 points) (Ezhov et al., 2019). Molecular genetic testing identified the previously reported *LDLR* variant NM_000527.5:c.940+3_940+6del (GRCh38: chr19:11107514 GGTGA>G) in the proband. The proband's father and sister were not tested for this variant as DNA samples were unavailable. The variant was absent from the general population in the gnomAD database and had previously been described in patients with familial hypercholesterolaemia (Rieck et al., 2020; Semenova et al., 2020; Miroshnikova et al., 2021; Nazarenko et al., 2023). According to ACMG/AMP criteria as specified by the ClinGen FH Variant Curation Expert Panel, c.940+3_940+6del was classified as a variant of uncertain significance (PM2, PP3, PP4).

The *LDLR* intronic variant NM_000527.5:c.941-3C>G (GRCh38: chr19:11110649 C>G) in intron 6 was identified in proband P02 and her mother. Proband P02 was a female born in 2001 (22 years of age at the time of examination); her maximum recorded TC was 7.8 mmol/L, and LDL-C, 5.5 mmol/L, exceeding the 95th percentile of reference values for women in this age group (Ezhov et al., 2019). Family history was strongly positive: her father had clinical coronary artery disease from the age of 45 years, experienced an acute myocardial infarction at age 50, and underwent coronary artery bypass grafting; her mother had TC levels up to 8 mmol/L and was receiving lipid-lowering therapy; her brother had marked hypercholesterolaemia with TC up to 10 mmol/L.

Table 4. Results of *in silico* splicing analysis of *LDLR* variants

Variant (Location in the *LDLR* gene)	Genomic position GRCh38.p14 (NC_000019.10)	Human Splicing Finder (HSF)	SpliceAI*	MaxEntScan**
c.940+3_940+6del (rs2077315885) Intron 6	19:11107514 GGTGA>G	Broken WT Donor Site: alteration of the WT Donor site, most probably affecting splicing (HSF, MaxEnt)	0.87 Donor loss	Var/Wt < 0 (–0.63) Substantial weakening of the donor splice site
		New Donor splice site: activation of a cryptic Donor site. Potential alteration of splicing (HSF)	0.45 Donor gain	
c.941-3C>G (rs2515989714) Intron 6	19:11110649 C>G	Broken WT Acceptor Site: alteration of the WT Acceptor site, most probably affecting splicing (MaxEnt)	0.37 Acceptor loss	Var/Wt < 0.0 (–1.18) Substantial weakening of the acceptor splice site
c.2389+5G>A (rs879255191) Intron 16	19:11128090 G>A	Broken WT Donor Site: alteration of the WT Donor site, most probably affecting splicing (HSF, MaxEnt)	0.41 Donor loss	Var/Wt = 0.50 Weakening of the donor splice site

* SpliceAI Δscores ≥0.80 are considered high-confidence predictions of splicing impact; ** the MaxEntScan Var/WT ratio <0.80 provides supportive evidence for pathogenicity, whereas the ratio ≥1.0 provides supportive evidence for a benign effect of the variant (Chora et al., 2022).

Ultrasound examination of the brachiocephalic arteries and arteries of the lower extremities revealed no atherosclerotic changes. The proband's DLCN score was 4 points, corresponding to "possible FH" (3–5 points). The c.941-3C>G variant was absent from the general population in gnomAD. Other family members were unavailable for genetic testing. This variant had previously been reported in individuals with familial hypercholesterolaemia in Spain (Martín-Campos et al., 2018) and in Russia (Meshkov et al., 2021). Based on ACMG/AMP criteria as adapted by the ClinGen FH Variant Curation Expert Panel, c.941-3C>G was classified as a VUS (PM2, PP3, PP1_Supporting).

In proband P03, a 43-year-old male at the time of examination, we had previously identified a variant located at the donor splice site, NM_000527.5:c.2389+5G>A (GRCh38: chr19:11128090 G>A), in intron 16 of the *LDLR* gene (Shakhtshneider et al., 2017). From childhood, the proband had arterial hypertension, and in 2011 he suffered an acute myocardial infarction. Laboratory testing revealed markedly elevated lipid levels: TC – 11.5 mmol/L, LDL-C – 10.12 mmol/L, HDL-C – 0.89 mmol/L, and TG – 1.08 mmol/L. Family history was strongly positive for premature coronary artery disease: the father had an acute myocardial infarction at 45 years of age, and three paternal uncles had infarctions between 50 and 53 years of age. The proband's older brother had hypercholesterolaemia and coronary artery disease and carried the same *LDLR* variant. The DLCN score was 17 points, corresponding to definite familial hypercholesterolaemia. Cascade screening identified the c.2389+5G>A variant in two of the proband's three daughters (aged 7–8 years), both of whom had hypercholesterolaemia with LDL-C levels of 4.28–4.69 mmol/L. In the gnomAD population database (version 4.1.1), the variant is reported at a frequency of 0.0001239 %, with no homozygotes observed. The variant has previously been described in individuals with familial hypercholesterolaemia (Semenova et al., 2020). According to ACMG/AMP criteria as specified by the ClinGen FH Variant Curation Expert Panel, c.2389+5G>A was classified as a VUS (PM2, PP1_Moderate, PP3, PP4).

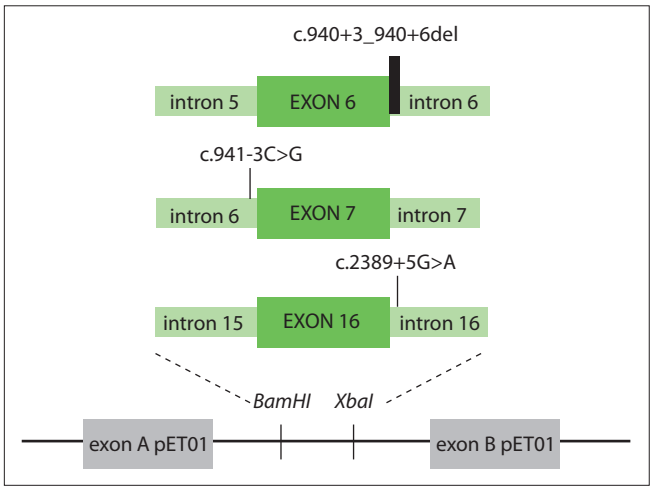


Fig. 1. Scheme of minigenes with the c.940+3_940+6del, c.941-3C>G, and c.2389+5G>A *LDLR* variants.

The pET01 fragments are colored in gray, the *LDLR* gene fragments, in green. *BamHI* and *XbaI* indicate the recognition sites for restriction endonucleases.

Based on *in silico* splicing predictions for the *LDLR* variants c.940+3_940+6del, c.941-3C>G, and c.2389+5G>A (Table 4), minigenes were constructed, each comprising a single exon and its flanking intronic regions (Fig. 1). For the minigene containing variant c.940+3_940+6del, the flanking intronic sequences were 116 and 96 bp in length, whereas for the c.941-3C>G construct, they were 51 and 79 bp. The shorter intronic segments in the latter construct reflect the fact that *LDLR* introns are composed predominantly of Alu repeat elements (~65 % of intronic sequence) (Amsellem et al., 2002).

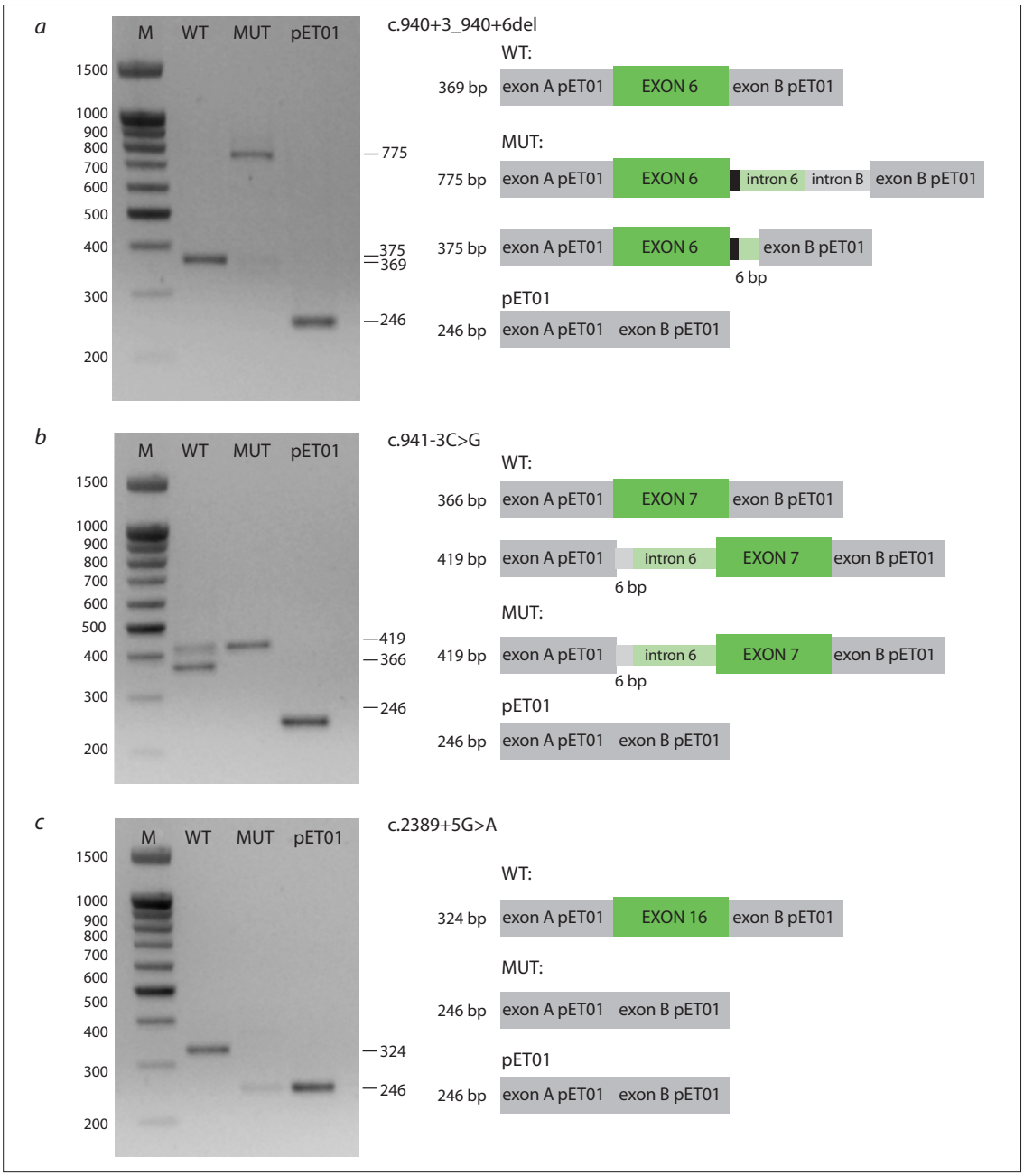


Fig. 2. Results of minigene assay (obtained on the HEK293 cell line) for *LDLR* variants c.940+3_940+6del, c.941-3C>G, c.2389+5G>A and the corresponding wild types (WT).

RT-PCR products were visualized on 3 % agarose gels, and schematic representations of the observed splicing patterns are shown below each gel. The pET01 fragments are colored in gray, the *LDLR* gene fragments, in green. *a* – minigenes: (MUT) c.940+3_940+6del and wild type (WT); *b* – minigenes: (MUT) c.941-3C>G and wild type (WT); *c* – minigenes: (MUT) c.2389+5G>A and wild type (WT); pET01 – pET01 vector without insertion; M – DNA size marker (100 + 1,500 bp).

The c.940+3_940+6del variant is a 4-nucleotide deletion at positions +1 to +4 of intron 6; *in silico* analysis predicted loss of the canonical donor splice site with activation of a cryptic donor site (Table 4). In the wild type (WT) minigene, a single product of 369 bp was detected, corresponding to correct splicing with complete removal of introns and retention of exon 6. In contrast, analysis of the c.940+3_940+6del (MUT) minigene revealed two major products of 775 and 375 bp. The 775 bp fragment reflects loss of the donor splice site and inclusion of intron 6 of *LDLR* together with the intron of the pET01 vector in the mRNA. The 375 bp fragment represents loss of the canonical donor site and activation of a cryptic donor site, leading to inclusion of six nucleotides derived from intron 6 (Fig. 2*a*). These *in vitro* data demonstrate a pathogenic effect of c.940+3_940+6del, resulting in aberrant splicing.

The c.941-3C>G variant is located at position −3 of intron 6 of *LDLR*, and *in silico* predictions for this variant were discordant (Table 4). In the wild-type minigene, two products were observed: a 366 bp product corresponding to correct splicing, and a 419 bp product containing a chimeric intron composed of six nucleotides from the pET01 intron and the entire intron 6 of *LDLR* (Fig. 2*b*). Retention of intron 6 in the wild-type minigene is likely related to specific features of the construct. In the minigene carrying the c.941-3C>G variant, a single fragment was detected, which also contained the same chimeric intron (six nucleotides from the pET01 intron and the full intron 6 of *LDLR*) (Fig. 2*b*). Thus, *in vitro* analysis confirmed that c.941-3C>G causes loss of the canonical acceptor splice site and activation of a cryptic splice site.

The c.2389+5G>A variant is located outside the canonical splice site at position +5 of intron 16. *In vitro* analysis confirmed the *in silico* prediction for c.2389+5G>A (Table 4), namely loss of the canonical donor splice site. In the minigene carrying this variant, the predominant product was 246 bp, corresponding in size to the empty pET01 vector (Fig. 2*c*). *In vitro* analysis confirmed the pathogenic effect of the c.2389+5G>A variant on splicing.

Considering the possible cell-type specificity of splicing regulation (Cooper, 2005), the experiment was repeated using the HeLa cell line. The results obtained in this cell line completely reproduced those obtained in HEK293 cells.

In vitro functional data were then incorporated into variant interpretation in accordance with ACMG/AMP pathogenicity criteria (Richards et al., 2015) and the gene-specific specifications of the ClinGen Familial Hypercholesterolemia Variant Curation Expert Panel (FH VCEP) (Chora et al., 2022). The minigene assay confirmed aberrant splicing for all three variants (c.940+3_940+6del, c.941-3C>G, and c.2389+5G>A), providing supportive functional evidence of pathogenicity (PS3_Supporting under FH VCEP specifications).

Variants c.940+3_940+6del and c.941-3C>G have been reported in patients with FH in several independent studies (Martín-Campos et al., 2018; Rieck et al., 2020; Semenova et al., 2020; Meshkov et al., 2021; Miroshnikova et al., 2021; Nazarenko et al., 2023). However, application of *LDLR*-specific FH VCEP criteria (Chora et al., 2022) indicates that the totality of accumulated evidence remains insufficient for reclassification to likely pathogenic. Accordingly, under the ACMG/AMP criteria, c.940+3_940+6del (PM2, PS3_Supporting, PP4) and c.941-3C>G (PM2, PS3_Supporting, PP1_Supporting) remain classified as variants of uncertain significance.

In contrast, c.2389+5G>A was reclassified from a VUS to likely pathogenic, supported by criteria PM2, PS3_Supporting, PP1_Moderate, and PP4 under the ACMG/AMP framework and FH VCEP specifications.

Discussion

In this study, we present an *in vitro* splicing analysis of three intronic *LDLR* variants – c.940+3_940+6del, c.941-3C>G, and c.2389+5G>A – using the exon-trap pET01 vector.

Canonical splice sites define exon–intron boundaries, are evolutionarily conserved, and are recognized by the spliceosome. Any variation in the sequence of these canonical sites may alter interactions between pre-mRNA and spliceosomal subunits (Abramowicz, Gos, 2018). Recognition of the 5′ splice site by U1 snRNP occurs via base pairing between the 5′ end of U1 snRNA and pre-mRNA over positions −3 to +6 relative to the exon–intron junction (Malard et al., 2022). Preservation of the invariant GT dinucleotide alone is therefore not a reliable predictor of donor site efficiency; a key determinant is the thermodynamics of U1 snRNA:5′ss duplex formation, in which the presence of a certain nucleotide at position +3 and further downstream makes a substantial contribution to duplex stability and donor site strength (Malard et al., 2022).

The c.940+3_940+6del variant results in deletion of four nucleotides at positions +1 to +4 of intron 6, affecting the canonical donor splice site ($delG_{+1}T_{+2}G_{+3}A_{+4}$). As a consequence of this rearrangement, the consensus GT is preserved at positions +1 and +2, whereas an altered motif ($G_{+1}T_{+2}C_{+3}T_{+4}$) is created at positions +3 and +4. Our *in vitro* analysis confirmed a deleterious effect of c.940+3_940+6del on pre-mRNA splicing, with two aberrant transcripts identified. In one transcript, six nucleotides of *LDLR* intron 6 were retained; *in silico* analysis predicts a frameshift and a premature termination codon at amino acid position 371.

In the second transcript, the entire intron 6 of *LDLR* together with the intron of pET01 were retained in the mRNA, which is likely to result in the synthesis of a protein with an altered amino acid sequence or in mRNA degradation. Additional studies at the protein level will be required to confirm these predicted consequences.

Similar intronic donor-site variants in *LDLR* have been reported previously. The intronic variant c.1358+3_1358+8del, which deletes six nucleotides at positions +1…+6 while preserving GT at +1 and +2, has been described in a proband with an FH phenotype. Although functional studies have not been performed, segregation data suggest that this variant is likely pathogenic (Marduel et al., 2010).

Variants located at intronic position +5 are frequently classified as VUS in routine analysis. However, the +5 position of the donor site participates in binding to the 5′ end of U1 snRNA and is one of the determinants of donor site strength; substitutions at this position can reduce spliceosomal recognition efficiency (Malard et al., 2022).

Several *LDLR* variants at intronic position +5 have previously been investigated. The c.1186+5G>A variant (intron 8) leads to loss of the donor splice site and inclusion of intron 8 in the mRNA, followed by nonsense-mediated mRNA decay (Etxebarria et al., 2012). In contrast, for c.1586+5G>A and c.2140+5G>A, only small amounts of mutant transcripts and no significant effect on protein activity were observed (Bourbon et al., 2009; Holla et al., 2009). In our study, substitution of G by A at position c.2389+5 resulted in exon 16 skipping from the transcript, and subsequent *in silico* analysis predicted loss of amino acid residues 771–796 and production of a truncated LDLR protein.

The acceptor splice site is characterized by highly conserved nucleotides at positions −2 and −1, which are recognized by the U2AF35 subunit, while the larger U2AF65 subunit binds to the upstream polypyrimidine tract. Variable positions near the AG dinucleotide, including −3, can modulate site strength by influencing U2AF heterodimer binding affinity to pre-mRNA (Jenkins et al., 2013; Voith von Voithenberg et al., 2016).

Although intronic position −3 is not itself conserved, its functional relevance has been demonstrated experimentally: mutations at −3 account for ~4 % of all reported acceptor-site defects in inherited disease genes and lead to activation of cryptic splice sites or intron retention *in vivo* (Vorechovský, 2006). In our study, the c.941-3C>G variant caused loss of the canonical acceptor site and inclusion of part of intron 6. Substitution of cytosine by guanine at −3 disrupts the pyrimidine-rich composition typical of the region adjacent to the acceptor AG dinucleotide.

Because U2AF65 recognizes the polypyrimidine tract upstream of AG, introduction of a purine in close proximity to the site likely weakens U2AF binding and promotes activation of a cryptic splice site. Depending on the surrounding nucleotide context, different splicing abnormalities may occur. For example, c.2312-3C>A in intron 15 has been shown to disrupt the 3′ acceptor site and cause skipping of exon 16 in mature mRNA in patients with FH (Liguori et al., 2001), whereas c.818-3C>G in intron 5 leads to inclusion of two intronic nucleotides in the mRNA, with a presumed premature stop codon at amino acid position 274 (Benito-Vicente et al., 2015). Our data are consistent with these observations.

According to R. Chong et al. (2019), among variants with a deleterious effect on splicing (3.8 % of all analyzed substitutions in the ExAC database), 83 % are located outside canonical ±1.2 positions and are evenly distributed across exons and introns. Ø.L. Holla et al. (2009) reported that functional analysis of 18 intronic variants demonstrated significant splicing effects for 14 of them (78 %). Together with our *LDLR* findings, these data emphasize that variants beyond canonical splice-site dinucleotides are often functionally significant but tend to remain classified as VUS in the absence of *in vitro* confirmation.

Functional *in vitro* assessment of variant effects on splicing is therefore a necessary, though not always sufficient, condition for reclassifying a VUS to likely pathogenic or likely benign. As shown in the present study, experimentally confirmed aberrant splicing strengthens the evidence for pathogenicity; however, final classification is determined by the aggregate ACMG/AMP criteria as applied within the gene-specific FH VCEP ClinGen framework (Chora et al., 2022). Thus, minigene assays represent an essential tool for verifying the pathogenicity of intronic *LDLR* variants located outside canonical splice-site dinucleotides, including for establishing a molecular diagnosis in patients with a clinical FH phenotype (Blakes et al., 2022).

Conclusion

In the present study, all three intronic *LDLR* variants – c.940+3_940+6del, c.941-3C>G, and c.2389+5G>A – were shown to exert a pronounced deleterious effect on pre-mRNA splicing in the minigene system, leading to loss of canonical splice sites, activation of cryptic sites, and formation of aberrant transcripts. These findings underscore the need for functional analysis to substantiate the pathogenicity of variants for which *in silico* splicing predictions are inconclusive or discordant.

