BGRS\SB-2018

The 11th International Conference

Biodiversity: Genomics and Evolution, BioGenEvo-2018

Symposium

BioGenEvo-2018
NOVOSIBIRSK, RUSSIA
21–24 AUGUST, 2018

CONF.BIONET.NSC.RU/BIOGENEVO

A.O. Kovalevsky Institute of Marine Biological Research
of Russian Academy of Sciences

Nikitsky Botanical Gardens – National Research Centre
of Russian Academy of Sciences

Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences

Novosibirsk State University

BIODIVERSITY: GENOMICS AND EVOLUTION (BioGenEvo-2018)

Symposium

Abstracts

21–24 August, 2018
Novosibirsk, Russia

Novosibirsk
ICG SB RAS
2018

УДК 57
В69

Biodiversity: Genomics and Evolution (BioGenEvo-2018) : Symposium (21–24 Aug. 2018, Novosibirsk, Russia); Abstracts / A.O. Kovalevsky Institute of Marine Biological Research, Russian Academy of Sciences; Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences. – Novosibirsk: ICG SB RAS, 2018. – 61 pp. – ISBN 978-5-91291-041-8.

Program Committee

Chairman

Leonid L. Moroz – Distinguished Professor of Neuroscience, Genetics, Biology, Chemistry University of Florida (USA), scientific adviser of Laboratory of Biodiversity and Functional Marine Genomics (IMBR RAS)

Members

A.D. Afonnikov, PhD, ICG SB RAS, Novosibirsk; IMBR RAS, Sevastopol, Russia
I.V. Dovgal, PhD, Dr.Sci., Professor, IMBR RAS, Sevastopol, Russia
N.A. Kolchanov, PhD, Dr.Sci., Professor, Academician of RAS, ICG SB RAS, Novosibirsk, Russia
I.V. Mitrofanova, PhD, Dr.Sci., NBG-NRC RAS, Yalta, Russia
Yu.K. Novozhilov, PhD, Dr.Sci., BIN RAS, St. Petersburg, Russia
Y.L. Orlov, PhD, Dr.Sci., Professor, ICG SB RAS, Novosibirsk; IMBR RAS, Sevastopol, Russia
Detlev Arendt, PhD, EMBL, Germany
Olga Krebs, PhD, Heidelberg Institute for Theoretical Studies, Germany
Alexey Kolodkin, PhD, University of Luxembourg, Luxembourg, Luxembourg
Gustav Pauley, PhD, professor, Florida Museum of Natural History, USA
Andrey Ptitsyn, PhD, Gloucester Marine Genomics Institute, USA
Tatiana Tatarinova, Prof., University of La Verne, CA, USA
Katrine Worsaae, PhD, Associate Professor, University of Copenhagen, Denmark
Maxim Zakhartsev, PhD, Centre for Integrative Genetics [CIGENE], Norwegian University of Life Sciences Ås, Norway
Yubo Zhang, PhD, Agricultural Genomics Institute at ShenZhen, CAAS, China

Organizing Committee

Chairman

Olga V. Krivenko, PhD, adviser of Laboratory of Biodiversity and Functional Marine Genomics (IMBR RAS)

Members

E.A. Vodyasova, IMBR RAS, Sevastopol, Russia
O.N. Kuleshova, IMBR RAS, Sevastopol, Russia
E.S. Chelebieva, PhD, IMBR RAS, Sevastopol, Russia
D.A. Rasskazov, ICG SB RAS, Novosibirsk, Russia
D.Yu. Romanova, PhD, IMBR RAS, Sevastopol, Russia
V.V. Starunov, St. Petersburg State University, St. Petersburg, Russia
O.N. Shchepinov, BIN RAS, St. Petersburg, Russia

Contact

e-mail: bio-fgmo@yandex.ru

ISBN 978-5-91291-041-8

Organizers

A.O. Kovalevsky Institute of Marine Biological Research of RAS

Nikitsky Botanical Gardens – National Research Centre of RAS

Federal Research Center Institute of Cytology and Genetics SB RAS

Novosibirsk State University (NSU)

The Federal Agency for Scientific Organizations (FASO Russia)

Sponsors

Russian Foundation for Basic Research
Grant No. 18-04-20048

Federal Agency for Scientific Organizations FASO Russia

Ministry of Education and Science of the Russian Federation (Minobrnauki of Russia) Grant No. ДНИТ 28.12487.2018/12.1

GOLD SPONSORS

HUAWEI

intel Software

SILVER SPONSORS

MP Biomedicals

Albiogen

Roche Diagnostics Rus Ltd.

Skygen, Ltd

Bioline, Llc

Khimexpert Ltd.

DIA-M, Ltd

BASIC SPONSORS

geneXplain GmbH

In Silico Biology
Journal of Biological Systems Modeling and Multi-Scale Simulation

IOS Press

Contents

DOI 10.18699/BioGenEvo-2018-01

Effect of stress-related hormones on host *Drosophila* fitness depends on endosymbiont *Wolbachia* genotype

N. Adonyeva*, E. Burdina, N. Gruntenko, I. Rauschenbach
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** e-mail: nadon@bionet.nsc.ru*

Key words: heat stress, juvenile hormone, 20-hydroxyecdysone, stress-resistance, *Drosophila*, *Wolbachia*

Motivation and Aim: It is the first time the influence of gonadotropins, juvenile hormone (JH) and 20-hydroxyecdysone (20E) on the heat stress resistance of wild-type *D. melanogaster* females infected with different genotypes of α-proteobacteria *Wolbachia pipientis* was investigated to verify the possibility that endosymbiont promotes adaptation to the host species, affecting the functioning of its protective systems.
Methods and Algorithms: To raise the JH level, we applied 0.1 μg JHIII (Sigma-Aldrich), dissolved in 0.5 μl acetone, to the abdomen of 4-day females, and 0.5 μl pure acetone to control individuals. In order to investigate the effect of level 20E changes on the resistance to heat stress of *D. melanogaster* females infected with different strains of *Wolbachia*, on the 4th day after eclosion the flies were placed in viales, the bottom and 1 cm of the walls of which were covered with filter paper, soaked with 0.5 ml of culture medium containing 0.5 % sucrose, 0.2 % yeast and 60 μg 20E (Sigma-Aldrich). In the control series in a solution of 20E was not added. On the 6th day after eclosion the JH or 20E-treated and control flies were subjected to thermal stress (38 °C, 2 hours 45 min) and 24 hours later the number of surviving flies were counted.
Results: It was found that the experimental increase in the level of the JH causes a decrease in resistance of 6-day females to heat stress, while the increase in the level of 20E – its increase, both in uninfected Wolbachia and infected with various strains of bacteria: *wMelCS*, *wMelPop* and *wMel*. However, the intensity of the response differs: in females infected with pathogenic strain *wMelPop*, the decline in survival at higher levels of the JH and its increase with increasing levels of 20E are more pronounced, while in females infected with genotype *wMelCS* is less pronounced than that in uninfected females and females infected with genotype *wMel*.
Conclusion: The obtained data indicate that the genotype of *wMelCS* causes a decrease, and *wMelPop* – an increase in the level of stress hormone, dopamine, in female *Drosophila*, since earlier we showed that an increase in the level of the JH in mature females increases the level of dopamine, while an increase in the level of 20E reduces it [1], and an increase/decrease in the level of dopamine, in turn, leads to a decrease/increase in the resistance of female *Drosophila* to heat stress [2].
Acknowledgements: Supported by the RFBR (16-04-00060) and State budget project No. 0324-2016-0002.

References

1. Gruntenko N.E., Rauschenbach I.Yu. (2008) Interplay of juvenile hormone, 20-hydroxyecdisone and biogenic amines under normal and stress conditions and its effect on reproduction. J. Insect. Physiol. 56:902-908.
2. Gruntenko N.E. et al. (2004) The effect of mutations altering biogenic amine metabolism in Drosophila on viability and the response to heat stress. Arch. Insect Biochem. Physiol. 55:55-67.

DOI 10.18699/BioGenEvo-2018-02

Genetic diversity of *Plantago major* L. local populations in the habitats of Nizhniy Tagil different in the level of technogenic load

E. Artemenko*, E. Zhuikova, I. Kiseleva
Ural Federal University, Yekaterinburg, Russia
** e-mail: lisaaartemenko@gmail.com*

Key words: *Plantago major*, genetic diversity, ISSR-analisis

Motivation and Aim: It has been shown that different types of pollutants can affecton genetic diversity of populations. Technogenic load can lead to changes in frequency of mutations, direction of selection of genotypes and/or genetic drift as a result of elimination of a large number of non-adapted individuals [1]. The aim is to study the level of genetic diversity of local populations of *Plantago major* growing in 5 location of Nizhny Tagil, differing in the level of technogenic load.
Methods and Algorithms: NA was isolated by the method of S. Porebski [2] with some modifications. Nine UBC primers was used for PCR. Visualization of PCR results was carried out by horizontal electrophoresis on a 1.2 % agarose gel in addition of ethidium bromide in 1 × TBE buffer. The result of all electrophoregrams were processed by the ImageJ program. The presence and absence matrices for each primer were processed in PAST [3] and JeneAlex [4] program.
Results: 52 samples of *P. major* were analyzedin the study; standardization of the ISSR protocol, quantitative and qualitative assessments of the NAwere done; the maximum, minimum and average number of bands per sample, habitat and primer, as well as indicators of genetic diversity: the percentage of polymorphic loci, effective alleles, expected heterozygosity, the Shannon's information index, the genetic distance and identity of Nei, the percentage of molecular variance within and among populations.
Conclusion: Genetic analysis of *P. major* subpopulations indicates that local populations in contaminated technogenic are less diverse genetically than population of background areas. Since genetic diversity is lower among plants in contaminated areas, despite their potentially higher level of mutational variability, the mutagenic effects of pollutants may not be the driving force behind the formation of genetic differences found between different populations.

References

1. Bae Ch.-H., Ryu J. (2012) Genetic diversity and relationship analysis of genus *Taraxacum* accessions collected in Korea. Korean J. Plant Res. 25(3):329-338.
2. Porebski S., Bailey L.G., Baum B.R. (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharides and polyphenol component. Plant Molecular Biology Reporter. 15:8-15.
3. Hammer O., Harper D.A.T., Ryan P.D. (2001) PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica. 4(1):9.
4. Peakall R., Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics. 28:2537-2539.

DOI 10.18699/BioGenEvo-2018-03

Genetic diversity of *Taraxacum officinale* Wigg. local populations in the habitats of Nizhniy Tagil different in the level of technogenic load

E. Artemenko*, E. Zhuikova, I. Kiseleva
Ural Federal University, Yekaterinburg, Russia
** e-mail: lisaaartemenko@gmail.com*

Key words: *Taraxacum officinale*, genetic diversity, ISSR-analisis

Motivation and Aim: Industrial pollutants are strong stressorsthat can cause degradation of genetic system in population [1]. The aim is to study the level of genetic diversity of local populations of *Taraxacum officinale* growing in the locations of Nizhny Tagil, differed in the level of technogenic impact.
Methods and Algorithms: NA was isolated by the method of S. Porebski [2] with some modifications. Eight UBC primers was used for PCR. Visualization of PCR results was carried out by horizontal electrophoresis on a 1.2 % agarose gel in addition of ethidium bromide in 1 × TBE buffer. The results of all electrophoregrams were processed by the ImageJ program. The presence and absence matrices for each primer were processed in PAST [3] and JeneAlex [4] programs.
Results: In our study, 92 samples of *T. officinale* were analyzed; standardization of the ISSR protocol, quantitative and qualitative assessments of the NA were done; the maximum, minimum and average number of bands per sample, habitat and primer, as well as indicators of genetic diversity: the percentage of polymorphic loci, effective alleles, expected heterozygosity, the Shannon`s information index, the genetic distance and identity of Nei, the percentage of molecular variance within and among populations .
Conclusion: The genetic analysis of five subpopulations of *T. officinale* revealed that there are no statisticallysignificant differences between these subpopulations. This indicates that in *T. officinale* species, adaptability to heavy metals in the environment is manifested probably by physiological level – acclimation, which does not change the genetic structure of populations.

References

1. Makhnev A.K. (2010) Problems of selection and conservation of forest genetic reserves in the areas of location of large industrial centers in the Urals. Coniferous boreal zone. (1-2):127-130 (in Russian).
2. Porebski S., Bailey L.G., Baum B.R. (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharides and polyphenol component. Plant Molecular Biology Reporter. 15:8-15.
3. Hammer Ø., Harper D.A.T., Ryan P.D. (2001) PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica. 4(1):9.
4. Peakall R., Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537-2539.

DOI 10.18699/BioGenEvo-2018-04

Repeated DNA sequences in genomes of species of the genus *Linum*

N. Bolsheva[1], I. Kirov[2], N. Melnikova[1]*, A. Dmitriev[1], G. Krasnov[1], A. Amosova[1], T. Samatadze[1], O. Yurkevich[1], S. Zoshchuk[1], T. Rozhmina[1, 3], A. Kudryavtseva[1], O. Muravenko[1]

[1] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

[2] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

[3] *All-Russian Research Institute for Flax, Torzhok, Russia*

** e-mail: mnv-4529264@yandex.ru*

Key words: genus *Linum*, flax genome, high-throughput sequencing, repeatome, 30-chromosome flaxes

Motivation and Aim: Members of different sections of the genus *Linum* are characterized by wide variability in size, morphology and number of chromosomes in karyotypes [1]. Since such variability is largely determined by the amount and composition of repeated sequences, we conducted a comparative study of the repeatomes of species from 4 sections forming a clade of blue-flowered flaxes.

Methods and Algorithms: Based on the results of high-throughput genome sequencing performed in this study, as well as available WGS data, bioinformatic analysis of repeated sequences from 12 flax samples was carried out [2].

Results: It was found that the genomes of closely related species, which have a similar karyotype structure, are also similar in the repeatome composition. In contrast, the repeatomes of karyologically distinct species differ significantly. No similar tandem-organized repeats have been identified, with the exception of one common repeat for 16- and 30-chromosome species of the sect. *Linum.* A number of mobile element families have been identified in genomes of all species, among which Athila, *Ty3/gypsy* LTR retrotransposon, was the most abundant. Genomes of 30-chromosome members of the sect. *Linum*, including the cultivated species *L. usitatissimum*, differ noticeably from other species by unusually high number of satellite DNA families, as well as their relative content.

Conclusion: Trends in decreasing the content of dispersed DNA repeats and increasing the content of satellite DNA during the evolution of blue-flowered flax genomes have been revealed. The extremely high content of satellite DNA in the genomes of 30-chromosome flaxes is probably due to their allotetraploid origin. The phylogenetic relationships between the investigated flax species, established on the basis of the similarity of the repeatomes, are in good agreement with the data obtained earlier with the help of other phylogenetic markers.

Acknowledgements: Supported by the RFBR (16-04-01239) and the Program of fundamental research for state academies for 2013–2020 years (Task 0103-2014-0008 Subprogram No. 53 General genetics).

References

1. Bolsheva N.L. et al. (2017) Evolution of blue-flowered species of genus *Linum* based on high-throughput sequencing of ribosomal RNA genes. BMC Evolutionary Biology. 17(Suppl 2):253.
2. Novak P. et al. (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads. Bioinformatics. 29(6):792-793.

