

INTRODUCTION

Open Access



Bioinformatics research at BGRS-2018

Tatiana V. Tatarinova^{1,2,3,4}, Ming Chen⁵ and Yuriy L. Orlov^{6,7,8*}

From 11th International Multiconference “Bioinformatics of Genome Regulation and Structure\Systems Biology” - BGRS\SB-2018

Novosibirsk, Russia. 20-25 August 2018

This thematic issue of BMC Bioinformatics continues the series of BioMed Central special post-conference journal issues presenting materials from the bioinformatics and systems biology summits BGRS\SB (Bioinformatics of Genome Regulation and Structure\Systems Biology). BGRS/SB conference is biannually conducted in Novosibirsk since 1998. In this issue we present five selected papers from the XIth BGRS\SB-2018 multi-conference (<http://conf.bionet.nsc.ru/bgrssb2018/en/>). This Special Issue is accompanied by other BioMed Central Special Issues collecting presented works in the fields of genomics, evolutionary biology, plant biology, and genetics, published as BMC Genomics, BMC Evolutionary Biology, BMC Systems Biology and BMC Plant Biology supplements [1–4]. Recent works at Belyaev Conference-2017 (“Belyaev Readings-2017” - <http://conf.bionet.nsc.ru/belyaev100/en>) in Novosibirsk, Russia were presented at BioMed Central as well [5–7]. BGRS\SB-2018 included several symposia: “Cognitive Sciences, Genomics and Bioinformatics” (SCGB-2018), “Systems biology and biomedicine” (SbioMed-2018), “Biodiversity: genomics and evolution” (BioGenEvo-2018), “Mathematical modeling and high-performance computing in bioinformatics, biomedicine and biotechnology” (MM-HPC-BBB-2018), “Systems biology of DNA repair processes and programmed cell death” (SbPCD-2018). Each session had presentations of bioinformatics applications.

Current special post-conference issue contains selected works on bioinformatics ranging from software for scientific literature mining to applications in plant biology. Below is a brief summary of the papers in this special issue.

Ivanisenko et al. [8] presented a new version of the popular ANDSYSTEM tool for automatic text mining of scientific publications, this time equipped with expanded functionality. Currently, there is a number of

commercial automated services allowing users to reconstruct molecular-genetic networks using the data automatically extracted from the texts of scientific publications, for example: STRING (<https://string-db.org>), Pathway Commons (<https://www.pathwaycommons.org/>), MetaCore (<https://portal.genego.com/>), and Ingenuity (<https://www.qiagenbioinformatics.com/>). Presented tool ANDSYSTEM reconstructs associative gene networks taking into account the tissue-specific gene expression. The system allows the reconstruction of combined gene networks, as well as performing the filtering of genes tissue-specific expression. As an example of the application of such filtering, gene network of the extrinsic apoptotic signaling pathway was analyzed. Note that previous publication of this tool was at BMC Systems Biology [9], and recent applications was published in BMC Medical Genomics by Saik et al. [10].

Ranajit Das and Priyanka Upadhyai [11] showed application of the Geographic Population Structure (GPS) and reAdmix algorithms [12, 13] to biogeographic analyses of captive gorillas. The Geographic Population Structure algorithm is an admixture based tool for inference of provenance and has been previously employed for the geo-localization of various human populations worldwide [14–17]. Given the strong correspondence between geography and genetics, a number of strategies have focused on the delineation of the precise geographic origin of human populations using high-resolution genetic data. Das and Upadhyai [11] applied the GPS tool for localization of the ancestral origins of wild and captive gorilla genomes, of unknown geographic source, available in the Great Ape Genome Project [18]. Determination of the source population of captive gorillas can provide valuable information to guide breeding programs and ensure their appropriate management at the population level. Finally, the authors’ findings shine light on the broader applicability of GPS for protecting the genetic integrity of endangered non-human species.