References

Abramowicz A., Gos M. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet*. 2018;59(3):253-268. doi 10.1007/s13353-018-0444-7

Ahmad Z., Agarwala A., Cuchel M., Barton Duell P., Hegele R.A., Hudgins L., Jamison A., … Soffer D.E., Warden B.A., Weintraub W.S., Williams L., Goldberg A.C. Update on familial hypercholesterolemia: an expert clinical consensus from the National Lipid Association. *J Clin Lipidol*. 2026;20(4):708-737. doi 10.1016/j.jacl.2026.01.011

Amsellem S., Briffaut D., Carrié A., Rabès J.P., Girardet J.P., Fredenrich A., Moulin P., Krempf M., Reznik Y., Vialettes B., de Gennes J.L., Brukert E., Benlian P. Intronic mutations outside of *Alu*-repeat-rich domains of the LDL receptor gene are a cause of familial hypercholesterolemia. *Hum Genet*. 2002;111(6):501-510. doi 10.1007/s00439-002-0813-4

Austin M.A., Hutter C.M., Zimmern R.L., Humphries S.E. Genetic causes of monogenic heterozygous familial hypercholesterolemia: a HuGE prevalence review. *Am J Epidemiol*. 2004;160(5):407-420. doi 10.1093/aje/kwh236

Benito-Vicente A., Alves A.C., Etxebarria A., Medeiros A.M., Martin C., Bourbon M. The importance of an integrated analysis of clinical, molecular, and functional data for the genetic diagnosis of familial hypercholesterolemia. *Genet Med*. 2015;17(12):980-988. doi 10.1038/gim.2015.14

Blakes A.J.M., Wai H.A., Davies I., Moledina H.E., Ruiz A., Thomas T., Bunyan D., … Douglas A.G.L.; Genomics England Research Consortium, Splicing and Disease Working Group; Whiffin N., Baralle D., Lord J. A systematic analysis of splicing variants identifies new diagnoses in the 100,000 Genomes Project. *Genome Med*. 2022;14(1):79. doi 10.1186/s13073-022-01087-x

Bourbon M., Duarte M.A., Alves A.C., Medeiros A.M., Marques L., Soutar A.K. Genetic diagnosis of familial hypercholesterolaemia: the importance of functional analysis of potential splice-site mutations. *J Med Genet*. 2009;46(5):352-357. doi 10.1136/jmg.2007.057000

Brown M.S., Goldstein J.L. Receptor-mediated endocytosis: insights from the lipoprotein receptor system. *Proc Natl Acad Sci USA*. 1979;76(7):3330-3337. doi 10.1073/pnas.76.7.3330

Chong R., Insigne K.D., Yao D., Burghard C.P., Wang J., Hsiao Y.E., Jones E.M., Goodman D.B., Xiao X., Kosuri S. A multiplexed assay for exon recognition reveals that an unappreciated fraction of rare genetic variants cause large-effect splicing disruptions. *Mol Cell*. 2019;73(1):183-194.e8. doi 10.1016/j.molcel.2018.10.037

Chora J.R., Iacocca M.A., Tichý L., Wand H., Kurtz C.L., Zimmermann H., Leon A., … Hegele R.A., Freiberger T., Knowles J.W., Bourbon M.; ClinGen Familial Hypercholesterolemia Expert Panel. The Clinical Genome Resource (ClinGen) Familial Hypercholesterolemia Variant Curation Expert Panel consensus guidelines for *LDLR* variant classification. *Genet Med*. 2022;24(2):293-306. doi 10.1016/j.gim.2021.09.012

Cooper T.A. Use of minigene systems to dissect alternative splicing elements. *Methods*. 2005;37(4):331-340. doi 10.1016/j.ymeth.2005.07.015

Etxebarria A., Palacios L., Stef M., Tejedor D., Uribe K.B., Oleaga A., Irigoyen L., Torres B., Ostolaza H., Martin C. Functional charac-

terization of splicing and ligand-binding domain variants in the LDL receptor. *Hum Mutat*. 2012;33(1):232-243. doi 10.1002/humu.21630

Ezhov M.V., Bazhan S.S., Ershova A.I., Meshkov A.N., Sokolov A.A., Kukharchuk V.V., Gurevich V.S., … Balakhonova T.V., Filippov A.E., Akhmejanov N.M., Aleksandrova O.Yu., Lipovetskiy B.M. Clinical guidelines for familial hypercholesterolemia. *Ateroscleroz*. 2019;15(1):58-98 (in Russian)

Freiberger T. Update on genetics of familial hypercholesterolemia. *Curr Opin Lipidol*. 2026;37(2):38-44. doi 10.1097/MOL.0000000000001027

Hobbs H.H., Russell D.W., Brown M.S., Goldstein J.L. The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. *Annu Rev Genet*. 1990;24:133-170. doi 10.1146/annurev.ge.24.120190.001025

Holla Ø.L., Nakken S., Mattingsdal M., Ranheim T., Berge K.E., Defesche J.C., Leren T.P. Effects of intronic mutations in the *LDLR* gene on pre-mRNA splicing: comparison of wet-lab and bioinformatics analyses. *Mol Genet Metab*. 2009;96(4):245-252. doi 10.1016/j.ymgme.2008.12.014

Hu P., Dharmayat K.I., Stevens C.A.T., Sharabiani M.T.A., Jones R.S., Watts G.F., Genest J., Ray K.K., Vallejo-Vaz A.J. Prevalence of familial hypercholesterolemia among the general population and patients with atherosclerotic cardiovascular disease: a systematic review and meta-analysis. *Circulation*. 2020;141(22):1742-1759. doi 10.1161/CIRCULATIONAHA.119.044795

Jenkins J.L., Agrawal A.A., Gupta A., Green M.R., Kielkopf C.L. U2AF65 adapts to diverse pre-mRNA splice sites through conformational selection of specific and promiscuous RNA recognition motifs. *Nucleic Acids Res*. 2013;41(6):3859-3873. doi 10.1093/nar/gkt046

Kulseth M.A., Berge K.E., Bogsrud M.P., Leren T.P. Analysis of LDLR mRNA in patients with familial hypercholesterolemia revealed a novel mutation in intron 14, which activates a cryptic splice site. *J Hum Genet*. 2010;55(10):676-680. doi 10.1038/jhg.2010.87

Liguori R., Bianco A.M., Argiriou A., Pauciullo P., Giannino A., Rubba P., De Simone V. LDL receptor cDNA sequence analysis in familial hypercholesterolemia patients: 5 novel mutations with high prevalence in families originating from southern Italy. *Hum Mutat*. 2001;17(5):433. doi 10.1002/humu.1122

Malard F., Mackereth C.D., Campagne S. Principles and correction of 5′-splice site selection. *RNA Biol*. 2022;19(1):943-960. doi 10.1080/15476286.2022.2100971

Marduel M., Carrié A., Sassolas A., Devillers M., Carreau V., Di Filippo M., Erlich D., … Junien C.; French ADH Research Network; Boileau C., Varret M., Rabès J.P. Molecular spectrum of autosomal dominant hypercholesterolemia in France. *Hum Mutat*. 2010;31(11):E1811-E1824. doi 10.1002/humu.21348

Martín-Campos J.M., Plana N., Figueras R., Ibarretxe D., Caixàs A., Esteve E., Pérez A., … Pintó X., Masana L., Julve J., Blanco-Vaca F.; Xarxa d'Unitats de Lípids i Arteriosclerosi (XULA). Autosomal dominant hypercholesterolemia in Catalonia: correspondence between clinical-biochemical and genetic diagnostics in 967 patients studied in a multicenter clinical setting. *J Clin Lipidol*. 2018;12(6):1452-1462. doi 10.1016/j.jacl.2018.09.002

Meshkov A., Ershova A., Kiseleva A., Zotova E., Sotnikova E., Petukhova A., Zharikova A., … 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* (*Basel*). 2021;12(1):66. doi 10.3390/genes12010066

Miroshnikova V.V., Romanova O.V., Ivanova O.N., Fedyakov M.A., Panteleeva A.A., Barbitoff Y.A., Muzalevskaya M.V., … Guseva D.M., Pchelina S.N., Glotov A.S., Zakharova E.Y., Glotov O.S. Identification of novel variants in the *LDLR* gene in Russian patients with familial hypercholesterolemia using targeted sequencing. *Biomed Rep*. 2021;14(1):15. doi 10.3892/br.2020.1391

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 Yu.V., Zakian S.M., Zakharova I.S. Calling and phasing of single-nucleotide and structural variants of the *LDLR* gene using Oxford Nanopore MinION. *Int J Mol Sci*. 2023;24(5):4471. doi 10.3390/ijms24054471

Nordestgaard B.G., Chapman M.J., Humphries S.E., Ginsberg H.N., Masana L., Descamps O.S., Wiklund O., … Stalenhoef A.F., Stroes E., Taskinen M.R., Tybjærg-Hansen A.; European Atherosclerosis Society Consensus Panel. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. *Eur Heart J*. 2013;34(45):3478-3490. doi 10.1093/eurheartj/eht273

Okonechnikov K., Golosova O., Fursov M. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012;28(8):1166-1167. doi 10.1093/bioinformatics/bts091

Reeskamp L.F., Hartgers M.L., Peter J., Dallinga-Thie G.M., Zuurbier L., Defesche J.C., Grefhorst A., Hovingh G.K. A deep intronic variant in *LDLR* in familial hypercholesterolemia. *Circ Genom Precis Med*. 2018;11(12):e002385. doi 10.1161/CIRCGEN.118.002385

Reeskamp L.F., Balvers M., Peter J., van de Kerkhof L., Klaaijsen L.N., Motazacker M.M., Grefhorst A., van Riel N.A.W., Hovingh G.K., Defesche J.C., Zuurbier L. Intronic variant screening with targeted next-generation sequencing reveals first pseudoexon in *LDLR* in familial hypercholesterolemia. *Atherosclerosis*. 2021;321:14-20. doi 10.1016/j.atherosclerosis.2021.02.003

Richards S., Aziz N., Bale S., Bick D., Das S., Gastier-Foster J., Grody W.W., Hegde M., Lyon E., Spector E., Voelkerding K., Rehm H.L. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. doi 10.1038/gim.2015.30

Rieck L., Bardey F., Grenkowitz T., Bertram L., Helmuth J., Mischung C., Spranger J., Steinhagen-Thiessen E., Bobbert T., Kassner U., Demuth I. Mutation spectrum and polygenic score in German patients with familial hypercholesterolemia. *Clin Genet*. 2020;98(5):457-467. doi 10.1111/cge.13826

Sambrook J., Russell D.W. Purification of nucleic acids by extraction with phenol:chloroform. *CSH Protoc*. 2006;2006(1):pdb.prot4455. doi 10.1101/pdb.prot4455

Segrest J.P., Jones M.K., De Loof H., Dashti N. Structure of apolipoprotein B-100 in low density lipoproteins. *J Lipid Res*. 2001;42(9):1346-1367. doi 10.1016/S0022-2275(20)30267-4

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

Shakhtshneider E.V., Ivanoshchuk D.E., Makarenkova K.V., Orlov P.S., Timoshchenko O.V., Bazhan S.S., Nikitin Yu.P., Voevoda M.I. Cascade genetic screening in diagnostics of heterozygous familial hypercholesterolemia: clinical case. *Rossijskii Kardiologicheskii Zhurnal = Russian Journal of Cardiology*. 2017;22(6):178-179. doi 10.15829/1560-4071-2017-6-178-179 (in Russian)

Shamsani J., Kazakoff S.H., Armean I.M., McLaren W., Parsons M.T., Thompson B.A., O'Mara T.A., Hunt S.E., Waddell N., Spurdle A.B.

A plugin for the Ensembl Variant Effect Predictor that uses MaxEntScan to predict variant spliceogenicity. *Bioinformatics*. 2019;35(13):2315-2317. doi 10.1093/bioinformatics/bty960

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

Voith von Voithenberg L., Sánchez-Rico C., Kang H.S., Madl T., Zanier K., Barth A., Warner L.R., Sattler M., Lamb D.C. Recognition of the 3′ splice site RNA by the U2AF heterodimer involves a dynamic population shift. *Proc Natl Acad Sci USA*. 2016;113(46):E7169-E7175. doi 10.1073/pnas.1605873113

Vorechovský I. Aberrant 3′ splice sites in human disease genes: mutation pattern, nucleotide structure and comparison of computational tools that predict their utilization. *Nucleic Acids Res*. 2006;34(16):4630-4641. doi 10.1093/nar/gkl535

Walker L.C., de la Hoya M., Wiggins G.A.R., Lindy A., Vincent L.M., Parsons M.T., Canson D.M., … Pesaran T., Karam R., Harrison S.M., Spurdle A.B.; ClinGen Sequence Variant Interpretation Working Group. Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: recommendations from the ClinGen SVI Splicing Subgroup. *Am J Hum Genet*. 2023;110(7):1046-1067. doi 10.1016/j.ajhg.2023.06.002

Conflict of interest. The authors declare no conflict of interest.
Received May 22, 2026. Revised June 22, 2026. Accepted June 29, 2026.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):814-823

doi 10.18699/vjgb-26-82

A novel variant c.1185_1196dup (p.(Gly396_Ser399dup)) in the *SLC26A4* gene associated with recessive hearing loss

M.V. Zytsar [1, 2#], V.Yu. Danilchenko [1, 2#], A.A. Bondar [3], K.E. Orishchenko [1, 2], O.L. Posukh [1, 2]

[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] Knorre Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
zytzar@bionet.nsc.ru

Abstract. Pathogenic variants in the *SLC26A4* gene (solute carrier family 26, member 4) are a common cause of inherited hearing loss. The *SLC26A4* gene encodes the transmembrane protein pendrin, a member of the SLC26 anion transporter family, with predominant expression in the inner ear, thyroid, and kidney tissues. Pathogenic *SLC26A4* variants cause nonsyndromic recessive hearing loss (DFNB4) and Pendred syndrome (sensorineural hearing loss and thyroid dysfunction). In many studies analyzing the *SLC26A4* gene, along with patients who have two recessive pathogenic *SLC26A4* variants (M2 patients) and an accurate molecular genetic diagnosis can be made, a group of patients with only one pathogenic *SLC26A4* variant (M1 patients) is often identified, which creates a diagnostic problem. The presence of M1 patients may reflect copy number variations (CNVs, deletions/duplications) in the *SLC26A4* gene that are not detected by routine diagnostic methods. The aim of the study is to search for deletions/duplications in the *SLC26A4* gene using the MLPA (Multiplex Ligation-dependent Probe Amplification) method in thirteen patients belonging to the indigenous people of the Tyva Republic (Southern Siberia), in whom only one pathogenic *SLC26A4* variant (M1 patients) was identified during the study of the etiology of hereditary hearing loss. As a result of MLPA analysis and molecular cloning in DNA samples of three M1 patients from two related Tuvinian families, a novel variant was revealed – tandem duplication of twelve nucleotides (c.1185_1196dup) in exon 10 of the *SLC26A4* gene. This variant is an in-frame insertion resulting in the inclusion of four additional amino acid residues Gly-Phe-Phe-Ser (p.(Gly396_Ser399dup)) in a highly conserved region of the pendrin protein. Most predictive programs predict the damaging effect of the c.1185_1196dup variant on the structure and function of the pendrin protein. The pathogenicity of the c.1185_1196dup variant is supported by its segregation with hearing pathology in the pedigree of patients in whom it was found in a compound heterozygous state with pathogenic *SLC26A4* variants, as well as its absence in a control sample of Tuvinians and in the world genomic databases. Functional *in vitro* studies are needed to confirm the potentially deleterious effects of the novel c.1185_1196dup (p.(Gly396_Ser399dup)) variant and its association with hearing pathology.

Key words: hearing loss; *SLC26A4*; pathogenic variants of the *SLC26A4* gene; MLPA method; indels

For citation: Zytsar M.V., Danilchenko V.Yu., Bondar A.A., Orishchenko K.E., Posukh O.L. A novel variant c.1185_1196dup (p.(Gly396_Ser399dup)) in the *SLC26A4* gene associated with recessive hearing loss. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(5):814-823. doi 10.18699/vjgb-26-82

Funding. This study was supported by the budget project FWNR-2026-0027 of the Institute of Cytology and Genetics SB RAS and the grant FSUS-2024-0018 of the Ministry of Science and Higher Education of the Russian Federation.

Новый вариант c.1185_1196dup (p.(Gly396_Ser399dup)) в гене *SLC26A4*, ассоциированный с рецессивно наследуемой тугоухостью

М.В. Зыцарь [1, 2#], В.Ю. Данильченко [1, 2#], А.А. Бондарь [3], К.Е. Орищенко [1, 2], О.Л. Посух [1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Институт химической биологии и фундаментальной медицины им. Д.Г. Кнорре Сибирского отделения Российской академии наук, Новосибирск, Россия
zytzar@bionet.nsc.ru

Аннотация. Патогенные варианты в гене *SLC26A4* (solute carrier family 26, member 4) – частая причина наследственной потери слуха. Ген *SLC26A4* кодирует трансмембранный белок пендрин из семейства анионных транспортеров SLC26, с преимущественной экспрессией в тканях внутреннего уха, щитовидной железы и почек.

# These authors contributed equally to this work.

Патогенные *SLC26A4*-варианты вызывают несиндромальную рецессивно наследуемую потерю слуха (DFNB4) и синдром Пендреда (сенсоневральная тугоухость и дисфункция щитовидной железы). Во многих исследованиях при анализе гена *SLC26A4*, наряду с пациентами, у которых обнаружено два рецессивных патогенных *SLC26A4*-варианта (M2-пациенты) и может быть поставлен точный молекулярно-генетический диагноз, часто выявляется группа пациентов только с одним патогенным *SLC26A4*-вариантом (M1-пациенты), что создает диагностическую проблему. Наличие M1-пациентов может отражать не обнаруженные рутинными методами диагностики вариации числа копий (CNV, делеции/дупликации в гене) *SLC26A4*. Цель работы – поиск делеций/дупликаций в гене *SLC26A4* методом MLPA (Multiplex Ligation-dependent Probe Amplification) у 13 пациентов, относящихся к коренному населению Республики Тыва (Южная Сибирь), у которых при изучении этиологии наследственной потери слуха был найден только один патогенный *SLC26A4*-вариант (M1-пациенты). В результате MLPA-анализа и молекулярного клонирования в образцах ДНК трех M1-пациентов из двух родственных тувинских семей обнаружен новый вариант – тандемная дупликация 12 нуклеотидов (c.1185_1196dup) в экзоне 10 гена *SLC26A4*, которая является внутрирамочной инсерцией, приводящей к включению четырех дополнительных аминокислотных остатков Gly-Phe-Phe-Ser (p.(Gly396_Ser399dup)) в высококонсервативном районе белка пендрин. Бóльшая часть предсказательных программ прогнозирует повреждающее воздействие варианта c.1185_1196dup на структуру и функцию белка пендрин. В пользу патогенности свидетельствует его сегрегация с патологией слуха в родословной пациентов, у которых он найден в компаунд-гетерозиготном состоянии с патогенными *SLC26A4*-вариантами, а также его отсутствие в контрольной выборке тувинцев и мировых базах геномных данных. Для подтверждения потенциально негативных эффектов нового варианта c.1185_1196dup (p.(Gly396_Ser399dup)) и его ассоциации с патологией слуха необходимы функциональные *in vitro* исследования.

Ключевые слова: потеря слуха; патогенные варианты гена *SLC26A4*; метод MLPA; инделы

Introduction

Hereditary hearing loss is a monogenic disorder characterized by exceptionally high genetic heterogeneity. It may occur as an isolated condition or as one of the clinical features of numerous distinct syndromes. According to the Hereditary Hearing Loss Homepage (https://hereditaryhearingloss.org, accessed April 8, 2026), pathogenic variants in over 150 different genes – including those located in mitochondrial DNA – are associated with nonsyndromic (isolated) forms of hearing loss. These forms exhibit diverse inheritance patterns, including autosomal dominant (DFNA loci), autosomal recessive (DFNB loci), and X-linked (DFNX loci). Pathogenic variants in the *GJB2* gene (gap junction protein beta-2, 13q12.11, OMIM #121011), which encodes the transmembrane protein connexin 26, are responsible for nonsyndromic autosomal recessive deafness type 1A (DFNB1A, OMIM #220290) and represent one of the most prevalent forms of hearing loss across many populations (del Castillo F., del Castillo I., 2017). Pathogenic variants in the *SLC26A4* gene (solute carrier family 26, member 4, 7q22.3, OMIM #605646) constitute the second most common cause of inherited hearing loss, at least in Asian populations (Park et al., 2003; Albert et al., 2006; Yuan et al., 2009; Du et al., 2013; Tsukada et al., 2015; Koohiyan, 2019; Lee et al., 2023).

The *SLC26A4* gene encodes protein pendrin, an anion transporter belonging to the SLC26 family. Pendrin is predominantly expressed in the inner ear, thyroid gland, and kidneys, where it mediates the transport of various anions (Everett et al., 1997; Mount, Romero, 2004). In the inner ear, pendrin maintains endolymph homeostasis via Cl^-/HCO_3^- anion exchange, whereas in thyroid cells it facilitates I^- efflux from follicular cells into the follicular lumen (Everett et al., 1997; Wangemann et al., 2007; Tesolin et al., 2021). Pathogenic variants in *SLC26A4* disrupt the structure and function of pendrin and can lead to either nonsyndromic recessive hearing loss (DFNB4, OMIM #600791) or Pendred syndrome (OMIM #274600), which is characterized by sensorineural hearing loss and thyroid dysfunction (goiter). An important clinical feature of *SLC26A4*-associated hearing loss is an enlarged vestibular aqueduct (EVA, OMIM #603545), along with other malformations of the inner ear; these are typically identified by computed tomography (CT) of the temporal bones.

Clinical manifestations in patients with pathogenic *SLC26A4* variants are highly variable. Hearing loss may be congenital or manifest later in life, with severity ranging from moderate impairment to profound deafness and often showing significant progression. Similarly, the extent of inner ear malformations and the onset and severity of thyroid dysfunction in Pendred syndrome vary widely (Wangemann et al., 2007; Griffith, Wangemann, 2011; Mey et al., 2019; Roesch et al., 2021; Honda, Griffith, 2022).