DOI 10.18699/BioGenEvo-2018-05

Genetic diversity of cultivated flax based on *CesA* genes

A. Dmitriev[1], T. Rozhmina[1, 2], G. Krasnov[1], A. Snezhkina[1], R. Novakovskiy[1], P. Kezimana[1], N. Bolsheva[1], O. Muravenko[1], A. Kudryavtseva[1], N. Melnikova[1]*
[1] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*
[2] *All-Russian Research Institute for Flax, Torzhok, Russia*
* *e-mail: mnv-4529264@yandex.ru*

Key words: cellulose synthase, *CesA*, *Linum usitatissimum,* flax, high-throughput sequencing, polymorphism, genetic diversity

Motivation and Aim: Cultivated flax (*Linum usitatissimum* L.) is a source of long bast fibres for textile industry, therefore, breeding for optimization of stem content in cellulose-rich bast fibres is conducted. Cellulose synthase genes (*CesA*) play an important role in cellulose biosynthesis [1, 2]. In our study, high-throughput sequencing of *CesA* genes was performed in flax cultivars and lines with varied cellulose content to evaluate genetic diversity of these genes.
Methods and Algorithms: DNA from 48 flax cultivars and lines with known cellulose content from the collection of the All-Russian Research Institute for Flax was extracted. DNA libraries for high-throughput sequencing of *CesA* genes on Illumina platform were prepared using two-stage PCR with primers designed by us. Obtained DNA libraries were quantified on Qubit 2.0 fluorometer, evaluated on Agilent 2100 Bioanalyzer, normalized, pooled, and sequenced on Illumina platform. For data analysis, CLC Genomics Workbench was used. Reads were mapped on reference sequences of *CesA* genes and SNP detection was performed.
Results: High-throughput sequencing of amplicons on Illumina platform allowed us to obtain thousands of reads related to *CesA* genes for each flax genotype that is important for proper polymorphism identification. Single nucleotide polymorphisms (SNPs) were revealed in studied cultivars and lines with varied cellulose content, and the assessment of their genetic diversity was performed.
Conclusion: In the present work, high-throughput sequencing of cellulose synthase genes was performed and polymorphisms within 48 flax cultivars and lines were identified. Our study is important for evaluation of polymorphism of *CesA* genes in cultivated flax, and contributes to determination of associations between particular *CesA* alleles and cellulose content that can be used in marker-assisted breeding.
Acknowledgements: This study was funded by the Presidium of the Russian academy of sciences, Program No. 41 “Biodiveristy of natural systems and biological resources of Russia”.

References

1. Chantreau M. et al. (2015) Functional analyses of cellulose synthase genes in flax (Linum usitatissimum) by virus-induced gene silencing. Plant Biotechnology Journal. 13(9):1312-1324.
2. Galinousky D. et al. (2017) Expression analysis of cellulose synthase and main cytoskeletal protein genes in flax (Linum usitatissimum L.). Cell Biology International. DOI: 10.1002/cbin.10837.

DOI 10.18699/BioGenEvo-2018-06

Evolutionary history of Native Americans drown by deep learning approach

O. Dolgova*, I. Maceda, O. Lao
Population Genomics Team, Centre Nacional d'Anàlisi Genòmica, Barcelona, Spain
** e-mail: olga.dolgova@cnag.crg.eu*

Key words: human population genomics, deep learning, Approximate Bayesian Computation, Native Americans

Motivation and Aim: Population history of present-day Native Americans has been a hot debated topic in last decades. The consensus view on the peopling of the Americas is that ancestors of modern Native Americans entered the Americas from Siberia via the Bering Land Bridge and this occurred at least ~14.6 thousand years ago (ka). However the number and timing of migrations into the Americas remain controversial, with conflicting interpretations based on anatomical and genetic evidence [1, 2]. Our aim is to detect the demographic processes along the evolutionary history of Native American populations, including multiple introgression and admixture events among ancient and even unknown archaic populations.
Methods and Algorithms: Quantifying the amount of admixture and the relationship of (sometimes) unknown ancestral populations is a complex task in human population genomics [3]. Here we present a novel approach based on coupling of Deep Learning with Approximate Bayesian Computation (ABC), which overcomes the common problems related with summary statistics (SS) redundancy. Six admixture models including introgression events from unknown populations were developed for subsequent simulation testing (100,000 simulations per model). Unfolded multidimentional site frequency spectrum (SFS) of 7,314 callable intergenic regions ~700 Mb long and free from CpG islands were computed and used as ABC SS. We trained 10 Artificial Neural Networks forming Deep Learning system of four layers in order to infer the informative SS and decrease the number of needed simulations. SFS estimation, model parameter computation and Deep Learning development were implemented in Java environment, while ABC was conducted using R scripts.
Results and conclusions: Positive correlations between replication and training datasets as an evidence of robustness of Neural Network results were strong and significant in most of the model/population combinations. Estimation of posterior probabilities from six models elucidated the footprints of archaic introgression at least in four Native American populations of Neanderthal-Denisovan ancestor (Pima, Quechua and Mixtec tribes) and *Homo erectus* (Chane tribe) nature.

References

1. Raghavan M. et al. (2015) POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. Science. 349(6250):aab3884.
2. Soglund P. et al. (2015) Genetic evidence for two founding populations of the Americas. Nature. 525:104-108.
3. Slatkin M. (2016) Statistical methods for analyzing ancient DNA from hominins. Current Opinion Genetics Development 41:72-76.

DOI 10.18699/BioGenEvo-2018-07

Evolutionary patterns of piRNA-generating clusters in human genome

O. Dolgova[1]*, R. Mulet[2], L. Llobet[3], S. Casillas[2], T. Vavouri[3]
[1] *Population Genomics Team, Centre Nacional d'Anàlisi Genòmica, Barcelona, Spain*
[2] *Institut de Biotecnologia i Biomedicina and Department de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Barcelona, Spain*
[3] *Josep Carreras Leukaemia Research Institute, Barcelona, Spain*
** e-mail: olga.dolgova@cnag.crg.eu*

Key words: piRNA, human genome, natural selection

Motivation and Aim: PIWI-interacting RNA (piRNA) is the largest class of small non-coding RNA molecules expressed the germline of animals. Their best understood function is genome defense against transposable elements [1]. Most piRNA-generating loci are found in clusters, from which piRNAs are transcribed. Our primary objective was to revisit previously published results, extending them to different populations and chromosome databases, and elucidate which selective forces are mainly acting on piRNA cluster regions in the human genome at the nucleotide level.
Methods and Algorithms: 1000 Genomes Project [2] in its GRCh37 version was chosen for this study as the main dataset, from which piRNA clusters and the nearest intergenic regions, as putatively neutral for all genetic comparisons, were extracted following a newly developed pipeline. Polymorphism, mutation rate and divergence from the chimpanzee genome were calculated for both piRNA clusters and intergenic regions in each of five super-populations, and tests for neutrality and selection were conducted using R and the PopGenome package.
Results: A strong difference in the levels of polymorphism was found between piRNA clusters and intergenic regions, the latter being more diverse. The values of alpha in the McDonald and Kreitman test were negative in 53–60 % and positive 40–47 % of significant results, suggesting recurrent directional selection fixing new advantageous alleles in some clusters.
Conclusion: Altogether, our results are in accordance with the previous findings [3, 4] that purifying selection is supposed to be the main force driving the evolution of piRNA clusters at the nucleotide level in humans. Nevertheless, directional selection also takes place in a substantial part of piRNA clusters, allowing the occurrence of an evolutionary "arm race" between these functional genomic regions and transposable elements.

References

1. Iwasaki Y.W., Siomi M.C., Siomi H. (2015) PIWI-Interacting RNA: Its Biogenesis and Functions. Annu Rev Biochem. 84:405-33.
2. 1000 Genomes Project Consortium et al. (2012) An integrated map of genetic variation from 1,092 human genomes. Nature. 471:56-65.
3. Lukic S., Chen K. (2011) Human piRNAs are under selection in Africans and repress transposable elements. Mol Biol Evol. 28:3061-3067.
4. Gould D.W., Lukic S., Chen K.C. (2012) Selective constraint on copy number variation in human piwi-interacting RNA Loci. PloS One. 7(10):e46611.

DOI 10.18699/BioGenEvo-2018-08

Evolution of unicellular eukaryotes and possible ways to multicellularity

I. Dovgal
Kovalevsky Institute of Marine Biological Research RAS, Sevastopol, Russia
e-mail: dovgal-1954@mail.ru

Key words: choanomonads, evolution, multicellularity, Metazoa

Motivation and Aim: Author's hypothesis on the possible continuity between the mechanisms of intercellular recognition in unicellular eukaryotes and the mechanisms of intercellular interactions in Metazoa is considered with the light of the latest data on the evolution, phylogeny and genomics of unicellular eukaryotes.
Results: E. Haeckel was first who proposed to combine both heterotrophic and autotrophic unicellular eukaryotes along with sponges and fungi in kingdom Protista, which was intermediate between kingdoms Planta and Animalia.
However, in accordance with more popular view the heterotrophic unicellulars were placed in kingdom Animalia as phylum Protozoa, while autotrophic in Planta. In turn, the relations within Protozoa were based on modes of locomotion, feeding and attended peculiarities of cell morphology.
In the subsequent, the assumption on the unicellular eukaryotes evolution have their bases in transmission electron microscopy (types of mitochondria crista) evidence.
However, the use of modern approaches has led to the rejection of morphological schemes, since the general evolutionary line of eukaryotic organisms has broken up into several molecular clusters, or supertaxa, that do not coincide with the morphological groups.
One of these supertaxa named Opisthoconta combines heterotrophic protists, including choanomonads, along with fungi and metazoans. As this takes place, the Choanomoada and Metazoa are sister groups, which confirms the views about the possible origin of Metazoa from the choanomonads, based on the similarity of the collar cell of choanomonads and sponge choanocyte morphology. Even if the Choanomonada are not the direct ancestors of sponges, the similarity possible reflects a general evolutionary trend, which allows tracing of possible ways of Metazoa origin.
Three general ways from unicellular eukaryotes to multicellular organization are recognized: complete or incomplete cell divisions within a common extracellular matrix, cellularization of a multinucleated cell and aggregations of cells attracted by chemical signals. The first of the ways, which gave rise to coloniality, seems most plausible. As this takes place, the cell differentiation into generating and transmitting signals in connection with increasing of colonies seems inevitable.
The mechanisms of regulation of protozoan ontogeny as well as mechanisms of intercellular interaction in these organisms may serve as a model for the formation of mechanisms of metazoans morphogenesis. This corroborated by information that in choanomonad *Salpingoeca rosetta* there are genes encode homologs of cell adhesion, neuropeptide, and glycosphingolipid metabolism, which previously found only in metazoans. In addition, the transcriptome analysis revealed that septins found in the species might regulate incomplete cytokinesis during colony development.
Acknowledgements: Supported by the grant of Ministry of Education of the Russian Federation (No. 14W03.31.0015).

DOI 10.18699/BioGenEvo-2018-09

Identification and study on *Lg*^{t} – a new cereal gene for liguleless leaf phenotype

A.E. Dresvyannikova[1, 2]*, A.F. Muterko[1], A.A. Krasnikov[3], N.P. Goncharov[1], N. Watanabe[4], O.B. Dobrovolskaya[1, 2]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Central Siberian Botanical Garden SB RAS, Novosibirsk, Russia*
[4] *College of Agriculture, Ibaraki University, Ami, Inashiki, Japan*
** e-mail: alinka.dresvyannikova@gmail.com*

Key words: leaf development; ligule, cereal, DArTseq, GBS

Motivation and Aim: The leaf of family Poaceae has a ligule and auricle between its distal and proximal parts. The formation of the distal-proximal axis of differentiation and the development of the organs of the ligule region occurs at the early stages of leaf development. A study of the developing and genetic regulation of the formation of the distal-proximal axis of leaf differentiation in the liguleless mutant *Zea mays L*. made it possible to identify new genes that controlled the development processes-*Lg1*, *Lg2*, *Lg3*/*Lg4*, *Lgn*, etc. A liguleless line of *Ae. tauschii* is an induced mutant (Lgt-mutant), whose phenotype is under control of the dominant gene *Lgt.*
Methods and Algorithms: Using 3933 polymorphic DArTseq markers, a high-throughput genotyping of F2 population from the cross Lgt-mutant/KU-2126 (KU-2126 is an accession with the wild-type leaf phenotype) was performed; highly saturated molecular-genetic maps of *Ae. Tauschii* were constructed.
Results: The *Lgt* gene was placed on the short arm of chromosome 5D by molecular-genetic mapping. The *Lgt* gene locates in the region of conserved synteny with bread wheat 5DS, rice *Os12*, sorghum *Sb2* and Brachypodium *Bd4* chromosomes. in silico mapping of the DArTseq markers, flanking the gene of interest – *Lgt*, on *Ae. tauschii* physical map allowed to establish the coordinates of *Lgt* on 5D pseudomolecule and to determine the list of the *Lgt* candidate genes. It was shown that the orthologs of genes *LIGULELESS 4*, *LIGULELESS NARROW 1* and *KNOX1*, whose dominant mutations cause the liguleless phenotype in cereals, locate on *Ae. tauschii* chromosomes 1D, 4D and 7D, respectively, and they cannot be considered as candidate genes for Lgt (5DS).
Conclusion: Thus, the *Lgt* gene is not an orthologue of the previously studied *Lg4*, *Lgn1* and *Knox1* cereal genes, whose dominant mutations cause the liguleless phenotype, and presents a new, previously unexplored cereal gene, involved in the genetic control of the development of the ligule region and the formation of the distal-proximal axis of differentiation. Using novel genetic models: near isogenic lines marked with the *Lgt* gene and recombinant lines at the next stage of research will help to shorten the list of candidate genes for *Lgt* and allow further isolation of this gene, to study its structural and functional organization.
Acknowledgements: This work was supported by grants of RSF (No. 16-16-10021) and the IC&G Budgetary Project No. 0324-2016-0001.

DOI 10.18699/BioGenEvo-2018-52

Opisthorchiidae triad: comparative genomics of the carcinogenic liver flukes using a draft genome of *Opisthorchis felineus*

N.I. Ershov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
e-mail: nikotinmail@mail.ru

Key words: *Opisthorchis felineus*, draft genome, comparative genomics, trans-splicing

The liver flukes *O. felineus*, *O. viverrini* and *C. sinensis* are distributed in the vast territories of Asia and Europe and pose a serious threat to the health of the population of endemic areas. Genomics of *O. felineus* have been studied significantly weaker than those of *O. viverrini* and *C. sinensis*. This lack of *O. felineus* genomic data is an obstacle to the development of comparative molecular biological approaches necessary to obtain new knowledge about the biology of these trematodes, to identify essential pathways linked to parasite-host interaction, to predict genes that might contribute to liver fluke pathogenesis. Here we report a draft genome assembly and annotation of *O. felineus* and results of comparative study of genomics and transcriptomics of the three Opisthorchiidae liver flukes. The overall genomic sequences of the three species show relatively high divergence from each other. At the same time, the repertoire of genes and their expression levels (at the whole-body scale) were highly consistent among the opisthorchiids. A comparison of the phylogenetic trees for the three studied opisthorchiids and the data on their synteny suggests that the *O. viverrini* genome was structurally remodeled after it had diverged from its common ancestor with *C. sinensis*. In addition, we have demonstrated that the products of almost half *O. felineus* genes contain the SL sequence, which suggests high importance of trans-splicing in the regulation of RNA processing in liver flukes. The study is also focused on some gene networks involved in regulation of host-parasite interaction. Availability of a genome for *O. felineus* should help support the development of comparative genomics, proteomics and other -omics studies necessary for understanding the parasites evolution and development of novel drugs and vaccines against liver flukes. This work was partly supported by the Russian Scientific Foundation (grant No. 18-74-00101).

