Dear colleagues, dear readers,

this issue of the journal focuses on bioinformatics. In the last decade, a rapid improvement of methods for decoding genomes resulted in an information explosion of such a power that genetics has become the largest source of data not only in world science, but also in all other aspects of human activity, including social networks. Studies looking into the human genome become more and more intensive common with the advent of large international projects. As of August 14, 2020, the 1000 Genomes Project (https://www.internationalgenome.org/) had sequenced 3202 genomes. The 100,000 Genomes Project (https://www.genomicsengland.co.uk/about-genomics-england/the-100000-genomes-project/) has sequenced the genomes of 85,000 patients with rare diseases/cancer.

As of July 31, 2020, the 1000 Bull Genomes Project (http://www.1000bullgenomes.com) had sequenced the genomes of more than 5000 animals in 200 cattle breeds and species. This resulted in the identification of more than 155 million genetic variants (SNPs and small deletions/insertions). As of August 1, 2019, the sheep genome sequence project SheepGenomesDB (https://sheepgenomesdb.org) had sequenced the genomes of 935 animals in 69 breeds and found more than 50 million genetic variants. As of November 9, 2020, the 1000 Goat Genome Project (http://www.goatgenome.org/vargoats_data_access.html) had collected data on 127,852,473 genetic variants identified in 1159 animals in 101 goat breeds.

Marker-oriented and genomic breeding as well as genomic editing called for the deciphering of the genomes of the main agricultural plants: wheat, maize, barley, rice, soybean, common bean, potato, a wide range of vegetables and fruits and others (http://plants.ensembl.org/species.html; http://www.plantgdb.org/prj/GenomeBrowser/). A large project seeking to study genetic variation in rice based on the sequencing of a 3000-strong collection of accessions from 89 countries (the 3,000 Rice Genomes Project. *Gigascience*. 2014;3:7. DOI 10.1186/2047-217X-3-7) has achieved completion. One of the ongoing projects is 10KP, a whole genome plant sequencing plan started in 2018 and aiming at the complete sequencing of 10,000 plants in the main clades of embryophytes, green algae and protists (with the exception of fungi) (Cheng S., Melkonian M., Smith S.A., Brockington S., Archibald J.M., Delaux P.M., Li F.W., Melkonian B., Mavrodiev E.V., Sun W., Fu Y., Yang H., Soltis D.E., Graham S.W., Soltis P.S., Liu X., Xu X., Wong G.K. 10KP: A phylodiverse genome sequencing plan. *Gigascience*. 2018; 7(3):1-9. DOI 10.1093/gigascience/giy013). The genomes sequenced to date include more than 36,000 whole viral genomes (https://www.ncbi.nlm.nih.gov/genome/browse/#!/viruses/), 163,645 whole bacterial genomes and 1886 whole Archaeal genomes (https://gold.jgi.doe.gov/distribution). More than 2590 fungal genomes have been sequenced under the 1000 Fungal Genomes Project (https://mycocosm.jgi.doe.gov/pages/fungi-1000-projects.jsf).

A tremendous wealth of protein sequence and annotation data been collected. The Universal Protein Resource (UniProt, https://www.uniprot.org/) is a database containing the descriptions of 563,082 experimentally confirmed primary protein structures; the TrEMBL database (https://www.uniprot.org/statistics/TrEMBL) contains more than 190 million amino acid sequences obtained by automated computer-aided genome annotation. Progress in methods for the physical and chemical studies of proteins has resulted in a fast-paced accumulation of data on their spatial structure (174,507 entries in the Protein Data Bank (PDB), https://www.rcsb.org/). Of invaluable importance is the information on protein structure contained in the mass spectrometry database Chemdata.nist.gov (https://chemdata.nist.gov/) including the descriptions of more than 100 million mass spectra of chemical peptides and metabolites from a range of tissues, biological liquids and cells.

At present, KEGG Pathway (manual annotation), STRING (https://string-db.org/), GeneMANIA (https://genemania.org/), Pathway Commons (https://www.pathwaycommons.org/) and other resources contain more than 70,000 gene regulatory networks, signal transduction pathways and metabolic pathways reconstructed to date.

As far as medicine is concerned, it will benefit from the gigantic bodies of data on human genetic variation: dbSNP (https://www.ncbi.nlm.nih.gov/) contains more than 72 million entries with human SNPs (of which ~24,000 are associated with various diseases) and Ensembl (http://www.ensembl.org/Homo_sapiens) contains more than 667 million entries with human SNPs.

The information explosion in genetics has become quite a challenge, for the rates at which genomic data are accumulated are well ahead of the rates at which these data can be analyzed in a computer-aided manner, and so most genome projects end up in formal assemblies with very rough annotations (or without them) (https://gold.jgi.doe.gov/). This suggests the fundamental importance of information technologies and bioinformatics for the storage, processing and analysis of genomic data for the benefit of addressing fundamental and applied problems in genetics, medicine, pharmacology, agriculture, biotechnology and biosafety.

The understanding and practical use of tremendously large bodies of exceptionally complex genetic experimental data asked for modern information technologies, efficient methods for computer-aided analysis of big data and mathematical modeling of biological systems and processes at different hierarchical levels of organization of living systems – from genomes, genes, proteins, metabolic pathways and gene regulatory networks, including cells and tissues, to whole organisms, populations and ecosystems. In this issue, the reader will find papers on such aspects of bioinformatics as computer-aided genomics and transcriptomics, computer-aided systems biology, computer-aided evolutionary biology and automated analysis of plant phenotypes.

*Nikolay A. Kolchanov,*

*Scientific Editor of the issue,*
*Full Member of the Russian Academy of Sciences,*
*Academic Advisor, the Institute of Cytology and Genetics*
*of the Siberian Branch of the Russian Academy of Sciences*