

## ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ ОБРАБОТКИ ДАННЫХ ПОЛНОГЕНОМНОГО СЕКВЕНИРОВАНИЯ МИКРООРГАНИЗМОВ\*

М. В. СПРИНДЖУК, Р. С. СЕРГЕЕВ, Ю. Е. ДЕМИДЧИК, О. М. ЗАЛУЦКАЯ, А. Е. СКРЯГИН, А. М. СКРЯГИНА

## SOFTWARE FOR PROCESSING OF DATA OF WHOLE-GENOME SEQUENCE ANALYSIS OF MICROORGANISMS

M. V. SPRINDZHUK, R. C. SERGEEV, YU. E. DEMIDCHIK, O. M. ZALUTSKAYA, A. E. SKRYAGIN, A. M. SKRYAGINA

ГНУ «Объединенный институт проблем информатики НАН Беларуси», г. Минск

Unified Research Institute of Information Technology Problems, Minsk, Belarus Republic

Представлен краткий обзор современного программного обеспечения, разработанного для целей анализа данных полных геномов микроорганизмов.

*Ключевые слова:* полногеномное секвенирование, секвенирование следующего поколения, биоинформатика.

The article presents the brief review of the current software developed for data analysis of microorganisms' whole genomes.

*Key words:* whole-genome sequence analysis, sequence analysis of the next generation, bioinformatics.

Развитие темы обработки данных полногеномного секвенирования геномов микро- и макроорганизмов связано с развитием технологии обработки и анализа нуклеиновых кислот, с быстрым и значительным ростом технических возможностей современных компьютеров и разработок технологий создания программного обеспечения для нужд биоинформатики.

Основное назначение информации о полногеномном составе микобактерий туберкулеза и других возбудителей инфекционных заболеваний, аналогичных по природе и патогенезу, – изучение филогенетической природы эпидемических вспышек особо опасных инфекций с целью выявления мишеней для своевременной эффективной антимикробной терапии [3, 27, 49, 53, 78, 79]. По мере увеличения объема знаний и достоверной информации, относящейся к технологии полногеномного секвенирования как темы «больших данных», можно ожидать, что будут раскрыты сравнительные различия различных штаммов микроорганизмов в ракурсе вирулентности, токсигенности, резистентности к различным факторам среды, онкогенности и т. п. Полногеномное секвенирование макроорганизмов и человека имеет схожие черты с этой технологией для более примитивных микроорганизмов, но еще более сложно в связи с более высоким уровнем геномной организации и еще большими объемами получаемых и анализируемых данных (рис.).



Рис. Общая методология обработки данных полногеномного секвенирования микроорганизмов для целей изучения антимикробной резистентности

Технологии изучения полных геномов тесно связаны с исследованиями, направленными на поиск ассоциаций генетического набора микро- и макроорганизма и его фенотипа (GWAS – genome wide association studies).

\* Исследование выполнено при финансовой поддержке ОИПИ, CRDF, СКИФ-СНГ.



В статье не рассматривается множество имеющихся программных продуктов мелкого масштаба и узкого назначения, интересующийся читатель, тем не менее, может найти такие нужные программные инструменты на сайтах [www.omicstools.com](http://www.omicstools.com), <http://sourceforge.net/directory/os:linux/freshness:recentlyupdated/?q=next%20generation%20sequencing>, [http://seqanswers.com/\\*\\*](http://seqanswers.com/**).

### Программы, полезные для научных исследований

#### 1. Программный комплекс NextGene

Программный комплекс NextGene (<http://www.softgenetics.com/NextGENe.html>) – наиболее гармонично развитая система для высокоскоростной обработки данных полногеномного секвенирования. Имеет понятный интуитивный интерфейс, способен конвертировать данные во все основные необходимые форматы. Основная функция данного программного обеспечения – находить мутации (одионые и множественные полиморфизмы, делеции и т. п.) относительно одного или нескольких референсных (ссылочных файлов). Эта программа генерирует многостраничный отчет в формате PDF, а также VCF (variant call format) файлы. Именно это программное обеспечение наиболее четко отвечает практическим потребностям ученого, изучающего мутации по данным полногеномного секвенирования микроорганизмов. Для обработки целых геномов туберкулезной палочки, по нашему опыту и мнению, это лучший выбор.

#### 2. Программный комплекс Ridom SeqSphere

Отличительная особенность программного обеспечения Ridom SeqSphere (<http://www.ridom.de/seqsphere/mlstplus/>) – функция нахождения мишеней мультилокусного типирования нуклеотидных последовательностей.

Этот программный комплекс отличается от других тем, что дает возможность пользователю

загружать целые геномы всех имеющихся в банке данных микроорганизмов [25, 26, 45, 65].

#### 3. Программное обеспечение LasergeneDNAStar

Это классическая высококачественная программа для обработки данных нуклеотидных последовательностей ДНК и РНК микроорганизмов. Разрабатывается с 70-х годов по настоящее время. Она состоит из множества модулей, наиболее часто используется SeqMap и SeqBuilder – для манипуляций с нуклеотидными последовательностями, выравнивания, нуклеотидного анализа, нахождения мутаций. Для эффективной работы с целыми геномами это программное обеспечение требует значительных ресурсов компьютера: RAM > 16 Гб, большой объем дисков, процессоры Intel Xeon и Intel Core i5-i7 последних поколений. Для быстрой эффективной и продуктивной работы целесообразно запускать эту программу на суперкомпьютере. Это программное обеспечение удобно для манипуляции относительно короткими нуклеотидными последовательностями [7, 9-11].