DOI 10.18699/BioGenEvo-2018-10

Biological diversity of the bacterial community of the Vostok bay (Japan Sea) by high-throughput sequencing

Y. Golozubova*, L. Buzoleva, E. Bogatyrenko
Far Eastern Federal University, Vladivostok, Russia
** e-mail: know-26@mail.ru*

Key words: Japan Sea, bacterial communities, anthropogenic pollution

Motivation and Aim: The biodiversity of the microorganisms of the seas is insufficiently studied. About 99 % of bacteria are noncultivated forms [1]. The application of the method of high-throughput sequencing allows obtaining the most complete data of the taxonomic structure and spatial distribution of bacterial communities. However, metagenome analysis of Japan Sea wasn't done.

Methods and Algorithms: The material for the work was samples of surface waters, selected in August 2015 in the Vostok Bay of the Japan Sea. A sample of sea water was taken with a syringe at a depth of 10–15 cm, and 1.5 l of sea water was filtered through membrane polycarbonate filters with a pore diameter of 0.22 μm (Millipore, USA), fixing in 80 % ethanol. The total DNA from the samples was isolated using the Amplipram DNA-sorb B kit according to the manufacturer's protocol with modification [2]. Taxonomic analysis was carried out based on the study of the V3-V4 variable region of the 16S rRNA gene. Amplification and metagenomic sequencing of 16S rRNA was performed on the MiSeq genomic sequencer of Illumina at the Genomics Center in Novosibirsk. The data obtained was analyzed using the Mothur software package, the Silva database.

Results: Analysis of the data showed that the prokaryotic community of the Vostok Bay was formed by the bacteria of the Proteobacteria (56 %) and Bacteroidetes (29 %), Actinobacteria (3 %) and Firmicutes (2 %) in the Vostok Bay. Our experiments also indicate that the families Flavobacteriaceaeae, Rhodobacteriaceae, occupy the dominant part of the microbial community.

Conclusion: Thus, with the use of metagenomic analysis in Japan Sea, the predominance of typical marine representatives of the phylum Proteobacteria, Bacteroidetes with prevalence of Alphaproteobacteria and Gammaproteobacteria.

Acknowledgements: Supported by the Russian Science Foundation (Agreement No. 14-50-00034).

References

1. Ravin N.V., Mardanov A.V., Skryabin K.G. (2015) Metagenomics as a tool for the investigation of uncultured microorganisms. Russian Journal Genetics. 51:431-439.
2. Belkova N.L. (2009) Modified technique for isolating total DNA from aqueous samples and ground extracts by the enzymatic lysis method. Molecular and genetic methods for the analysis of microbial communities. Diversity of microbial communities in inland water reservoirs of Russia: Printhouse. 53-63.

DOI 10.18699/BioGenEvo-2018-11

Host *Drosophila* fitness and hormonal status depends on the genotype of *Wolbachia* symbiont

N.E. Gruntenko*, N.V. Adonyeva, Y.Y. Ilinsky, E.V. Burdina, O.V. Andreenkova, R.A. Bykov, I.Yu. Rauschenbach
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** e-mail: nataly@bionet.nsc.ru*

Key words: systems biology, aging, hormone metabolism, symbiont, *Drosophila*

The idea of study presented was to find out if the effect of prokaryotic symbionts *Wolbachia pipientis* on the host *Drosophila* fitness and hormonal status depends on the genotype of the bacteria. For this purpose we used five conplastic strains carrying the nuclear background of interbreeded wild type Bi90 line and cytoplasmic backgrounds with different genotype variants of *Wolbachia*. Bi90 line treated with tetracycline for 3 generations was used as a control group. We demonstrated that only two out of five investigated *Wolbachia* variants promote changes in *Drosophila* fitness and metabolism of two hormones, dopamine and juvenile hormone. *wMel*, *wMel2 wMel4* genotypes of *Wolbachia* do not show any effect on *Drosophila* stress resistance, the intensity of dopamine metabolism and fecundity, whereas *wMelCS* genotype increases first two traits and decreases the fecundity in the beginning of the oviposition increasing it later. Young *Drosophila* females with *wMelCS* genotype demonstrate the increased level of juvenile hormone metabolism compared to females with *wMel* the control ones, while mature females show the opposite pattern, what corresponds well with the changes in their fecundity. The obtained results are quite surprising because genotypes of wMel group predominate in the nature populations all over the world and wMelCS variants are very rare. Especially interesting that pathogenic *wMelPop* variant of *wMelCS Wolbachia* genotype increases fecundity and decreases stress resistance, the intensity of dopamine and juvenile hormone metabolism in young females, while increases juvenile hormone metabolism and decreases the fecundity, stress resistance and the dopamine metabolism in mature females. This is also quite unexpected, because the strong negative influence of *wMelPop* on host metabolism and stress resistance starts much earlier than mass death of flies carrying this pathogenic strain.
Acknowledgements: The study was supported by the RFBR grant No. 16-04-00060 and State Budgeted Project No. 0324-2018-0016.

DOI 10.18699/BioGenEvo-2018-12

Diel cycle of the tropical air microbiome

E.S. Gusareva*, E. Acerbi, K.J.X. Lau, A. Wong, S. Kolundžija, P.B.N. Vasantha, R. Purbojati, I. Luhung, J.N. Houghton, D. Miller, N.P.E. Gaultier, C.E. Heinle, M. Clare, V.K. Vettath, C. Kee, S.B.Y. Lim, C. Chenard, W.J. Phung, K.K. Kushwaha, A.P. Nee, A. Putra, D. Panicker, M. Koh, Y.Z. Hwee, S.R. Lohar, H.L. Kim, L. Yang, A. Uchida, D. Moses, A.C. Junqueira, S.C. Schuster

Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, 60 Nanyang Drive, Singapore

** e-mail: egusareva@ntu.sdu.sg*

Key words: air microbiome, airborne microbial communities, air microbial ecosystem

Motivation and Aim: Air as an ecosystem for microbial organisms is vastly underexplored. Estimates of microbial diversity and abundance historically have relied on cultivation or amplification of gene markers. We here show that airborne microbial biomass from the Tropics can be explored by metagenomic sequencing, even when sampling short time intervals.

Results: In our five time series experiments covering 13 months period, we collected and sequencing 795 air samples that implies analyzing 1.89 billion reads of 250 bp length. Although in our analysis we could identify less than 20 % of the airborne microbial taxa, the diversity of the yet identified microbes were comparable to the human gut microbiome where about 60 % of the microbial community is identified. Thus the taxonomic diversity of the air ecosystem is potentially large. Unlike terrestrial or aquatic environments, air is dominated by DNA of eukaryotes, particularly fungi of the order Asco- and Basidiomycota. Moreover, air communities remain remarkably stable over periods of time of up to 13 months, but show significant taxonomic variation within a 24 hours time scale. The observed recurrent diel fluctuations of species diversity and total biomass were shown to be driven by temperature/relative humidity changes, air CO_2 concentration and rain events/lightning. The discovered dial cycle of the air microbial community was in fact so reproducible over the investigated 13-month period that a single air sample can be assigned to the actual time slot of the day it was collected, when using our data as a reference.

Conclusion: In the atmosphere microbial life is most abundantly near the planet's surface where light, temperature, CO_2 and humidity are optimal. A gaseous ocean that supports life not in liquid water, but nevertheless allows for the formation of a complex ecosystem that follows the same diel patterns, as those described for the microbial communities of the world's marine systems.

Acknowledgements: This study was supported by Academic Research Fund (AcRF) Tier 3, Singapore Ministry of Education (grant number: MOE2013-T3-1-013).

DOI 10.18699/BioGenEvo-2018-13

Searching for signatures of cold adaptation in human *TRP* genes

A.V. Igoshin[1]*, K.V. Gunbin[1], N.S. Yudin[1, 2], M.I. Voevoda[1, 2, 3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Institute of Internal and Preventive Medicine – branch of ICG SB RAS, Novosibirsk, Russia*
** e-mail: igoshin@bionet.nsc.ru*

Key words: cold adaptation, *TRP* genes, minimum winter temperature, SNPs, human population

Motivation and Aim: Since his origin from sub-Saharan Africa 200,000 years ago, modern human eventually spread across the Earth. These migrations of *Homo sapiens* have led to adaptation to local environments. The advent of high-throughput next-generation sequencing and microarray technologies has enabled researchers to discover genetic bases of such adaptations. Here, applying three different statistical approaches, we try to detect signatures of cold adaptation in seven human *TRP* genes (*TRPA1*, *TRPM8*, *TRPV1*, *TRPV2*, *TRPV3*, *TRPV4*, and *TRPM3)* as being immediately involved in thermoregulation in mammals.
Methods and Algorithms: 407 SNP IDs were obtained from dbSNP database by querying gene's names. Information about minimum winter temperature was taken from NCEP-NCAR database using coordinates from study by H. Cann et al. [1]. Genotypic data was taken from HGDP database, and then imputed with fastPhase software. Bayesian linear model (BLM) implemented in Bayenv2 software accounts for sample size and covariance of allele frequencies across populations, but not robust to outliers and less sensitive to nonlinear relationships. Spearman's rank correlation test (SRC) from Bayenv2 uses allele frequencies standardized to have no covariance. It is less powerful than BLM, but more robust to outliers and can detect nonlinear relationships. BayScenv approach associates F_{st} distances with an environmental variable, but does not account for covariance. For BLM, covariance matrix was obtained by running 100,000 iterations of Markov Chain Monte Carlo (MCMC) algorithm and averaging across 20 matrix estimates. Then, for each SNP we ran 200,000 MCMC iterations. Because of relatively low reproducibility of Bayenv2, we repeated this procedure 5 times. Bayesian factor averaged across all repetitions was used as a final estimate for each SNP. The same rules were also used for SRC. BayScenv model was run with default options except that parameter p was set to 0.25. Two repetitions were done.
Results: At the intersection of top 1 % of all three methods, we found SNP rs17617922, in *TRPM3* gene that has already been reported to be associated with minimum winter temperature among human populations by whole genome scan [2]. This SNP does not have clear evidences supporting its functional effect, so it is possible that rs17617922 is linked to some causal variant. In addition, we found SNP rs8065080 in *TRPV1* gene being at the top 1 % of SRC and BLM results. It is interesting that this SNP leads to amino acid substitution. Furthermore, individuals homozygous for allele present at high frequency in regions with cold climate have higher tolerance to cold [3].
Conclusion: SNPs rs17617922 in *TRPM3* gene and rs8065080 in *TRPV1* gene are probably associated with human adaptation to cold climate. Further studies based on detecting selective sweeps are needed.

References

1. Cann H. et al. (2002) A human genome diversity cell line panel. Science. 296(5566):261-262.
2. Günther T., Coop G.(2013) Robust identification of local adaptation from allele frequencies. Genetics. 195:205-220.
3. Binder A. et al. (2011) Transient receptor potential channel polymorphisms are associated with the somatosensory function in neuropathic pain patients. PLoS One. 6:e17387.

DOI 10.18699/BioGenEvo-2018-14

Comparative genome analysis of related *Lymantria dispar* nucleopolyhedrovirus isolates differing in virulence

Yu. Ilinsky[1, 2, 3], E. Lunev[2], S. Toshchakov[2, 4], J. Podgwaite[5], V. Martemyanov[6, 7]*
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Immanuel Kant Baltic Federal university, Kaliningrad, Russia*
[3] *Novosibirsk State University, Novosibirsk, Russia*
[4] *Winogradsky Institute of Microbiology, Moscow, Russia*
[5] *USDA Forest Service, Hamden, USA*
[6] *Institute of systematics and ecology of animals SB RAS, Novosibirsk, Russia*
[7] *National Research Tomsk State University, Tomsk, Russia*
** e-mail: martemyanov79@yahoo.com*

Key words: nucleopolyhedrovirus, Lymantria dispar, NGS, genome, virulence

Motivation and Aim: Nucleopolyhedroviruses (NPVs, Baculoviridae) are specific viruses of insects that are used for pest control. High variation in virus virulence depends upon genetic factors [1]. Here we compare genomes of closely related genetic variants of *Lymantria dispar* multiple NPV (LdMNPV) that differ in virulence.
Methods and Algorithms: Five viral genotypes, differing in virulence, were isolated from the standard strain of NPV in the biopesticide "Gypchek" [2]. Complete genomes of these five genotypes were sequenced using Illumina technology by paired end sequencing of fragment genomic libraries, and assembled by SPAdes 3.9.0 [3]. Initial assemblies consisted of 1–3 scaffolds with a total assembly length of 159–174 kbp. Genome coverage was in range of 200x–850x. Closing of scaffold gaps was performed by Sanger sequencing of amplicons. To determine genetic relatedness among studied LdMNPV variants a phylogenetic tree was reconstructed based on a set of core loci. Comparative analysis of candidate genes was performed to reveal genetic determinates of variation in virulence.
Results: Phylogenetic analyses indicated recent divergence of LdMNPV-studied isolates from a common ancestor. We found a number of nonsynonymous nucleotide substitutions and *indels* in many genes. However, we could not find a locus that could be considered as the main effector of observed variation in virulence. This fact indicated a complex genetic nature of virulence variation among closely related virus isolates.
Conclusion: Variation in virulence among related LdMNPV isolates can be explained by the complex effect of different loci.
Acknowledgements: the study was supported by RSF (grant No. 17-46-07002).

References

1. Martemyanov V.V. et al. (2017) A comparison of the adaptations of strains of Lymantria dispar multiple nucleopolyhedrovirus to hosts from spatially isolated populations. Journal of Invertebrate Pathology. 146:41-46.
2. Podgwaite J.D. et al. (2013) Potency of Nucleopolyhedrovirus Genotypes for European and Asian Gypsy Moth (Lepidoptera: Lymantriidae). Journal Entomological Science. 48(4):332-344.
3. Bankevich A. et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Computational Biology. 19(5):455-477.