Screening for pathogenic *SLC26A4* variants is an important part of molecular genetic diagnostics for hearing impairment, particularly in patients with EVA and other anomalies of the hearing organ. Nevertheless, the contribution of *SLC26A4* to the etiology of hearing loss remains incompletely characterized. There is considerable variability across studies in the proportion of patients with biallelic pathogenic *SLC26A4* variants (M2 patients) and those with only a single pathogenic variant (M1 patients) (Azaiez et al., 2007; Choi B. et al., 2009; Chattaraj et al., 2013; Pique et al., 2014; Rose et al., 2017; Smits et al., 2022; Tian et al., 2024). Moreover, a substantial number of individuals exhibiting clinical features typical of *SLC26A4*-related hearing loss do not show any detectable pathogenic variants in this gene (M0 patients). This variability in diagnostic yields can be caused by several

factors: heterogeneity in patient cohorts (differing in size and inclusion criteria); potential contributions from other genes or environmental factors (Yang et al., 2007, 2009; Pique et al., 2014; Chattaraj et al., 2017; Chao et al., 2019; Smits et al., 2022); and differences in the methodologies used to analyze the *SLC26A4* gene. These range from targeted screening of common pathogenic variants to comprehensive analysis of coding and flanking intronic regions or the entire gene sequence via Sanger sequencing and/or next-generation sequencing (NGS) technologies. Notably, the relatively high proportion of M1 patients reported in many studies may reflect the presence of undetected pathogenic variants or copy number variations (CNVs) in the *SLC26A4* gene.

CNVs (deletions or duplications ranging from 50–100 bp to several megabases) are well-established causes of numerous genetic disorders, including hereditary hearing loss (Zhang et al., 2009; Stankiewicz, Lupski, 2010; Shearer et al., 2014; Rentas, Abou Tayoun, 2021; Royer-Bertrand et al., 2021).

Among the methods developed for detecting such deletions and duplications, Multiplex Ligation-dependent Probe Amplification (MLPA) is considered one of the most promising approaches (Schouten et al., 2002; Stuppia et al., 2012). The technique relies on hybridization of two probe segments to the target DNA sequence, followed by ligation and PCR amplification with universal fluorescently labeled primers. Each amplified product has a distinct length, enabling simultaneous detection of multiple fragments by capillary electrophoresis. The signal intensity of each PCR product correlates with the copy number of the corresponding DNA fragment; thus, comparison with reference samples allows identification of copy number alterations in specific genomic regions.

To date, several large deletions affecting one or more exons of the *SLC26A4* gene have been identified using various molecular genetic methods, including MLPA (Park et al., 2003; Hu H. et al., 2007; Pera et al., 2008; Anwar et al., 2009; Pique et al., 2014; Pang et al., 2015; Cirello et al., 2018; Lin et al., 2019; Tian et al., 2024). Although most studies report that the frequency of CNVs in *SLC26A4*-associated hearing loss is low (typically a few percent or less), CNV testing is recommended as part of the diagnostic workflow for patients with unresolved etiologies, particularly among M1 individuals in whom only a single pathogenic *SLC26A4* variant has been identified.

In a previous study investigating the etiology of hereditary hearing loss in the indigenous population of the Tyva Republic (Southern Siberia), a notably high proportion of Tuvinian patients (62/220, 28.2 %) was found to carry biallelic pathogenic *SLC26A4* variants in either a homozygous or compound heterozygous state (M2 patients), thereby enabling a molecular diagnosis of recessively inherited hearing loss due to *SLC26A4* pathogenic variants. Additionally, 13 patients were identified as M1 carriers, harboring only one pathogenic *SLC26A4* variant (Danilchenko et al., 2021). The aim of the present study is to investigate deletions and duplications in the *SLC26A4* gene using the MLPA method in a group of M1 patients who carry a single pathogenic variant of this gene.

Materials and methods

Samples analyzed

MLPA analysis was carried out on a group of 13 Tuvinian patients (M1 patients) in whom only a single recessive pathogenic variant of the *SLC26A4* gene had been identified during the initial molecular screening (Danilchenko et al., 2021). Additionally, a control sample comprising 129 ethnically matched (Tuvinians) healthy individuals without hearing impairment was used to screen for the novel c.1185_1196dup variant.

Experimental methods

Multiplex Ligation-dependent Probe Amplification (MLPA). Deletions and duplications in the *SLC26A4* gene were investigated using MLPA with the SALSA MLPA Probemix P280-B4 SLC26A4 kit (MRC-Holland, Netherlands). The kit includes 24 probes: 21 targeting each exon of the *SLC26A4* gene and 3 designed to detect point mutations c.707T>C, IVS8+1G>A (c.1001+1G>A), and c.1246A>C. MLPA was performed according to the manufacturer's protocol. Amplified MLPA products were separated by capillary electrophoresis on a 3130XL Genetic Analyzer (Applied Biosystems, USA). Data analysis was conducted using Coffalyser. Net software (https://www.mrcholland.com/).

Molecular cloning. DNA samples from patients harboring the novel c.1185_1196dup variant (exon 10) of the *SLC26A4* gene were used as templates for molecular cloning. Exon 10 was amplified using specific primers (5′-ACC TCTAGAAGATGGTATGGCGT-3′ and 5′-GCTAGGATCC GTGCGAGCCTTCCT-3′), which contain *XbaI* and *BamHI* restriction endonuclease recognition sites at their 5′-ends, respectively. PCR products and the pUC18 plasmid vector were digested with *XbaI* and *BamHI* (New England Biolabs, USA). Subsequently, PCR products were purified using Agencourt Ampure XP magnetic beads (Beckman Coulter, USA), and the pUC18 vector was isolated from a 1 % agarose gel with the QIAEX II Gel Extraction Kit (Qiagen, Netherlands). The purified PCR fragment was ligated into the prepared pUC18 vector using T4 DNA ligase (SibEnzyme, Russia). The ligation mixture was purified on Cleanup S-Cap columns (Eurogen, Russia) and used to transform *E. coli* XL1-blue cells via electroporation. Recombinant clones carrying the target DNA fragment were selected by PCR colony screening 14–16 h after transformation, using primers 5′-GTGGAATTGTGAGCGG ATAAC-3′ and 5′-TCCCAGTCACGACGTTGTAAA-3′. The nucleotide sequence of the insert was confirmed by Sanger sequencing.

Sanger sequencing. Exon 10 of the *SLC26A4* gene was amplified from DNA samples of patients using primers 5′-GAGCAGATATAGCATTTGATGATGAGATG-3′ and 5′-CTCGGTGCGAGCCTTCCT-3′. PCR products were purified with Agencourt Ampure XP magnetic beads, followed by cycle sequencing with the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA). Sanger reaction products were purified by gel filtration on Sephadex G-50

medium columns (GE Healthcare, USA) and analyzed on an ABI 3130XL Genetic Analyzer (Applied Biosystems, USA).

Screening of the novel c.1185_1196dup variant in the control sample. The c.1185_1196dup variant in exon 10 was screened in the control sample by amplifying the corresponding DNA fragment with primers 5′-GACCTCTCT CAGATGGTATGGCGT-3′ and 5′-CTCGGT GCG AGC C TTCCT-3′. PCR products were resolved on a 4 % agarose gel, allowing discrimination between wild-type (395 bp) and heterozygous (395 and 407 bp) alleles.

All genetic engineering procedures were performed at the Cell Technologies Sector of the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (Novosibirsk). Sanger sequencing and MLPA analyses were carried out at the SB RAS Genomics Core Facility (D.G. Knorre Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences, Novosibirsk).

Bioinformatics analysis

The functional impact of the novel c.1185_1196dup variant on the structure and function of the pendrin protein was predicted using the following bioinformatics tools: MutationTaster2021 (https://www.genecascade.org/MutationTaster2021/) (Steinhaus et al., 2021), PROVEAN (http://provean.jcvi.org) (Choi Y. et al., 2012, 2015), SIFT (https://siftdna.org/www/SIFT_indels2.html) (Hu J. et al., 2013), VEST-Indel / CRAVAT (https://cravat.us/CRAVAT/) (Douville et al., 2015), and MutPred-Indel (https://mutpred2.mutdb.org/mutpredindel/) (Pagel et al., 2019). The secondary structure of pendrin was modeled using the I-TASSER platform (https://aideepmed.com/I-TASSER/) (Zheng et al., 2021, 2026; Zhou et al., 2022). Structural visualization of the pendrin protein was performed with DeepView / Swiss-PdbViewer v4.1.0 (http://www.expasy.org/spdbv/) (Guex and Peitsch, 1997).

Results

MLPA analysis targeting the *SLC26A4* gene was performed in a group of 13 Tuvinian patients previously classified as M1 (i. e., carrying a single recessive pathogenic *SLC26A4* variant). MLPA enables the detection of copy number alterations, including deletions and duplications, in specific genomic regions. In three of 13 M1 patients, a similar copy number alteration was observed in the region of exon 10: the normalized ratio for probe 09251-L09442 (SALSA MLPA Probemix P280-B4 SLC26A4 kit) was approximately 0.5 (Fig. 1*a*). To further characterize the DNA samples from these three patients, fragments of exon 10 were subjected to molecular cloning into the pUC18 plasmid vector, followed by Sanger sequencing. Sequence analysis of the cloned fragments revealed a 12-nucleotide duplication (AGGATTCTTCTC) on one allele, corresponding to the c.1185_1196dup variant (Fig. 1*b*).

The observed reduction in the MLPA signal (~0.5) for the 09251-L09442 probe in samples carrying the c.1185_1196dup variant is consistent with a heterozygous deletion (Fig. 1*a*). The 09251-L09442 probe (184 bp) contains two oligonucleotides (5′-TCTCAGGATTCT-3′ and 5′-TCTCTTGTTTTG-3′) that hybridize adjacently to the target sequence in *SLC26A4* (TCTCAGGATTCT-TCTCTTGTTTTG), enabling their subsequent ligation (Fig. 1*c*). Notably, the duplicated sequence (AGGATTCTTCTC) in the c.1185_1196dup variant overlaps with the probe sequence (Fig. 1*c*). In the heterozygous state of the c.1185_1196dup variant, the probe signal is detected only from the wild-type (wt) allele. This is because the 12-nucleotide duplication increases the distance between the two probe segments, thereby preventing their ligation and subsequent amplification. Consequently, the absence of a signal from the mutant allele results in an apparent reduction of the target copy number to ~0.5, masking the true nature of the variant as a duplication (Fig. 1*a*). This observation exemplifies a well-documented limitation of the MLPA method: genetic variants that overlap with probe binding sites can interfere with hybridization and/or ligation, leading to false-positive fluorescent signals and potentially erroneous interpretation of the results. In such cases, it is recommended to continue research using other methods (Schouten et al., 2002; Stuppia et al., 2012; Varga et al., 2012). In the present study, MLPA initially identified three DNA samples with a copy number alteration (~0.5) in exon 10 of *SLC26A4*. Subsequent detailed characterization by molecular cloning and Sanger sequencing revealed that this alteration represented a heterozygous 12-nucleotide duplication (c.1185_1196dup). Further analysis demonstrated that c.1185_1196dup is an in-frame insertion in exon 10, resulting in the incorporation of four additional amino acid residues (Gly-Phe-Phe-Ser) within the ninth transmembrane domain (TM9, amino acid positions 385–405, UniProt: https://www.uniprot.org/uniprotkb/O43511/entry) of the pendrin protein (p.(Gly396_Ser399dup)) (Fig. 1*d*).

Three patients from two related Tuvinian families, who had previously been classified as monoallelic carriers (M1) for *SLC26A4*, were reclassified following the identification of the c.1185_1196dup variant. Their refined genotypes are as follows: III-1 – c.[1185_1196dup];[1545T>G], III-2 – c.[1185_1196dup];[1545T>G], and IV-2 – c.[1185_1196dup];[2027T>A] (Fig. 1*e*). Notably, pedigree analysis (Fig. 1*e*) revealed the co-segregation of three distinct pathogenic *SLC26A4* alleles: the novel c.1185_1196dup variant (exon 10); the unique c.1545T>G variant (exon 14), previously reported only in Tuvinian individuals; and the c.2027T>A variant (exon 17), which is relatively frequent in Tuvinians but rare elsewhere (Danilchenko et al., 2021).

Screening for the c.1185_1196dup variant in a control sample of 129 Tuvinian individuals without hearing impairment did not detect this variant (data not shown).

To evaluate the potential functional consequences of the c.1185_1196dup (p.(Gly396_Ser399dup)) variant on the structure and function of the pendrin protein, several bioinformatics tools were applied (see the Table). They were chosen by the need to apply algorithms specifically adapted for indel (insertions/deletions) prediction, as most available tools are

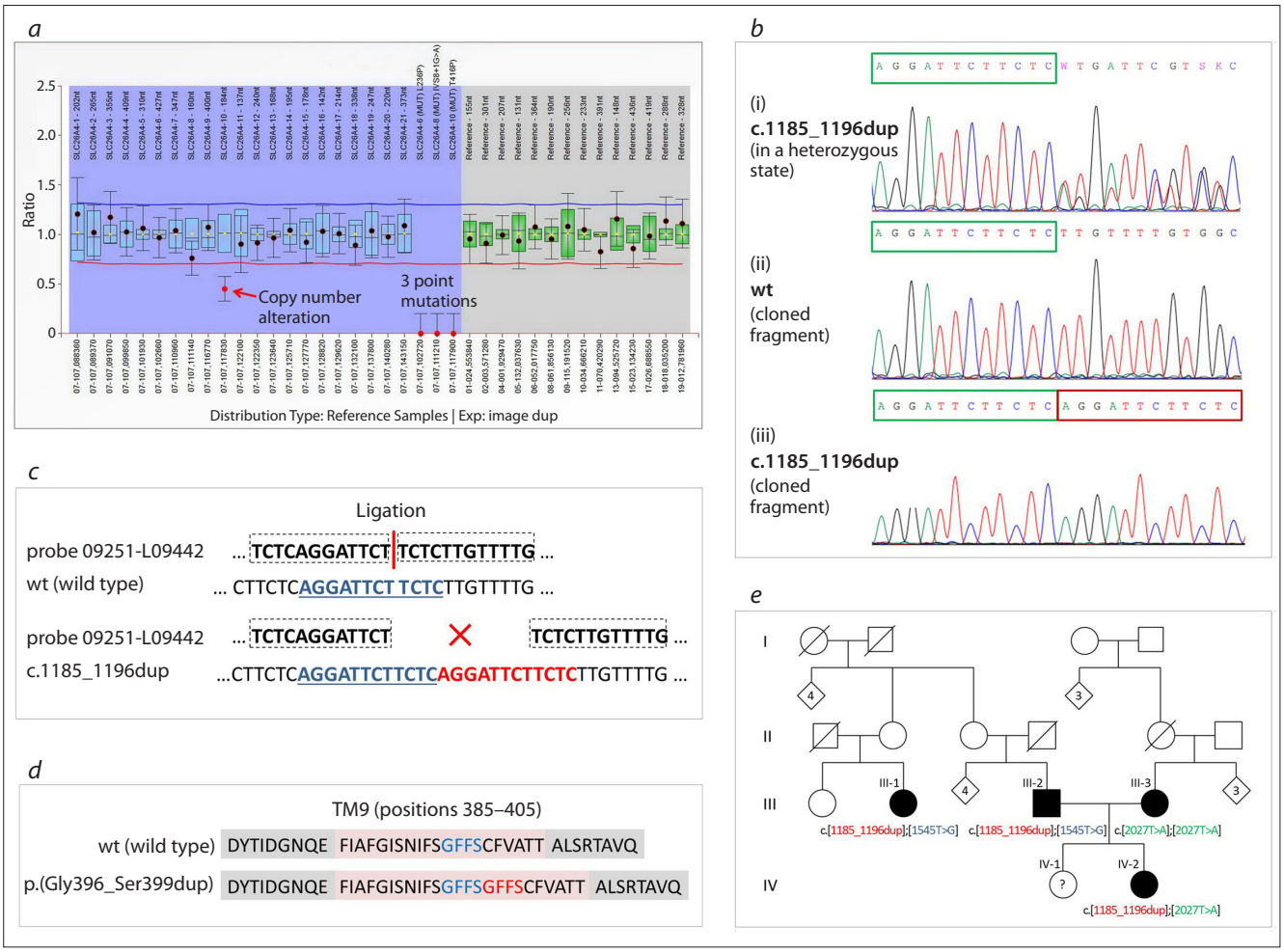


Fig. 1. Identification and molecular genetic analysis of a novel variant c.1185_1196dup in the *SLC26A4* gene.

a – MLPA profile of one of the DNA samples with a copy number alteration in the region of exon 10 (shown by the red arrow) (see explanation in the text); *b* – Sanger sequencing traces revealing the c.1185_1196dup variant: (i) directly sequenced DNA fragment with the heterozygous c.1185_1196dup variant in exon 10; (ii) cloned wild-type (wt) fragment of exon 10; (iii) cloned fragment of exon 10 carrying the c.1185_1196dup variant; *c* – sequence fragments of probe 09251-L09442 (the *SLC26A4* exon 10 region) from the SALSA MLPA Probemix P280-B4 SLC26A4 kit and the target region in the *SLC26A4* gene: the wild-type (wt) is highlighted in blue, the 12-nucleotide insertion is shown in red (see explanation in the text); *d* – amino acid sequence of the pendrin protein region including transmembrane domain 9 (TM9); GFFS – four amino acid residues Gly-Phe-Phe-Ser: the wild-type sequence (wt) is shown in blue, the duplicated sequence resulting from c.1185_1196dup is highlighted in red; *e* – pedigree of Tuvinian families with individuals III-1, III-2, and IV-2 having compound heterozygous *SLC26A4* genotypes. Variant c.1185_1196dup (exon 10) is highlighted in red, variant c.1545T>G (exon 14) – in blue, variant c.2027T>A (exon 17) – in green. Black symbols indicate individuals with hearing loss. "?" denotes unknown hearing status (individual IV-1).

Bioinformatic assessment of the functional effect of the novel c.1185_1196dup variant in the *SLC26A4* gene

Predictive programs	Potential effect assessment (Threshold value)	References
Mutation Taster21 (https://www.genecascade.org/MutationTaster2021/)	Benign	Steinhaus et al., 2021
PROVEAN (http://provean.jcvi.org)	Deleterious −10.34 (< −2.5)	Choi Y. et al., 2012, 2015
SIFT (https://siftdna.org/www/SIFT_indels2.html)	Damaging 0.826 (> 0)	Hu J. et al., 2013
VEST-indel (https://cravat.us/CRAVAT/)	Pathogenic 0.99 (pathogenic ≥ 0.5, benign < 0.5)	Douville et al., 2015
MutPred-Indel (https://mutpred2.mutdb.org/mutpredindel/)	Pathogenic 0.81601 (pathogenic ≥ 0.5, benign < 0.5)	Pagel et al., 2019

primarily optimized for single nucleotide variants (SNVs). The majority of the applied prediction algorithms indicated a damaging effect of the c.1185_1196dup variant on the structure and function of pendrin (see the Table).

Discussion

In the present study, MLPA analysis of three M1 patients from two related Tuvinian families led to the identification of a novel variant in the *SLC26A4* gene: NM_000441.2:c.1185_1196dup. This variant represents a 12-nucleotide tandem duplication in exon 10, classified as an in-frame insertion. It results in the incorporation of four additional amino acid residues (Gly-Phe-Phe-Ser) in transmembrane domain 9 (TM9) of the pendrin protein, designated as p.(Gly396_Ser399dup) (Fig. 1 and 2). Variant c.1185_1196dup was not detected in the control sample of Tuvinians (n = 129), and was also not registered in any of the world genomic databases (dbSNP, gnomAD, ClinVar).

Currently, the total number of all identified variations in the *SLC26A4* gene sequence (21 exons), spanning approximately 57 kb of genomic DNA (GRCh38, NC_000007.14, 7:107,660,828-107,717,809), exceeds 37,000 (the Deafness Variation Database: https://deafnessvariationdatabase.org/genes/SLC26A4, May 2026). However, only approximately 6–7 % (~2,500 variants) have been classified according to current guidelines (Richards et al., 2015) as “pathogenic”, “likely pathogenic”, “benign”, or “likely benign”; the pathogenicity status of the vast majority remains uncertain. More than 800 variants with established pathogenicity (“pathogenic”

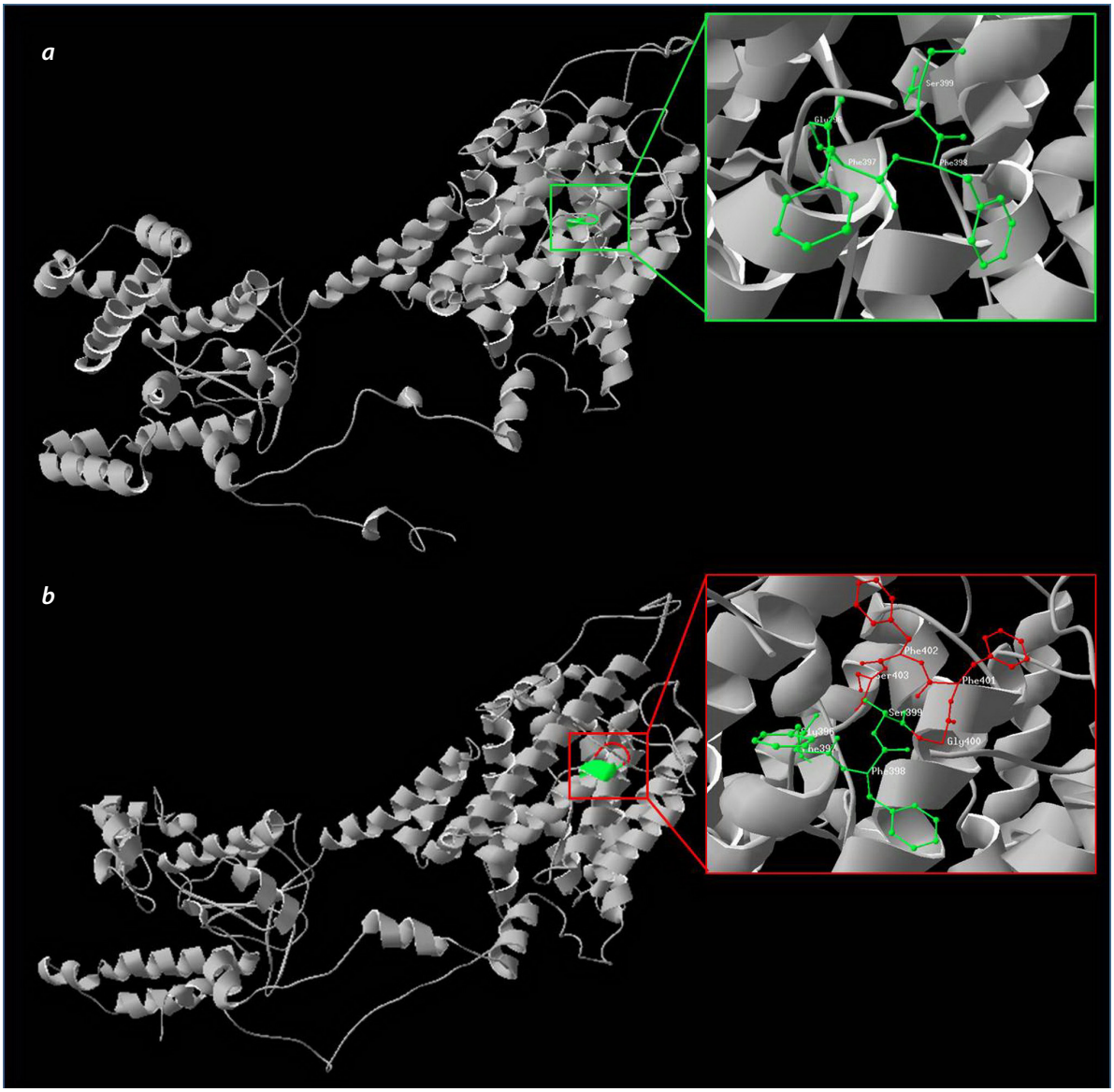


Fig. 2. Schematic representation of the pendrin protein secondary structure and the fragments of its amino acid sequence.

a – wild-type amino acid sequence 396Gly-397Phe-398Phe-399Ser (shown in green); *b* – duplicated sequence Gly396_Ser399dup (shown in red).