#### 4. Программное обеспечение CLC Genomics

CLC Genomics Workbench (<http://www.clcbio.com/products/clc-main-workbench/>) – кросс-платформенное настольное приложение с графическим интерфейсом пользователя. Предназначено для анализа и визуализации NGS (next generation sequencing – секвенирование следующего поколения) данных, включает передовую технологию и алгоритмы, также поддерживает стандартный технологический процесс обработки данных NGS. Полная платная коммерческая версия данной программы включает большой спектр функций для предметных областей геномики, транскриптомики и эпигеномики. Эта программа удобна для нуклеотидного анализа и визуализации цепей нуклеотидов, а также для построения филогенетических деревьев. CLC Genomics Workbench также способен выполнять выравнивание данных полных геномов бактерий, искать белки и нуклеотидные последовательности

**\*\*Программное обеспечение для выравнивания нуклеотидных последовательностей полных геномов:**

Eland: [http://illumina.ucr.edu/illumina\\_docs/](http://illumina.ucr.edu/illumina_docs/)

velvet: <http://www.ebi.ac.uk/~zerbino/velvet/>

Maq (Mapping and Assembly with Quality): <http://maq.sourceforge.net/maq-man.shtml>

SOAP alignment tool: <http://soap.genomics.org.cn/>

SeqMap: <http://biogibbs.stanford.edu/~jiangh/SeqMap/>

Bowtie: <http://bowtie-bio.sourceforge.net/>

Novoalign: <http://www.novocraft.com>

BWA – Burrows-Wheeler Alignment Tool: <http://maq.sourceforge.net/bwa-man.shtml>

BWA-SW (general purpose local alignment tool): <http://i.cs.hku.hk/~ckwong3/bwtsw/>

Zoom: <http://www.bioinformaticsolutions.com/products/zoom/index.php>

SHRiMP: <http://compbio.cs.toronto.edu/shrimp/>

SHORE (1001genomes.org): <http://1001genomes.org/downloads/shore.html>

CloudBurst (Highly Sensitive Short Read Mapping with MapReduce): <http://apps.sourceforge.net/mediawiki/cloudburst-bio/index.php?title=CloudBurst>



в базах данных генерации филогенетических деревьев. Для эффективной работы эта программа требует очень больших ресурсов компьютера: оперативная память (RAM – random access memory) не менее 32 Гб [35].

#### 5. Язык программирования R

Пакет R Rsubread – основная библиотека для языка R, предназначенная для обработки данных полногеномного секвенирования. Практический интерес представляют функции для сравнения VCF-файлов и реализация алгоритмов для выравнивания нуклеотидных последовательностей.

#### 6. Программа UGENE

Это программное обеспечение (<https://ru.wikipedia.org/wiki/UGENE>) заслуживает особого внимания, так как оно имеет открытый исходный код C++ с Qt-интерфейсом, что предоставляет уникальную возможность изучать реализацию алгоритмов и яркий пример архитектуры биоинформатического программного обеспечения.

Ниже представлены основные возможности этого программного продукта:

1. Создание, редактирование и аннотирование нуклеотидных и белковых последовательностей.

2. Быстрый поиск в последовательности.

3. Множественное выравнивание последовательностей: ClustalW [20, 37, 39, 50, 52, 72, 76], ClustalO, MUSCLE, Kalign, MAFFT, T-Coffee [15, 16, 33, 43, 46-48, 60, 71, 80].

4. Создание и редактирование биоинформационной базы данных с общим доступом.

5. Поиск в онлайн-базах данных: NCBI, PDB, UniProtKB/Swiss-Prot [5, 6, 18, 21, 31, 41, 66], UniProtKB/TrEMBL, серверы DAS.

6. Онлайн- и локальный BLAST-поиск.

7. Поиск открытых рамок считывания.

8. Рестрикционный анализ со встроенной базой данных ферментов рестрикции REBASE.

9. Интегрированный пакет Primer3 для дизайна полимеразной цепной реакции (ПЦР) праймеров.

10. Аннотирование плазмид.

11. Клонирование *in silico*.

12. Выравнивание на геном с помощью Bowtie, BWA или UGENE Genome Aligner.

13. Визуализация выравненных коротких прочтений с помощью UGENE Assembly Browser.

14. Поиск геномных вариаций с помощью SAMtool [17, 32, 38, 57].

15. Обработка сырых данных NGS.

16. Анализ RNA-Seq данных с помощью TopHat [54, 75] и инструментов Cufflinks [54, 74].

17. Анализ ChIP-Seq данных с помощью MACS, CEAS и других инструментов.

18. Поиск гомологов с HMMER2 и HMMER3.

19. Работа с хроматограммами.

20. Поиск сайтов связывания транскрипционных факторов с использованием весовых матриц или алгоритма SITECON.

21. Поиск повторов в последовательности ДНК: прямых, обратных, tandemных.

22. Локальное выравнивание последовательности с использованием оптимизированной версии алгоритма Смита – Ватермана.

23. Построение филогенетических деревьев (с помощью PHYLIP Neighbor Joining [1, 40, 59, 64], MrBayes или PhyML Maximum Likelihood) и редактирование деревьев.

24. Комбинирование различных алгоритмов в вычислительную схему с помощью дизайнера вычислительных схем.

25. Сборки контигов (CAP3).

26. Отображение 3D-структуры белков для форматов PDB и MMDB formats, поддержка стереоэффекта.

27. Предсказание вторичной структуры белка с помощью алгоритмов GOR IV и PSIPRED.

28. Конструирование точечных графиков для ДНК последовательностей.

29. Выравнивание мРНК (Spidey) [82].

30. Поиск комплексных сигналов с Expert Discovery [77].

31. Поиск шаблона результатов различных алгоритмов в нуклеотидной последовательности с помощью дизайнера запросов.

32. ПЦР *in silico*.

33. Spades de novo ассемблер [22, 24, 51, 77, 81].

#### 7. Программа GeneiousPro

GeneiousPro предоставляет интеллектуальную программную платформу для биологов-биоинформатиков. Сложные данные, такие как целые бактериальные геномы или человеческие хромосомы, могут храниться, визуализироваться и анализироваться в среде единственного настольного приложения, которое объединяет различные функции и инструменты и гарантирует, что исследования легко повторимы. Это программа отличается от других интеллектуальными веб-агентами для поиска и загрузки информации, способностью генерировать высококачественные отчеты с яркими рисунками. Она также способна визуализировать различные биоинформатические данные и имеет стандартные функции ClustalW [20, 37, 39, 50, 52, 72, 76], MrBayes [4, 29, 42, 62, 63, 83], EMBOSS, Mauve, MUSCLE, PAUP\*, PFam, Uniprot [22, 51], BLAST, PhyML.