DOI 10.18699/BioGenEvo-2018-15

Polymorphism in genes related to fatty acid composition in *Linum usitatissimum*

P. Kezimana[1, 2], A. Dmitriev[1], T. Rozhmina[1, 3], R. Novakovskiy[1]*, E. Romanova[2], A. Kudryavtseva[1], N. Melnikova[1]

[1] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*
[2] *Peoples' Friendship University of Russia* (*RUDN University*), *Moscow, Russia*
[3] *All-Russian Research Institute for Flax, Torzhok, Russia*
** e-mail: 0legovich46@mail.ru*

Key words: fatty acids, FAD genes, SAD genes, polymorphism, flax, *Linum usitatissimum*

Motivation and Aim: Flax (*Linum usitatissimum* L.) is highly recommended in human health because of its alpha-linolenic acid (ALA) content, as it has been reported that the ALA component of flax oil (omega-3 fatty acid) improves bone and cardio-vascular health [1]. Fatty acid (FA) biosynthesis in flax involves several consecutive steps governed by different gene families, such as stearoyl-acyl carrier protein desaturase (SAD) genes and FA desaturase (FAD) genes [2]. Despite the fact that these genes were identified, information on their polymorphism is restricted. Therefore, the aim of our study was to determine polymorphism of *SAD* and *FAD* genes in flax by sequencing these genes from 192 flax genotypes with different composition of FAs.

Methods and Algorithms: Seeds were obtained from the All-Russian Research Institute for Flax. DNA was isolated from 5-day-old seedlings using a CTAB method. We designed the primers for amplification of target genes by MethyMer and prepared DNA libraries using a two-stage PCR. We analyzed the concentration of obtained libraries on Qubit 2.0 fluorometer and their quality – on Agilent 2100 Bioanalyzer. Sequencing of *FAD* and *SAD* genes was performed on Illumina platform. For bioinformatics analysis, CLC Genomics Workbench was used. Obtained reads were mapped on reference sequences of *SAD* and *FAD* genes for polymorphism identification.

Results: High-throughput sequencing of *SAD* and *FAD* genes from 192 flax genotypes with diverse FA composition was performed. High coverage of DNA sequences was achieved, we obtained no less than 3000 reads for each sample. This therefore enabled the accurate assessment of levels of polymorphism and the identification of single nucleotide polymorphisms (SNPs).

Conclusion: High-throughput sequencing can be successfully used to analyze the level of polymorphism of flax desaturase genes. Identified SNPs will assist in estimation of genetic diversity, phylogenetic analysis, identification of functional polymorphisms, and marker-assisted selection.

Acknowledgements: This work was performed within the framework of the Program of fundamental research for state academies for 2013-2020 years (No. 01201363819) and was funded by RFBR according to the research project 17-29-08036.

References

1. Kim Y., Ilich J.Z. (2011) Implications of dietary α-linolenic acid in bone health. Nutrition. 27:1101-1107.
2. You F.M. et al. (2014) Genome-wide identification and characterization of the gene families controlling fatty acid biosynthesis in flax (*Linum usitatissimum* L.). Journal Proteomics Bioinformatics. 7:310-326.

DOI 10.18699/BioGenEvo-2018-16

Comparative analysis of transcriptome from different sympatric morphs of Dolly Varden *Salvelinus malma* from the Kronotskoe Lake

E.A. Konorov[1], V.O. Burskaya[2], I.V. Artyushin[1], V.A. Scobeyeva[1], G.N. Markevich[3], N. Osman[4], S.V. Nuzhdin[4]

[1] *Moscow State University, Moscow, Russia*
[2] *Skolkovo Institute of Science and Technology, Moscow, Russia*
[3] *Kronotsky Nature Reserve, Yelizovo, Russia*
[4] *Molecular and Computational Biology, University of Southern California, Los Angeles, USA*
** e-mail: skobei-khanum@yandex.ru*

Key words: sympatric morphs, *Salvelinus malma*, transcriptome

Motivation and Aim: Some different morphs of Dolly Varden *Salvelinus malma* inhabit Lake Kronotskoe (Russia, Kamchatka). They differ more or less in head proportions, feeding, timing, and place of spawning, but can not be separated by mtDNA markers [1]. We obtained transcriptome from 5 morphs and tried to establish phylogenetic relations among morphs and compare the expression level of assembled transcripts.
Methods and Algorithms: Reads from each individual were assembled separately with Trinity. Only 17 specimen were assembled with reasonable quality and were taken for analysis. All transcripts were blasted against Salmo salar mRNA and only sequences with 1 reciprocal hit were selected and filtered for median coverage 40. We found 130 genes common for all 17 specimen and built phylogenetic tree with STAR. Differential expression were measured with edgeR. Reads were mapped on the assembly of specimen with best quality, transcript abundance were estimated with utilites from Trinity package and salmon. 117 transcripts were found over- or underexpressed (FDR < 0.005, logFC < 2) and were taken for analysis of GO encrichment.
Results: 117 transcipts differentially expressed between morphs were found, most of them with indefined onthology. Transrcipts form two groups – one is overexpressed in dwarf morph and the other – in white, nosute and big-mouth forms. No transcripts of suitable quality were obtained from long-headed morph. Phylogenetic analysis of 130 trancripts produced moderately supported tree with separate branch including most transcripts of dwarf morph.
Conclusion: Only dwarf morph of Kronotskoe lake Dolly Warden demonstrates moderate genetic separation.

References

1. Senchukova A.L., Mugue N.S., Pavlov S.D. et al. (2013) On the origin of charrs of the genus Salvelinus of the Kronotskoe Lake and their relationships with other charr populations of the Kamchatka peninsula. J. Ichthyol. 53:840.

DOI 10.18699/BioGenEvo-2018-17

Evolutionary study of low complexity glycine-arginine rich domains

A. Kotyurgin[1]*, A. Alexeevski[1, 2]
[1] *Lomonosov Moscow State University, A.N. Belozersky Institute of Physico-Chemical Biology and Faculty of Bioengineering and Bioinformatics, Moscow, Russia*
[2] *Scientific Research Institute for System Analysis RAS, Moscow, Russia*
** e-mail: alekoksan@gmail.com*

Key words: low-complexity regions, glycine-arginine rich domains, neutral evolution

Motivation and Aim: Low complexity glycine-arginine rich (GAR, also known as RGG) domains are abundantly presented in many RNA-binding proteins [1]. The study of evolution of GAR domains is complicated by the impossibility to use standard methods of phylogeny due to low complexity corruption of alignment. Previously, it was discovered that GAR domains of the mammalian fibrillarin have some specific features that make it possible to classify GAR domains by taxa at the order rank. The aim of this work is to study of the diversity, occurrence and evolution of GAR domains.
Methods and Algorithms: We created python script to find all GAR-containing proteins in the Uniref100 database (126152540 sequences) and describe their functions using such databases as Gene Ontology, Pfam and eggNOG. While standard methods of phylogeny do not work, we used taxonomic cladograms with an indicated proportion of GAR-containing proteins in different taxa to study the evolution of GAR domains. Such characteristics as frequencies of amino acids and specific number of glycine repeats per length were counted.
Results: 141608 GAR-containing proteins were found in Uniref100 database. Annotation on GO, Pfam and eggNOG databases confirmed that the most common activities of GAR-containing proteins are RNA binding, helicase activity and other functions associated with nucleic acid. Cladograms analysis of the most represented COGs showed that GAR domains are not ubiquitous in COGs and could emerge independently during evolution. Frequency distribution of amino acids seems to be result of the interaction of two modes of evolution, such as neutral evolution determined by the most probable path of non-synonymous mutations and directional selection according to physicochemical properties of residues (hydrophobicity, affinity to nucleic acids). The first type of evolution is assumed in the cases of alanine and serine, the second in the case of phenylalanine, proline, tyrosine and non-alanine hydrophobic residues.
Conclusion: A functional association of GAR-containing proteins with the metabolism of nucleic acid was shown on a much larger sample than in any previous study. Multiple independent emergency of GAR domains during evolution seems to be the most typical scenario. Amino acid composition of GAR is presumably determined by both neutral evolution and directional selection. We intend to complete sequence based GAR classification compatible with their evolution.
Acknowledgements: the work was partially supported by RSF grant 16-14-10319.

References

1. Thandapani P. et al. (2013) Defining the RGG/RG motif. Molecular Cell. 50(5):613-23.

DOI 10.18699/BioGenEvo-2018-18

Phylogenetic analysis of genes encoding the enzymes of plant amino acids catabolism in representatives of the genus *Methylobacterium*

M. Krohaleva[1]*, D. Fedorov[2], G. Ekimova[2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Biochemistry and Physiology of Microorganisms RAS, Pushchino, Russia*
** e-mail: krohaleva-mariya@mail.ru*

Key words: *Methylobacterium*, phytosymbiosis, ACC-deaminase, D-cysteine desulfhydrase

Motivation and Aim: Aerobic methylobacteria use oxidized and substituted methane derivatives as the sources of carbon and energy. Methylobacteria can stimulate growth and development of plants due to biosynthesis of phytohormones and vitamins. One of the mechanisms of the beneficial effects of bacteria is the reduction of level of the "stress" hormone ethylene in plant tissues due to the activity of 1-aminocyclopropane-1-carboxylate (ACC) deaminase [1]. An effective approach of assessing the ability of bacteria to phytosymbiosis is the search of genes whose products can stimulate the growth of plants, using PCR with degenerate primers. In the case of the ACC deaminase gene (*acdS*), such analysis is complicated by the presence in the genomes of some bacteria of a gene encoding D-cysteine desulfhydrase (*dcyD*), an enzyme whose amino acid sequence exhibits a high level of identity with the sequences of ACC deaminases. The activity of this enzyme probably increases the resistance of plants to phytopathogens due to the formation of hydrogen sulphide, which functions as a fungicide [2].
Methods and Algorithms: 48 species of the genus *Methylobacterium* were studied. The following methods were used: DNA extraction, PCR, cloning and sequencing of DNA fragments. For amino acid and nucleotide sequences phylogenetic analysis neighbor joining/UPGMA methods were used realized in the MEGA6 [3].
Results: Taking together with genomic data, either *acdS* or *dcyD* were found and/or sequenced in 29 of the 48 type strains of different species of the genus *Methylobacterium*. Of these, 18 strains possess the *acdS* gene, whereas 13 strains – the *dcyD* gene, two species have two genes simultaneously and, therefore, can stimulate the growth and development of plants due to the activity of both enzymes.
Conclusion: For the first time it is shown that the distribution of the genes for *acdS* and *dcyD Methylobacterium* species is subject to certain phylogenetic regularities, since two groups of species containing the gene *acdS* and one gene of *dcyD* have been found.
Acknowledgements: The work was supported by the grant of the Russian Fund for Basic Research RFBR (16-04-00381_a).

References

1. Glick B.R. et al. (2007) Promotion of plant growth by ACC deaminase-producing soil bacteria. European Journal Plant Pathology. 119(3):329-339.
2. Papenbrock J. et al. (2007) Characterization of cysteine-degrading and H2S-releasing enzymes of higher plants-fromr the field to the test tube and back. Plant Biology. 9(5):582-588.
3. Tamura K. et al. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology Evolution. 30(12):2725-2729.

DOI 10.18699/BioGenEvo-2018-19

Single-cell bioinformatics: practical implementations

O. Kuleshova
A.O. Kovalevsky Institute of Marine Biological Research of RAS, Sevastopol, Russia
e-mail: v_olgo4ka@inbox.ru

Key words: single-cell analysis, bioinformatics, single-cell software

Motivation and Aim: Rapid development of the next-generation sequencing (NGS) technologies observed the last 15 years. High potential of the NGS methods for a single cell RNA-seq (scRNA-seq) has been showed [1]. This revolutionary analysis gives additional information relative to RNA-seq to bulk sample by analyzing expression profiles at the cellular level. These technologies have opened up opportunities for understanding biological processes at a fundamentally new level of the cellular heterogeneity detections, cell lines tracering, subpopulations identification and clarification of the cell-specific characteristics [2]. Modern scRNA-seq technologies require a development of the special approaches for the big-data sets analysis. In this regard, rapid improving of analytical techniques of the processing of increasing data stream, as well as the implementing software packages, has observed. So, current task is to study and analyze software methods and tools for scRNA-seq bioinformatic analysis.
Methods and Algorithms: There are two main stages in the bioinformatics analysis of the scRNA-seq data: primary and secondary analysis. Primary analysis includes (1) data preprocessing (adapter trimming, paired-end data processing, nucleotide quality filter), (2) constraction of expression matrix including reads quality control, alignment, mapping, quantification. Secondary analysis includes data data preprocessing (filtering and normalization), estimation of the differential expression, clustering, dimensionality reduction (visualization), subpopulation detection, pseudo-time construction. The primary stage is implemented using a set of specialized program packets (e. g., Trimmomatic, FASTQC, STAR, HiSeq or functional analogues), or automated pipelines (e. g., CellRanger, zUMIs, scPipe, Dr.seq2). For normalization, it could be used one of the packages: MAST, SCDE, Monocle2, BCSeq et al. Evaluation of differential expression could be performed using MAST, SCDE, Monocle 2, D3E, DESeq, edgeR, Seurat, Scanpy et al. Most of the variant clustering and visualization methods are implemented in the software packages mentioned above. Pseudo-time construction is available, for example, in the Monocle2 and Scanpy packages. Subpopulation detection could be performed using packages SCUBA, SCENT, Scanpy et al.
Conclusion: Bioinformatic tools for scRNA-seq data analysis are very rapidly evolving to meet the needs for the analysis of the huge increasing data flows. Currently, secondary analysis packages Seurat and Scanpy could be view as the most optimal. However, bioinformatic tools improve rapidly and constantly, that is why continuous monitoring, evaluation and selection of software packages for specific tasks are required.
Acknowledgements: Supported by the Ministry of Education and Science of the Russian Federation grant No. 14.W03.31.001

References

1. D'Argenio V. (2018) The High-Throughput Analyses Era: Are We Ready for the Data Struggle? High-Throughput. 7:8.
2. Poirion O.B., Zhu X., Ching T., Garmire L. (2016) Single-Cell Transcriptomics Bioinformatics and Computational Challenges. Front. Genet. 7:163.

DOI 10.18699/BioGenEvo-2018-20

Finding shifts in the evolution of mitochondrial metabolism

A.A. Kuzminkova[1]*, K.Yu. Popadin[1, 2], K.V. Gunbin[1, 3]
[1] *School of Life Science, Immanuel Kant Federal Baltic University, Kaliningrad, Russia*
[2] *Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland*
[3] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** e-mail: kuzminckova@yandex.ru*

Key words: adaptive evolution, mitochondria-nucleus interaction, proteins and gene networks evolution

Motivation and Aim: It is well known that metabolic rates change significantly during Vertebrata evolution. However, the evolutionary time points and the molecular basis of these metabolic shifts are unknown.
Methods and Algorithms: Because mitochondria is strongly involved in the maintenance of basal metabolism, we focused on the molecular co-evolution of mitochondrial proteins with more than 500 nuclear proteins tightly associated with mitochondria (from MitoMiner4.0 database). In order to find the inner branches of Vertebrata tree with such changes, the evolution of protein structure was analyzed based on about 3000 species with completely sequenced mtDNAs and based on more than 100 species with completely sequenced nuclear genomes. Taking into account various known drawbacks in the evolutionary analyze of protein structures (the technical impossibility to reconstruct ancestral protein structures in all inner nodes of tree; heterotachy; inequality in various taxa sampling, etc.) we invented and implemented our software pipeline. It is based on the comparative analysis of 3 sets of heterotachy-aware and species-tree informed phylogenetic trees that were reconstructed using 50 % jackknife of multiple alignments of (1) protein sequences, (2) residue solvent accessibilities and (3) residue secondary structures. For the reconstruction of trees (2) and (3) 3D protein structure properties were predicted using SCRATCH-1D v.1.1.
Results: We observed highly significant accelerations in protein structure evolution at several inner tree branches of mitochondrial and nuclear proteins that tightly linked with well-known evolutionary innovations in maintenance of basal metabolism. The tree branches with accelerations in nuclear proteins evolution concentrated on Mammalia clade at either relatively recent divergences, e.g. Primates, Rodentia, Chiroptera, Carnivora, or evolutionary old divergences such as Mammalia or Monotremata. At the same time, mitochondrial proteins also have evolutionary accelerations at relatively recent Primates and Carnivora divergences, while the spectra of evolutionary old divergences is much more extensive. Among the nuclear proteins that enriched with accelerated structure evolution events, we found the Succinate Dehydrogenase Complex Assembly Factors, mitochondrion morphogenesis proteins, ADP/ATP translocases. It is also interesting that nuclear proteins enriched with accelerated structure evolution on Primate divergence related with mitochondrial nucleoid while the other divergences demonstrated acceleration of nuclear proteins of mitochondrial matrix and/or membrane.
Conclusion: Thus, the vast majority of observed accelerations in structure evolution of mitochondrial and mitochondria-related nuclear proteins are most likely driven by adaptive functional changes because they are tightly associated with changes in environment and physiology at the course of evolution.
Acknowledgements: This study was supported by the 5 Top 100 Project at the I. Kant Baltic Federal University.