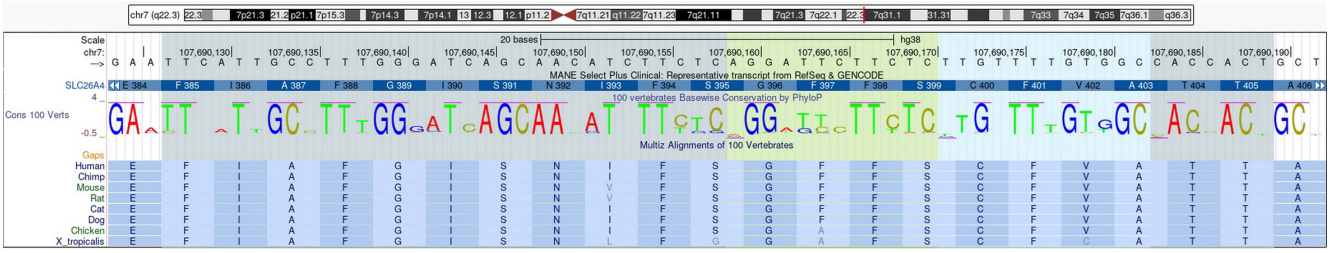


Fig. 3. Multiple sequence alignment of the pendrin protein generated using the UCSC Genome Browser (https://genome.ucsc.edu/).

The transmembrane segment TM9 (amino acid positions 385–405, according to UniProt: https://www.uniprot.org/uniprotkb/O43511/entry) is shaded in gray. The wild-type nucleotide and amino acid sequences are shown in green; the insertion site corresponding to the c.1185_1196dup variant is highlighted in blue.

and "likely pathogenic") are linked to a broad phenotypic spectrum of hearing impairments. Notably, small insertions and deletions (one to several nucleotides) constitute a substantial proportion (~25 %) of pathogenic variants in the coding regions of *SLC26A4*. Most of these are frameshift variants that introduce premature stop codons and lead to truncated, nonfunctional proteins. In contrast, in-frame indels are considerably rarer and have been reported only in isolated cases of hearing loss (the Deafness Variation Database: https://deafnessvariationdatabase.org/genes/SLC26A4, May 2026). These variants do not disrupt the reading frame but alter the length of the pendrin protein.

The precise molecular structure of human pendrin, a member of the SLC26 family of anion transporters (SLC26A1–11), has not yet been determined. However, structural insights into human pendrin have been advanced by cryo-electron microscopy (cryo-EM) studies of orthologous proteins in several mammalian species (*Mus musculus*, *Sus scrofa*), which show significant conservation with human pendrin (Walter et al., 2019; Liu et al., 2023; Takahashi, Homma, 2024; Wang et al., 2024). Current structural models indicate that human pendrin functions as a homodimer, with each protomer comprising 14 transmembrane α-helices (TM1–TM14) and a C-terminal cytoplasmic STAS domain (Sulfate Transporter and Anti-Sigma factor antagonist domain), whose precise function remains incompletely understood. The transmembrane region adopts an inverted repeat structure formed by TM1–TM7 and TM8–TM14. The core domain (TM1–TM4 and TM8–TM11) harbors two anion-binding sites and is directly responsible for substrate transport, whereas a smaller scaffold (or gate) domain (TM5–TM7 and TM12–TM14) provides structural support and functional coordination for the core domain (Liu et al., 2023; Wang et al., 2024).

The p.(Gly396_Ser399dup) variant resulting from c.1185_1196dup introduces an insertion within TM9, a transmembrane segment located in a central position (385–405) in the pendrin protein (780 amino acid residues). TM9 and its adjacent segments are highly conserved regions of the protein (Fig. 3), as evidenced by the high density of pathogenic and likely pathogenic *SLC26A4* variants in these regions (the Deafness Variation Database: https://deafnessvariationdatabase.org/genes/SLC26A4, May 2026).

Thus, the "shift" of the amino acid sequence in the case of the p.(Gly396_Ser399dup) variant, which occurs in the transmembrane segment TM9 and leads to an increase in the length of the pendrin protein from 780 to 784 amino acid residues, can probably change the secondary structure of the protein, which entails a disruption or complete loss of the anion transport function.

Unlike frameshift indels, which typically result in premature translation termination and loss of protein function, in-frame indels alter the length of a protein while leaving the underlying amino acid sequence intact. The larger the size of such indels and the more evolutionarily conserved the amino acid residues within the indel, the more significant their negative effect on protein structure and function may be. The functional impact of such variants is often challenging to predict, because most bioinformatics tools are optimized for missense substitutions rather than indels, although algorithms for in-frame indel prediction are constantly being improved (Pagel et al., 2019; Cannon et al., 2023; Abderrazzaq et al., 2026). According to established variant interpretation guidelines (Richards et al., 2015), in-frame indels located in non-repetitive, evolutionarily conserved regions and resulting in the synthesis of a protein of altered length, may be assigned a moderate pathogenicity criterion (PM4). Most of the predictive programs used to evaluate the novel variant c.1185_1196dup (p.(Gly396_Ser399dup)) predict its damaging effect on the structure and function of the pendrin protein (see the Table). The pathogenicity of this variant is also supported by its segregation with hearing loss in three Tuvinian patients, in whom it was found in a compound heterozygous state with other known pathogenic variants of the *SLC26A4* gene (Fig. 1*e*), as well as its absence in world genomic databases (dbSNP, gnomAD, ClinVar). A limitation of the present study is the absence of functional *in vitro* assays to directly assess the impact of p.(Gly396_Ser399dup) on the structure and functions of protein pendrin. In the future, we plan to conduct such studies based on experimental platforms designed taking into account the transmembrane localization of the pendrin protein and its functioning as an anion transporter.

Conclusion

The molecular diagnosis of patients harboring a single identified pathogenic recessive allele of the *SLC26A4* gene (M1 patients) remains a significant challenge, as conventional molecular genetic approaches may fail to detect all underlying variants. In this study, reanalysis of the *SLC26A4* locus by MLPA in a cohort of 13 M1 patients enabled the identification of a novel variant, c.1185_1196dup, in three individuals. The c.1185_1196dup variant represents an in-frame 12-nucleotide insertion resulting in the incorporation of four additional amino acid residues Gly-Phe-Phe-Ser (p.(Gly396_Ser399dup)) within a highly conserved transmembrane domain (TM9) of the pendrin protein. The combined data on the novel c.1185_1196dup variant, including its segregation with hearing loss, consistent prognostic estimates obtained by various bioinformatics tools, and its absence in global databases, support its probable pathogenicity. However, functional *in vitro* studies are needed to confirm the potential negative effects of novel c.1185_1196dup (p.(Gly396_Ser399dup)) variant.

References

Abderrazzaq H., Singh M., Babb L., Bergquist T., Brenner S.E., Pejaver V., O'Donnell-Luria A., Radivojac P., ClinGen Computational Working Group and ClinGen Variant Classification Working Group. Calibration of in-frame indel variant effect predictors for clinical variant classification. *bioRxiv* [Preprint]. 2026. doi 10.64898/2026.04.15.718599

Albert S., Blons H., Jonard L., Feldmann D., Chauvin P., Loundon N., Sergent-Allaoui A., … Petit C., Garabédian E.N., Couderc R., Marlin S., Denoyelle F. *SLC26A4* gene is frequently involved in non-syndromic hearing impairment with enlarged vestibular aqueduct in Caucasian populations. *Eur J Hum Genet*. 2006;14(6):773-779. doi 10.1038/sj.ejhg.5201611

Anwar S., Riazuddin S., Ahmed Z.M., Tasneem S., Ateeq-ul-Jaleel, Khan S.Y., Griffith A.J., Friedman T.B., Riazuddin S. *SLC26A4* mutation spectrum associated with DFNB4 deafness and Pendred's syndrome in Pakistanis. *J Hum Genet*. 2009;54(5):266-270. doi 10.1038/jhg.2009.21

Azaiez H., Yang T., Prasad S., Sorensen J.L., Nishimura C.J., Kimberling W.J., Smith R.J. Genotype-phenotype correlations for *SLC26A4*-related deafness. *Hum Genet*. 2007;122(5):451-457. doi 10.1007/s00439-007-0415-2

Cannon S., Williams M., Gunning A.C., Wright C.F. Evaluation of *in silico* pathogenicity prediction tools for the classification of small in-frame indels. *BMC Med Genomics*. 2023;16(1):36. doi 10.1186/s12920-023-01454-6

Chao J.R., Chattaraj P., Munjal T., Honda K., King K.A., Zalewski C.K., Chien W.W., Brewer C.C., Griffith A.J. *SLC26A4*-linked CEVA haplotype correlates with phenotype in patients with enlargement of the vestibular aqueduct. *BMC Med Genet*. 2019;20(1):118. doi 10.1186/s12881-019-0853-4

Chattaraj P., Reimold F.R., Muskett J.A., Shmukler B.E., Chien W.W., Madeo A.C., Pryor S.P., Zalewski C.K., Butman J.A., Brewer C.C., Kenna M.A., Alper S.L., Griffith A.J. Use of *SLC26A4* mutation testing for unilateral enlargement of the vestibular aqueduct. *JAMA Otolaryngol Head Neck Surg*. 2013;139(9):907-913. doi 10.1001/jamaoto.2013.4185

Chattaraj P., Munjal T., Honda K., Rendtorff N.D., Ratay J.S., Muskett J.A., Risso D.S., Roux I., Gertz E.M., Schäffer A.A., Friedman T.B., Morell R.J., Tranebjærg L., Griffith A.J. A common *SLC26A4*-linked haplotype underlying non-syndromic hearing loss with enlargement of the vestibular aqueduct. *J Med Genet*. 2017;54(10):665-673. doi 10.1136/jmedgenet-2017-104721

Choi B.Y., Madeo A.C., King K.A., Zalewski C.K., Pryor S.P., Muskett J.A., Nance W.E., Butman J.A., Brewer C.C., Griffith A.J. Segregation of enlarged vestibular aqueducts in families with non-diagnostic *SLC26A4* genotypes. *J Med Genet*. 2009;46(12):856-861. doi 10.1136/jmg.2009.067892

Choi Y., Sims G.E., Murphy S., Miller J.R., Chan A.P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One*. 2012;7(10):e46688. doi 10.1371/journal.pone.0046688

Choi Y., Chan A.P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015;31(16):2745-2747. doi 10.1093/bioinformatics/btv195

Cirello V., Giorgini V., Castronovo C., Marelli S., Mainini E., Sironi A., Recalcati M.P., Pessina M., Giardino D., Larizza L., Persani L., Finelli P., Russo S., Fugazzola L. Segmental maternal UPD of chromosome 7q in a patient with Pendred and Silver Russell syndromes-like features. *Front Genet*. 2018;9:600. doi 10.3389/fgene.2018.00600

Danilchenko V.Y., Zytsar M.V., Maslova E.A., Bady-Khoo M.S., Barashkov N.A., Morozov I.V., Bondar A.A., Posukh O.L. Different rates of the *SLC26A4*-related hearing loss in two indigenous peoples of Southern Siberia (Russia). *Diagnostics* (*Basel*). 2021;11(12):2378. doi 10.3390/diagnostics11122378

Del Castillo F.J., Del Castillo I. DFNB1 non-syndromic hearing impairment: Diversity of mutations and associated phenotypes. *Front Mol Neurosci*. 2017;10:428. doi 10.3389/fnmol.2017.00428

Douville C., Masica D.L., Stenson P.D., Cooper D.N., Gygax D.M., Kim R., Ryan M., Karchin R. Assessing the pathogenicity of insertion and deletion variants with the variant effect scoring tool (VEST-indel). *Hum Mutat*. 2015;37(1):28-35. doi 10.1002/humu.22911

Du W., Guo Y., Wang C., Wang Y., Liu X. A systematic review and meta-analysis of common mutations of *SLC26A4* gene in Asian populations. *Int J Pediatr Otorhinolaryngol*. 2013;77(10):1670-1676. doi 10.1016/j.ijporl.2013.07.023

Everett L.A., Glaser B., Beck J.C., Idol J.R., Buchs A., Heyman M., Adawi F., Hazani E., Nassir E., Baxevanis A.D., Sheffield V.C., Green E.D. Pendred syndrome is caused by mutations in a putative sulphate transporter gene (PDS). *Nat Genet*. 1997;17(4):411-422. doi 10.1038/ng1297-411

Griffith A.J., Wangemann P. Hearing loss associated with enlargement of the vestibular aqueduct: mechanistic insights from clinical phenotypes, genotypes, and mouse models. *Hear Res*. 2011;281(1-2):11-17. doi 10.1016/j.heares.2011.05.009

Guex N., Peitsch M.C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997;18(15):2714-2723. doi 10.1002/elps.1150181505

Honda K., Griffith A.J. Genetic architecture and phenotypic landscape of *SLC26A4*-related hearing loss. *Hum Genet*. 2022;141(3-4):455-464. doi 10.1007/s00439-021-02311-1

Hu H., Wu L., Feng Y., Pan Q., Long Z., Li J., Dai H., Xia K., Liang D., Niikawa N., Xia J. Molecular analysis of hearing loss associated with enlarged vestibular aqueduct in the mainland Chinese: a unique *SLC26A4* mutation spectrum. *J Hum Genet*. 2007;52(6):492-497. doi 10.1007/s10038-007-0139-0

Hu J., Ng P.C. SIFT Indel: predictions for the functional effects of amino acid insertions/deletions in proteins. *PLoS One*. 2013;8(10):e77940. doi 10.1371/journal.pone.0077940

Koohiyan M. A systematic review of *SLC26A4* mutations causing hearing loss in the Iranian population. *Int J Pediatr Otorhinolaryngol*. 2019;125:1-5. doi 10.1016/j.ijporl.2019.06.012

Lee Y.H., Tsai C.Y., Lu Y.S., Lin P.H., Chiang Y.T., Yang T.H., Hsu J.S., Hsu C.J., Chen P.L., Liu T.C., Wu C.C. Revisiting genetic epidemiology with a refined targeted gene panel for hereditary hearing impairment in the Taiwanese population. *Genes* (*Basel*). 2023;14(4):880. doi 10.3390/genes14040880

Lin Y.H., Wu C.C., Lin Y.H., Lu Y.C., Chen C.S., Liu T.C., Chen P.L., Hsu C.J. Targeted next-generation sequencing facilitates genetic diagnosis and provides novel pathogenetic insights into deafness with enlarged vestibular aqueduct. *J Mol Diagn*. 2019;21(1):138-148. doi 10.1016/j.jmoldx.2018.08.007

Liu Q., Zhang X., Huang H., Chen Y., Wang F., Hao A., Zhan W., … Li G.L., Zhang W., Shu Y., Sun L., Chen Z. Asymmetric pendrin homodimer reveals its molecular mechanism as anion exchanger. *Nat Commun*. 2023;14(1):3012. doi 10.1038/s41467-023-38303-0

Mey K., Muhamad A.A., Tranebjaerg L., Rendtorff N.D., Rasmussen S.H., Bille M., Cayé-Thomasen P. Association of *SLC26A4* mutations, morphology, and hearing in pendred syndrome and NSEVA. *Laryngoscope*. 2019;129(11):2574-2579. doi 10.1002/lary.27319

Mount D.B., Romero M.F. The SLC26 gene family of multifunctional anion exchangers. *Pflugers Arch*. 2004;447(5):710-721. doi 10.1007/s00424-003-1090-3

Pagel K.A., Antaki D., Lian A., Mort M., Cooper D.N., Sebat J., Iakoucheva L.M., Mooney S.D., Radivojac P. Pathogenicity and functional impact of non-frameshifting insertion/deletion variation in the human genome. *PLoS Comput Biol*. 2019;15(6):e1007112. doi 10.1371/journal.pcbi.1007112

Pang X., Chai Y., He L., Chen P., Wang X., Li L., Jia H., Wu H., Yang T. A 7666-bp genomic deletion is frequent in Chinese Han deaf patients with non-syndromic enlarged vestibular aqueduct but without biallelic *SLC26A4* mutations. *Int J Pediatr Otorhinolaryngol*. 2015;79(12):2248-2252. doi 10.1016/j.ijporl.2015.10.015

Park H.J., Shaukat S., Liu X.Z., Hahn S.H., Naz S., Ghosh M., Kim H.N., … Usami S.I., Nance W.E., Wilcox E.R., Riazuddin S., Griffith A.J. Origins and frequencies of *SLC26A4* (PDS) mutations in east and south Asians: global implications for the epidemiology of deafness. *J Med Genet*. 2003;40(4):242-248. doi 10.1136/jmg.40.4.242

Pera A., Villamar M., Viñuela A., Gandía M., Medà C., Moreno F., Hernández-Chico C. A mutational analysis of the *SLC26A4* gene in Spanish hearing-impaired families provides new insights into the genetic causes of Pendred syndrome and DFNB4 hearing loss. *Eur J Hum Genet*. 2008;16(8):888-896. doi 10.1038/ejhg.2008.30

Pique L.M., Brennan M.L., Davidson C.J., Schaefer F., Greinwald J. Jr., Schrijver I. Mutation analysis of the *SLC26A4*, *FOXI1* and *KCNJ10* genes in individuals with congenital hearing loss. *PeerJ*. 2014;2:e384. doi 10.7717/peerj.384

Rentas S., Abou Tayoun A. Utility of droplet digital PCR and NGS-based CNV clinical assays in hearing loss diagnostics: current status and future prospects. *Expert Rev Mol Diagn*. 2021;21(2):213-221. doi 10.1080/14737159.2021.1887731

Richards S., Aziz N., Bale S., Bick D., Das S., Gastier-Foster J., Grody W.W., Hegde M., Lyon E., Spector E., Voelkerding K., Rehm H.L.; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. doi 10.1038/gim.2015.30

Roesch S., Rasp G., Sarikas A., Dossena S. Genetic determinants of non-syndromic enlarged vestibular aqueduct: A review. *Audiol Res*. 2021;11(3):423-442. doi 10.3390/audiolres11030040

Rose J., Muskett J.A., King K.A., Zalewski C.K., Chattaraj P., Butman J.A., Kenna M.A., Chien W.W., Brewer C.C., Griffith A.J. Hearing loss associated with enlarged vestibular aqueduct and zero or one mutant allele of *SLC26A4*. *Laryngoscope*. 2017;127(7):E238-E243. doi 10.1002/lary.26418

Royer-Bertrand B., Cisarova K., Niel-Butschi F., Mittaz-Crettol L., Fodstad H., Superti-Furga A. CNV detection from exome sequencing data in routine diagnostics of rare genetic disorders: opportunities and limitations. *Genes* (*Basel*). 2021;12(9):1427. doi 10.3390/genes12091427

Schouten J.P., McElgunn C.J., Waaijer R., Zwijnenburg D., Diepvens F., Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res*. 2002;30(12):e57. doi 10.1093/nar/gnf056

Shearer A.E., Kolbe D.L., Azaiez H., Sloan C.M., Frees K.L., Weaver A.E., Clark E.T., Nishimura C.J., Black-Ziegelbein E.A., Smith R.J. Copy number variants are a common cause of non-syndromic hearing loss. *Genome Med*. 2014;6(5):37. doi 10.1186/gm554

Smits J.J., de Bruijn S.E., Lanting C.P., Oostrik J., O'Gorman L., Mantere T.; DOOFNL Consortium; … Hoischen A., Pegge S.A.H., Neveling K., Pennings R.J.E., Kremer H. Exploring the missing heritability in subjects with hearing loss, enlarged vestibular aqueducts, and a single or no pathogenic *SLC26A4* variant. *Hum Genet*. 2022;141(3-4):465-484. doi 10.1007/s00439-021-02336-6

Stankiewicz P., Lupski J.R. Structural variation in the human genome and its role in disease. *Annu Rev Med*. 2010;61:437-455. doi 10.1146/annurev-med-100708-204735

Steinhaus R., Proft S., Schuelke M., Cooper D.N., Schwarz J.M., Seelow D. MutationTaster2021. *Nucleic Acids Res*. 2021;49(W1):W446-W451. doi 10.1093/nar/gkab266

Stuppia L., Antonucci I., Palka G., Gatta V. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int J Mol Sci*. 2012;13(3):3245-3276. doi 10.3390/ijms13033245

Takahashi S., Homma K. The molecular principles underlying diverse functions of the SLC26 family of proteins. *J Biol Chem*. 2024;300(5):107261. doi 10.1016/j.jbc.2024.107261