#### 8. Genematrix

<http://www.genomatix.de/cgi-bin/RegionMiner/regionminer.pl?s=1f31c9c2b86387342c4a2276003861d6>



Этот веб-комплекс имеет много функций для работы с различными форматами биоинформатических данных полногеномного секвенирования. Практическое значение имеет функция сравнения VCF-файлов.

#### 9. Веб-комплекс Galaxy

DNA Galaxy является попыткой произвести с открытым исходным кодом программное обеспечение для биоинформатики альтернативой дорогому программному обеспечению Lasergene DNASTar. Его особенности в настоящее время включают Blast- и Entrez-поиск, интерфейс для CLUSTAL и последовательность редактора. Она требует java1.5 <http://ru.downloadv.com/free-download/animated-galaxy>. Особенность веб-сервиса для полногеномного секвенирования бактериального генома – пакетный анализ VCF-файлов с возможностью отсылать результаты вычисления на e-mail заказчика. Компания представляет бесплатный испытательный срок на 8 сут [2].

#### 10. BioPython, BioPerl, BioJava, BioRuby

Эти библиотеки имеют происхождение от BioPerl [14, 68, 69]. Они неодинаковы между собой по функциональности и представляют широкий арсенал для манипуляции и визуализации данных нуклеотидных последовательностей различной природы происхождения. Стоит отметить, что эти ресурсы более удобны для коротких последовательностей, а не для NGS-данных [12, 14, 23, 28, 30, 55, 68-70].

#### 11. GATK (Genome Association Tool Kit)

Данный открытый для бесплатного использования программный комплекс имеет большой арсенал Java- и Perl-утилит для нахождения и анализа геномных вариантов (variant call) [44, 84]. Эта система является стандартом и эталоном для сравнения других, коммерческих и бесплатных программ, в которых реализована функциональность генерации и анализа VCF-файлов.

#### 12. PLINK

Это программа, написанная на языке C++, вызывается из командной строки и не имеет интерфейса (<http://pngu.mgh.harvard.edu/%E2%88%BCpurcell/dist/plink-1.07-x86/>) [8, 19, 56, 58, 67]. Утилита предназначена для вычисления статистики одиночных нуклеотидных полиморфизмов, контроля качества данных и популяционной стратификации.

#### 13. Геномные браузеры

Наиболее популярный и удобный для визуализации геномный браузер – это IGV (integrative genomics viewer) [13, 61, 73]. Его особенно удобно устанавливать прямо из менеджера программного обеспечения Linux Ubuntu 14.04. Программа удобна для просматривания целых геномов, редактирования, поиска и визуализации

нужных регионов геномов. Полный список геномных браузеров можно просмотреть по ссылке: [http://en.wikipedia.org/w/index.php?title=Special:Book&bookcmd=download&collection\\_id=9a5e06d735ceb1db15e78a95c66f8b22c2fbb7b7&writer=rd2latex&return\\_to=Genome+browser](http://en.wikipedia.org/w/index.php?title=Special:Book&bookcmd=download&collection_id=9a5e06d735ceb1db15e78a95c66f8b22c2fbb7b7&writer=rd2latex&return_to=Genome+browser).

## Заключение

Анализ «больших данных» в биоинформатике обработки данных полных геномов требует применения высококачественной современной дорогостоящей техники и коммерческого программного обеспечения. Практически единственной альтернативой платным программным продуктам являются ресурсы языков R, Python, Java, Ruby, C++, Perl и программа Ugene.

Возможные пути улучшения программного обеспечения для анализа полных геномов микроорганизмов:

1) реализация возможности сопоставлять идентификаторы-имена нуклеотидных последовательностей из различных баз данных;

2) осуществление способности программного обеспечения автоматически загружать информацию о значимых в данном вычислительном эксперименте белках и генах (описание, публикации по теме и т. п.);

3) улучшение интерфейсов программного обеспечения в плане возможности перетаскивать данные в окна, становиться в нужную директорию, скрывать лишнее и т. п.;

4) расширение возможностей яркой полноценной визуализации вычислительных экспериментов циркулярными диаграммами, понятными таблицами, графиками и графами;

5) интеграции программных комплексов и существующих веб-платформ для высокоскоростной надежной работы;

6) расширение поисковых возможностей программ в плане возможности быстро находить в отчетах мутированные гены, сопоставлять информацию множественных экспериментов с возможностью генерации удобочитаемых отчетов.

## ЛИТЕРАТУРА

1. Abdennadher N., Boesch R. Porting PHYLIP phylogenetic package on the desktop GRID platform XtremWeb-CH // *Stud. Health Technol. Inform.* – 2007. – Vol. 126. – P. 55-64.
2. Aranguren M. E. et al. OPPL-Galaxy, a Galaxy tool for enhancing ontology exploitation as part of bioinformatics workflows // *J. Biomed. Semantics.* – 2013. – Vol. 4, № 1. – P. 2.
3. Bakker H. C. et al. A whole-genome single nucleotide polymorphism-based approach to trace and identify outbreaks linked to a common *Salmonella enterica* subsp. *enterica* serovar Montevideo pulsed-field gel electrophoresis type // *Appl. Environ. Microbiol.* – 2011. – Vol. 77, № 24. – P. 8648-8655.