DOI 10.18699/BioGenEvo-2018-21

Recombination landscapes in eight avian species

L. Malinovskaya[1, 2]*, N. Torgunakov[2], A. Torgasheva[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: l.malinovskaya@g.nsu.ru*

Key words: avian chromosomes, recombination nodules, synaptonemal complex, MLH1

Motivation and Aim: Recombination is a key process of sexual reproduction. Shuffling the alleles, it increases genetic variation in populations and provides the raw material for natural selection. Recombination rate shows considerable interspecies variation, tightly linked with karyotypic variation in mammals. Relatively constant diploid chromosome number and inter-chromosomal conservation make birds an ideal model to study on evolution of recombination. However, up to date recombination landscapes have been described in a handful of bird species. Here we present a first description of the recombination characteristics of eight avian species from three orders.
Methods and Algorithms: Pachytene chromosome spreads were prepared from ovaries of nestling females and testes of adult males. To visualize recombination nodules and synaptonemal complex (SC) in pachytene cells we used immunolocalization of SYCP3, the main protein of the lateral element of the SC, MLH1, mismatch repair protein and centromere proteins.
Results: We found no significant variation of recombination rate across examined species. Recombination rate varied from 46.6 ± 2.9 MLH1 foci per autosomal set in male sand martin ($2n = 80$) to 55.1 ± 5.2 in female European pied flycatcher ($2n = 80$). Interestingly, Eurasian hobby ($2n = 50$) with the lower chromosome number had recombination rate similar to that in other species ($2n \sim 80$). We observed two different patterns of crossing over event distribution along the macrobivalents. The first one was characterized by recombination nodules more or less evenly distributed along the chromosome arms in male Eurasian Hobby, female barn swallow and European pied flycatcher. The second one was highly polarized with strong peaks at telomeric regions in male common swift, female great tit, female barn swallow, female and male sand and pale martins. In most cases, a chromosome morphology did not affect these patterns.
Conclusion: Our findings confirm the suggestion that birds have the higher level of recombination rate than mammals. Low interspecies variation of recombination rate may result from the relatively stable avian karyotype. However, we have found that reduction of chromosome number does not lead to decrease of recombination rate. This indicate that recombination rate might be constrained in birds by other factors. More data from different taxa are required to draw statistically supported conclusions about the evolution of recombination in birds.
Acknowledgements: The study was supported by the RFBR (17-29-08019, 18-04-00924) and the Federal Agency for Scientific Organizations via the Institute of Cytology and Genetics (Grant No. 0324-2018-0019).

DOI 10.18699/BioGenEvo-2018-22

Transition transversion ratio in mtDNA is higher in long-versus short-lived mammalians: effects of ROS and replication?

A.G. Mikhaylova[1]*, A.A. Mikhaylova[1], K. Ushakova[1], E. Tretyakov[1], A. Yurchenko[2], D. Knorre[3], I. Mazunin[1], A. Reymond[5], K. Gunbin[1, 4], K. Popadin[1, 5]
[1] *The School of Life Sciences, Immanuel Kant Baltic Federal University, Kaliningrad, Russia*
[2] *Institute of Biodiversity Animal Health and Comparative Medicine, University of Glasgow, Glasgow, UK*
[3] *The A.N. Belozersky Institute Of Physico-Chemical Biology, MSU, Moscow, Russia*
[4] *The Institute of Cytology and Genetics of the SB RAS, Novosibirsk, Russia*
[5] *Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland*
** e-mail: polar_song@mail.ru*

Key words: mtDNA, transition/transversion ratio, mutation spectres, ROS, mtDNA replication

Motivation and Aim: Transition/transversion ratio (ts/tv) in animal mtDNAs significantly differs between various taxa; however, no universal explanation has been suggested.
Methods and Algorithms: Using four-fold degenerative synonymous fixations of mammalian mtDNAs here we reconstructed mutation spectrum for more than 300 species. In order to do so we built an original pipeline implemented in Perl and Python. The pipeline intended to build of intraspecies mutation spectra (IMS) for each species under analysis by reconciliation of intraspecies sequences on the basis of ancestral sequence reconstruction (using parsimony and/or ML) in each inner tree nodes. After the IMS building we normalized IMS by genome-wide nucleotide content. This work we analyzed species if it have at least 30 synonymous fixations along intraspecies tree (>300 mammalian species were selected). At the last step we used phylogenetic comparative methods implemented in R for the comparison of the IMS evolution with species generation time (age of the maturation of a female taken from AnAge database).
Results: The average mutation signature is very similar with mutation signature derived from somatic mitochondrial mutations in human cancers: two common types of substitutions (G->A and T->C transitions, light mtDNA strand notations) demonstrating strong strand asymmetry (occurring mainly on a heavy mtDNA strand). Comparing mutation spectra of long- versus short-lived mammals we observed a gradient: species-specific ts/tv increases with generation time and this correlation is robust to numerous potential confounders, such as nucleotide content, phylogenetic inertia and types of analyzed mitochondrial genes.
Conclusion: Our findings might be explained by two, non-mutually exclusive hypotheses (i) short-lived species have increased basal metabolic rate and thus can suffer from the increased burden of ROS, manifesting itself by G:C->T:A transversions; (ii) long-lived species have prolonged replication of mtDNA and thus accumulate more C->T and A->G transitions occurring on single-stranded heavy strand during replication [2]. Our findings are in line with previously observed correlations between mitochondrial nucleotide content and mammalian lifespan [1] and emphasize that at least some of them are driven by purely mutagenic not selective forces.

References

1. Lehman G. et al. (2008) Rejuvenation Res. 11(2):409-417.
2. Faith J.J., Pollock D.D. (2003) Genetics. 165(2):735-745.

DOI 10.18699/BioGenEvo-2018-23

Haplotype analysis of the HFE gene among patients with different forms of tick-borne encephalitis

S. Mikhailova[1]*, A. Barkhash[1], I. Kozlova[2], I. Borischuk[3], N. Yudin[1], O. Zaitseva[4], L. Pozdnyakova[5], M. Voevoda[1]

[1] *Federal Research Centre Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

[2] *Federal State Public Scientific Institution "Scientific Centre for Family Health and Human Reproduction Problems", Irkutsk, Russia*

[3] *Irkutsk Regional Infectious Clinical Hospital, Irkutsk, Russia*

[4] *Scientific Research Institute of Medical Problems of the North, Federal Research Center "Krasnoyarsk Science Center" SB RAS, Krasnoyarsk, Russia*

[5] *City Infectious Clinical Hospital No. 1, Novosibirsk, Russia*

** e-mail: mikhail@bionet.nsc.ru*

Key words: tick-borne encephalitis, haplotype analysis, HFE gene

Motivation and Aim: Human *HFE* gene is located in short arm of human chromosome 6, 4 megabases from the major histocompatibility complex on the telomeric side. This human-genome locus encodes for human leukocyte antigens and is characterized by significant linkage disequilibrium and a high polymorphism level at the same time. It has been shown that some intragenic *HFE* haplotype frequencies are race specific. We suppose the sharp phylogenetic difference in this locus could be explained by natural selection under pathogen pressure. Tick-borne encephalitis (TBE) was probably one of such endemic infections in North Eurasia.

Methods and Algorithms: Haplotype analysis for the rs1799945, rs1800730, rs1800562, rs2071303, rs1800708, and rs1572982 was performed in 166 Russian patients with different clinical forms of TBE and in a control population group (356 individuals). The case sample consists of 128 non-immunized (44 with fever, 49 with meningitis, and 35 with severe central nervous system disease) and 38 immunized Russian patients with TBE.

Results: We did not reveal any genetic difference among immunized patients with different forms of TBE. Frequency of the TTA haplotype of the *HFE* gene – in the groups of non-immunized patients with severe central nervous system disease – are higher than such frequencies in TBE patients with fever, with meningitis, and in Russian population (0.2 vs. 0.1, 0.15, and 0.14, respectively). Previously, TTA haplotype frequency was shown to be 0.02-0.07 in East Siberian native populations. Significant differences in the TTG/TTG genotype frequency were found between the sample of TBE patients with fever and population cohort (0.38 vs. 0.22, respectively, $P = 0.022$), as well as in the TTG allele frequency between the sample of TBE patients with fever and with severe central nervous system disease (0.57 vs. 0.4, respectively, $P = 0.026$).

Conclusion: TTA haplotype of the *HFE* gene is possibly associated with severe form of TBE and could be under selection pressure in North Eurasia. Additional studies with TBE patients are required for further validation of the results.

Acknowledgements: Supported partially by the Russian Science Foundation (Grant No. 16-15-00127) and by project 0324-2018-0016 of the Federal Research Center "Institute of Cytology and Genetics" SB RAS.

DOI 10.18699/BioGenEvo-2018-24

Genetic footprints of medieval nomads on the crossroads of civilizations in Southern Russian

A. Mikheyev, E. Batyeva, V. Klyuchniov, Yu. Orlov, N. Moshkov, I. Dmitrievsky, T. Lorentz, T. Tatarinova*
University of La Verne, CA, USA
** e-mail: tatiana.tatarinova@gmail.com*

Key words: human genetics, population genetics, genome, paleogenetics, evolution

Investigations of human origin sparkles numerous debates. Emergence of consumer genetics and genotyping of thousands of individuals around the globe did nothing to calm these debates – just the opposite. A person (or a group of people) who have nothing in common genetically with the ethnic group with which they identify culturally, will naturally, contest and oppose the findings. It is human to be nervous about our genomes. We do not see it, but we have a copy of this genome in every cell of our body. And this invisible but powerful thing makes us who we are, determines kinship, sets limitations and pre-determines illnesses. It is very human to fear such ghostly unharnessed power. The potential to control human genome is even scarier to many people. Genetics has the power to rewrite family and country histories, uncover infidelities, and challenge ownership.

We have conducted anthropological and genetic analysis of three skeletons from the burial sites in Southern Russia, dated early-middle 9th century AD. In this talk, we will present our analysis pipeline, talk about the hurdles and lessons we learned from overcoming them, and formulate a hypothesis about the trace of early medieval nomads in genomes of contemporary eastern Europeans.

DOI 10.18699/BioGenEvo-2018-25

Massive inter-phylum lateral gene transfer from *Planctomycetes*: the case of TIGR02604 family of the putative glycoside hydrolases

D.G. Naumoff
Winogradsky Institute of Microbiology, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia
e-mail: daniil_naumoff@yahoo.com

Key words: glycoside hydrolase, β-galactosidase, TIGR02604, *Planctomycetes*, *Verrucomicrobia,* protein evolution, lateral gene transfer, paralogue, orthologue, CAZy

Motivation and Aim: Genome sequencing has revealed that lateral gene transfer is a major evolutionary process in bacteria. Genes of the metabolic enzymes and particularly various glycoside hydrolases are among the most intensively transferred ones. Over 150 glycoside hydrolase families are currently recognized in the Carbohydrate-Active Enzymes database (http://www.cazy.org/). However, a wide diversity of the putative glycoside hydrolases remains unclassified. This is especially true for members of the families which are significantly underrepresented in the well-studied bacterial phyla (*Proteobacteria*, *Firmicutes*, and *Actinobacteria*). One particular example is TIGR02604 family, which is overrepresented in *Planctomycetes* and *Verrucomicrobia* (http://www.jcvi.org/cgi-bin/tigrfams/index.cgi). The only experimentally characterized member of this family (GenPept, AGW45552.1) has the β-galactosidase activity [EC 3.2.1.23]. Detailed phylogenetic analysis of its closest homologues is the purpose of this work.
Methods and Algorithms: Protein sequences were retrieved from the NCBI database. Multiple sequence alignment of TIGR02604-proteins was made in BioEdit program. The phylogenetic trees were built using programs of PHYLIP package.
Results: Eighteen proteins from TIGR02604 family were identified during genome annotation (GenBank, CP019082.1) of the recently described planctomycete from wetlands, *Paludisphaera borealis*. Pairwise comparison allowed us to group them into ten subfamilies based on the sequence similarity. Four of these proteins belong to the same subfamily as the β-galactosidase mentioned above. Screening of the database revealed about 200 members of this subfamily and clear overrepresentation of proteins from *Planctomycetes*. No representatives from *Proteobacteria*, *Firmicutes*, *Actinobacteria*, *Cyanobacteria*, *Chloroflexi*, and *Spirochaetes* were found. Proteins from *Acidobacteria*, *Bacteroidetes*, and *Verrucomicrobia* composed multiple stable clusters on the subfamily phylogenetic tree suggesting massive lateral gene transfer.
Conclusion: Phylogenetic analysis of a new family of putative β-galactosidases revealed multiple lateral gene transfer events from *Planctomycetes* to *Acidobacteria*, *Bacteroidetes*, and *Verrucomicrobia*. Poorly studied bacterial phyla are an important source of new enzymes for science and industry.
Acknowledgements: This analysis was performed in frame of the budgetary support (research topic No. 01201350930).