Tesolin P., Fiorino S., Lenarduzzi S., Rubinato E., Cattaruzzi E., Ammar L., Castro V., Orzan E., Granata C., Dell'Orco D., Morgan A., Girotto G. Pendred syndrome, or not Pendred syndrome? That is the question. *Genes* (*Basel*). 2021;12(10):1569. doi 10.3390/genes12101569

Tian Y., Liu M., Lu Y., Zhao X., Yan Z., Sun Y., Ma J., Tang W., Wang H., Xu H. Exonic deletions and deep intronic variants of the *SLC26A4* gene contribute to the genetic diagnosis of unsolved patients with enlarged vestibular aqueduct. *Hum Mutat*. 2024;2024:8444122. doi 10.1155/2024/8444122

Tsukada K., Nishio S.Y., Hattori M., Usami S. Ethnic-specific spectrum of *GJB2* and *SLC26A4* mutations: their origin and a literature review. *Ann Otol Rhinol Laryngol*. 2015;124(Suppl. 1):61S-76S. doi 10.1177/0003489415575060

Varga R.E., Mumtaz R., Jahic A., Rudenskaya G.E., Sánchez-Ferrero E., Auer-Grumbach M., Hübner C.A., Beetz C. MLPA-based evidence for sequence gain: pitfalls in confirmation and necessity for exclusion of false positives. *Anal Biochem*. 2012;421(2):799-801. doi 10.1016/j.ab.2011.12.002

Walter J.D., Sawicka M., Dutzler R. Cryo-EM structures and functional characterization of murine *Slc26a9* reveal mechanism of uncoupled chloride transport. *eLife*. 2019;8:e46986. doi 10.7554/eLife.46986

Wang L., Hoang A., Gil-Iturbe E., Laganowsky A., Quick M., Zhou M. Mechanism of anion exchange and small-molecule inhibition of pendrin. *Nat Commun*. 2024;15(1):346. doi 10.1038/s41467-023-44612-1

Wangemann P., Nakaya K., Wu T., Maganti R.J., Itza E.M., Sanneman J.D., Harbidge D.G., Billings S., Marcus D.C. Loss of cochlear HCO_3^- secretion causes deafness via endolymphatic acidification

and inhibition of Ca^{2+} reabsorption in a Pendred syndrome mouse model. *Am J Physiol Renal Physiol*. 2007;292(5):F1345-F1353. doi 10.1152/ajprenal.00487.2006

Yang T., Vidarsson H., Rodrigo-Blomqvist S., Rosengren S.S., Enerback S., Smith R.J. Transcriptional control of *SLC26A4* is involved in Pendred syndrome and nonsyndromic enlargement of vestibular aqueduct (*DFNB4*). *Am J Hum Genet*. 2007;80(6):1055-1063. doi 10.1086/518314

Yang T., Gurrola J.G. II, Wu H., Chiu S.M., Wangemann P., Snyder P.M., Smith R.J. Mutations of *KCNJ10* together with mutations of *SLC26A4* cause digenic nonsyndromic hearing loss associated with enlarged vestibular aqueduct syndrome. *Am J Hum Genet*. 2009;84(5):651-657. doi 10.1016/j.ajhg.2009.04.014

Yuan Y., You Y., Huang D., Cui J., Wang Y., Wang Q., Yu F., Kang D., Yuan H., Han D., Dai P. Comprehensive molecular etiology analysis of nonsyndromic hearing impairment from typical areas in China. *J Transl Med*. 2009;7:79. doi 10.1186/1479-5876-7-79

Zhang F., Gu W., Hurles M.E., Lupski J.R. Copy number variation in human health, disease, and evolution. *Annu Rev Genomics Hum Genet*. 2009;10:451-481. doi 10.1146/annurev.genom.9.081307.164217

Zheng W., Zhang C., Li Y., Pearce R., Bell E.W., Zhang Y. Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. *Cell Rep Methods*. 2021;1(3):100014. doi 10.1016/j.crmeth.2021.100014

Zheng W., Wuyun Q., Li Y., Liu Q., Zhou X., Peng C., Zhu Y., Freddolino L., Zhang Y. Deep-learning-based single-domain and multidomain protein structure prediction with D-I-TASSER. *Nat Biotechnol*. 2026;44(4):641-653. doi 10.1038/s41587-025-02654-4

Zhou X., Zheng W., Li Y., Pearce R., Zhang C., Bell E.W., Zhang G., Zhang Y. I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. *Nat Protoc*. 2022;17(10):2326-2353. doi 10.1038/s41596-022-00728

Conflict of interest. The authors declare no conflict of interest.
Received May 27, 2026. Revised July 8, 2026. Accepted July 14, 2026.

MEDICAL GENETICS
Original article

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):824-829

doi 10.18699/vjgb-26-83

The *APOE* ε2 allele is associated with risk of neovascular age-related macular degeneration in a Russian cohort

O.S. Kozhevnikova [1], I.F. Nikulich [2, 3], A.S. Derbeneva [2, 3], V.A. Devyatkin [1], N.G. Kolosova [1], A.Zh. Fursova [1, 2, 3]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] State Novosibirsk Regional Clinical Hospital, Novosibirsk, Russia
[3] Novosibirsk State Medical University, Novosibirsk, Russia
anzhellafursova@yandex.ru

Abstract. Age-related macular degeneration (AMD) is a progressive disease that leads to blindness and is the cause of visual impairment in an aging population. Apolipoprotein E (ApoE) is one of the key proteins in lipoprotein metabolism. ApoE is expressed in the retina and is found in drusen in AMD patients. ApoE isoforms are significant genetic determinants of AMD risk (ε2 > ε3 > ε4), wherein the effects of *APOE* alleles vary by ethnic group. *APOE* genotype may also influence response to treatment. Here we examine the effect of *APOE* alleles on the risk of neovascular AMD and response to anti-VEGF treatment in patients from the population of Western Siberia. Genotyping of the rs429358 and rs7412 polymorphisms of the *APOE* gene was carried out in patients with neovascular AMD (n = 315) and the control group (n = 317) using allele-specific PCR. To assess the association of *APOE* alleles with treatment response, 275 patients were categorized into three groups of treatment response (good, weak, no response) after three loading injections of aflibercept. We revealed a significant association between the ε2 allele and neovascular AMD: carriers of ε2 had a 1.5-fold higher risk of AMD compared with carriers of other allelic variants (OR = 1.536; 95% CI: 1.075–2.195). The results indicate a moderate association of the ε2 allele with the development of neovascular AMD in the Russian population. However, no association was found between *APOE* genotype and short-term response to anti-VEGF therapy.
Key words: age-related macular degeneration; *APOE*; rs429358; rs7412; response to anti-VEGF

For citation: Kozhevnikova O.S., Nikulich I.F., Derbeneva A.S., Devyatkin V.A., Kolosova N.G., Fursova A.Zh. The *APOE* ε2 allele is associated with risk of neovascular age-related macular degeneration in a Russian cohort. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(5):824-829. doi 10.18699/vjgb-26-83

Funding. This work was supported by Russian Science Foundation grant No. 26-25-20004.

Аллель *APOE* ε2 ассоциирован в российской когорте с риском неоваскулярной возрастной макулярной дегенерации

О.С. Кожевникова [1], И.Ф. Никулич [2, 3], А.С. Дербенёва [2, 3], В.А. Девяткин [1], Н.Г. Колосова [1], А.Ж. Фурсова [1, 2, 3]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Государственная Новосибирская областная клиническая больница, Новосибирск, Россия
[3] Новосибирский государственный медицинский университет, Новосибирск, Россия
anzhellafursova@yandex.ru

Аннотация. Возрастная макулярная дегенерация (ВМД) – прогрессирующее нейродегенеративное заболевание сетчатки, которое становится основной причиной снижения и потери зрения людьми старшего возраста. Аполипопротеин Е (АпоЕ) – один из ключевых белков липидного метаболизма, который экспрессируется в сетчатке и обнаруживается в друзах у пациентов с ВМД. Изоформы АпоЕ являются значимыми генетическими детерминантами риска ВМД (ε2 > ε3 > ε4), при этом эффекты аллелей *APOE* различаются в зависимости от этнической группы. Генотип *APOE* также может влиять на ответ на лечение. В данном исследовании мы оценили влияние аллелей *APOE* на риск неоваскулярной ВМД и ответ на анти-VEGF-терапию у пациентов из населения Западной Сибири. Генотипирование полиморфизмов rs429358 и rs7412 гена *APOE* проведено у пациентов с неоваскулярной ВМД (n = 315) и в контрольной группе (n = 317) с использованием аллель-специфической ПЦР. Для оценки связи аллелей *APOE* с ответом на лечение 275 пациентов были разделены на три группы по степени ответа на лечение (хороший, слабый, отсутствие ответа) после трех нагрузочных инъекций афлиберцепта. Мы выявили значимую ассоциацию между аллелем ε2 и неоваскулярной ВМД: у носителей ε2 риск развития ВМД был в 1.5 раза выше по сравнению с носителями других аллельных вариантов (ОШ = 1.536; 95 % ДИ: 1.075–2.195). Результаты указывают на умеренную связь аллеля ε2 с развитием неоваскулярной ВМД в российской популяции. Однако не обнаружено ассоциации между генотипами *APOE* и краткосрочным ответом на анти-VEGF-терапию.
Ключевые слова: возрастная макулярная дегенерация; *APOE*; rs429358; rs7412; ответ на анти-VEGF-терапию

Introduction

Age-related macular degeneration (AMD) is a progressive multifactorial blinding disease, representing the leading cause of visual impairment in an aging population. AMD is characterized by atrophy of the retinal pigment epithelium (RPE), accumulation of drusen, inflammation, and photoreceptor loss in the macular region of the eye. Genetic susceptibility, along with age and lifestyle, constitute the primary risk factor of AMD, and can influence the disease phenotype and treatment outcome. Among multiple genes and/or single nucleotide polymorphisms (SNPs) associated with AMD, genome-wide association studies have identified lipid metabolism pathway genes, contributions of rare and common variants of which are highlighted in a large genome-wide association study (Fritsche et al., 2016; Fernández-Vega et al., 2020).

Accordingly, numerous data confirm the involvement of lipid metabolism violations in the pathogenesis of AMD including: (1) drusen, the hallmark histopathologic feature of AMD, are composed of lipids; (2) polymorphisms of genes involved in lipid homeostasis are associated with increased odds of AMD; (3) metabolomics studies show that patients with AMD have alterations in metabolites from lipid pathways, and (4) alterations in serum lipid profiles as a reflection of systemic dyslipidemia are associated with AMD (Lin et al., 2022). Apolipoprotein E (ApoE) is a transport protein of lipid and cholesterol, playing a pivotal role in lipid metabolism in the peripheral circulation, in the brain and in the re-innervation process following nervous system injury. ApoE has the ability to bind to membrane receptors, mediating lipid transfer between circulating lipoproteins and tissues. Additionally, ApoE nonspecifically binds to lipophilic inflammatory components, contributing to their clearance and participating in the innate immune response (Huebbe, Rimbach, 2017).

Three major isoforms are known to exist in the ApoE protein. ApoE2, ApoE3, and ApoE4 are the products of three alleles (ε2, ε3, and ε4) at a single gene locus defined by two SNPs (rs429358 and rs7412): ApoE3 has cysteine in position 112, and arginine, in position 158, ApoE2 has cysteine in both positions, and ApoE4 has arginine in both positions (Arg112, Cys158 configuration is extremely rare). There are six genotypes in the general population: three homozygous (*APOE* ε2/ε2, *APOE* ε3/ε3, and *APOE* ε4/ε4) and three heterozygous (*APOE* ε2/ε3, *APOE* ε2/ε4, and *APOE* ε3/ε4). The ε3 allele is the most common and is found in 77 % of the general population. Meanwhile, the ε2 allele is found in 8 % of the general population, and the ε4 allele is found in 15 % of the general population. However, the frequency of *APOE* alleles varies between different ethnic groups (Chan et al., 2023).

The *APOE* gene is a major genetic risk factor for development of both AMD and late-onset Alzheimer's disease (AD) – the leading cause of dementia in the elderly. Despite the different clinical manifestations of these diseases, growing evidence suggests that they share common pathways in their pathogenesis including inflammation, oxidative stress, and impaired autophagy. It is also worth noting that ApoE is a ubiquitous component of drusen, a hallmark of early AMD, and abundantly present in amyloid plaques in the AD brain (Wisniewski, Drummond, 2020). There are many data indicating that ApoE variants that have a protective association against AD are also associated with a simultaneous high risk of AMD. Therefore, it is known that the ε4 allele reduces the risk of AMD and at the same time is the major risk factor for AD. On the contrary, individuals with the ε2 allele have a higher risk of AMD but a reduced risk of AD (Liuska et al., 2023).

There are, however, conflicting results. Recently, Fernández-Vega et al. (2020) described a protective role the *APOE* ε2 allele has against wet AMD in the Northern Spanish population. A meta-analysis integrating several recent large-sample studies to verify the effect of *APOE* polymorphisms on AMD subtypes revealed a significantly protective role of ε4 on each subtype of AMD, but no supportive evidence of the association of ε2 with AMD (Xiying et al., 2017). On the contrary, according to the data of Gupta and Chawla (2025), the *APOE* ε4 allele may be a risk factor for AMD in the Indian population. At the same time in a large, longitudinal cohort of more than 100 000 Europeans, Rasmussen et al. (2023) revealed that the ε2 allele has an effect on AMD and confirmed an increased risk of AMD in people with this genotype. No evidence was received to support an association of *APOE* gene polymorphisms with early or exudative AMD in the Chinese (Sun et al., 2011) and Brazilian populations (Viturino et al., 2021). According to data of Adams et al. (2012), in the Australian population, associations of ε2ε2 and ε2ε4 with early AMD varied by smoking status.

These results demonstrate the importance of potential modifiers of association between *APOE* polymorphism and AMD and highlight its significant effects on this association that should be taken into account in further studies. Thus, the association of the ε2 allele with AMD is controversial, and their roles in different subtypes of the disease need to be properly explored in different populations for a better understanding of AMD pathophysiology. Our goal was to study the effect of *APOE* alleles on the risk of developing AMD and response to anti-VEGF treatment in patients of Russian ethnicity from the Western Siberian population.

Materials and methods

Study population and ethical approval. This study was conducted at the Ophthalmology Department of the Novosibirsk Regional Clinical Hospital (clinical assessment) and the Institute of Cytology and Genetics, SB RAS (genetic analysis). The study was performed in accordance with the tenets of the Declaration of Helsinki and was approved by the local Ethics Committee. All participants provided written informed consent.

The study included 315 patients diagnosed with neovascular age-related macular degeneration (nAMD) and a control group of 317 individuals without AMD. Inclusion criteria for the nAMD group were: age over 50 years, a diagnosis of nAMD with active macular neovascularization confirmed by the presence of subretinal and/or intraretinal fluid on optical coherence tomography (OCT), and provision of informed consent. A subgroup of 275 treatment-naïve patients was selected for the anti-VEGF therapy response analysis. Exclu-

Table 1. The frequency of genotypes and alleles of the *APOE* gene in the nAMD and control groups

Group	Genotypes						Alleles		
	ε2/ε2	ε2/ε3	ε2/ε4	ε3/ε3	ε3/ε4	ε4/ε4	ε2	ε3	ε4
nAMD									
n	8	60	7	199	38	3	83	496	51
%	2.94	22.06	2.57	73.16	13.97	1.10	0.13	0.79	0.08
Control									
n	4	43	6	225	36	3	57	529	48
%	1.59	17.13	2.39	89.64	14.34	1.20	0.09	0.83	0.08

sion criteria were: a history of laser coagulation, previous intravitreal pharmacotherapy, geographic atrophy, signs of intraocular inflammation, or any prior vitreoretinal surgery.

The control group consisted of individuals over 50 years of age with no clinical signs of AMD (e. g., soft or hard drusen, geographic atrophy, RPE detachment, macular fibrosis, or pigmentary abnormalities). Patients with senile cataract were included in the control group.

Clinical examination and anti-VEGF treatment protocol. Peripheral blood samples were collected from all participants in the morning after fasting into EDTA-containing vacuum tubes. Patients in the therapy subgroup (n = 275) received three loading doses of aflibercept (0.5 mg/0.05 mL) administered intravitreally at 4-week intervals. Injections were performed under topical epibulbar anesthesia (Alcaine) using a 27-gauge needle inserted at least 3 mm from the limbus. Treatment efficacy was assessed through clinical and instrumental examination before the initiation of therapy and after the third loading dose. This included: best-corrected visual acuity (BCVA) measured using an ETDRS (Early Treatment Diabetic Retinopathy Study) chart, central retinal thickness (CRT) measured by optical coherence tomography (OCT). The treatment response was classified by clinicians at the Novosibirsk Regional Clinical Hospital as optimal (good), partial (weak), or no response, based on the dynamics of BCVA and CRT changes, approximating the criteria proposed by Amoaku et al. (2015).

DNA extraction and *APOE* genotyping. Genomic DNA was isolated from whole blood using a diaGene (Russia) DNA extraction kit according to the manufacturer's protocol. DNA samples were diluted to a concentration of 8 ng/μL and stored at –20 °C. *APOE* genotyping for the rs429358 and rs7412 polymorphisms was performed using the allele-specific PCR with the TaqMan probes method described by Zhong et al. (2016). The genotyping assay comprised three separate reactions to identify the ε2, ε3, and ε4 alleles. Each 25 μL PCR reaction mixture contained: 12.5 μL of 2× BioMaster HS-qPCR mix (Biolabmix, Russia), containing HS-Taq DNA polymerase, dNTPs, PCR buffer, and Mg^{2+}, 0.5 μg of each forward and reverse *APOE* primer, *APOE* TaqMan probe, genomic DNA template, water to a final volume of 25 μL. The amplification protocol was as follows: initial polymerase activation at 95 °C for 5 min, 40 cycles of denaturation at 95 °C for 15 s, and annealing/extension at 64 °C for 1 min.

Statistical analysis was performed using StatTech v. 4.0.6 (StatTech LLC, Russia). The normality of quantitative data distribution was assessed using the Kolmogorov–Smirnov test. Data are presented as mean (M) ± standard deviation (SD) or median (Me) with interquartile range (Q1–Q3) and 95% confidence intervals (95% CI), as appropriate. Categorical data are described using absolute values and percentages. Comparisons of percentages in contingency tables were performed using Pearson's chi-square test. A p-value of < 0.05 was considered statistically significant.

Results

APOE genotype and allele distribution

Genotyping of the *APOE* gene polymorphisms (rs429358 and rs7412) was performed in 315 patients with neovascular age-related macular degeneration (nAMD) and 317 control subjects using the allele-specific PCR method described by Zhong et al. (2016). The genotype distributions for both polymorphisms were in Hardy–Weinberg equilibrium in all groups.

The frequencies of the resulting ApoE isoforms (ε2, ε3, ε4) are summarized in Table 1. The overall allelic and genotypic frequencies of the rs429358 and rs7412 variants did not differ significantly between the nAMD and control groups in a logistic regression analysis adjusted for age and sex.

However, a subsequent analysis focusing on the ApoE isoforms revealed a significant association. The frequency of the ε2 allele was significantly higher in nAMD patients compared to the control group (0.13 vs. 0.09, p = 0.018, likelihood ratio chi-square test). Carriers of the ε2 allele had a 1.5-fold increased risk of nAMD compared to carriers of other allelic variants (OR = 1.536; 95% CI: 1.075–2.195) (Table 2). These results indicate a moderate association of the *APOE* ε2 allele with nAMD development in the Russian population.

Table 2. Analysis of the association of carriage of the ε2 allele with the risk of AMD

Group	Allele *APOE*			Odds ratio	CI (95%)
	ε2	ε3	ε4		
nAMD, n = 315	83	547		1.536	(1.075–2.195)
Control, n = 317	57	577			

Association between *APOE* genotype and treatment response

The association between *APOE* genotype and variability in treatment response was assessed in a cohort of 275 nAMD patients (275 eyes). Patient clinical characteristics are presented in Table 3. The majority of patients presented with type 1 (occult) macular neovascularization (52.4 %, $n = 144$), followed by type 2 (44.0 %, $n = 121$) and type 3 (3.6 %, $n = 10$).

Following three loading doses of anti-VEGF therapy, patients were categorized based on treatment response: 221 (80.4 %) had an optimal response, 41 (14.9 %) had a partial response, and 13 (4.7 %) were non-responders. No statistically significant associations were found between the type of response and patient sex or age ($p > 0.05$).

At diagnosis, the median central retinal thickness (CRT) was 294 μm (IQR: 238–399 μm). After three loading doses, 269 patients (97.8 %) showed a reduction in CRT, with a median decrease of 22 % to 236 μm (IQR: 213–267 μm). A positive dynamic in best-corrected visual acuity (BCVA) was observed in 167 eyes (60.7 %). The median BCVA improved from 32 letters (IQR: 30–50) at baseline to 60 letters (IQR: 35–75) in the optimal response group after loading doses. A significant increase in BCVA and a reduction in CRT were observed in the early treatment phase ($p < 0.05$). However, at baseline, there were no statistically significant differences in BCVA ($p = 0.306$) or CRT ($p = 0.928$) between patients who later demonstrated an optimal response and those with a partial or absent response.

We hypothesized that *APOE* genotype might influence interindividual variability in response to therapy. However, Pearson's chi-square analysis showed no significant association between the studied *APOE* alleles or genotypes and the type of response to anti-VEGF therapy (Table 4).

