

Уважаемые коллеги, дорогие читатели! Настоящий выпуск журнала имеет биоинформатическую направленность. В последнее десятилетие в результате стремительного совершенствования методов расшифровки геномов произошёл информационный взрыв такой силы, что генетика стала самым большим источником данных не только в мировой науке, но и во всех других сферах человеческой деятельности, включая социальные сети. Всё шире разворачиваются исследования генома человека в рамках крупных международных проектов. В проекте «1000 геномов» (<https://www.internationalgenome.org/>) на 14.08.2020 просеквенировано 3202 генома. В проекте «100 000 геномов» (<https://www.genomicsengland.co.uk/about-genomics-england/the-100000-genomes-project/>) проанализированы геномы 85 000 пациентов с редкими заболеваниями/раком.

В рамках проекта «1000 геномов быков» (<http://www.1000bullgenomes.com>) на 31.07.2020 было отсеквенировано свыше 5000 животных, относящихся более чем к 200 породам и видам крупного рогатого скота. В результате в геномах этих животных идентифицировано более 155 млн генетических вариантов (ОНП и небольшие инсерции/делетии). На 01.08.2019 в проекте по секвенированию геномов овец SheepGenomesDB (<https://sheepgenomesdb.org>) было отсеквенировано 935 животных, относящихся к 69 породам, у которых идентифицировано более 50 млн генетических вариантов. Проект «1000 геномов коз» (http://www.goatgenome.org/vargoats_data_access.html) на 09.11.2020 содержит данные о 127852473 генетических вариантах, выявленных у 1159 животных, относящихся к 101 породе коз.

Маркёр-ориентированная и геномная селекция, а также геномное редактирование потребовали расшифровки геномов основных сельскохозяйственных растений: пшеницы, кукурузы, ячменя, риса, сои, фасоли, картофеля, широкого спектра овощных и фруктовых культур и других (<http://plants.ensembl.org/species.html>; <http://www.plantgdb.org/prj/GenomeBrowser/>). Выполнен крупный проект по исследованию генетической изменчивости генома риса на основе секвенирования коллекции из 3000 образцов этого сельскохозяйственного растения из 89 стран (The 3,000 rice

genomes project. *Gigascience*. 2014;3:7. DOI 10.1186/2047-217X-3-7). С 2018 г. осуществляется проект по полному секвенированию геномов 10000 растений, относящихся к основнымкладам эмбриофитов, зелёных водорослей и протистов (за исключением грибов) (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). Расшифровано более 36000 полных геномов вирусов (<https://www.ncbi.nlm.nih.gov/genome/browse/#!/viruses/>), секвенировано 163 645 полных геномов бактерий и 1886 полных геномов архей (<https://gold.jgi.doe.gov/distribution>). Прочитано более 2590 геномов грибов (проект «1000 геномов грибов», <https://mycoscosm.jgi.doe.gov/pages/fungi-1000-projects.jsf>).

Получены огромные объёмы информации по структуре белков. В базе данных UniProt <https://www.uniprot.org/> (хранилище данных аминокислотных последовательностей) содержится описание 563 082 экспериментально подтверждённых первичных структур белков, а в базе TrEMBL (<https://www.uniprot.org/statistics/TrEMBL>) хранится более 190 млн аминокислотных последовательностей, полученных на основе автоматической компьютерной аннотации геномов. Совершенствование методов физико-химического изучения белков обеспечило стремительное накопление сведений по их пространственным структурам (174 507 записей в базе данных PDB <https://www.rcsb.org/>). Ценнейшая информация о структуре белков представлена в базе масс-спектрометрических данных Chemdata.nist.gov (<https://chemdata.nist.gov/>), включающей описание свыше 100 млн масс-спектров химических пептидов и метаболитов из различных тканей, биологических жидкостей и клеток.

Следует отметить, что к настоящему времени реконструировано более 70 000 геномных сетей, путей передачи сигналов и метаболических путей, представленных в базе KEGG Pathway (ручная аннотация), базах данных систем STRING (<https://string-db.org/>), GeneMANIA (<https://genemania.org/>), Pathway Commons (<https://www.pathwaycommons.org/>) и других.

Особенно значимы для медицины гигантские объёмы информации по генетической изменчивости человека: в базе dbSNP (<https://www.ncbi.nlm.nih.gov/>) на текущий момент содержится более 72 млн записей об SNP в геномах человека (из них ~24 тыс. связаны с развитием заболеваний), а в базе Ensembl (http://www.ensembl.org/Homo_sapiens) приведено более 667 млн записей об SNP в геномах человека.

Информационный взрыв в генетике стал грандиозным вызовом, поскольку темпы накопления геномных данных на порядок опережают возможности их компьютерного анализа, из-за чего большая часть геномных проектов заканчивается их формальной сборкой с очень поверхностной аннотацией (или даже без неё) (<https://gold.jgi.doe.gov/>).

Всё это свидетельствует о фундаментальной значимости информационных технологий и биоинформатики для хранения, обработки и анализа геномных данных в интересах решения фундаментальных и прикладных задач генетики, медицины, фармакологии, сельского хозяйства, биотехнологии и биобезопасности.

Понимание и практическое применение огромных объёмов генетических экспериментальных данных исключительно высокой сложности потребовало разработки современных информационных технологий, эффективных методов компьютерного анализа больших данных

и математического моделирования биологических систем и процессов на различных иерархических уровнях организации живых систем, начиная с геномов, генов, белков, метаболических путей и геновых сетей, включая клетки и ткани и заканчивая целостными организмами, популяциями и экосистемами. В этом номере вниманию читателей предложены статьи, посвящённые конкретным исследованиям по таким направлениям биоинформатики, как компьютерная геномика и транскриптомика, системная компьютерная биология, эволюционная компьютерная биология и автоматический анализ фенотипов растений.

*Научный редактор выпуска
академик Н.А. Колчанов
научный руководитель ФИЦ ИЦиГ СО РАН*