DOI 10.18699/BioGenEvo-2018-26

Complex life cycle: a set of phenotypes on the single genome

M. Nesterenko[1], V. Starunov[1, 2], S. Shchenkov[1, 3]*, A. Dobrovolskij[1], K. Khalturin[1, 4]
[1] *Saint-Petersburg State University, St. Petersburg, Russia*
[2] *Zoological Institute of RAS, St. Petersburg, Russia*
[3] *Kovalevsky Institute of Marine Biological Research of RAS, Russia*
[4] *Okinawa Institute of science and technology, Okinawa, Japan*
** e-mail: sergei.shchenkov@gmail.com*

Key words: complex life cycle, RNA-seq, comparative transcriptomics, trematodes

Motivation and Aim: Complex life cycles are characteristic to different groups of invertebrates, free-living as well as parasitic. During the single life cycle implementation several generations contrasting in morphology, physiology and behavior alternate successively and regularly. In the life cycles with heterogony different generations and phases are formed on the basis of single genome. In the trematodes life cycle (which are endoparasites with heterogony) amphimictic generation alternates parthenogenetic ones, and both of them include free-living and parasitic phases. The aim of our research is to reveal the molecular basics of different phenotype formation during trematodes life cycle. As an object of study, we chose two trematodes species from Psilostomatidae family, *Sphaeridiotrema pseudoglobulus* and *Psilotrema simillimum*.
Methods and Algorithms: Transcriptomes of rediae, cercariae and maritae of both investigated species were sequenced using Illumina Hiseq 2500 instrument and assembled *de novo* by Trinity, SOAPdenovo-Trans and TransABySS. Contigs were clustered (CD-HIT-EST) and only best ones were selected by TransRate. Web-resources FunctionAnnotator and KEGG were used for annotation. The close sequences identification between transcriptomes were performed by reciprocal best hit search and custom Python scripts. To determine the domain architecture of predicted amino acids sequences, hidden Markov models (hmmscan versus Pfam-A database) and custom Python parser were used. Comparison of sets of highly expressed genes both among life cycle phases, and between two species were performed by TROM package for R.
Results: Selected assembling strategy allowed us to obtain high-quality transcriptome for each life cycle phase of both species. Approximately 50 % of the assembled sequences had analogs in NCBI NR protein database, however domain architecture of protein was predicted for contigs with annotation as well as for many unannotated sequences. Among homologous and orthologues sequences at least 1/5 part demonstrate variation in the domain architecture. The corresponding phases of the two trematode species show more significant overlap in highly expressed gene sets, than the phases within a single life cycle.
Conclusion: In current biology genome-phenotype relationship is still weakly studied, however it becomes clear, that not only gene expression or its absence determines the traits. Present results suggest that during the implementation of a complex life cycle both the change in the gene expression level and creation of transcripts isoforms are important.
Acknowledgements: Scientific research were performed at the Research park of St. Petersburg State University “Biobank”, “Center for Molecular and Cell Technologies”, “Computing Center”, and “Center for Culture Collection of Microorganisms”. The research was founded by grant MK-2105.2017.4.

DOI 10.18699/BioGenEvo-2018-27

Comparison of evolutionary rates of the regions and nucleotide substitutions in the *Allium* plastomes

D.O. Omelchenko[1, 2]*, A.A. Krinitsina[1, 2], M.D. Logacheva[1, 2], M.S. Belenikin[3], E.A. Konorov[1, 4], S.V. Kuptsov[1], A.P. Seregin[1], A.S. Speranskaya[1, 2]
[1] *Lomonosov Moscow State University, Moscow, Russia*
[2] *Skolkovo Institute of Science and Technology, Moscow Region, Russia*
[3] *Moscow Institute of Physics and Technology, Moscow Region, Dolgoprudny, Russia*
[4] *Vavilov Institute of General Genetics, Moscow, Russia*
** e-mail:*

Key words: *Allium*, evolution, high-throughput sequencing, chloroplast genome

Motivation and Aim: Genus *Allium* includes several economically important species, some of them of the oldest cultivated crops, along with some rare and endemic species. Still, complete sequences of *Allium* genomes are majorly underrepresented (only *A. cepa*, *A. sativum* and *A. obliquum* has complete chloroplast DNA records in GenBank). Modern molecular taxonomy of the genus is based on nuclear internal transcribed spacers (ITS) sequences analysis. Here we performed sequencing, assembly and comparative analysis on evolution and diversity of the complete sequences of the chloroplast genomes of *Allium* species, representing all three evolutionary lines of this genus.
Methods and Algorithms: The plastomes of 11 *Allium* species (*A. ursinum*, *A. paradoxum*, *A. zebdanense*, *A. victorialis*, *A. macleanii*, *A. fistulosum*, *A. nutans*, *A. platyspathum*, *A. obliquum*, *A. shoenoprasum*, *A. pskemense*) were sequenced on the Illumina MiSeq and assembled *de novo* (using various assemblers, including Velvet and MIRA4). Sequences were annotated by DOGMA and GeSeq, followed by manual review and correction of annotated features where necessary. Phylogenetic analysis was performed with Mr. Bayes and phangorn R package. Length of branches was calculated, and natural selection was detected by PAML and HyPhy. Evolutionary rates were calculated using Erable.
Results: Besides expected differences represented by small indels in intergenic spacers, there are conspicuous characteristic large deletions and nonsynonymous substitutions in some genes in the *Allium* plastomes, some of them specific to the evolutionary lines. *A. paradoxum* plastome showed most interesting results that differ it from other *Allium* species – 4.9 kbp inversion of *rpl32-ndhE* region with impairment of *ndhG* and *ndhF* genes in SSC. Evolutionary analysis showed that some genes commonly used for phylogenetic analysis (e. g., *matK*) have less constant evolutionary rate within the genus *Allium* compared to intergenic regions.
Conclusion: Analysis of *Allium* plastomes revealed prominent species specific and even evolutionary line specific differences in sequence and gene set. We assume the effect of evolution on various parts of *Allium* plastomes differs in character between species of this genus.
Acknowledgements: Supported by the RFBR grant No. 18-04-01203.

DOI 10.18699/BioGenEvo-2018-28

Analysis of transcription binding and developmental genes regulated by Zic3 factor in zebrafish

Y.L. Orlov[1,2,3]*, C.L. Winata[1], I. Kondrychyn[1], S.S. Kovalev[2], A.V. Tsukanov[2]
[1] *Genome Institute of Singapore, Singapore*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[3] *A.O. Kovalevsky Institute of Marine Biology Research RAS, Sevastopol, Russia*
** e-mail: orlov@bionet.nsc.ru*

Key words: genomics, bacteria, microbiology, sequencing, environments, GC content, bioinformatics

Motivation and Aim: Zebrafish *(D.rerio)* is a model organism for neurobiology with growing number of genome sequencing experiments. We aimed analysis of transcription regulation in development based on ChIP-seq and RNA-seq experiments [1]. Object of the study Zic3 belongs to a family of transcription factors known for their role in early embryonic patterning. In the vertebrates, loss of Zic3 function is known to disrupt gastrulation, left-right patterning, and neurogenesis. Zic genes are the vertebrate homologues of the Drosophila odd-paired gene, which is involved in early embryonic patterning. However, molecular events downstream of this transcription factor were poorly characterized as well as transcription factor binding in the genome.
Methods and Algorithms: Here we use the zebrafish as a model to study the developmental role of Zic3 in vivo. Sequencing of the 8 hpf (hours post fertilization) ChIP sample generated and the 24 hpf ChIP sample generated about 20 mln reads, about 51% of which were mappable. Genomic regions of significant enrichment representing Zic3-binding sites (peaks) were identified using the peak-calling algorithm QuEST.
Results and conclusion: Using a combination of two genomics approaches – ChIP-seq and microarray, we identified Zic3 targets, which include genes from the Nodal and Wnt pathways, and uncovered a previously unrecognized link between Zic3 and the non-canonical Wnt pathway in gastrulation and left-right patterning. Only a minority of Zic3 binding sites were found within promoter regions. We show for the first time cis-regulation of several of these target genes by Zic3. Binding site analysis of Zic3 revealed a biased distribution towards distal intergenic regions, indicative of a long distance regulatory mechanism; some of these binding sites were highly conserved during evolution and were functional enhancers. Our study establishes the zebrafish as an excellent model for genome-wide study of a transcription factor in vivo.
Acknowledgements: The research has been supported by the Ministry of Education and Science of the Russian Federation grant No. 14.W03.31.0015.

References

1. Winata C.L., Kondrychyn I., Kumar V., Srinivasan K.G., Orlov Y., Ravishankar A., Prabhakar S., Stanton L.W., Korzh V., Mathavan S. (2013) Genome-wide analysis reveals Zic3 interaction with distal regulatory elements of stage specific developmental genes in zebrafish. PLoS Genetics, 9(10): e1003852.

DOI 10.18699/BioGenEvo-2018-29

Biodiversity and genomics of diatoms

D. Romanova*, E. Nevrova
The A.O. Kovalevsky Institute of Marine Biological Research of RAS, Sevastopol, Russia
** e-mail: driaromanova@yandex.ru*

Key words: diatoms, biodiversity, genomics

Motivation and Aim: Diatoms are single-cell photosynthetic algae that fix up to 20 % of total marine primary productivity [1]. Biodiversity of diatoms is huge: *a.* 12000 diatom species have already been described, and *a.* 100000 are assumed to exist [2]. In this regard, joint analysis on the species richness assessment and genomes survey of Bacillariophyta are the crucial both for development of the measures to maintain the sustainable functioning of marine ecosystem chains and conservation of the gene pool. Based on the published sources and own surveys, the current taxonomic richness of Black Sea benthic diatoms was evaluated. Updated diatom inventory holds 1094 species and intraspecific taxa, pooled in 953 species, 149 genera, 61 families, 32 order and 3 classes Bacillariophyta [3]. At present, genomes of 8 diatom species have been obtained only [4–6]. The mainstream of diatom research is elicitation of genes responsible for the silicon transport during the valves morphogenesis, the uptake of high-affinity iron, biosynthetic enzymes, a complete urea cycle. An explanation of these mechanisms will help to understand the wide prevalence of diatoms, which able to survive even in soils, ice and hot springs. Another important aim is the study of diatom genetic regulatory elements that determine gene expression and the light response regulation [7]. According to some data, there is a mechanism for switching from phototrophic to heterotrophic feeding under limited light regime.

Conclusion: Combined taxonomical evaluation of diatom diversity and its genomes should be employed for necessity of environmental security measures and conservation of marine flora biodiversity at the modern transformation and anthropogenic development of the Black Sea and World Ocean shores. The obtained in future research results will allow to consider the diatom evolution, ecology and metabolism processes within the framework of the biodiversity concept.

Acknowledgements: Supported by the Ministry of Education and Science of the Russian Federation grant No. 14.W03.31.0015.

References

1. The Diatom World. Seckbach J., Kociolek J.P. (eds) (2011) Dordrecht, Heidelberg, London, New-York: Springer. 533 p.
2. Mann D.G., Vanormelingen P. (2013) An inordinate fondness? The number, distributions and origins of diatom species. Journal Eukaryotic Microbiology. 60:414-420.
3. Nevrova E.L. (2015) Benthic Diatoms (Bacillariophyta) in the Black Sea: diversity and structure of taxocenes of varios biotopes : Dr. Sci. dissertation, Moscow. 442 p.
4. Armbrust E.V. et al. (2004) The Genome of the Diatom *Thalassiosira pseudonana*: Ecology, Evolution, and Metabolism. Science. 306(5693):79-86.
5. Bowler C. et al. (2004) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. Nature. 456:239-244.
6. Tang X., Bi G. (2016) Complete mitochondrial genome of *Fistulifera solaris* (Bacillariophycidae). Mitochondrial DNA A DNA Mapp Seq Anal. 27(6):4405-4406.
7. Lepetit B. et al. (2017) The diatom *Phaeodactylum tricornutum* adjusts NPQ capacity in response to dynamic light via fine-tuned Lhcx and xanthophyll cycle pigment synthesis. New Phytologist. 214(1): 1469-8137.

DOI 10.18699/BioGenEvo-2018-30

Microbial diversity in the hot spring Faust Lake, Kunashir Island

A. Rozanov[1]*, A. Korzhuk[1, 2], S. Peltek[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** e-mail: sibiryak.n@gmail.com*

Key words: metagenomic, hot spring, microbial communities, genome

Motivation and Aim: Metagenomics is the study of genetic material recovered directly from environmental samples. This approach makes it possible to obtain information about uncultivated microorganisms. Microbial communities of extreme ecosystems have an applied and fundamental interest. Applied interest is mainly focused on searching for new protein sequences. From the fundamental point of view communities of extreme ecosystems are interesting as simpler models, in comparison with communities of usual ecosystems. But the greatest interest now is caused by the fact that communities of extreme ecosystems are living chronicles that contain information about microbial communities of long-past geological epochs.
Kunashir is one of the southern islands of the Kurile-Kamchatka volcanic belt located in Russia. The large amount of organic remains from the environment that fall into water sources are favorable for the microflora grows. Its makes hot springs on the island were interestin for microbiologist. Due to the closeness of the territory, the geothermal springs of the southern Kuril Islands, were not studied by microbiology methods and with the metagenomic approaches.
Methods and Algorithms: In the summer of 2017, researchers from the Laboratory of Molecular Biotechnologies of the IC&G SB RAS organized the expedition to collect material for metagenomic research. To date, we have performed a metagenomic sequencing of microbial communities from the bottom sediments of Faust Lake. Libraries for sequencing were prepared by Ph.D. Vasilyev G.V. in the Center for Genomic Research of the IC&G SB RAS. Sequencing was carried out by Genetics and Reproductive Medicine Center “Genetico” (Moscow) on the Illumina NovaSeq 6000.
Results: The data allows extracting most of the genomic sequences for prokaryotes with abundance greater than 0.1 % of the total number of microorganisms. During the treatment, we obtained data about a fairly complete sequence of new Cyanidioschyzon species. It is unicellular haploid red alga adapted to high sulfur acidic hot spring environments. We have data about genomes of new bacterial species with different complitnes. Previously, similar work in Russia was not performed.
Conclusion: The results obtained in the work will give a new material for phylogeography of microorganisms and also will allow obtaining data on the genomes of previously unknown microorganisms.
Acknowledgements: Supported by the fundamental research program of the SB RAS II.1 No. 0324-2018-0039.

DOI 10.18699/BioGenEvo-2018-31

Metagenomic analysis of metabolically active microbial communities of Salenoye Lake # 48 in the Novosibirsk region

A. Rozanov[1]*, A. Shipova[1, 2], A. Bryanskaya[1], E. Lazareva[3], O. Taran[4], S. Peltek[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Institute of Geology and Mineralogy SB RAS, Novosibirsk, Russia*
[4] *Boreskov Institute of Catalysis SB RAS, Novosibirsk, Russia*
** e-mail: sibiryak.n@gmail.com*

Key words: metagenomic, biodiversity of microorganisms, salt lakes, microbial communities

Motivation and Aim: Microbial communities of saline lakes are highly productive in terms of production of organic matter. This is primarily due to the fact that the accumulation of high concentrations of organic matter in cells is necessary to counteract the high osmotic force of the medium. Saline lakes located in the south of western Siberia are interesting for studying for several reasons: they are in a territory with a temperate climate and are subject to constant fluctuations in the concentration of salts and temperatures; they are remote from the most studied saline lakes, which allows us to obtain data about new species. The aim of this study was to investigate phylogenetic and metabolic properties of most metabolic active parts of saline lakes ecosystems.
Methods and Algorithms: As a model object we took the Solenoye Lake # 48 (Bagan district, Novosibirsk region). We have performed a metagenomic sequencing of microbial communities from of microbial mat and top layer of sediments. Libraries for sequencing were prepared by Ph.D. Vasilyev G.V. in the Center for Genomic Research of the ICG SB RAS. Sequencing was carried out by Genetics and Reproductive Medicine Center "Genetico" (Moscow) on the Illumina NovaSeq 6000. The following resources were used for data processing and analysis: fastQC, trmamatics, metaSPAdes, metaQuast, MaxBin, MGrast.
Results: We performed a metagenomic sequencing of the two most active microbial communities of this lake. The total amount of data was 50 GB for each point. This allows us to obtain genomes of microorganisms with a high degree of coverage, a sequence of viral genetic material. In contrast to the 16s rRNA of metagenomic analysis, this approach allowed to obtain information on species diversity without distortion caused by the selection of primers and PCR. In particular, we were able to obtain information on the presence in the microbial communities of diatom algae of the family Phaeodactylaceae. In the report, we will provide information on the phylogenetic diversity and metabolic possibilities of the microbial communities under investigation, as well as on the genomes of microorganisms that will be extracted by that time.
Acknowledgements: The reported study was funded by RFBR and Novosibirsk region according to the research project No. 17-44-540815, and by ICG SB RAS budget project No. VI.58.1.8.