Table 3. Clinical characteristics of patients

Parameter	Clinical characteristics
Number of patients, *n*	275
Age, years	72 [64; 78]
Gender, *n* (%)	
Women	194 (70.5)
Men	81 (29.5)
MNV type, *n* (%)	
MNV 1	144 (52.4)
MNV 2	121 (44.0)
MNV 3	10 (3.6)
BCVA, letters	50 [30; 62]
BCVA after 3 IVI, letters	60 [35; 70]
BCVA gain after 3 IVI	
Positive	167 (60.7)
No gain	83 (30.2)
Negative	25 (9.1)
CRT initial, mkm	303 [264; 366]
CRT after treatment, mkm	240 [215; 281]
CRT change after 3-IVI, *n* (%)	
Increase	6 (2.2)
Decrease	269 (97.8)
Type of response to treatment, *n* (%)	
No response	13 (4.7)
Weak response	41 (14.9)
Good response	221 (80.4)

Discussion

This is the first study to investigate the association of *APOE* ε2, ε3, and ε4 alleles with both the risk of AMD and the response to anti-VEGF therapy in a Russian population from Western Siberia. Our findings confirm a significant association between the *APOE* ε2 allele and an increased risk of developing neovascular AMD. However, we found no statistically significant link between *APOE* polymorphisms and the short-term anatomical and functional response to anti-VEGF treatment after the loading phase. Previously, the ApoE ε2/ε3/ε4 isoforms have been reported to be associated with the risk of developing AMD: in several studies, the ε4 allele was associated with a decreased risk, and the ε2 allele, with an increased risk (Adams et al., 2012).

Apolipoprotein E is a key protein in lipoprotein metabolism. It exists in three major isoforms, defined by amino acid substitutions at positions Cys112Arg and Arg158Cys, which are determined by two single-nucleotide polymorphisms (rs429358

Table 4. Distribution of *APOE* genotypes and alleles in groups with different therapeutic outcomes of anti-VEGF therapy

Parameter		Type of response to treatment			*p*-value
		No response	Weak	Good	
APOE genotypes, *n* (%)	ε2/ε2	0 (0.0)	0 (0.0)	7 (100.0)	0.326
	ε2/ε3	1 (2.0)	8 (15.7)	42 (82.4)	
	ε2/ε4	1 (12.5)	3 (37.5)	4 (50.0)	
	ε3/ε3	8 (4.7)	28 (16.3)	136 (79.1)	
	ε3/ε4	3 (8.8)	2 (5.9)	29 (85.3)	
	ε4/ε4	0 (0.0)	0 (0.0)	3 (100.0)	
APOE alleles, *n* (%)	ε2	2 (15.4)	11 (26.8)	53 (24.0)	0.392
	ε3	8 (61.5)	28 (68.3)	136 (61.5)	
	ε4	3 (23.1)	2 (4.9)	32 (14.5)	

and rs7412). These ApoE isoforms are encoded by three alleles: ε2 (rs429358*T/rs7412*T), ε3 (rs429358*T/rs7412*C), and ε4 (rs429358*C/rs7412*C). ApoE is expressed in the retina and has been identified within drusen in AMD patients (Abyadeh et al., 2023). In numerous population-based studies, the *APOE* ε4 variant has been associated with a protective effect against AMD, whereas the ε2 allele has been linked to an increased risk of the disease (Hu et al., 2021).

The pivotal role of anti-VEGF agents in managing neovascular AMD is undeniable, having revolutionized the prevention of severe vision loss. Nonetheless, a subset of patients, approximately 10–30 % (Fursova et al., 2023), exhibit a suboptimal response and continue to experience vision deterioration. The underlying causes for this treatment resistance remain unclear, with genetic predisposition considered a potential determinant (Kozhevnikova et al., 2023).

Given the high prevalence of AMD and the substantial socio-economic burden associated with long-term anti-VEGF therapy, the ability to predict treatment outcomes is of paramount clinical importance. We hypothesized that *APOE* genotype could influence the considerable individual variability in treatment response. This premise is supported by several international studies investigating the association between ApoE and AMD biomarkers (Gupta, Chawla, 2025). However, an analysis of the relationship between *APOE* genotype, the risk of neovascular AMD, and the response to anti-VEGF therapy had not been previously conducted within a Russian population.

We focused on the short-term response following the initial loading doses of anti-VEGF therapy, as this period is considered a critical window for assessing the maximum potential morphological and functional response. Analyzing this phase was hypothesized to best reveal the extent of genetic influence on subsequent long-term outcomes. While some studies have demonstrated that the most significant improvements occur within the first three months of treatment, our investigation in the West Siberian cohort did not identify *APOE* genotype as a significant predictor of outcome at this stage. This lack of association concerning the treatment response contrasts with findings from other studies. For instance, research by Wickremasinghe et al. (2011) and Bakbak et al. (2016) reported that the *APOE* ε4 allele was linked to better visual outcomes following anti-VEGF treatment. The disparity between our results and those in the literature could be attributed to several factors. The ethnic variation in *APOE* allele and genotype distribution is well-documented, suggesting that the relationship between *APOE* and AMD pathophysiology may be population-specific (Ramadan et al., 2019).

Furthermore, the influence of ApoE might become apparent only over a longer observation period or under different treatment regimens, such as personalized dosing protocols, as suggested by other authors. Another critical consideration is the treatment initiation timeline. A significant delay often exists between diagnosis and the start of therapy, which can drastically diminish the potential for visual improvement and might obscure underlying pharmacogenetic associations (Kozhevnikova et al., 2023).

The biological plausibility of ApoE's role in AMD is supported by its involvement in lipid metabolism. ApoE is detected in drusen, the earliest hallmarks of AMD, and is crucial for lipid transport across cell membranes, being highly expressed by the retinal pigment epithelium (RPE) (Anderson et al., 2001). Evidence from murine models indicates that the ε2 allele is associated with age-related accumulation of subretinal phagocytes and exacerbated choroidal neovascularization, while the ε4 allele shows protective effects (Levy et al., 2015). This aligns with population studies where the ε4 allele often confers a protective effect against AMD development, and ε2 increases risk, a pattern consistent with our findings regarding disease susceptibility (Adams et al., 2012).

APOE polymorphisms can influence treatment outcomes in AMD patients. In patients with neovascular AMD, the presence of the *APOE* ε4 allele conferred significantly better visual outcomes after anti-VEGF treatment (Wickremasinghe et al., 2011). The results of a study by Bakbak et al. (2016) also showed improvement of visual activity in AMD patients carrying the ε4 allele and indicated an association between the ε4 allele and visual improvement in AMD patients using ranibizumab, whereas no association with treatment was observed for the ε2 allele.

Conclusion

Our study revealed a statistically significant association between the ε2 allele of *APOE* and the development of neovascular AMD in the Russian population of Western Siberia. No association was found between *APOE* genotype and short-term response to anti-VEGF therapy (after three intravitreal injections). Considering the rising prevalence of AMD, further pharmacogenetic studies are essential to elucidate potential genetic markers that could guide personalized treatment strategies.

References

Abyadeh M., Gupta V., Paulo J.A., Sheriff S., Shadfar S., Fitzhenry M., Amirkhani A., Gupta V., Salekdeh G.H., Haynes P.A., Graham S.L., Mirzaei M. Apolipoprotein ε in brain and retinal neurodegenerative diseases. *Aging Dis*. 2023;14(4):1311-1330. doi 10.14336/AD.2023.0312-1

Adams M.K., Simpson J.A., Richardson A.J., English D.R., Aung K.Z., Makeyeva G.A., Guymer R.H., Giles G.G., Hopper J., Robman L.D., Baird P.N. Apolipoprotein E gene associations in age-related macular degeneration: the Melbourne Collaborative Cohort Study. *Am J Epidemiol*. 2012;175(6):511-518. doi 10.1093/aje/kwr329

Amoaku W.M., Chakravarthy U., Gale R., Gavin M., Ghanchi F., Gibson J., Harding S., Johnston R.L., Kelly S.P., Lotery A., Mahmood S., Menon G., Sivaprasad S., Talks J., Tufail A., Yang Y. Defining response to anti-VEGF therapies in neovascular AMD. *Eye* (*Lond.*). 2015;29(6):721-731. doi 10.1038/eye.2015.48

Anderson D.H., Ozaki S., Nealon M., Neitz J., Mullins R.F., Hageman G.S., Johnson L.V. Local cellular sources of apolipoprotein E in the human retina and retinal pigmented epithelium: implications for the process of drusen formation. *Am J Ophthalmol*. 2001;131(6):767-781. doi 10.1016/s0002-9394(00)00961-2

Bakbak B., Ozturk B.T., Zamani A.G., Gonul S., Iyit N., Gedik S., Yıldırım M.S. Association of apolipoprotein E polymorphism with intravitreal ranibizumab treatment outcomes in age-related macular degeneration. *Curr Eye Res*. 2016;41(6):862-866. doi 10.3109/02713683.2015.1067325

Chan M.L., Meyer O.L., Farias S.T., Whitmer R.A., Rajan K., Olichney J., Johnson D., Mungas D. APOE effects on late life cognitive trajectories in diverse racial/ethnic groups. *J Int Neuropsychol Soc*. 2023;29(2):126-135. doi 10.1017/S1355617722000030

Fernández-Vega B., García M., Olivares L., Álvarez L., González-Fernández A., Artime E., Fernández-Vega Cueto A., Cobo T., Coca-Prados M., Vega J.A., González-Iglesias H. The association study of lipid metabolism gene polymorphisms with AMD identifies a protective role for APOE-E2 allele in the wet form in a Northern Spanish population. *Acta Ophthalmol*. 2020;98(3):e282-e291. doi 10.1111/aos.14280

Fritsche L.G., Igl W., Bailey J.N., Grassmann F., Sengupta S., Bragg-Gresham J.L., Burdon K.P., … Lee K.E., Hunter D.J., Sardell R.J., Mitchell P., Heid I.M. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet*. 2016;48(2):134-143. doi 10.1038/ng.3448

Fursova A.Zh., Derbeneva A.S., Vasilyeva M.A., Nikulich I.F., Tarasov M.S., Gamza Yu.A., Chubar N.V., Gusarevich O.G., Dmitrieva E.I., Telegina D.V., Kozhevnikova O.S. Current data on the age-related macular degeneration pathophysiology: focus on growth factors and neurotrophins. *Adv Gerontol*. 2023;13:105-115. doi 10.1134/S2079057024600162

Gupta D., Chawla S. Association of apolipoprotein E (APOE) gene polymorphism with age-related macular degeneration (AMD) in Indian patients. *Cureus*. 2025;17(6):e85441. doi 10.7759/cureus.85441

Hu M.L., Quinn J., Xue K. Interactions between apolipoprotein E metabolism and retinal inflammation in age-related macular degeneration. *Life* (*Basel*). 2021;11(7):635. doi 10.3390/life11070635

Huebbe P., Rimbach G. Evolution of human apolipoprotein E (APOE) isoforms: gene structure, protein function and interaction with dietary factors. *Ageing Res Rev*. 2017;37:146-161. doi 10.1016/j.arr.2017.06.002

Kozhevnikova O.S., Fursova A.Z., Derbeneva A.S., Nikulich I.F., Devyatkin V.A., Kolosova N.G. Pharmacogenetic association between allelic variants of the autophagy-related genes and anti-vascular endothelial growth factor treatment response in neovascular age-related macular degeneration. *Biomedicines*. 2023;11(11):3079. doi 10.3390/biomedicines11113079

Levy O., Lavalette S., Hu S.J., Housset M., Raoul W., Eandi C., Sahel J.A., Sullivan P.M., Guillonneau X., Sennlaub F. APOE isoforms control pathogenic subretinal inflammation in age-related macular degeneration. *J Neurosci*. 2015;35(40):13568-13576. doi 10.1523/JNEUROSCI.2468-15.2015

Lin J.B., Halawa O.A., Husain D., Miller J.W., Vavvas D.G. Dyslipidemia in age-related macular degeneration. *Eye* (*Lond*). 2022;36(2):312-318. doi 10.1038/s41433-021-01780-y

Liuska P.J., Rämö J.T., Lemmelä S., Kaarniranta K., Uusitalo H., Lahtela E., Daly M.J., Harju M., Palotie A., Turunen J.A.; for the FinnGen Study†. Association of *APOE* haplotypes with common age-related ocular diseases in 412,171 individuals. *Invest Ophthalmol Vis Sci*. 2023;64(14):33. doi 10.1167/iovs.64.14.33

Ramadan A., Foda B.M., Sabry N.A., Refaat O., Saleh A.A., Fawzy A. Association analysis of ApoE gene polymorphisms among Egyptian patients with Alzheimer' disease. *Gene Rep*. 2019;14:65-71. doi.org/10.1016/j.genrep.2018.10.017

Rasmussen K.L., Tybjærg-Hansen A., Nordestgaard B.G., Frikke-Schmidt R. Associations of Alzheimer disease-protective APOE variants with age-related macular degeneration. *JAMA Ophthalmol*. 2023;141(1):13-21. doi 10.1001/jamaophthalmol.2022.4602

Sun E., Lim A., Liu X., Snellingen T., Wang N., Liu N. Apolipoprotein E gene and age-related macular degeneration in a Chinese population. *Mol Vis*. 2011;17:997-1002

Viturino M.G., Neto J.M., Bajano F.F., Costa S.M., Roque A.B., Borges G.F., Ananina G., Rim P.H., Medina F.M., Costa F.F., Vasconcellos J.P., Melo M.B. Evaluation of *APOE* polymorphisms and the risk for age-related macular degeneration in a Southeastern Brazilian population. *Exp Biol Med* (*Maywood*). 2021;246(10):1148-1155. doi 10.1177/1535370220985466

Wickremasinghe S.S., Xie J., Lim J., Chauhan D.S., Robman L., Richardson A.J., Hageman G., Baird P.N., Guymer R. Variants in the *APOE* gene are associated with improved outcome after anti-VEGF treatment for neovascular AMD. *Invest Ophthalmol Vis Sci*. 2011;52(7):4072-4079. doi 10.1167/iovs.10-6550

Wisniewski T., Drummond E. APOE-amyloid interaction: therapeutic targets. *Neurobiol Dis*. 2020;138:104784. doi 10.1016/j.nbd.2020.104784

Xiying M., Wenbo W., Wangyi F., Qinghuai L. Association of apolipoprotein E polymorphisms with age-related macular degeneration subtypes: an updated systematic review and meta-analysis. *Arch Med Res*. 2017;48(4):370-377. doi 10.1016/j.arcmed.2017.08.002

Zhong L., Xie Y.Z., Cao T.T., Wang Z., Wang T., Li X., Shen R.C., Xu H., Bu G., Chen X.F. A rapid and cost-effective method for genotyping apolipoprotein E gene polymorphism. *Mol Neurodegener*. 2016;11:2. doi 10.1186/s13024-016-0069-4

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

Received May 13, 2026. Revised June 14, 2026. Accepted June 25, 2026.

Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2026;30(5):830-838

doi 10.18699/vjgb-26-84

Identification of potential target genes of the trematode transport RNA-derived small RNAs among differentially expressed genes of human monocytes

E.A. Lishai [1, 2], M.Y. Pakharukova [1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
lishai.ekaterina@gmail.com

Abstract. RNA interference is an evolutionarily conserved mechanism of post-transcriptional gene expression regulation present in all living organisms. It is highly relevant to multiple areas of molecular medicine, including diagnostics and novel immunomodulatory approaches. Recent studies have increasingly highlighted the role of extracellular vesicles and their cargo in parasite-host interactions. The trematode *Opisthorchis felineus*, which parasitizes the human biliary tract, causes opisthorchiasis, a disease associated with chronic inflammation and other hepatobiliary injuries. The long-term survival of the parasite is likely linked to modulation of host immune response. Nevertheless, the mechanism by which liver fluke vesicles affect human immune cells remains unclear. The aim of this study was to identify potential human gene targets of tRNA-derived small RNAs and to assess their enrichment among differentially expressed genes (DEGs) in the transcriptome of THP-1 human monocytes. Target prediction for eight most abundant *O. felineus* tRNA fragments identified 1,299 potential target genes in human monocytes. Seven cDNA libraries prepared from THP-1 monocytes before and after treatment with vesicles from adult *O. felineus* were sequenced using paired-end 2 × 150 bp sequencing (DNBseq, BGI), yielding 43.8 million reads per library. Among 1484 DEGs, 159 predicted targets of tRNA-derived small RNAs showed significant expression changes upon vesicle treatment. These genes were associated with pathways involved in cell-cycle regulation and cell death, including DNA integrity control (M16357), G1/S transition (M2074), apoptosis (M15303), and programmed cell death (M27436), as well as immune-related pathways, including the monocyte pathway (M4956), mononuclear macrophage differentiation (M25144), MHC-I antigen presentation (M1066), and proteasome-mediated antigen processing (M1070). These findings suggests that *O. felineus* tRNA-derived small RNAs may contribute to the regulation of these processes in human monocytes.

Key words: opisthorchiasis; tRNA; exosome-like vesicles; RNA interference; transcriptomics; human monocytes

For citation: Lishai E.A., Pakharukova M.Y. Identification of potential target genes of the trematode transport RNA-derived small RNAs among differentially expressed genes of human monocytes. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2026;30(5):830-838. doi 10.18699/vjgb-26-84

Funding. This work was supported by the Russian Science Foundation (grant No. 24-44-00048).

Acknowledgements. Bioinformatic analyses were performed using the computational resources of the Bioinformatics Core Facility at the Institute of Cytology and Genetics SB RAS (project FWNR-2026-0027).

Выявление потенциальных генов-мишеней малых некодирующих фрагментов тРНК трематод среди дифференциально экспрессирующихся генов моноцитов человека

Е.А. Лишай [1, 2], М.Ю. Пахарукова [1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
lishai.ekaterina@gmail.com

Аннотация. РНК-интерференция – эволюционно консервативный механизм посттранскрипционной регуляции генов, представленный у всех видов живых организмов, в настоящее время актуален для многих направлений молекулярной медицины, включая диагностику и новые подходы к иммуномодуляции. В последнее время появляется все больше данных об участии внеклеточных везикул и их компонентов в контексте взаимодействия между паразитом и хозяином. Трематода *Opisthorchis felineus*, паразитирующая в желчных протоках человека, вызывает описторхоз, что сопровождается хроническим воспалением и рядом неблагоприятных последствий.

Длительное выживание паразита, вероятно, связано с иммуномодуляцией хозяина. Однако механизм влияния везикул трематод на моноциты человека неизвестен. Целью настоящей работы было найти потенциальные гены-мишени фрагментов тРНК трематод в геноме человека и проанализировать их представленность среди дифференциально экспрессирующихся генов (ДЭГ) в транскриптоме моноцитов человека линии THP-1. В результате поиска генов-мишеней для восьми наиболее представленных фрагментов тРНК *O. felineus* было выявлено 1299 потенциальных генов, которые могут быть мишенями в моноцитах человека. Семь кДНК библиотек клеточной культуры моноцитов человека линии THP-1 до и после обработки выделенными везикулами взрослой особи *O. felineus* были просеквенированы методом парных прочтений 2×150 пар оснований (технология DNBseq, BGI, Китай), получено около 43.8 млн прочтений на библиотеку. Среди 1484 ДЭГ в моноцитах было обнаружено 159 генов-мишеней фрагментов тРНК везикул, экспрессия которых значимо менялась при обработке экзосомо-подобных везикул (ЭПВ). Выявленные потенциальные гены-мишени относились к путям, связанным с регуляцией клеточного цикла и клеточной гибели (контроль целостности ДНК (M16357), переход из фазы G1 в фазу S (M2074), апоптоз (M15303) и программируемая клеточная гибель (M27436)), иммунного ответа (путь моноцитов (M4956), дифференцировка мононуклеарных макрофагов (M25144), презентация антигена через MHC-1 (M1066), процессинг антигена через протеасому (M1070)). Вероятно, фрагменты тРНК трематоды *O. felineus* могут вносить вклад в регуляцию этих процессов в моноцитах хозяина.

Ключевые слова: описторхоз; тРНК; экзосомо-подобные везикулы; РНК-интерференция; транскриптомика; моноциты человека

Introduction

RNA interference is an evolutionarily conserved and widely distributed mechanism of post-transcriptional gene regulation across living organisms. Increasing interest in this field of genetics and molecular biology reflects the growing potential of non-coding RNAs for regulating target gene expression, supporting diagnostics, and enabling novel immunomodulatory strategies. Many small non-coding RNAs that repress target gene expression at the post-transcriptional level are transported within extracellular vesicles. However, the contribution of small RNAs to interspecies communication, including parasite–host interactions, remains incompletely understood. A relevant model for studying parasite RNA–host interactions is the epidemiologically important trematode *Opisthorchis felineus* (Rivolta, 1884).

O. felineus parasitises the bile ducts of fish-eating mammals, including humans. Chronic infection is associated with persistent inflammation and biliary intraepithelial neoplasia (Xia et al., 2015; Mordvinov et al., 2021; Lishai et al., 2024). The long-term survival of trematodes within a host is likely due to their ability to manipulate the host's immune system. Consistent with this, helminth infections have been associated with attenuated manifestations of autoimmune diseases, such as Crohn's disease and asthma (Araújo et al., 2004; Saltykova et al., 2014).

Immune modulation, as well as parasite-driven effects on other host processes, including carcinogenesis, requires active communication between the parasite and the host. One mechanism that may mediate such communication involves extracellular vesicles that are internalised by host cells through distinct endocytic pathways. For example, human cholangiocytes internalise *O. felineus* exosome-like vesicles (ELVs) via clathrin-dependent endocytosis (Pakharukova et al, 2023). These ELVs carry diverse bioactive molecules, including small non-coding RNAs, such as tRNA-derived small RNAs.

tRNA-derived small RNAs (tsRNAs), generated from precursor or mature tRNAs, are increasingly recognised as a novel class of regulatory molecules implicated in various diseases (Prehn, Jirström, 2020; Chen et al., 2024; Strømme et al., 2024). tsRNAs have been detected in extracellular vesicles, including exosome-like vesicles secreted by multiple parasites, and may modulate host cell functions (Artuyants et al., 2020; Fontenla et al., 2022; Kusakisako et al., 2023).

Monocytes are key cells of the innate immune system and participate in the early response to invading pathogens (Passos et al, 2017). Parasite-derived extracellular vesicles have been shown to affect monocyte polarisation and cytokine production in the host (Silverman et al., 2010; Lertjuthaporn et al., 2022; Norouzi et al., 2022; Khowawisetsut et al., 2023). Nevertheless, the responses of human monocytes to ELVs released by *O. felineus* remain poorly characterised. Furthermore, it is unknown whether tsRNAs transported by *O. felineus* ELVs can directly influence human monocytes.