4. Bao J. et al. Efficient implementation of MrBayes on multi-GPU // *Mol. Biol. Evol.* – 2013. – Vol. 30, № 6. – P. 1471-1479.
5. Boutet E. et al. UniProtKB/Swiss-Prot // *Methods Mol. Biol.* – 2007. – Vol. 406. – P. 89-112.
6. Braconi Quintaje S., Orchard S. The annotation of both human and mouse kinomes in UniProtKB/Swiss-Prot: one small step in manual annotation, one giant leap for full comprehension of genomes // *Mol. Cell. Proteomics.* – 2008. – Vol. 7, № 8. – P. 1409-1419.
7. Burland T. G. Dnastar's Lasergene sequence analysis software // *Methods Mol. Biol.* – 2000. – Vol. 132. – P. 71-91.
8. Chang C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets // *Gigasience.* – 2015. – Vol. 4. – P. 7.
9. Clewley J. P. Geneman of lasergene // *Methods Mol. Biol.* – 1997. – Vol. 70. – P. 189-196.
10. Clewley J. P. Macintosh sequence analysis software. DNAStar's LaserGene // *Mol. Biotechnol.* – 1995. – Vol. 3, № 3. – P. 221-224.
11. Clewley J. P., Arnold C. MEGALIGN. The multiple alignment module of lasergene // *Methods Mol. Biol.* – 1997. – Vol. 70. – P. 119-129.
12. Cock P. J. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics // *Bioinformatics.* – 2009. – Vol. 25, № 11. – P. 1422-1423.
13. Coletta A. et al. InSilico DB genomic datasets hub: an efficient starting point for analyzing genome-wide studies in GenePattern, Integrative Genomics Viewer, and R/Bioconductor // *Genome Biol.* – 2012. – Vol. 13, № 11. – P. R104.
14. Crabtree J. et al. Circleator: flexible circular visualization of genome-associated data with BioPerl and SVG // *Bioinformatics.* – 2014. – Vol. 30, № 21. – P. 3125-3127.
15. Di Tommaso P. et al. Cloud-Coffee: implementation of a parallel consistency-based multiple alignment algorithm in the T-Coffee package and its benchmarking on the Amazon Elastic-Cloud // *Bioinformatics.* – 2010. – Vol. 26, № 15. – P. 1903-1904.
16. Di Tommaso P. et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension // *Nucleic Acids Res.* – 2011. – Vol. 39, № 1. – P. W13-W17.
17. Etherington G. J. et al. Bio-samtools 2: a package for analysis and visualization of sequence and alignment data with SAMtools in Ruby // *Bioinformatics.* – 2015. – № 1. – P. 1-12.
18. Famiglietti M. L. et al. Genetic variations and diseases in UniProtKB/Swiss-Prot: the ins and outs of expert manual curation // *Hum. Mutat.* – 2014. – Vol. 35, № 8. – P. 927-935.
19. Fan S. B. et al. Using pLink to analyze cross-linked peptides // *Curr. Protoc. Bioinformatics.* – 2015. – Vol. 49. – P. 1-8.
20. Fukami-Kobayashi K., Saito N. How to make good use of CLUSTALW // *Tanpakushitsu Kakusan Koso.* – 2002. – Vol. 47, № 9. – P. 1237-1239.
21. Gaudet P. et al. Collaborative annotation of genes and proteins between UniProtKB/Swiss-Prot and dictyBase // *Database (Oxford).* – 2009. – Vol. 2009. – P. bap016.
22. Golosova O. et al. Unipro UGENE NGS pipelines and components for variant calling, RNA-seq and ChIP-seq data analyses // *Peer. J.* – 2014. – Vol. 2. – P. e644.
23. Goto N. et al. BioRuby: bioinformatics software for the Ruby programming language // *Bioinformatics.* – 2010. – Vol. 26, № 20. – P. 2617-2619.
24. Guo C. et al. Ugene, a newly identified protein that is commonly overexpressed in cancer and binds uracil DNA glycosylase // *Cancer Res.* – 2008. – Vol. 68, № 15. – P. 6118-6126.
25. Harmsen D. et al. RIDOM: comprehensive and public sequence database for identification of *Mycobacterium* species // *BMC Infect. Dis.* – 2003. – Vol. 3. – P. 26.
26. Harmsen D. et al. RIDOM: Ribosomal differentiation of medical micro-organisms database // *Nucleic Acids Res.* – 2002. – Vol. 30, № 1. – P. 416-417.
27. Harris S. R. et al. Read and assembly metrics inconsequential for clinical utility of whole-genome sequencing in mapping outbreaks // *Nat. Biotechnol.* – 2013. – Vol. 31, № 7. – P. 592-594.
28. Holland R. C. et al. BioJava: an open-source framework for bioinformatics // *Bioinformatics.* – 2008. – Vol. 24, № 18. – P. 2096-2097.
29. Huelsenbeck J. P., Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees // *Bioinformatics.* – 2001. – Vol. 17, № 8. – P. 754-755.
30. Jayadeepa R. M. et al. Review and research analysis of computational target methods using BioRuby and in silico screening of herbal lead compounds against pancreatic cancer using R programming // *Curr. Drug. Metab.* – 2014. – Vol. 15, № 5. – P. 535-543.
31. Jungo F. et al. The UniProtKB/Swiss-Prot Tox-Prot program: A central hub of integrated venom protein data // *Toxicon.* – 2012. – Vol. 60, № 4. – P. 551-557.
32. Kaisers W. et al. rbamtools: an R interface to samtools enabling fast accumulative tabulation of splicing events over multiple RNA-seq samples // *Bioinformatics.* – 2015. – Vol. 31, № 10. – P. 1663-1664.
33. Katoh K. et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform // *Nucleic Acids Res.* – 2002. – Vol. 30, № 14. – P. 3059-3066.
34. Kim D., Salzberg S. L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts // *Genome Biol.* – 2011. – Vol. 12, № 8. – P. R72.
35. Kim K. U. et al. Comparison of functional gene annotation of *Toxascaris leonina* and *Toxocara canis* using CLC genomics workbench // *Korean. J. Parasitol.* – 2013. – Vol. 51, № 5. – P. 525-530.
36. Kleindienst R. et al. Highly efficient refractive Gaussian-to-tophat beam shaper for compact terahertz imager // *Appl. Opt.* – 2010. – Vol. 49, № 10. – P. 1757-1763.
37. Kohli D. K., Bachhawat A. K. CLOURE: Clustal Output Reformatter, a program for reformatting ClustalX/ClustalW outputs for SNP analysis and molecular systematics // *Nucleic Acids Res.* – 2003. – Vol. 31, № 13. – P. 3501-3502.
38. Li H. et al. The Sequence Alignment/Map format and SAMtools // *Bioinformatics.* – 2009. – Vol. 25, № 16. – P. 2078-2079.
39. Li K. B. ClustalW-MPI: ClustalW analysis using distributed and parallel computing // *Bioinformatics.* – 2003. – Vol. 19, № 12. – P. 1585-1586.
40. Lim A., Zhang L. WebPHYLP: a web interface to PHYLIP // *Bioinformatics.* – 1999. – Vol. 15, № 12. – P. 1068-1069.
41. Lima T. et al. HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot // *Nucleic. Acids Res.* – 2009. – Vol. 37. – P. 471-478.
42. Ling C. et al. MrBayes tgMC(3): a tight GPU implementation of MrBayes // *PLoS One.* – 2013. – Vol. 8, № 4. – P. e60667.
43. Magis C. et al. T-Coffee: Tree-based consistency objective function for alignment evaluation // *Methods Mol. Biol.* – 2014. – Vol. 1079. – P. 117-129.
44. McCormick R. E. et al. RIG: Recalibration and Interrelation of Genomic Sequence Data with the GATK // *G3 (Bethesda).* – 2015. – Vol. 5, № 4. – P. 655-665.
45. Mellmann A. et al. Evaluation of RIDOM, MicroSeq, and Genbank services in the molecular identification of *Nocardia* species // *Int. J. Med. Microbiol.* – 2003. – Vol. 293, № 5. – P. 359-370.
46. Notredame C. Computing multiple sequence/structure alignments with the T-coffee package // *Curr. Protoc. Bioinformatics.* – 2010. – Vol. 3. – P. 1-25.
47. Notredame C. et al. T-Coffee: A novel method for fast and accurate multiple sequence alignment // *J. Mol. Biol.* – 2000. – Vol. 302, № 1. – P. 205-217.
48. Notredame C., Suhre K. Computing multiple sequence/structure alignments with the T-coffee package // *Curr. Protoc. Bioinformatics.* – 2004. – Vol. 3. – P. 3-8.
49. Octavia S. et al. Delineating community outbreaks of *Salmonella enterica* serovar Typhimurium by use of whole-genome sequencing: insights into genomic variability within an outbreak // *J. Clin. Microbiol.* – 2015. – Vol. 53, № 4. – P. 1063-1071.
50. Ogden T. H., Rosenberg M. S. Alignment and topological accuracy of the direct optimization approach via POY and traditional phylogenetics via ClustalW + PAUP\* // *Syst. Biol.* – 2007. – Vol. 56, № 2. – P. 182-193.
51. Okonechnikov K. et al. Unipro UGENE: a unified bioinformatics toolkit // *Bioinformatics.* – 2012. – Vol. 28, № 8. – P. 1166-1167.
52. Oliver T. et al. Using reconfigurable hardware to accelerate multiple sequence alignment with ClustalW // *Bioinformatics.* – 2005. – Vol. 21, № 16. – P. 3431-3432.
53. Pantosti A. Whole-genome sequencing may be key to abating hospital-based methicillin-resistant *Staphylococcus aureus* outbreaks // *J. Pediatr.* – 2013. – Vol. 162, № 5. – P. 1079-1080.
54. Pollier J. et al. Analysis of RNA-Seq data with TopHat and Cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cultures // *Methods Mol. Biol.* – 2013. – Vol. 1011. – P. 305-315.
55. Prlic A. et al. BioJava: an open-source framework for bioinformatics in 2012 // *Bioinformatics.* – 2012. – Vol. 28, № 20. – P. 2693-2695.