Fedor Kazantsev and co-authors [19] presented an application in plant biology - the database on Molecular

* Correspondence: orlov@bionet.nsc.ru

⁶Institute of Cytology and Genetics SB RAS, 630090 Novosibirsk, Russia

⁷Novosibirsk State University, 630090 Novosibirsk, Russia

Full list of author information is available at the end of the article



Identification of Genes for Resistance in Wheat (abbreviated as MIGREW). Wheat is one of the leading crops worldwide. The wheat pathogen complex affecting the plant organs containing chlorophyll is represented by the following species: *Puccinia triticina* causing leaf rust, *Puccinia graminis* causing stem rust, and *Blumeria graminis* causing powdery mildew. Population structure of fungal pathogens depends on environment and wheat ecotype. Taking into account the evolution of host-pathogen interactions, genetic diversity of wheat and fungus must be monitored. MIGREW database is developed using classical Model-View-Controller architecture. The model layer is the PostgreSQL database containing sixteen tables. The Controller layer is the Java application designed using *spring.io* libraries that performs REST API access to the data. The MIGREW database has been developed to present in single web-based interface the information on fungi-wheat objects keeping the data available for users with different requests, breeders and plant pathologists.

Kuzmin et al. [20] presented a result of a challenging project – assembly of the Siberian larch nuclear genome. Conifers have large genomes (~12–30 Gbp, which is 4–10 times larger than the human genome) containing ~80% of repetitive DNA. Using a new stepwise de novo assembling method presented in the paper, the genome of Siberian larch, *Larix sibirica* Ledeb. was for the first time completely assembled using de novo assembler by the CLC Assembly Cell. Sequencing and computational difficulties make it the first larch genome, and the sixth conifer genome assembly. The approach presented by Kuzmin et al. paves a road for assembling of very large genomes with a reasonable computing time and without engaging huge computing resources. The assemblies produced by this approach are of reasonable quality allowing their annotation and further use.

Evgenia Bondar and co-authors [21] analyzed chloroplast genome of the Siberian larch. Illumina sequencing reads were processed using the Bowtie2 [22] mapping program and assembled with the SPAdes genomic assembler [23]. Genome annotation was performed using the RAST server [24–26]. GMATo program was used for the SSRs search, and the Bowtie2 and UGENE [27] programs for the SNPs detection. This is the first effort to sequence and assemble the complete chloroplast genome sequence of Siberian larch. This assembly provides a reference for chloroplast resequencing and search for additional genetic markers using population samples. It will be useful for further phylogenetic and gene flow studies in conifers.

Therefore, this issue includes reports of recent bioinformatics application in text mining, animal genetics, database and algorithm development for computational plant sciences. BGRS\SB-2018 multi-conference had several parallel symposia, sessions and workshops, including First Sino-Russian Workshop on Integrative Bioinformatics and Systems Biology (<http://conf.bionet.nsc.ru/srw2018/en/>)

and international Round table on education in bioinformatics. Other related computational biology works are presented in parallel BioMed Central issues by 2018. The conference was completed by Young Scientists School “Systems Biology and Bioinformatics” (SBB-2018) (<http://conf.bionet.nsc.ru/bgrssb2018/en/school/>). BioMed Central previously had published special issues by materials of SBB Schools. We invite our readers worldwide to attend our next event - Systems Biology and Bioinformatics Young Scientists School in summer 2019 in Novosibirsk, Russia.

Acknowledgements

We are grateful to Professors N.A. Kolchanov and N.A. Kochetov for organization of the multi-conference and providing platform for international bioinformatics research. We thank the Russian Foundation of Basic Research for the conference organization support, Zhejiang bioinformatics Society, China, for logistic support of conference participants, Institute of Cytology and Genetics and Novosibirsk State University for hosting the conference. Round table on education in bioinformatics was supported by the project “Investigation, analysis and complex independent expertise of projects of the National technological initiatives, including the accompanying of projects of “road map” “NeuroNet” (state assignment 28.12487.2018/12.1 of the Ministry of Higher Education and Science of the Russian Federation) and grant #14.W03.31.0015.