Therefore, this study aimed to investigate the effects of *O. felineus* ELVs (*in vivo*) and their associated tsRNAs (*in silico*) on human THP-1 monocytes. Specifically, the work included transcriptome sequencing of THP-1 monocytes following exposure to *O. felineus* ELVs and identification of predicted tsRNA target genes among the differentially expressed genes.

Materials and methods

Animals. Two-month-old golden hamsters (*Mesocricetus auratus*) provided by the Animal Facility of the Institute of Cytology and Genetics SB RAS (Novosibirsk, Russia). *O. felineus* metacercariae were isolated from ide (*Leuciscus idus*) caught in the Ob River near Novosibirsk, using previously described procedures (Pakharukova et al., 2018, 2019). Each hamster was infected orally with 75 metacercariae. Adult flukes were recovered from the bile ducts two months post-infection and washed 10 times in sterile physiological saline, as described previously. All animal experiments were approved by the Bioethics Committee of the Institute of Cytology and Genetics SB RAS (protocol No. 42, May 25, 2018).

Isolation of exosome-like vesicles. Exosome-like vesicles (ELVs) were isolated from the culture medium of 300 adult *O. felineus* flukes (RPMI-1640 medium supplemented with 1 % glucose, 100 U/mL penicillin G, 100 µg/mL streptomy-

O. felineus tRNA derived fragments and their relative abundance in *O. felineus* ELVs

tRNA fragment name	Sequence (5′–3′)	Abundance in *O. felineus* ELVs (%)
Asp-GTC-1	5′-TCCTCGGTGGTATAGTGGTGAGTATCTCCGCCTGTC-3′	13.82
Asp-GTC-2	5′-TCCTCGGTATAGTGGTGAGTATCTCCGCCTGTC-3′	7.16
Asp-GTC-3	5′-TCCTCGGGTATAGTGGTGAGTATCTCCGCCTGTC-3′	2.86
His-GTG-1	5′-GGCCGTCGTAGTCTAGTGGTTAGGACATTGCGTT-3′	2.04
Ile-AAT-1	5′-GGGTCTGTAGCTCAGTTGGTTAGAGCGCAC-3′	4.70
Ile-AAT-2	5′-GGGTCTGTAGCTCAGTTGGTTAGAGCGCACC-3′	3.45
Ile-AAT-3	5′-GGGTCTGTAGCTCAGTTGGTTAGAGCGCA-3′	3.91
Tyr-GTA-1	5′-GGGTCGTTAGCTCAGCTGGTAGAGCAGCG-3′	3.74

cin, and 0.25 μg/mL amphotericin B) as described previously (Pakharukova et al, 2023).

Prediction of target genes for *O. felineus* tRNA fragments. Previously characterised tRNA-derived small RNAs from *O. felineus* ELVs were used in this study (Lishai, Pakharukova, 2026). These tsRNAs were identified by sequencing small RNA libraries generated from *O. felineus* ELVs. Small RNA libraries were mapped to predicted tRNA genes in the *O. felineus* genome (assembly GCA_004794785.1) using tRNAscan-SE 2.0 (v.2.0.9) (Chan et al, 2021). Among 4,523 predicted tRNA loci, 4,276 pseudogenes were removed, leaving 247 sequences classified as bona fide tRNA genes. For each tRNA isotype, the most abundant tRNA sequence was selected. This yielded 50 representative tRNA sequences, which were then used as references for re-mapping the small RNA libraries to define the most abundant tsRNA fragments (see the Table).

Target prediction for tsRNAs was performed using tRFtarget 2.0 (Ningshan et al, 2024) (http://trftarget.net/), which combines RNAhybrid (Rehmsmeier et al, 2004) and IntaRNA (Mann et al, 2017) and reports binding sites predicted by both algorithms with a minimum free energy below –15 kcal/mol. Human 3′ untranslated regions (3′UTRs) of protein-coding genes were retrieved from Ensembl Genes 115 (Human genes, GRCh38.p14) (https://www.ensembl.org/biomart/martview/330951e98fb906b5999fa97914e6118c).

Transcripts shorter than seven nucleotides or with undefined sequence were excluded. The final dataset contained 196,902 transcripts (89 % of the initial 220,194) corresponding to 19,082 genes.

For functional enrichment analysis in human monocytes, genes not expressed in monocytes or blood cells were removed using expression data from the Human Protein Atlas (https://www.proteinatlas.org/humanproteome/single+cell/immune+cell/monocytes). The list of excluded genes is provided in Table S1[1]. Enrichment analyses were performed using the R packages msigdb v.1.18.0 (Bhuva et al, 2025) (Gene Ontology (GO) and MSigDB curated collections (MsigDB_Curated)) and clusterProfiler v.4.18.4 (Yu et al, 2012). Visualisations were generated using ggplot2 v.4.0.1 (https://ggplot2.tidyverse.org/), LibreOffice Calc v.24.2.7.2 (https://www.libreoffice.org/discover/calc/), and Inkscape v.1.4.2 (https://inkscape.org/ru/).

cDNA library preparation and RNA sequencing. Human THP-1 monocytes were cultured in RPMI-1640 medium (Gibco, USA) supplemented with 10 % fetal bovine serum (FBS; Gibco, USA), 100 U/mL penicillin (Sigma Aldrich, USA), 10 μg/mL streptomycin (Sigma Aldrich, USA), and 29.2 μg/mL L-glutamine (Gibco, USA). To generate macrophage-like cells, THP-1 monocytes were treated with 100 ng/ml phorbol 12-myristate 13-acetate (PMA; Solarbio, China) for 24 h, followed by a 24-h rest period in PMA-free medium. Total RNA was extracted from untreated cells and from cells exposed to *O. felineus* ELVs (10 μg/mL, 24 h) using the Aurum Total RNA Extraction Kit (Bio-Rad, USA), according to the manufacturer's protocol. RNA integrity and concentration were assessed using an Agilent 2100 Bioanalyzer with the Total RNA Nano Kit (Agilent Technologies, USA).

Seven cDNA libraries were prepared in total: four from untreated THP-1 cells and three from THP-1 cells treated with *O. felineus* ELVs. Libraries were constructed using the NEBNext Poly(A) mRNA Magnetic Isolation Module and the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs, USA), following the manufacturer's instructions. Size selection was performed using Agencourt AMPure XP beads (Beckman Coulter, USA). After PCR enrichment and quality control, libraries were sequenced on a DNBseq platform (BGI, China) using 2 × 150 bp paired-end reads. Library quality was evaluated using FastQC v.0.12.1 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequencing data are available in the NCBI BioProject repository under accession PRJNA1429255.

Bioinformatic analysis of transcriptome data. Reads were aligned to the human reference genome GRCh38.p14 (GCA_000001405.29) using STAR v.2.7.10a with default parameters (Dobin et al, 2013).

Differential gene expression between THP-1 cells treated with ELVs and untreated controls was analysed using the

[1] Supplementary Tables S1–S7 are available at: https://vavilov.elpub.ru/jour/manager/files/Suppl_Lishai_Engl_30_5.xlsx

Wald test implemented in DESeq2 v.1.50.2. (Love et al., 2014). Genes were considered differentially expressed if they showed an absolute fold change greater than 2 and a Benjamini–Hochberg-adjusted p-value < 0.05. Low-abundance genes with a total count <24 across all libraries were removed prior to analysis.

Variance-stabilising transformation was applied to the count matrix using the vst() function (blind = FALSE) in DESeq2. Principal component analysis (PCA) was then performed on the transformed expression matrix using the prcomp() function in the R stats package v.4.5.2.

Pathway enrichment analysis was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Molecular Signatures Database (MSigDB) collections. KEGG annotations were obtained via org.Hs.eg.db v.3.22.0, and enrichment analyses were carried out with clusterProfiler v.4.18.4 and msigdb v.1.18.0. Results were visualised using LibreOffice Calc v.24.2.7.2 and Inkscape v.1.4.2. Functional enrichment for differentially expressed genes that were also predicted tsRNA targets was performed using the same pipeline.

To assess whether tsRNA target genes were significantly enriched among differentially expressed genes in THP-1 monocytes, a Pearson chi-square test with Yates' continuity correction was applied using the chisq.test() function from the R stats package v.4.5.2.

Interaction networks between tsRNA fragments and differentially expressed genes were constructed and manually curated in Inkscape v.1.4.2.

Results

O. felineus exosome-like vesicles alter gene expression and activate immune- and cell cycle-related pathways in THP-1 monocytes

On average, 43.8 million paired-end reads were obtained per RNA-seq library. Mapping to the human reference genome yielded approximately 37 million uniquely mapped reads per library (84.63 %), while an additional 3.7 million reads per library (8.63 %) mapped to multiple genomic locations.

Principal component analysis revealed clear clustering of samples according to treatment status, separating ELV-treated from untreated THP-1 cells (Fig. 1*a*). A total of 1,484 genes were identified as differentially expressed (absolute fold change >1.5, adjusted p-value <0.05), including 392 downregulated and 1,092 upregulated genes (Fig. 1*b*; Table S2). Among the most strongly upregulated genes (log2 fold change >4.6; >25-fold increase) were *JCHAIN*, *LINC03063*, *CD8B2*, *MLC1*, and *RNASE2*. In contrast, the expression of *LOC124902900*, *HBA1*, *HBA2*, *LOC124900173*, and *LOC112268061* was downregulated more than 6.9-fold (log2 fold change <–2.78).

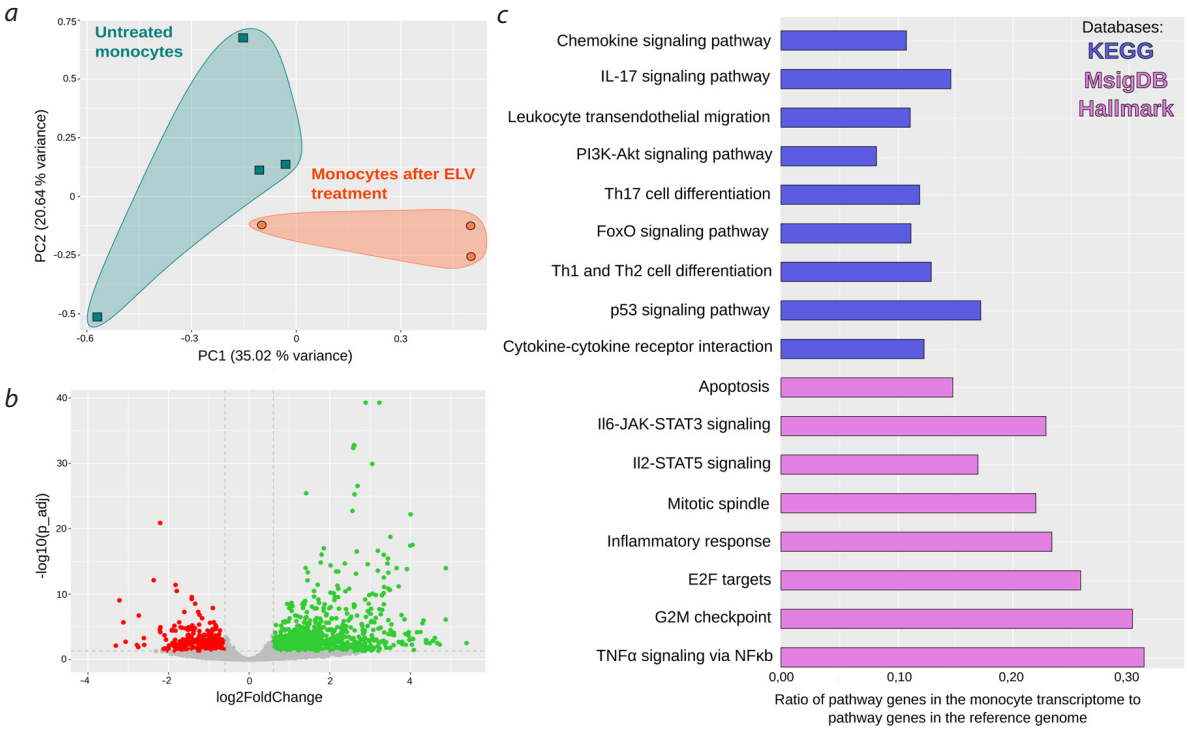


Fig. 1. Transcriptomic profile of THP-1 monocytes treated with *O. felineus* exosome-like vesicles.

a – principal component analysis (PCA) of RNA seq data showing separation of samples according to treatment with *O. felineus* exosome-like vesicles (ELVs) versus untreated THP-1 monocytes; *b* – Volcano plot of differentially expressed genes (DEGs) in ELV treated cells compared with untreated monocytes. Genes with adjusted p value < 0.05 (–log10(padj) > 1.3) and absolute fold change > 1.5 are highlighted: upregulated genes (log2 fold change > 0.6) are shown in green and downregulated genes (log2 fold change < –0.6) in red; *c* – functional enrichment analysis of DEGs using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the Hallmark gene sets from the Molecular Signatures Database (MSigDB).

Functional enrichment analysis of differentially expressed genes was performed using KEGG and the Hallmark collection from MSigDB. Among upregulated genes, 43 KEGG pathways were significantly enriched (adjusted $p < 0.05$), whereas nine KEGG pathways were enriched among downregulated genes. For the Hallmark gene sets, eight pathways were significantly enriched in the upregulated gene set and one pathway in the downregulated gene set (adjusted $p < 0.05$). Full pathway lists and statistics are provided in Table S2.

Among the upregulated DEGs, enrichment was observed for pathways related to immune processes, including leukocyte transendothelial migration (hsa04670), Th17 cell differentiation (hsa04659), Th1 and Th2 cell differentiation (hsa04658), and inflammatory response (M5932). Pathways associated with cell cycle regulation and cell death were also enriched, such as mitotic spindle assembly (M5893), the G2/M checkpoint (M5901), apoptosis (M5902), and the p53 signalling pathway (hsa04115) (Fig. 1*c*).

Prediction of *O. felineus* tsRNA target genes among genes expressed in human monocytes

For target prediction, 3′ untranslated region (3′UTR) sequences of human protein-coding genes were used (196,902 3′UTRs corresponding to 19,082 genes). Genes were retained for further analysis if at least one of their 3′UTR transcripts contained a predicted binding site for at least one *O. felineus* tRNA-derived fragment. This yielded 2,145 target genes among human protein-coding genes. Because the focus of this study was the potential impact of *O. felineus* tsRNAs on human monocytes, genes not expressed in blood cells or monocytes were removed from the target list. After this filtering, 1,299 genes remained as potential targets of *O. felineus* tsRNAs in human monocytes. The largest number of predicted targets was found for the tsRNA Asp-GTC-3 (522 genes), whereas the smallest number was observed for His-GTG-1 (95 genes) (Fig. 2*a*; Table S3).

The largest number of enriched pathways was observed for target genes of the tsRNA Ile-AAT-1: 15 Hallmark/curated MSigDB gene sets and 26 Gene Ontology (GO) terms. In contrast, target genes of Asp-GTC-1 and Tyr-GTA-1 showed enrichment for the smallest number of pathways (two GO terms for 172 Asp-GTC-1 targets and two MSigDB pathways for 195 Tyr-GTA-1 targets). Notably, the largest target set (Asp-GTC-3) and one of the smallest sets (His-GTG-1) were associated with a similar number of enriched pathways (18 for Asp-GTC-3 and 17 for His-GTG-1).

In human monocytes, tsRNA target genes were enriched for pathways linked to cell cycle control and cell death, including cell cycle regulation (M16357, GO:0044774, GO:0044819) and programmed cell death (M27436, M15303). Additional enrichment was observed for immune and inflammatory processes, such as neutrophil degranulation (M27620), antigen presentation via MHC class I (M1066), proteasome-mediated

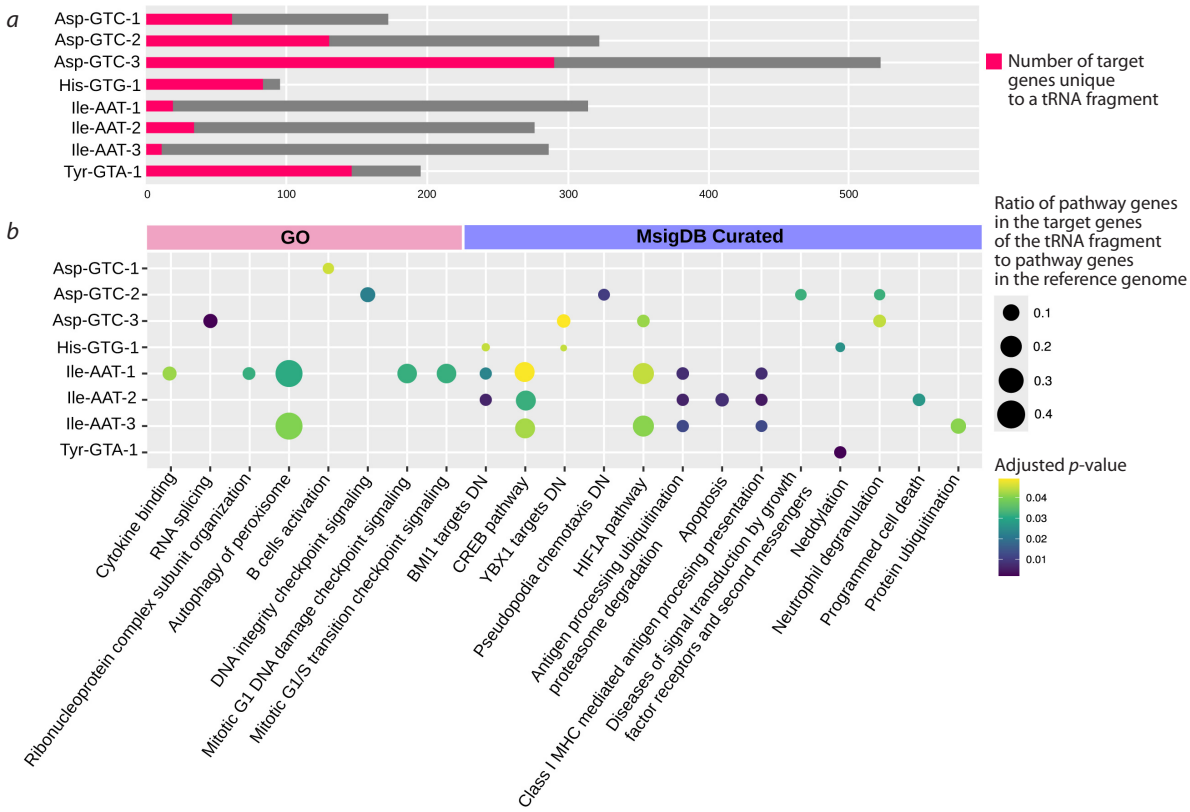


Fig. 2. Predicted *O. felineus* tsRNA target genes among transcripts expressed in human monocytes.

a – number of predicted target genes for each *O. felineus* tRNA derived fragment. Genes with a binding site for only a single tsRNA are shown in pink, whereas genes targeted by multiple tsRNAs are shown in grey; *b* – functional enrichment of predicted target genes for each tsRNA using Gene Ontology (GO) and the curated collection of the Molecular Signatures Database (curated MSigDB).

antigen processing (M1070), B-cell activation (M10657), and cytokine binding (M18255). These data are summarised in Table S3.

Taken together, these results indicate that *O. felineus* ELV-derived tsRNAs have predicted binding sites in host genes whose products participate in key cellular processes. By modulating the expression of such genes through tsRNA–mRNA interactions, liver flukes may be able to influence the activity and functional state of host immune cells.

Prediction of tsRNA target genes among DEGs in THP-1 monocytes

Although THP-1 cells are a widely used model for studying monocyte-mediated immune responses, substantial transcriptional differences exist between primary human monocytes and THP-1 cells (Bosshart, Heinzelmann, 2016; Liu Y. et al., 2021). Therefore, when assessing the overlap between *O. felineus* tsRNA targets and differentially expressed genes in ELV-treated THP-1 monocytes, all 2,145 predicted tsRNA target genes were considered, rather than only those verified as expressed in monocytes. A significant enrichment of tsRNA target genes was observed among DEGs (Pearson chi-square test with Yates' correction = 6.145, $p = 0.014$). Of the 1,484 DEGs identified in THP-1 cells after ELV treatment, 159 genes were predicted targets of O. felineus tsRNAs. These genes were used to construct a tsRNA–mRNA interaction network (Fig. 3; Table S4).