56. Purcell S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses // *Am. J. Hum. Genet.* – 2007. – Vol. 81, № 3. – P. 559-575.
57. Ramirez-Gonzalez R. H. et al. Bio-samtools: Ruby bindings for SAMtools, a library for accessing BAM files containing high-throughput sequence alignments // *Source Code Biol. Med.* – 2012. – Vol. 7, № 1. – P. 6.
58. Renteria M. E. et al. Using PLINK for Genome-Wide Association Studies (GWAS) and data analysis // *Methods Mol. Biol.* – 2013. – Vol. 1019. – P. 193-213.
59. Retief J. D. Phylogenetic analysis using PHYLIP // *Methods Mol. Biol.* – 2000. – Vol. 132. – P. 243-258.
60. Rius J. et al. A user-friendly web portal for T-Coffee on supercomputers // *BMC Bioinformatics.* – 2011. – Vol. 12. – P. 150.
61. Robinson J. T. et al. Integrative genomics viewer // *Nat. Biotechnol.* – 2011. – Vol. 29, № 1. – P. 24-26.
62. Ronquist F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space // *Syst. Biol.* – 2012. – Vol. 61, № 3. – P. 539-542.
63. Ronquist F., Huelsenbeck J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models // *Bioinformatics.* – 2003. – Vol. 19, № 12. – P. 1572-1574.
64. Ropelewski A. J. et al. MPI-PHYLIP: parallelizing computationally intensive phylogenetic analysis routines for the analysis of large protein families // *PLoS One.* – 2010. – Vol. 5, № 11. – P. e13999.
65. Rothganger J. et al. Ridom TraceEdit: a DNA trace editor and viewer // *Bioinformatics.* – 2006. – Vol. 22, № 4. – P. 493-494.
66. Schneider M. et al. The UniProtKB/Swiss-Prot knowledgebase and its Plant Proteome Annotation Program // *J. Proteomics.* – 2009. – Vol. 72, № 3. – P. 567-573.
67. Shi H. et al. Using Fisher's method with PLINK «LD clumped» output to compare SNP effects across Genome-wide Association Study (GWAS) datasets // *Int. J. Mol. Epidemiol. Genet.* – 2011. – Vol. 2, № 1. – P. 30-35.
68. Stajich J. E. An Introduction to BioPerl // *Methods Mol. Biol.* – 2007. – Vol. 406. – P. 535-548.
69. Stajich J. E. et al. The Bioperl toolkit: Perl modules for the life sciences // *Genome Res.* – 2002. – Vol. 12, № 10. – P. 1611-1618.
70. Talevich E. et al. Bio.Phylo: a unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython // *BMC Bioinformatics.* – 2012. – Vol. 13. – P. 209.
71. Taly J. F. et al. Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures // *Nat. Protoc.* – 2011. – Vol. 6, № 11. – P. 1669-1682.
72. Thompson J. D. et al. Multiple sequence alignment using ClustalW and ClustalX // *Curr. Protoc. Bioinformatics.* – 2002. – Vol. 2. – P. Unit 2.3.
73. Thorvaldsdottir H. et al. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration // *Brief Bioinform.* – 2013. – Vol. 14, № 2. – P. 178-192.
74. Trapnell C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks // *Nat. Protoc.* – 2012. – Vol. 7, № 3. – P. 562-578.
75. Trapnell C. et al. TopHat: discovering splice junctions with RNA-Seq // *Bioinformatics.* – 2009. – Vol. 25, № 9. – P. 1105-1111.
76. Vangala R. K. et al. BioParishodhana: A novel graphical interface integrating BLAST, ClustalW, primer3 and restriction digestion tools // *Bioinformation.* – 2012. – Vol. 8, № 13. – P. 639-643.
77. Vaskin Y. Y. et al. ExpertDiscovery and UGENE integrated system for intelligent analysis of regulatory regions of genes // *In Silico Biol.* – 2011. – Vol. 11, № 3-4. – P. 97-108.
78. Walker T. M. et al. Contact investigations for outbreaks of *Mycobacterium tuberculosis*: advances through whole genome sequencing // *Clin. Microbiol. Infect.* – 2013. – Vol. 19, № 9. – P. 796-802.
79. Walker T. M. et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study // *Lancet Infect. Dis.* – 2013. – Vol. 13, № 2. – P. 137-146.
80. Wallace I. M. et al. M-Coffee: combining multiple sequence alignment methods with T-Coffee // *Nucleic. Acids Res.* – 2006. – Vol. 34, № 6. – P. 1692-1699.
81. Wang L. T. et al. Functional interaction of Ugene and EBV infection mediates tumorigenic effects // *Oncogene.* – 2011. – Vol. 30, № 26. – P. 2921-2932.
82. Wheelan S. J. et al. Spidey: a tool for mRNA-to-genomic alignments // *Genome Res.* – 2001. – Vol. 11, № 11. – P. 1952-1957.
83. Zhou J. et al. MrBayes on a graphics processing unit // *Bioinformatics.* – 2011. – Vol. 27, № 9. – P. 1255-1261.
84. Zhu P. et al. OTG-snp caller: an optimized pipeline based on TMAP and GATK for SNP calling from ion torrent data // *PLoS One.* – 2014. – Vol. 9, № 5. – P. e97507.