DOI 10.18699/BioGenEvo-2018-32

Challenges of *in vitro* conservation of Citrus genetic resources

L.S. Samarina[1]*, V.I. Malyarovskaya[1], R.S. Rakhmangulov[1], Y.L. Orlov[2], O.B. Dobrovolskaya[2]

[1] *Russian Research Institute of Floriculture and Subtropical Crops, Sochi, Russia*

[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*

** e-mail: samarinalidia@gmail.com*

Key words: Citrus, contamination, micropropagation, culture media, true-to-type, endophytes

Motivation and Aim: To preserve *in vitro* collections of plant genetic recourses including elite cultivars of tree crops, it is necessary to use vegetative buds and meristems as explants that are genetically identical to the stock plant. However, the stage of aseptic culture initiation of woody vegetative buds is associated with such challenges as a high contamination rate *in vitro*, a slow growth, and the appearance of secondary-infected explants. The main challenges in creating *in vitro* backup collection of Citrus cultivars are a high degree of fungal contamination of vegetative buds and the subsequent decrease in the growth potential of plantlets *in vitro*. It is well known, woody plants, in particular Citrus species, are in close cohabitation with fungal microorganisms. Surface sterilization of explants does not relieve tissue from internal infection, on the cultural medium the hyphae of the fungus leave the plant tissues and proliferate on culture media, which inhibits the development of explants *in vitro* [1]. In this regard, the aim of our work is to study the effectiveness of various tools of decontamination of vegetative Citrus explants for initiation *in vitro* culture and development of a reliable *in vitro* conservation.

Methods and algorithms: Pre-cultivation techniques, pre-treatment of cuttings with fungicides, gradual sterilization, addition of antibiotics to the nutrient medium, as well as micro-grafting were tested for establish efficient tissue culture initiation of elite lemon cultivars of collection FGBSI RRIFSC (Sochi, Russia). The studies were carried out on *Citrus limon* (L.) Burm cultivars. Axillary buds with shoot segment of 0.5–0.7 cm long were taken as explants.

Results: The highest rate of sterile explants 32–42 %, was obtained by pre-cultivation the cuttings in a incubating chamber at 22 °C, followed by pretreatment with fungicides, using gradual sterilization and adding tetracycline 400 mg/l to the culture medium. However, after the third passage of subculture and conservation, the viability of the plantlets was reduced, leaves dropped and plantlets died within 6 months of in vitro conservation. Using *in vitro* micrografting technique, it was possible to overcome the problem of viability losses of bud explants, but this method is extremely labor- and time-consuming, only 1 of 300 manipulations was successful.

Conclusion: Thus, reliable preservation of valuable citrus genotypes in collection *in vitro* is currently problematic due to close cohabitation with fungal microorganisms and low efficiency of micropropagation as a whole.

References

1. Samarina L.S. (2013) Optimization of micropropagation and conservation protocols for lemon cultivars *in vitro*. PhD Thesis / Russian State Agrarian University. Moscow.

DOI 10.18699/BioGenEvo-2018-33

Hidden diversity of myxomycetes: problems and perspectives

O.N. Shchepin[1, 3]*, Y.K. Novozhilov[1], I.S. Prikhodko[2], M. Schnittler[3]
[1] *Komarov Botanical Institute of the Russian Academy of Sciences, Saint Petersburg, Russia*
[2] *St. Petersburg State University, Saint Petersburg, Russia*
[3] *Institute of Botany and Landscape Ecology, EMA University of Greifswald, Greifswald, Germany*
** e-mail: ledum_laconicum@mail.ru*

Key words: myxomycetes, hidden diversity, species concepts

Motivation and Aim: Myxomycetes (= Myxogastria, plasmodial slime molds) represent a monophyletic group of free-living amoeboid protists within the supergroup Amoebozoa that are characterized by a unique life cycle with an alternation of uninucleate myxamoebae/swarm cells, multinucleate plasmodia and fruiting bodies (sporocarps) filled with airborne spores. Since Linnean times about 1000 species were validly described within five orders based almost exclusively on morphological characters of the sporocarps. However, DNA sequence-based studies revealed a paraphyly of many taxonomical groups within the class as well as an unexpected extent of hidden diversity. The direction of further research in the field of myxomycete taxonomy and diversity studies using modern approaches should be outlined.
Results and discussion: In recent years, a number of molecular phylogenetic studies revealed two problems of the morphological approach. 1. The long established five-order system does not reflect correctly the phylogenetic relationships within the class. This could be solved by creating a system based on the analysis of multiple gene markers, e. g. transcriptomic or genomic data. At the moment, transcriptomic data are available for only four myxomycetes species and genomic data for one. 2. Many morphospecies seem to comprise several, reproductively isolated, biospecies; others are even paraphyletic [1]. Application of DNA barcoding [2] and 18S amplicon metagenomics [3] provided evidence that diversity assessments based on morphological determination of sporocarps underestimate the diversity of the group. The presence of phylogenetic clades consisting of OTUs not matching to any known species, as well as several reports of myxamoebal strains with unknown species identity isolated from unusual habitats lend evidence for a significant, yet undescribed diversity of myxomycetes which presumably never or rarely form sporocarps. However, the incompleteness of the currently existing database of reference sequences does not allow us to distinguish such species reliably.
Acknowledgements: Supported by the RFBR (15-29-02622, 18-04-01232 A), by Presidium of UB RAS (12-M-14-2009, 12-П-4-1067) and by German DFG (SCHN 1080/2, RTG 2010).

References

1. Schnittler M., Shchepin O.N., Dagamac N.H., Borg Dahl M. et al. (2017) Barcoding myxomycetes with molecular markers: challenges and opportunities. Nova Hedwigia. 104(1-3):323-341.
2. Shchepin O.N., Dagamac N.H., Sanchez O.M., Novozhilov Y.K. et al. (2017) DNA barcoding as a tool for identification of plasmodia and sclerotia of myxomycetes (Myxogastria) appearing in moist chamber cultures. Mycosphere. 8(10):1904-1913.
3. Borg Dahl M., Brejnrod A.D., Unterseher M., Hoppe T. et al. (2017) Genetic barcoding of dark-spored myxomycetes (Amoebozoa) – identification, evaluation and application of a sequence similarity threshold for species differentiation in NGS studies. Molecular Ecology Resources. 18(2):306-318.

DOI 10.18699/BioGenEvo-2018-34

Polymorphism in the promoter region of the squalene synthase gene in different amaranth species

A.B. Shcherban*, A.I. Stasyuk, E.A. Salina
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** e-mail: atos@bionet.nsc.ru*

Key words: amaranth, squalene, squalene synthase gene, promoter, polymorphism

Motivation and Aim: Grain amaranth is potentially an important source of squalene, a natural 30-carbon organic compound that has numerous commercial purposes such as lubrication and protection of the skin (in cosmetics), as an adjuvant in vaccines (in medicine), etc. According to previous reports, the squalene concentration in amaranth oil is many times higher than that in other plant oils and varies between different species of amaranth [1]. The squalene synthase (SQS) gene may play a significant role in regulation of squalene biosynthesis. However, structural polymorphism of this gene among amaranth species differing in squalene content have not yet been studied.
Methods and Algorithms: Based on the known coding sequence of the SQS gene from *A. cruentus* [2], we downloaded the corresponding nucleotide sequence of *A. hypochondriacus* (including the promoter region ~1 kb long), using the high-quality draft genome sequence of this species available at GoGe (id40120; (https://genomevolution.org/coge/). The predicted amino acid SQS sequences from both species showed a high level of identity, especially, within the most conservative functional domains. We designed specific primers to the promoter region of the SQS gene of *A. hypochondriacus* and used them to amplify and sequence this region from 95 accessions related to 23 amaranth species. Multiple sequence alignments and the subsequent phylogenetic analysis were carried out using the ClustalW program and MEGA4 software. The putative cis- regulatory elements in the gene promoter were searched using database PlantPAN 2.0 (http://plantpan2.itps.ncku.edu.tw).
Results: A high level of interspecific polymorphism was revealed inside the promoter of the SQS gene. The most significant variation that could affect a probable transcriptional regulatory sites was found immediately upstream the putative TATA-box located at position -168 from ATG-codon. Based on the alignment of the nucleotide SQS promoter sequences the neighbor-joining tree was constructed.
Conclusion: The results obtained imply a major impact of the SQS promoter on the regulation of the gene expression and allow to proceed to the next stage, the study of association between the promoter polymorphism and the SQS expression patterns which in turn may determine production of squalene in different amaranth species.

References

1. He H.P. et al. (2002) Extraction and purification of squalene from *Amaranthus* grain. J. Agric. Food Chem. 50:368-372.
2. Park Y.-J. et al. (2016) Molecular cloning, expression and characterization of a squalene synthase gene from grain Amaranth (*Amaranthus cruentus* L.). Japan Agricultural Research Quarterly. 50(4):307-317.

DOI 10.18699/BioGenEvo-2018-35

Inferring phylogeny among cryptic lineages of *Eisenia nordenskioldi nordenskioldi* (Lumbricidae) based on transcriptomic data

S.V. Shekhovtsov[1, 2]*, N.I. Ershov[1], G.V. Vasiliev[1], S.E. Peltek[1]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Biological Problems of the North FEB RAS, Magadan, Russia*
** e-mail: shekhovtsov@bionet.nsc.ru*

Key words: earthworms, transcriptomics, phylogeny, cryptic lineages

Motivation and Aim: Cryptic diversity is widespread among little-studied taxa with poor morphologic diversity. In earthworms it is manifested by the presence of strongly diverged (10–20 % of sequence divergence for the COI gene) mitochondrial haplotypes within otherwise morphologically solid species, and even within single populations. Such situation was found in *Eisenia nordenskioldi* (Eisen, 1879) a widespread Siberian species, in which at least 14 cryptic mitochondrial lineages were found [1].
It proved hard to find out if nuclear genomes of these cryptic lineages are have a similar level of sequence divergence, because many earthworms are polyploids, and thus construction of universal primers, DNA amplification and sequencing often fail. In this study we obtained transcriptomic data for five genetic lineages of *E. n. nordenskioldi*, the pigmented subspecies of *E. nordenskioldi*, as well as from the congeneric *E. andrei*, and attempted to reveal phylogenetic relationships among them.
Methods and Algorithms: Total RNA was extracted from living specimens, poly-A fraction was isolated, reverse transcribed and sequenced using the IonTorrent platform. Reads were assembled using the Trinity software. ORFs and corresponding proteins were predicted using TransDecoder, and CD-HIT was utilized to remove duplicated sequences (95 % similarity threshold). Orthogroups of proteins were defined using ProteinOrtho and aligned using Clustal-Omega aligner. Alignments were strictly trimmed with TrimAl, concatenated and converted to NEXUS format. MrBayes software was used to select an appropriate model (Jones) and infer phylogeny.
Results: We constructed a Bayesian phylogenetic tree for protein sequences of 1185 orthologous sequences with the total alignment length of 190 081 a. a. *E. n. nordenskioldi* proved to be monophyletic. Cryptic lineages of *E. n. nordenskioldi* had 7.4–7.7 % sequence divergence from the outgroup (p-distance), while intraspecific divergence was as low as 2.2 % between lineages 9 and 7 (from Magadan and Bashkiria, respectively), up to 3.9–5.3 % among the rest of the lineages. Intraspecific divergence within *E. n. nordenskioldi* is thus sufficiently high, comparable to that between *E. n. nordenskioldi* and strongly differing congeneric *E. andrei*.
Conclusion: The obtained results suggest that cryptic lineages of *E. nordenskioldi* can indeed be considered as separate species.
Acknowledgements: Supported by the RFBR (18-04-00507_a) and by the State Assignment No. 0324-2018-0017.

References

1. Shekhovtsov S.V., Berman D.I., Bazarova N.E., Bulakhova N.A., Porco D., Peltek S.E. (2016) Cryptic genetic lineages in *Eisenia nordenskioldi pallida* (Oligochaeta, Lumbricidae). European Journal Soil Biology. 75:151-156.

DOI 10.18699/BioGenEvo-2018-36

An intron of the *hsp90* gene as a new promising phylogenetic marker for the genus *Carex* L.

I.N. Shekhovtsova[1], S.V. Shekhovtsov[2, 3]*, S.E. Peltek[2]
[1] *Central Siberian Botanical Garden SB RAS, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[3] *Institute of Biological Problems of the North FEB RAS, Magadan, Russia*
** e-mail: shekhovtsov@bionet.nsc.ru*

Key words: *Carex*, sedges, DNA barcoding, phylogeny

Motivation and Aim: *Carex* L. (family Cyperaceae) is one of the most species-rich genera of higher plants. It contains over 2000 species. Many aspects of *Carex* taxonomy are yet unclear, so DNA sequence data is promising as a supplement to traditional morphological comparison. Traditional DNA barcoding markers, i.e. various plastid genes and intergenic spacers and the nuclear ribosomal gene cluster are often insufficiently informative, which makes the development of new gene markers an important task.
Methods and Algorithms: We tested a set of universal primers from [1] on sedges belonging to sections *Vesicariae* Meinsh., *Paludosae* Fries ex Kük., and *Carex*.
Results: Most amplifications were either unsuccessful or yielded conservative sequences with very few substitutions. However, in the case of the *hsp90* gene we obtained an intron approximately 450 bp long flanked by highly conservative exons. We designed a set of universal primers for this intron and checked amplification consistency on a set of *Carex* species. Amplification was successful in ~80% of specimens. The *hsp90* intron was found to harbor more characteristic sequence substitutions that conventional plastid and nuclear ribosomal markers. For the section *Vesicariae*, a portion of the *hsp90* intron 314–366 bp long (size difference due to indels) contained 11 parsimony-informative substitutions and 5 taxon-specific indels, while a part of the *matk* gene 591 bp long had only three parsimony-informative substitutions, and the ITS2 spacer 438 bp long had five. We should note that sequencing was impeded by the presence of two poly-T tracts, both 7–9 bp long. However, high sequence divergence of the *hsp90* intron compensates for its comparatively low reliability, and its short length makes is very convenient for the study of herbarium specimens that often have degraded DNA.
Conclusion: The *hsp90* intron is a promising marker for DNA barcoding and phylogenetic studies of sedges.
Acknowledgements: Supported by the State Assignment No. 0324-2018-0017.