The guest editors of the special issue are grateful to the conference committee members and reviewers who helped in the articles editing and issue preparation: Ancha Baranova (George Mason University, USA), Guoliang Li (Huazhong Agricultural University, China), Matteo Barberis (University of Amsterdam, Netherlands), Hongjun Chen (Zhejiang University, Hangzhou, China), Andrey Ptitsyn (Gloucester Marine Genomics Institute, MA, USA), Anna Kudryavtseva (Engelhardt Institute of Molecular Biology of the RAS, Russia), Fatima Cilingir (University of La Verne, USA), Leonid Brodsky (University of Haifa, Israel), Vadim Nimaev (Research Institute of Clinical and Experimental Lymphology SB RAS, Novosibirsk, Russia), Todd Lorenz (University of La Verne, CA, USA), Olga Zolotareva (Bielefeld University, Germany), Mikhail Pyatnitskiy (Orekhovich Institute of Biomedical Chemistry, Moscow, Russia), Nick Alexandrov (San Diego Supercomputer Center, University of California, USA), Nina Oparina (Karolinska Institute, Sweden).

Funding

Publication of this article was not covered by sponsorship.

About this supplement

This article has been published as part of *BMC Bioinformatics Volume 20 Supplement 1, 2019: Selected articles from BGRS\SB-2018: bioinformatics*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-20-supplement-1>.

Authors' contributions

TT and YO are guest editors of the special post-conference issues. TT, MC and YO are Program Committee members of BGRS\SB-2018 conference and First Sino-Russian workshop on integrative bioinformatics and systems biology. All the authors read, revised and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Biology, University of La Verne, La Verne, CA, USA. ²Department of Fundamental Biology and Biotechnology, Siberian Federal University, 660074 Krasnoyarsk, Russia. ³Vavilov Institute of General Genetics RAS, Moscow, Russia. ⁴A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia. ⁵Zhejiang University, Hangzhou 310058, China. ⁶Institute of Cytology and Genetics SB RAS, 630090 Novosibirsk, Russia. ⁷Novosibirsk State University, 630090

Novosibirsk, Russia. ⁸A.O. Kovalevsky Institute of Marine Biological Research of RAS, 119334 Moscow, Russia.