## REFERENCES

1. Abdennadher N., Boesch R. Porting PHYLIP phylogenetic package on the desktop GRID platform XtremWeb-CH. *Stud. Health Technol. Inform.*, 2007, vol. 126, pp. 55-64.
2. Aranguren M.E. et al. OPPL-Galaxy, a Galaxy tool for enhancing ontology exploitation as part of bioinformatics workflows. *J. Biomed. Semantics*, 2013, vol. 4, no. 1, pp. 2.
3. Bakker H.C. et al. A whole-genome single nucleotide polymorphism-based approach to trace and identify outbreaks linked to a common *Salmonella enterica* subsp. *enterica* serovar Montevideo pulsed-field gel electrophoresis type. *Appl. Environ. Microbiol.*, 2011, vol. 77, no. 24, pp. 8648-8655.
4. Bao J. et al. Efficient implementation of MrBayes on multi-GPU. *Mol. Biol. Evol.*, 2013, vol. 30, no. 6, pp. 1471-1479.
5. Boutet E. et al. UniProtKB/Swiss-Prot. *Methods Mol. Biol.*, 2007, vol. 406, pp. 89-112.
6. Braconi Quintaje S., Orchard S. The annotation of both human and mouse kinomes in UniProtKB/Swiss-Prot: one small step in manual annotation, one giant leap for full comprehension of genomes. *Mol. Cell. Proteomics*, 2008, vol. 7, no. 8, pp. 1409-1419.
7. Burland T.G. Dnastar's Lasergene sequence analysis software. *Methods Mol. Biol.*, 2000, vol. 132, pp. 71-91.
8. Chang C.C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 2015, vol. 4, pp. 7.
9. Clewley J.P. Geneman of lasergene. *Methods Mol. Biol.*, 1997, vol. 70, pp. 189-196.
10. Clewley J.P. Macintosh sequence analysis software. DNASTar's LaserGene. *Mol. Biotechnol.*, 1995, vol. 3, no. 3, pp. 221-224.
11. Clewley J. P., Arnold C. MEGALIGN. The multiple alignment module of lasergene. *Methods Mol. Biol.*, 1997, vol. 70, pp. 119-129.
12. Cock P.J. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 2009, vol. 25, no. 11, pp. 1422-1423.
13. Coletta A. et al. InSilico DB genomic datasets hub: an efficient starting point for analyzing genome-wide studies in GenePattern, Integrative Genomics Viewer, and R/Bioconductor. *Genome Biol.*, 2012, vol. 13, no. 11, pp. R104.
14. Crabtree J. et al. Circleator: flexible circular visualization of genome-associated data with BioPerl and SVG. *Bioinformatics*, 2014, vol. 30, no. 21, pp. 3125-3127.
15. Di Tommaso P. et al. Cloud-Coffee: implementation of a parallel consistency-based multiple alignment algorithm in the T-Coffee package and its benchmarking on the Amazon Elastic-Cloud. *Bioinformatics*, 2010, vol. 26, no. 15, pp. 1903-1904.
16. Di Tommaso P. et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Reseach*, 2011, vol. 39, no. 1, pp. W13-W17.
17. Etherington G.J. et al. Bio-samtools 2: a package for analysis and visualization of sequence and alignment data with SAMtools in Ruby. *Bioinformatics*, 2015, no. 1, pp. 1-12.