Genes whose expression was significantly altered in THP-1 monocytes following treatment with *O. felineus* exosome-like vesicles and that were also predicted tsRNA targets were mapped onto a tsRNA–mRNA interaction network. These target genes were enriched for pathways involved in cell cycle control and cell death, including DNA integrity checkpoint (M16357), G1/S transition (M2074), apoptosis (M15303), and programmed cell death (M27436), as well as immune response pathways such as the monocyte pathway (M4956),

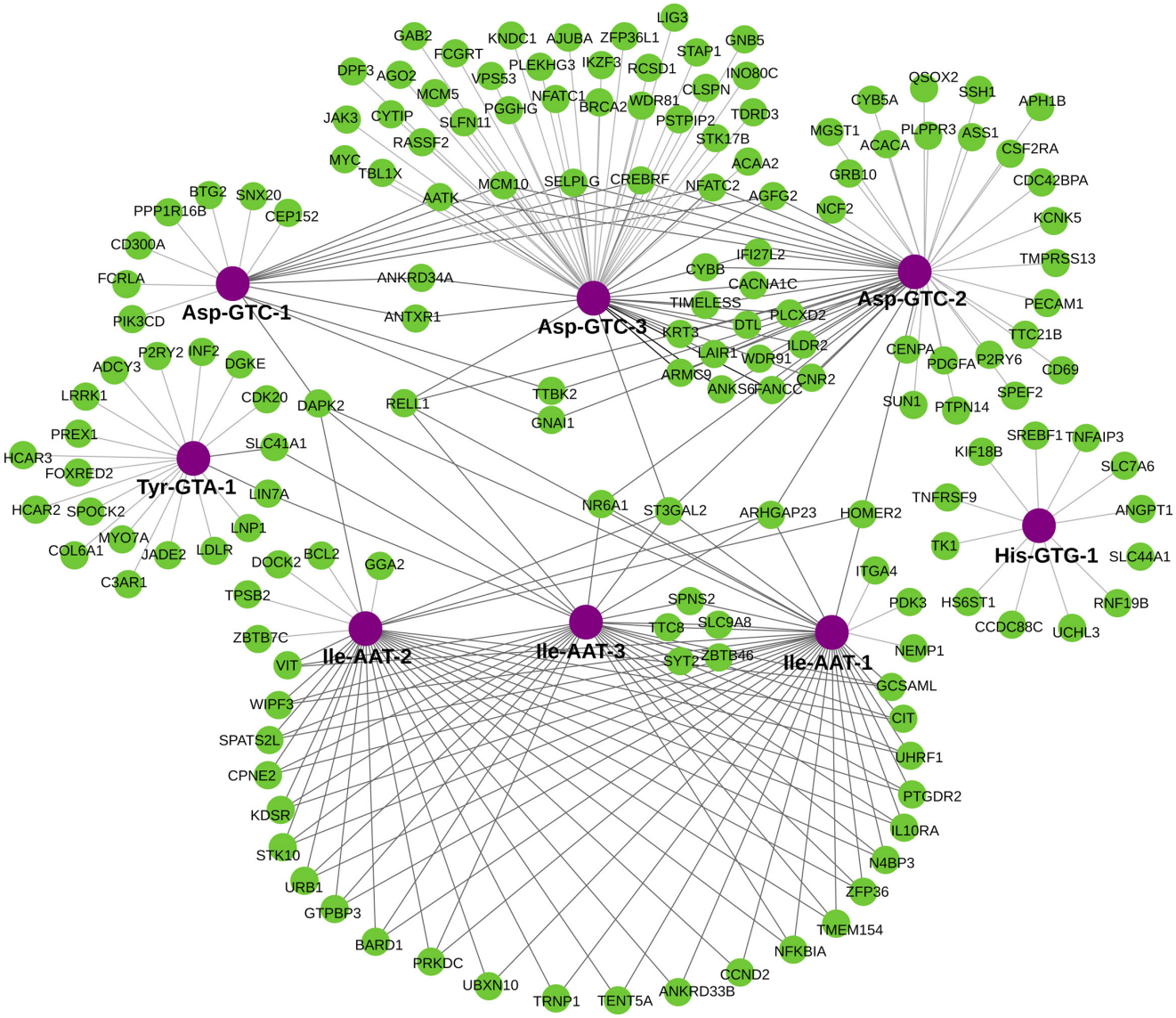


Fig. 3. Interaction network between *O. felineus* tsRNAs and differentially expressed genes in THP-1 monocytes.

mononuclear macrophage differentiation (M25144), antigen presentation via MHC class I (M1066), and proteasome-mediated antigen processing (M1070). Detailed gene and pathway annotations are provided in Table S4.

Discussion

Exosomes and exosome-like vesicles (EVs) are key and evolutionarily conserved mediators of intercellular communication in many organisms, yet their role in parasite–host interactions has only recently begun to be systematically explored. EVs contain diverse bioactive molecules, including proteins, small and long non-coding RNAs, mRNAs, and lipids (Zhang et al., 2022; Pakharukova et al., 2023).

Parasite-derived exosomes can be internalised by host cells and alter both gene expression and the secretion of effector proteins. For example, exosomes from *Trichomonas vaginalis* stimulated increased production of the pro-inflammatory cytokines IL-6 and IL-8 in human cervical epithelial cells Ect1 E6/E7 (Twu et al., 2013). Mouse RAW264.7 macrophages internalised exosomes from *Toxoplasma gondii*, leading to increased secretion of IL-12, TNFα and IFNγ together with reduced production of the anti-inflammatory cytokine IL-10 (Li Y. et al., 2018).

In *Schistosoma japonicum*, exosomes modulated the expression of genes involved in Rap1, TNF and Toll-like receptor signalling as well as cytokine–cytokine receptor interaction pathways. Treatment of RAW264.7 macrophages with exosomal microRNAs enhanced TNFα production (Liu J. et al., 2019). In line with these findings, treatment of PMA-differentiated human THP-1 monocytes with *O. felineus* exosome-like vesicles (ELVs) in the present study resulted in increased expression of genes associated with TNFα signalling and pro-inflammatory responses, including *TNF*, *CCL2*, *CXCL2*, *CXCL3*, *IL6ST*, *NFKB2*, *REL*, and *RELB*.

Interestingly, exosomes from *Fasciola hepatica* did not enhance TNFα production in THP-1 monocytes but instead increased the levels of anti-inflammatory mediators such as TGF-β and NOS2 (Sánchez-López et al., 2023). These divergent responses to ELVs from *F. hepatica* versus *O. felineus* are consistent with studies showing that excretory–secretory products and secreted proteins from *F. hepatica* drive macrophages towards an anti-inflammatory M2 phenotype (Flynn et al., 2007; Figueroa-Santiago, Espino, 2014), whereas exposure to adult *Opisthorchis viverrini* and excretory–secretory products from *Clonorchis sinensis* – species closely related to *O. felineus* – promotes a pro-inflammatory M1 phenotype (Hongsrichan et al., 2014; Kim et al., 2017). However, the specific molecular components within the excretory–secretory products of these trematodes that account for the distinct host immune responses remain unknown.

tRNA-derived small RNAs (tsRNAs) represent a distinct class of small non-coding RNAs and have been identified in exosomes from several parasitic species (Fontenla et al., 2022; Kusakisako et al., 2023; Natali et al., 2023). Nevertheless, their roles in parasite–host communication remain poorly characterised. To date, there are no experimental data directly demonstrating the effects of parasite-derived tsRNAs from exosomes on host cells *in vivo* or *in vitro*.

At the same time, tsRNAs are actively studied as disease biomarkers and as novel regulatory molecules, including in immune cells. In colorectal cancer, elevated levels of specific tsRNAs (tRF-3022b, tRF-3030b and tRF-5008b) have been detected in plasma exosomes, and reducing the levels of these fragments induces cell-cycle arrest and apoptosis in tumor cells; tRF-3022b can also reduce M2 macrophage polarisation by binding to macrophage migration inhibitory factor (MIF) (Lu et al., 2022). In THP-1 cells, the 5′-His-GUG tsRNA is elevated in monocytes and decreases upon differentiation into macrophages, and macrophages transfected with this fragment exhibit reduced survival following exposure to apoptotic stimuli (Jayaram et al., 2025).

Parasite-derived tsRNAs from *O. felineus* ELVs possess predicted binding sites within the 3′UTRs of human genes, although their functional impact is likely to be cell-type specific. In human monocytes, the predicted tsRNA targets identified in this study were enriched for pathways related to cell cycle control, cell death, and immune responses. These observations suggest that trematodes may influence such processes in host immune cells by secreting ELVs containing tsRNAs.

To assess the potential impact of *O. felineus* tsRNAs on the expression of their target genes, tsRNA targets were intersected with DEGs in THP-1 monocytes after ELV treatment, revealing significant enrichment. For some predicted targets, expression levels decreased, whereas others were upregulated. This pattern may reflect non-canonical modes of gene regulation analogous to those described for microRNAs, which can also enhance gene expression under certain conditions (Valinezhad et al., 2014).

Target genes whose expression changed significantly in ELV-treated THP-1 monocytes were associated with cell-cycle and immune pathways, which substantially overlapped with pathways enriched among all tsRNA targets expressed in monocytes and blood cells. Together, these findings support the idea that parasite-derived tsRNAs may contribute to the regulation of these processes in host monocytes by modulating the expression of a subset of genes.

Further functional studies are required to confirm the involvement of parasite tsRNAs in regulating these pathways. In particular, *in vivo* experiments using synthetic tsRNA analogues will be necessary to validate their direct effects on host gene expression and immune cell function. Moreover, because *O. felineus* excretory–secretory products and exosome-like vesicles contain a broad spectrum of bioactive molecules (Pakharukova et al., 2023), synergistic effects of multiple components on host cells are likely. Thus, it will be important to investigate other classes of regulatory molecules present in *O. felineus* ELVs, including microRNAs, to obtain a comprehensive view of how this parasite modulates the host immune response.

Conclusion

The role of exosome-like vesicles (ELVs) released by the liver fluke *Opisthorchis felineus* in shaping the host immune response remains poorly understood. In particular, the contribution of small non-coding RNAs carried by these vesicles, such as tRNA-derived fragments (tsRNAs), has not been

characterised in detail, even though tsRNAs are now actively investigated as disease biomarkers and as a novel class of regulatory molecules.

In this study, transcriptomic profiling of human THP-1 monocytes treated with *O. felineus* ELVs was combined with *in silico* prediction of *O. felineus* tsRNA target genes among transcripts expressed in human monocytes and with the identification of tsRNA targets within the set of differentially expressed genes in ELV-treated THP-1 cells.

Further validation *in vivo* is required to confirm these findings. Future studies are necessary to be focused on: 1) assessment of proliferation, migration and cell-cycle dynamics in host cells exposed to synthetic analogues of *O. felineus* tsRNAs; 2) verification of predicted target tsRNA binding sites using reporter system; and 3) quantification of target gene expression at both the mRNA level (quantitative real-time PCR) and the protein level (Western blot analysis).

References

Araújo M.I., Hoppe B.S., Medeiros M. Jr, Carvalho E.M. *Schistosoma mansoni* infection modulates the immune response against allergic and auto-immune diseases. *Mem Inst Oswaldo Cruz.* 2004;99(5): 27-32. doi 10.1590/s0074-02762004000900005

Artuyants A., Campos T.L., Rai A.K., Johnson P.J., Dauros-Singorenko P., Phillips A., Simoes-Barbosa A. Extracellular vesicles produced by the protozoan parasite *Trichomonas vaginalis* contain a preferential cargo of tRNA-derived small RNAs. *Int J Parasitol.* 2020; 50(14):1145-1155. doi 10.1016/j.ijpara.2020.07.003

Bhuva D., Smyth G., Garnham A. msigdb: an ExperimentHub package for the Molecular Signatures Database (MSigDB). R package version 1.18.0. 2025. doi 10.18129/B9.bioc.msigdb

Bosshart H., Heinzelmann M. THP-1 cells as a model for human monocytes. *Ann Transl Med*. 2016;4(21):438. doi 10.21037/atm.2016.08.53

Chan P.P., Lin B.Y., Mak A.J., Lowe T.M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res*. 2021;49(16):9077-9096. doi 10.1093/nar/gkab688

Chen W., Peng W., Wang R., Bai S., Cao M., Xiong S., Li Y., Yang Y., Liang J., Liu L., Yazdani H.O., Zhao Y., Cheng B. Exosome-derived tRNA fragments tRF-GluCTC-0005 promotes pancreatic cancer liver metastasis by activating hepatic stellate cells. *Cell Death Dis*. 2024;15(1):102. doi 10.1038/s41419-024-06482-3

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

Figueroa-Santiago O., Espino A.M. *Fasciola hepatica* fatty acid binding protein induces the alternative activation of human macrophages. *Infect Immun.* 2014;82(12):5005-5012. doi 10.1128/IAI.02541-14

Flynn R.J., Irwin J.A., Olivier M., Sekiya M., Dalton J.P., Mulcahy G. Alternative activation of ruminant macrophages by *Fasciola hepatica*. *Vet Immunol Immunopathol.* 2007;120(1-2):31-40. doi 10.1016/j.vetimm.2007.07.003

Fontenla S., Langleib M., de la Torre-Escudero E., Domínguez M.F., Robinson M.W., Tort J. Role of *Fasciola hepatica* small RNAs in the interaction with the mammalian host. *Front Cell Infect Microbiol*. 2022;11:812141. doi 10.3389/fcimb.2021.812141

Hongsrichan N., Intuyod K., Pinlaor P., Khoontawad J., Yongvanit P., Wongkham C., Roytrakul S., Pinlaor S. Cytokine/chemokine secretion and proteomic identification of upregulated annexin A1 from peripheral blood mononuclear cells cocultured with the liver fluke *Opisthorchis viverrini*. *Infect Immun*. 2014;82(5):2135-2147. doi 10.1128/IAI.00901-13

Jayaram A., Galande S., Sharbati S., Deshpande D., Pawar K. Profile of tRNA-derived short non-coding RNAs during monocyte differentiation and their role in macrophage survival. *RNA Biol.* 2025;22(1): 1-9. doi 10.1080/15476286.2025.2525886

Khowawisetsut L., Vimonpatranon S., Lekmanee K., Sawasdipokin H., Srimark N., Chotivanich K., Pattanapanyasat K. Differential effect of extracellular vesicles derived from *Plasmodium falciparum*-infected red blood cells on monocyte polarization. *Int J Mol Sci*. 2023;24(3): 2631. doi 10.3390/ijms24032631

Kim E.M., Kwak Y.S., Yi M.H., Kim J.Y., Sohn W.M., Yong T.S. *Clonorchis sinensis* antigens alter hepatic macrophage polarization *in vitro* and *in vivo*. *PLoS Negl Trop Dis*. 2017;11(5):e0005614. doi 10.1371/journal.pntd.0005614

Kusakisako K., Nakao R., Katakura K. Detection of parasite-derived tRNA and rRNA fragments in the peripheral blood of mice experimentally infected with *Leishmania donovani* and *Leishmania amazonensis* using next-generation sequencing analysis. *Parasitol Int.* 2023;93:102716. doi 10.1016/j.parint.2022.102716

Lertjuthaporn S., Somkird J., Lekmanee K., Atipimonpat A., Sukapirom K., Sawasdipokin H., Tiewcharoen S., Pattanapanyasat K., Khowawisetsut L. Extracellular vesicles from *Naegleria fowleri* induce IL-8 response in THP-1 macrophage. *Pathogens*. 2022;11(6): 632. doi 10.3390/pathogens11060632

Li N., Yao S., Yu G., Lu L., Wang Z. tRFtarget 2.0: expanding the targetome landscape of transfer RNA-derived fragments. *Nucleic Acids Res*. 2024;52(5):345-350. doi 10.1093/nar/gkad815

Li Y., Liu Y., Xiu F., Wang J., Cong H., He S., Shi Y., Wang X., Li X., Zhou H. Characterization of exosomes derived from *Toxoplasma gondii* and their functions in modulating immune responses. *Int J Nanomedicine*. 2018;13:467-477. doi 10.2147/IJN.S151110

Lishai E., Pakharukova M. The liver fluke *Opisthorchis felineus* exosomal tRNA-derived small RNAs as potential mediators of host manipulation. *Biomolecules*. 2026;16(2):244. doi 10.3390/biom16020244

Lishai E.A., Zaparina O.G., Kapushchak Y.K., Sripa B., Hong S.J., Cheng G., Pakharukova M.Y. Comparative liver transcriptome analysis in hamsters infected with food-borne trematodes *Opisthorchis felineus*, *Opisthorchis viverrini*, or *Clonorchis sinensis*. *PLoS Negl Trop Dis.* 2024;18(12):e0012685. doi 10.1371/journal.pntd.0012685

Liu J., Zhu L., Wang J., Qiu L., Chen Y., Davis R.E., Cheng G. *Schistosoma japonicum* extracellular vesicle miRNA cargo regulates host macrophage functions facilitating parasitism. *PLoS Pathog*. 2019; 15(6):e1007817. doi 10.1371/journal.ppat.1007817

Liu Y., Li H., Czajkowsky D.M., Shao Z. Monocytic THP-1 cells diverge significantly from their primary counterparts: a comparative examination of the chromosomal conformations and transcriptomes. *Hereditas*. 2021;158(1):43. doi 10.1186/s41065-021-00205-w

Love M.I., Huber W., Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550. doi 10.1186/s13059-014-0550-8

Lu S., Wei X., Tao L., Dong D., Hu W., Zhang Q., Tao Y., Yu C., Sun D., Cheng H. A novel tRNA-derived fragment tRF-3022b modulates cell apoptosis and M2 macrophage polarization via binding to cytokines in colorectal cancer. *J Hematol Oncol*. 2022;15(1):176. doi 10.1186/s13045-022-01388-z

Mann M., Wright P.R., Backofen R. IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res.* 2017;45(1):435-439. doi 10.1093/nar/gkx279

Mordvinov V.A., Minkova G.A., Kovner A.V., Ponomarev D.V., Lvova M.N., Zaparina O., Romanenko S.A., Shilov A.G., Pakharukova M.Y. A tumorigenic cell line derived from a hamster cholangiocarcinoma associated with *Opisthorchis felineus* liver fluke infection. *Life Sci*. 2021;277:119494. doi 10.1016/j.lfs.2021.119494

Natali L., Luna Pizarro G., Moyano S., de la Cruz-Thea B., Musso J., Rópolo A.S., Eichner N., Meister G., Musri M.M., Feliziani C., Touz M.C. The exosome-like vesicles of *Giardia* assemblages A,

B, and E are involved in the delivering of distinct small RNA from parasite to parasite. *Int J Mol Sci*. 2023;24(11):9559. doi 10.3390/ijms24119559

Norouzi M., Pirestani M., Arefian E., Dalimi A., Sadraei J., Mirjalali H. Exosomes secreted by *Blastocystis* subtypes affect the expression of proinflammatory and anti-inflammatory cytokines (TNFα, IL-6, IL-10, IL-4). *Front Med*. 2022;9:940332. doi 10.3389/fmed.2022.940332

Pakharukova M.Y., Pakharukov Y.V., Mordvinov V.A. Effects of miconazole/clotrimazole and praziquantel combinations against the liver fluke *Opisthorchis felineus in vivo* and *in vitro*. *Parasitol Res*. 2018;117(7):2327-2331. doi 10.1007/s00436-018-5895-6

Pakharukova M.Y., Samsonov V.A., Serbina E.A., Mordvinov V.A. A study of tribendimidine effects *in vitro* and *in vivo* on the liver fluke *Opisthorchis felineus. Parasit Vectors*. 2019;12(1):23. doi 10.1186/s13071-019-3288-z

Pakharukova M.Y., Savina E., Ponomarev D.V., Gubanova N.V., Zaparina O., Zakirova E.G., Cheng G., Tikhonova O.V., Mordvinov V.A. Proteomic characterization of *Opisthorchis felineus* exosome-like vesicles and their uptake by human cholangiocytes. *J Proteomics*. 2023;283-284:104927. doi 10.1016/j.jprot.2023.104927

Passos L.S., Gazzinelli-Guimarães P.H., Oliveira Mendes T.A., Guimarães A.C., Silveira Lemos D.D., Ricci N.D., Gonçalves R., Bartholomeu D.C., Fujiwara R.T., Bueno L.L. Regulatory monocytes in helminth infections: insights from the modulation during human hookworm infection. *BMC Infect Dis.* 2017;17(1):253. doi 10.1186/s12879-017-2366-0

Prehn J.H.M., Jirström E. Angiogenin and tRNA fragments in Parkinson's disease and neurodegeneration. *Acta Pharmacol Sin*. 2020;41(4):442-446. doi 10.1038/s41401-020-0375-9

Rehmsmeier M., Steffen P., Hochsmann M., Giegerich R. Fast and effective prediction of microRNA/target duplexes. *RNA*. 2004;10(10):1507-1517. doi 10.1261/rna.5248604

Saltykova I.V., Ogorodova L.M., Bragina E.Y., Puzyrev V.P., Freidin M.B. *Opisthorchis felineus* liver fluke invasion is an environmental factor modifying genetic risk of atopic bronchial asthma. *Acta Trop*. 2014;139:53-56. doi 10.1016/j.actatropica.2014.07.004

Sánchez-López C.M., González-Arce A., Soler C., Ramírez-Toledo V., Trelis M., Bernal D., Marcilla A. Extracellular vesicles from the trematodes *Fasciola hepatica* and *Dicrocoelium dendriticum* trigger different responses in human THP-1 macrophages. *J Extracell Vesicles*. 2023;12(4):e12317. doi 10.1002/jev2.12317

Silverman J.M., Clos J., Horakova E., Wang A.Y., Wiesgigl M., Kelly I., Lynn M.A., McMaster W.R., Foster L.J., Levings M.K., Reiner N.E. Leishmania exosomes modulate innate and adaptive immune responses through effects on monocytes and dendritic cells. *J Immunol*. 2010;185(9):5011-5022. doi 10.4049/jimmunol.1000541

Strømme O., Heck K.A., Brede G., Lindholm H.T., Otterlei M., Arum C.J. tRNA-derived fragments as biomarkers in bladder cancer. *Cancers.* 2024;16(8):1588. doi 10.3390/cancers16081588

Twu O., de Miguel N., Lustig G., Stevens G.C., Vashisht A.A., Wohlschlegel J.A., Johnson P.J. *Trichomonas vaginalis* exosomes deliver cargo to host cells and mediate host:parasite interactions. *PLoS Pathog*. 2013;9(7):e1003482. doi 10.1371/journal.ppat.1003482

Valinezhad Orang A., Safaralizadeh R., Kazemzadeh-Bavili M. Mechanisms of miRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation. *Int J Genomics*. 2014;2014:970607. doi 10.1155/2014/970607

Xia J., Jiang S.C., Peng H.J. Association between liver fluke infection and hepatobiliary pathological changes: a systematic review and meta-analysis. *PloS One*. 2015;10(7):e0132673. doi 10.1371/journal.pone.0132673

Yu G., Wang L.G., Han Y., He Q.Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287. doi 10.1089/omi.2011.0118

Zhang X., Duan S., Li X., Ding J., Zuo L., Sun B., Zhang X., Jiang X., Gao Y., Hu X., Han S. Differences in the secretory exosomes of *Clonorchis sinensis* adults at different incubation times. *Acta Trop*. 2022;234:106604. doi 10.1016/j.actatropica.2022.106604

Conflict of interest. The authors declare no conflict of interest.
Received October 31, 2025. Revised March 4, 2026. Accepted April 24, 2026.

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, Белый список (уровень 1), 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 выхода в свет 25.08.2026. Формат 60 × 84 $^{1}/_{8}$. Уч.-изд. л. 17.6.