Published: 5 February 2019

References

- Orlov YL, Baranova AV, Markel AL. Computational models in genetics at BGRS\SB-2016: introductory note. *BMC Genet.* 2016;17(Suppl 3):155.
- Orlov YL, Baranova AV, Hofstadt R, Kolchanov NA. Computational genomics at BGRS\SB-2016: introductory note. *BMC Genomics.* 2016; 17(Suppl 14):996.
- Baranova AV, Orlov YL. Evolutionary biology at BGRS\SB-2016. *BMC Evol Biol.* 2017;17(Suppl 1):21.
- Orlov YL, Baranova AV, Salina EA. Computational plant bioscience at BGRS\SB-2016: introductory note. *BMC Plant Biol.* 2016;16(Suppl 3):243.
- Orlov YL, Baranova AV, Herbeck YE. Evolutionary biology at Belyaev conference - 2017. *BMC Evol Biol.* 2017;17(Suppl 2):260.
- Orlov YL, Baranova AV, Tatarinova TV, Kolchanov NA. Genetics at Belyaev conference - 2017: introductory note. *BMC Genet.* 2017;18(Suppl 1):116.
- Orlov YL, Baranova AV, Chen M, Salina EA. Plant biology at Belyaev conference - 2017. *BMC Plant Biol.* 2017;17(Suppl 2):257.
- Ivanisenko VA, Demenkov PS, Ivanisenko TV, Mishchenko EL, Saik OV. A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. *BMC Bioinformatics.* 2019. <https://doi.org/10.1186/s12859-018-2567-6>.
- Ivanisenko VA, Saik OV, Ivanisenko NV, Tiys ES, Ivanisenko TV, Demenkov PS, Kolchanov NA. ANDSystem: an associative network discovery system for automated literature mining in the field of biology. *BMC Syst Biol.* 2015; 9(Suppl 2):S2.
- Saik OV, Demenkov PS, Ivanisenko TV, Bragina EY, Freidin MB, Goncharova IA, Dosenko VE, Zolotareva OI, Hofstaedt R, Lavrik IN, et al. Novel candidate genes important for asthma and hypertension comorbidity revealed from associative gene networks. *BMC Med Genet.* 2018;11(Suppl 1):15.
- Das R, Upadhyai P. Application of the geographic population structure (GPS) algorithm for biogeographical analyses of wild and captive gorillas. *BMC Bioinformatics.* 2019. <https://doi.org/10.1186/s12859-018-2568-5>.
- Kozlov K, Chebotarev D, Hassan M, Triska M, Triska P, Flegontov P, Tatarinova TV. Differential evolution approach to detect recent admixture. *BMC Genomics.* 2015;16(Suppl 8):S9.
- Elhaik E, Tatarinova T, Chebotarev D, Piras IS, Maria Calo C, De Montis A, Atzori M, Marini M, Tofanelli S, Francalacci P, et al. Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun.* 2014;5:3513.
- Triska P, Chekanov N, Stepanov V, Khusnutdinova EK, Kumar GPA, Akhmetova V, Babalyan K, Boulygina E, Kharkov V, Gubina M, et al. Between Lake Baikal and the Baltic Sea: genomic history of the gateway to Europe. *BMC Genet.* 2017;18(Suppl 1):110.
- Chekalina E, Rubanovich A, Tatarinova TV, Kasianov A, Bender N, Chekalina M, Staub K, Koepke N, Ruhl F, Bruskin S, et al. Changes in biological pathways during 6,000 years of civilization in Europe. *Mol Biol Evol.* 2018.
- ArunKumar G, Tatarinova TV, Duty J, Rollo D, Syama A, Arun VS, Kavitha VJ, Triska P, Greenspan B, Wells RS, et al. Genome-wide signatures of male-mediated migration shaping the Indian gene pool. *J Hum Genet.* 2015; 60(9):493–9.
- Flegontov P, Changmai P, Zidkova A, Logacheva MD, Altinisik NE, Flegontova O, Gelfand MS, Gerasimov ES, Khrameeva EE, Konovalova OP, et al. Genomic study of the Ket: a paleo-Eskimo-related ethnic group with significant ancient north Eurasian ancestry. *Sci Rep.* 2016;6:20768.
- Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, et al. Great ape genetic diversity and population history. *Nature.* 2013;499(7459):471–5.
- Kazantsev FV, Skolotneva ES, Kelbin VN, Salina EA, Lashin SA. MIGREW: database on molecular identification of genes for resistance in wheat. *BMC Bioinformatics.* 2019. <https://doi.org/10.1186/s12859-018-2569-4>.
- Kuzmin DA, Feranchuk SI, Sharov W, Cybin AN, Makolov SV, Putintseva YA, Oreshkova NV, Krutovsky KV. Stepwise large genome assembly approach: a case of Siberian larch (*Larix sibirica* Ledeb.) *BMC Bioinformatics.* 2019. <https://doi.org/10.1186/s12859-018-2570-y>.
- Bondar EI, Putintseva YA, Oreshkova NV, Krutovsky KV. Siberian larch (*Larix sibirica* Ledeb.) chloroplast genome and development of polymorphic chloroplast markers. *BMC Bioinformatics.* 2019. <https://doi.org/10.1186/s12859-018-2571-x>.
- Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9(4):357–9.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J computational biology : a journal of computational molecular cell biology.* 2012;19(5):455–77.
- Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics.* 2008;9:386.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. The RAST server: rapid annotations using subsystems technology. *BMC Genomics.* 2008;9:75.
- Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F. Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harb Protoc.* 2010;2010(1):pdb prot5368.
- Okonechnikov K, Golosova O, Fursov M, the UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics.* 2012;28(8):1166–7.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