18. Famiglietti M.L. et al. Genetic variations and diseases in UniProtKB/Swiss-Prot: the ins and outs of expert manual curation. *Hum. Mutat.*, 2014, vol. 35, no. 8, pp. 927-935.
19. Fan S.B. et al. Using pLink to Analyze Cross-Linked Peptides. *Curr. Protoc. Bioinformatics*, 2015, vol. 49, pp. 1-8.
20. Fukami-Kobayashi K., Saito N. How to make good use of CLUSTALW. *Tanpakushitsu Kakusan Koso*, 2002, vol. 47, no. 9, pp. 1237-1239.
21. Gaudet P. et al. Collaborative annotation of genes and proteins between UniProtKB/Swiss-Prot and dictyBase. Database (Oxford). 2009, vol. 2009, pp. bap016.
22. Golosova O. et al. Unipro UGENE NGS pipelines and components for variant calling, RNA-seq and ChIP-seq data analyses. *Peer. J.*, 2014, vol. 2, pp. e644.
23. Goto N. et al. BioRuby: bioinformatics software for the Ruby programming language. *Bioinformatics*, 2010, vol. 26, no. 20, pp. 2617-2619.
24. Guo C. et al. Ugene, a newly identified protein that is commonly overexpressed in cancer and binds uracil DNA glycosylase. *Cancer Res.*, 2008, vol. 68, no. 15, pp. 6118-6126.
25. Harmsen D. et al. RIDOM: comprehensive and public sequence database for identification of Mycobacterium species. *BMC Infect. Dis.*, 2003, vol. 3, pp. 26.
26. Harmsen D. et al. RIDOM: Ribosomal differentiation of medical micro-organisms database. *Nucleic Acids Research*, 2002, vol. 30, no. 1, pp. 416-417.
27. Harris S.R. et al. Read and assembly metrics inconsequential for clinical utility of whole-genome sequencing in mapping outbreaks. *Nat. Biotechnol.*, 2013, vol. 31, no. 7, pp. 592-594.
28. Holland R.C. et al. BioJava: an open-source framework for bioinformatics. *Bioinformatics*, 2008, vol. 24, no. 18, pp. 2096-2097.
29. Huelsenbeck J.P., Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 2001, vol. 17, no. 8, pp. 754-755.
30. Jayadeepa R.M. et al. Review and research analysis of computational target methods using BioRuby and in silico screening of herbal lead compounds against pancreatic cancer using R programming. *Curr. Drug. Metab.*, 2014, vol. 15, no. 5, pp. 535-543.
31. Jungo F. et al. The UniProtKB/Swiss-Prot Tox-Prot program: A central hub of integrated venom protein data. *Toxicon*. 2012, vol. 60, no. 4, pp. 551-557.
32. Kaisers W. et al. rbamtools: an R interface to samtools enabling fast accumulative tabulation of splicing events over multiple RNA-seq samples. *Bioinformatics*, 2015, vol. 31, no. 10, pp. 1663-1664.
33. Katoh K. et al. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 2002, vol. 30, no. 14, pp. 3059-3066.
34. Kim D., Salzberg S.L. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol.*, 2011, vol. 12, no. 8, pp. R72.
35. Kim K.U. et al. Comparison of functional gene annotation of *Toxascaris leonina* and *Toxocara canis* using CLC genomics workbench. *Korean J. Parasitol.*, 2013, vol. 51, no. 5, pp. 525-530.
36. Kleindienst R. et al. Highly efficient refractive Gaussian-to-tophat beam shaper for compact terahertz imager. *Appl. Opt.*, 2010, vol. 49, no. 10, pp. 1757-1763.
37. Kohli D.K., Bachhawat A.K. CLOURE: Clustal Output Reformatter, a program for reformatting ClustalX/ClustalW outputs for SNP analysis and molecular systematics. *Nucleic Acids Res.*, 2003, vol. 31, no. 13, pp. 3501-3502.
38. Li H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 2009, vol. 25, no. 16, pp. 2078-2079.
39. Li K.B. ClustalW-MPI: ClustalW analysis using distributed and parallel computing. *Bioinformatics*, 2003, vol. 19, no. 12, pp. 1585-1586.
40. Lim A., Zhang L. WebPHYLP: a web interface to PHYLIP. *Bioinformatics*, 1999, vol. 15, no. 12, pp. 1068-1069.
41. Lima T. et al. HAMAP: a database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot. *Nucleic. Acids Res.*, 2009, vol. 37, pp. 471-478.
42. Ling C. et al. MrBayes tgMC(3): a tight GPU implementation of MrBayes. *PLoS One*, 2013, vol. 8, no. 4, pp. e60667.
43. Magis C. et al. T-Coffee: Tree-based consistency objective function for alignment evaluation. *Methods Mol. Biol.*, 2014, vol. 1079, pp. 117-129.
44. McCormick R.E. et al. RIG: Recalibration and Interrelation of Genomic Sequence Data with the GATK. G3 (Bethesda). 2015, vol. 5, no. 4, pp. 655-665.
45. Mellmann A. et al. Evaluation of RIDOM, MicroSeq, and Genbank services in the molecular identification of *Nocardia* species. *Int. J. Med. Microbiol.*, 2003, vol. 293, no. 5, pp. 359-370.
46. Notredame C. Computing multiple sequence/structure alignments with the T-coffee package. *Curr. Protoc., Bioinformatics*, 2010, vol. 3, pp. 1-25.
47. Notredame C. et al. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, 2000, vol. 302, no. 1, pp. 205-217.
48. Notredame C., Suhre K. Computing multiple sequence/structure alignments with the T-coffee package. *Curr. Protoc., Bioinformatics*, 2004, vol. 3, pp. 3-8.
49. Octavia S. et al. Delineating community outbreaks of *Salmonella enterica* serovar Typhimurium by use of whole-genome sequencing: insights into genomic variability within an outbreak. *J. Clin. Microbiol.*, 2015, vol. 53, no. 4, pp. 1063-1071.
50. Ogden T.H., Rosenberg M.S. Alignment and topological accuracy of the direct optimization approach via POY and traditional phylogenetics via ClustalW + PAUP\*. *Syst. Biol.*, 2007, vol. 56, no. 2, pp. 182-193.
51. Okonechnikov K. et al. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, 2012, vol. 28, no. 8, pp. 1166-1167.
52. Oliver T. et al. Using reconfigurable hardware to accelerate multiple sequence alignment with ClustalW. *Bioinformatics*, 2005, vol. 21, no. 16, pp. 3431-3432.
53. Pantosti A. Whole-genome sequencing may be key to abating hospital-based methicillin-resistant *Staphylococcus aureus* outbreaks. *J. Pediatr.* 2013, vol. 162, no. 5, pp. 1079-80.
54. Pollier J. et al. Analysis of RNA-Seq data with TopHat and Cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cultures. *Methods Mol. Biol.*, 2013, vol. 1011, pp. 305-315.
55. Prlic A. et al. BioJava: an open-source framework for bioinformatics in 2012. *Bioinformatics*, 2012, vol. 28, no. 20, pp. 2693-2695.
56. Purcell S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, 2007, vol. 81, no. 3, pp. 559-575.
57. Ramirez-Gonzalez R.H. et al. Bio-samtools: Ruby bindings for SAMtools, a library for accessing BAM files containing high-throughput sequence alignments. *Source Code Biol. Med.*, 2012, vol. 7, no. 1, pp. 6.
58. Renteria M.E. et al. Using PLINK for Genome-Wide Association Studies (GWAS) and data analysis. *Methods Mol. Biol.*, 2013, vol. 1019, pp. 193-213.
59. Retief J.D. Phylogenetic analysis using PHYLIP. *Methods Mol. Biol.*, 2000, vol. 132, pp. 243-258.
60. Rius J. et al. A user-friendly web portal for T-Coffee on supercomputers. *BMC Bioinformatics*, 2011, vol. 12, pp. 150.
61. Robinson J.T. et al. Integrative genomics viewer. *Nat. Biotechnol.*, 2011, vol. 29, no. 1, pp. 24-26.

62. Ronquist F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.*, 2012, vol. 61, no. 3, pp. 539-542.
63. Ronquist F., Huelsenbeck J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 2003, vol. 19, no. 12, pp. 1572-1574.
64. Ropelewski A.J. et al. MPI-PHYLIP: parallelizing computationally intensive phylogenetic analysis routines for the analysis of large protein families. *PLoS One*, 2010, vol. 5, no. 11, pp. e13999.
65. Rothganger J. et al. RIdom TraceEdit: a DNA trace editor and viewer. *Bioinformatics*, 2006, vol. 22, no. 4, pp. 493-494.
66. Schneider M. et al. The UniProtKB/Swiss-Prot knowledgebase and its Plant Proteome Annotation Program. *J. Proteomics*, 2009, vol. 72, no. 3, pp. 567-573.
67. Shi H. et al. Using Fisher's method with PLINK «LD clumped» output to compare SNP effects across Genome-wide Association Study (GWAS) datasets. *Int. J. Mol. Epidemiol., Genet.* 2011, vol. 2, no. 1, pp. 30-35.
68. Stajich J.E. An Introduction to BioPerl. *Methods Mol. Biol.*, 2007, vol. 406, pp. 535-548.
69. Stajich J.E. et al. The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.*, 2002, vol. 12, no. 10, pp. 1611-1618.
70. Talevich E. et al. Bio.Phylo: a unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. *BMC Bioinformatics*, 2012, vol. 13, pp. 209.
71. Taly J.F. et al. Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. *Nat. Protoc.*, 2011, vol. 6, no. 11, pp. 1669-1682.
72. Thompson J.D. et al. Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics*, 2002, vol. 2, pp. Unit 2.3.
73. Thorvaldsdottir H. et al. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.*, 2013, vol. 14, no. 2, pp. 178-192.
74. Trapnell C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.*, 2012, vol. 7, no. 3, pp. 562-578.
75. Trapnell C. et al. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 2009, vol. 25, no. 9, pp. 1105-1111.
76. Vangala R.K. et al. BioParishodhana: A novel graphical interface integrating BLAST, ClustalW, primer3 and restriction digestion tools. *Bioinformatics*, 2012, vol. 8, no. 13, pp. 639-643.
77. Vaskin Y.Y. et al. ExpertDiscovery and UGENE integrated system for intelligent analysis of regulatory regions of genes. *In Silico. Biol.*, 2011, vol. 11, no. 3-4, pp. 97-108.
78. Walker T.M. et al. Contact investigations for outbreaks of *Mycobacterium tuberculosis*: advances through whole genome sequencing. *Clin. Microbiol. Infect.*, 2013, vol. 19, no. 9, pp. 796-802.
79. Walker T.M. et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect. Dis.*, 2013, vol. 13, no. 2, pp. 137-146.
80. Wallace I.M. et al. M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Res.*, 2006, vol. 34, no. 6, pp. 1692-1699.
81. Wang L.T. et al. Functional interaction of Ugene and EBV infection mediates tumorigenic effects. *Oncogene*, 2011, vol. 30, no. 26, pp. 2921-2932.
82. Wheelan S.J. et al. Spidey: a tool for mRNA-to-genomic alignments. *Genome Res.*, 2001, vol. 11, no. 11, pp. 1952-1957.
83. Zhou J. et al. MrBayes on a graphics processing unit. *Bioinformatics*, 2011, vol. 27, no. 9, pp. 1255-1261.
84. Zhu P. et al. OTG-snp caller: an optimized pipeline based on TMAP and GATK for SNP calling from ion torrent data. *PLoS One*, 2014, vol. 9, no. 5, pp. e97507.

#### ДЛЯ КОРРЕСПОНДЕНЦИИ:

Спринджук М. В.  
ГНУ «Объединенный институт проблем информатики  
НАН Беларуси»,  
220012, г. Минск, ул. Сурганова, д. 6.  
E-mail: msprindzhuk@mail.ru