References

1. Steele P.R., Guisinger-Bellian M., Linder C.R., Jansen R.K. (2008) Phylogenetic utility of 141 low-copy nuclear regions in taxa at different taxonomic levels in two distantly related families of rosids. Molecular Phylogenetics Evolution. 48(3):1013-1026.
2. Shekhovtsov S.V., Shekhovtsova I.N., Peltek S.E. (2012) Phylogeny of Siberian species of *Carex sect. Vesicariae* based on nuclear and plastid markers. Nordic Journal Botany. 30(3):343-351.

DOI 10.18699/BioGenEvo-2018-37

The genome and transcriptome of a freshwater bryozoan *Cristatella mucedo*

V. Starunov[1, 2], A. Predeus[3], Y. Barbitov[3], A. Maltseva[1], V. Kutyumov[1], E. Vodiasova[4], E. Chelebieva[4], D. Romanova[4], A. Ostrovsky[1], L. Moroz[4]
[1] *Saint-Petersburg State University, St. Petersburg, Russia*
[2] *Zoological Institute of RAS, St. Petersburg, Russia*
[3] *Bioinformatics institute, St. Petersburg, Russia*
[4] *Kovalevsky Institute of Marine Biological Research of RAS, Sevastopol, Russia*
** e-mail: staunov@gmail.com*

Key words: Bryozoa, *Cristatella mucedo*, genome, transcriptome

Motivation and Aim: The freshwater bryozoan *Cristatella mucedo* is a colonial organism from the widespread, but poorly studied phylum Bryozoa. Colonies of bryozoans consist of numerous zooid modules, which possess their own brains, digestive systems, and trapping apparatuses. The phylogenetic position of bryozoans and their relationship with other animal phyla are still rather disputable. Unlike other bryozoans, the colonies of *C. mucedo* are transparent and do not have mineral skeleton, which makes experimental procedures easier. Moreover, the *C. mucedo* colonies are capable to active directional crawling which is a unique case for colonial invertebrates. Mechanisms of the colony integration are still unknown. The genomic and transcriptomic studies are expected to uncover the molecular basis of integrative mechanisms, uniting individual zooids into a colony.
Methods and Algorithms: We performed whole-genome sequencing with long and short reads. Long reads were obtained with Oxford Nanopore Minion instrument; short paired-end reads were obtained with Illumina MiSeq And Hiseq 2500. The assembly was done using Canu and Abruijn software, with subsequent polishing by Nanopolish and Pilon. The RNA sequencing was performed with Illumina HiSeq 2500 instrument and the transcriptomes were assembled with Trinity. The annotation was done by Dammit and OrthoDB, Rfam, and Pfam databases. We also performed single cell RNA sequencing on the young *C. mucedo* colonies using microfluidic device 10-x Genomics for capturing.
Results: The genome size of *C. mucedo* is about 600–650 Mb. The hybrid genome assembling with both long and short reads was suggested to be an efficient for eukaryotic genomes and allowed us to build a very high-quality assembly with N50 between 1.1–1.3 Mb. The BUSCO score was more than 90 % from "core metazoa" set. The bulk transcriptome was also of very high quality with BUSCO score more than 97 %, and allowed to determine and analyze several important groups of proteins such as neuropeptides, cell adhesion molecules, immune-related molecules, development toolkit genes etc. The single cell sequencing allowed to determine 10 different cellular clusters and find out cluster-specific genes.
Conclusion: Our results make *C. mucedo* a new emerging model to study different integrative and developmental processes in colonial organisms. This project provides powerful instruments for further molecular, biochemical, neurobiological and evolutionary researches.
Acknowledgements: Scientific research were performed at the Research park of St. Petersburg State University "Biobank", "Center for Molecular and Cell Technologies", "Computing Center", and "Center for Culture Collection of Microorganisms". The research were supported by the Ministry of Education and Science of the Russian Federation grant No. 14.W03.31.0015, and by RSF grant 18-14-00086.

DOI 10.18699/BioGenEvo-2018-38

Revised molecular phylogeny of Acrididae family

I. Sukhikh[1]*, K. Ustyantsev[1], V. Vavilova[1], A. Blinov[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Immunology and Physiology UB RAS, Yekaterinburg, Russia*
** e-mail: igor3419@gmail.com*

Key words: Acrididae, phylogeny, grasshoppers, mitochondrial DNA, ribosomal DNA

Motivation and Aim: Acrididae is the biggest family of Acridoidea superfamily, consisting of more than 6000 species out of 12000 species in the Caelifera suborder. Different authors distinguish different number of subfamilies, reaching the limit of 33. It is worth noting that the few subfamilies consists of small number of genus and species. For a long time, taxonomy, the renovation of phylogenetic relationships and understanding of the evolution of Acrididae family were based mainly on comparison analysis of key morphological structures of recent and fossil species. Currently, one of the most effective methods for establishing phylogenetic relationships is the analysis of mitochondrial and ribosomal DNA. For grasshoppers, there are 4 most widely used phylogenetic markers – mitochondrial genes COI, COII, Cytb and NADH5, as well as sequences of nuclear (18S, 28S, including ITS sequences) and mitochondrial (12S, 16S) clusters of ribosomal genes. In the present study, we used markers with the most overlapping species and aimed to include nuclear and mitochondrial DNA.
Methods and Algorithms: New nucleotide sequences of COI, COII and ITS2 markers of Acrididae species were obtained using Sanger sequencing and uploaded in GenBank NCBI database. Phylogenetic trees were built using Maximum likelihood and Bayesian methods.
Results: In present study, we performed phylogenetic analysis of more than 240 species of 14 subfamilies of Acrididae family, based on complete mitochondrial DNA sequences, concatenated DNA sequences of CytB, COII and COI genes, concatenated DNA sequences of COII and COI genes and DNA sequences of ITS2 region. Phylogenetic tree constructed using complete mitochondrial sequences served as a basis for the analysis. Sequences of studied Acrididae species divide into two major groups, which in turn divides into several clusters, as well as four separate branches, each consisting of species of single subfamily (Catantopinae, Oxyinae, Spathosterninae and Proctolabinae). First group consists of representatives of three subfamilies: Acridinae, Gomphocerinae and Oedipodinae; second group includes 8 subfamilies: Catantopinae, Calliptaminae, Conophyminae, Cyrthacantacridinae, Eyprepocnemidinae, Hemiacridinae, Melanoplinae and Pezotettiginae. Five out of 14 subfamilies (Acridinae, Oedipodinae, Gomphocerinae, Oxyinae and Catantopinae) found to be polyphyletic in the present study. Geographical distribution of Acrididae species in polyphyletic branches in most cases correlate with their positions on the phylogenetic tree.
Conclusion: The results obtained show that the current systematic of Acrididae family requires revision of at least five subfamilies, which were found to be polyphyletic. All experimentally obtained sequences of Acrididae species are available in the NCBI GenBank database under accession numbers KX272717-KX272739 for COI gene, KX272670-KX272716 for COII gene and KX289534-KX289579 for ITS2 region.
Acknowledgements: The project was supported by the Russian Foundation for Basic Research (Grant No. 17-04-01615).

DOI 10.18699/BioGenEvo-2018-39

Homologous series and parallel evolution problem

V. Suslov*, M. Ponomarenko, D. Rasskazov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** e-mail: valya@bionet.nsc.ru*

Key words: parallel evolution, transcription factor, Vavilov's law of homologous series

Motivation and Aim: The routine events of contradictions between the classic and molecular phylogenies of any taxa, irrespective of which reconstruction is closer to the truth, translocate the parallel evolution problem from the rare and facile event of convergence evolution associated with selection of similar environment to the homologous series phenomenon (HS), in which adaptation to a certain environment is only one of the reasons for the parallelism. HS arise due to the evolution of any trait in a limited space of possibilities (LSP). Various LSP allows to talk about the bouquet of the laws of HS. Current evolutionary synthesis incorporated only for those for whom the mechanisms of evolution of LSP were identified: Mueller-Haeckel and von Baer laws (explanation is the hourglass of ontogenesis) and HS for inherited traits and non-inherited modification built by Lamarck and explained by genocoping according to Schmalhausen, Lukin or Turesson. Vavilov's law of HS remained outside synthesis because according to STE selection worked with traits, not HS. In 2016, we have shown that this is not so due to traits functional overlap of depending on which the HS must converge, diverge or stabilize even in a stable environment as autoadaptation [1]. Here we continued the search for the laws of HS evolution.
Methods: SNP_TATA_Comparator (http://beehive.bionet.nsc.ru) and the sample of TATA-boxes and composite elements (CE) with TATA-boxes used as [1].
Results: It is shown that clear HS of CE with TATA-boxes are revealed if the in result of the functional overlap between two components of CE overlap of sequences leads to the emergent appearance of a new function. In the function any mutations in any of the components of CE can not be compensated by any substitutions in the other component of CE. Thus, the HS can stabilize not only the minimization of the functional overlapping of traits (the classical HS of Vavilov and Sobolev) [1], but also the emergence of a new function as a by-effect of functional overlap. Otherwise, the HS may be lost by triggering the autoadaptive mechanism of the compensatory divergence [1].
Conclusion: The work of the Meyen school showed [2] but not explained that clear HS are well identified on the molecular-biological (suborganismal) and ecological (or supraorganismal) levels, that is, where both mechanisms of stabilization of the HS mast work best. Complex behavior and stress that appear on the organismal level open up wide opportunities for both compensation and overlapping of functions, which makes the preserve of the HS in evolution on organismal level is problematic.
Acknowledgements: Supported by the grant 0324-2018-0017 and IP 0324-2018-0021.

References

1. Suslov V.V. et al. (2016) Vavilov's homologous series as evolutionary force. BGRS/SB'2016:307.
2. Chebanov S.V. (2017) Meronomy of S.V. Meyen: to the 40th anniversary of formulation. Lethaea rossica. 14:64-92.

DOI 10.18699/BioGenEvo-2018-40

Analysis of genome size, CG content and characteristics of microorganisms' environment habitats

V.V. Suslov[1], A.V. Tsukanov[1,2], Y.L. Orlov[2,3]*
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[3] *A.O. Kovalevsky Institute of Marine Biology Research RAS, Sevastopol, Russia*
** e-mail: orlov@bionet.nsc.ru*

Key words: genomics, bacteria, microbiology, sequencing, environments, GC content, bioinformatics

Motivation and Aim: We aimed to analyze microorganisms habitats and environment dependence by genome size and context characteristics [1]. Research on evolution of early unicellular organisms relies on complete genome sequencing data. Rapid growth of data banks allows us reexamine sequence features necessary for minimal genome size and minimal gene set as unit of evolution. Several theoretical and experimental studies have endeavored to derive the minimal set of genes that are necessary and sufficient to sustain a functioning cell (unicellular organism). Minimal genome size rise questions about relation of environment and genome nucleotide characteristics.
Methods and Algorithms: We have downloaded from NCBI ftp site thousands of complete and assembled whole genome sequences of bacteria and achaea. We used original software for GC content estimation [2] comparing genome size with gene number and GC content separately by groups of organisms.
Results: For whole genome assemblies we found high linear correlation between genome size and gene number (0.8), that been expected, and, more interestingly, between genome size and GC content (0.46). Thus, larger genome size is strongly related to higher fraction of G and C nucleotides. We have considered grouping of species by habitat and environment and found same common trend of correlation between genome size and GC content. But we found bias for some groups. Fusobacteria and planctomycetes have lower GC content that might be expected. *Deinococcus-thermus* have higher GC content.
Conclusion: Overall correlation of GC content to genome size follows the same trend in archaeal and bacterial groups. Comparing oxygen requirement, genome size and GC content we found that aerobic organism have larger average genome size and corresponding GC content. It is interesting to note, that smaller genome size is associated with specialized habitat, hyperthermophilic temperature and microaerophilic oxygen requirement. In all cases it is related to lower GC content. Thus, genome content has restriction for organism adaptation.
Acknowledgements: The research has been supported by the Ministry of Education and Science of the Russian Federation grant No. 14.W03.31.0015. Computing was done at Siberian Supercomputer center SB RAS.

References

1. Suslov V.V., Afonnikov D.A., Podkolodny N.L., Orlov Y.L. (2013) Genome features and GC content in prokaryotic genomes in connection with environmental evolution. Paleontological Journal. 47(9):1056-1060.
2. Babenko V.N., Bogomolov A.G., Babenko R.O., Galieva E.R., Orlov Y.L. (2018) CpG islands' clustering uncovers early development genes in the human genome. Computer Science and Information Systems. 15(2):473-485.

DOI 10.18699/BioGenEvo-2018-41

Sequencing technologies for 3D chromosomes contacts analysis and computer tools

O. Thierry[1], S.S. Kovalev[2], A.V. Tsukanov[2], Y.L. Orlov[2,3]*
[1] *University of Bordeaux, France*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[3] *A.O. Kovalevsky Institute of Marine Biology Research RAS, Sevastopol, Russia*
** e-mail: orlov@bionet.nsc.ru*

Key words: genomics, bacteria, microbiology, sequencing, environments, GC content, bioinformatics

Motivation and Aim: The aim of this work was to review existing computer tools for 3D genome structure data analysis and spatial topological domains, especially for the Hi-C method. Chromatin interactions play a critical role for gene expression regulation. Series of post-genome technologies have been developed to study the total binding of transcription factors in genome, such as chromatin immunoprecipitation arrays (ChIP-Seq) and Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) [1]. Several methods have been developed in order to study the spacial proximity between genomic regions, based on microscopy or high throughput sequencing methods (HTS) as Chromosome Conformation Capture (3C) and chromatin immunoprecipitation. Identification of genome-wide distal chromatin interactions that lead the regulatory elements to their target genes may provide novel insights into the study of transcription regulation [2]. Microscopy methods involve fluorescence in situ hybridation (FISH), a single-cell assay with optically labeled probes that hybridize to complementary regions of chromosomes, allowing direct measuring physical distance between them. We reviewed computer tools for 3D genome organization analysis.
Methods and Algorithms: As all the HTS methods, the Hi-C generates a huge amount of data which need to be analyzed, targeting the whole genome, leading to the development of various software and tools with different effectiveness and accuracy. We used to review different categories of software developed for 3D genomics experiments support – data processing, construction of contact maps, statistical estimates, visuzalization. Note that the visualization can be done through four main process: heatmap, 3-dimensional representation, circular representation and network.
Results and conclusion: We tested program for analysis of ChIA-PET experimental data. With the rapidly increasing resolution of Hi-C datasets, the size of the chromatin contact map will soon exceed the memory capacity of general computers. The same problem related to ChIA-PET and subsequent data integration has to be solved by software development.
Acknowledgements: The research has been supported by the Ministry of Education and Science of the Russian Federation grant #14.W03.31.0015. Computing was done at Siberian Supercomputer center SB RAS.

References

1. M.J. Fullwood et al. (2009) An oestrogen-receptor-alpha-bound human chromatin interactome, Nature, 462(7269):58-64.
2. G. Li et al. (2014) Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application, BMC Genomics, 15(Suppl 12):S11.

DOI 10.18699/BioGenEvo-2018-42

Endosymbiotic bacteria *Wolbachia* in Siberian populations of Acrididae grasshopper (Orthoptera)

A. Tikhomirova[1]*, M. Yudina[1, 2], G. Yurlova[2], R. Bykov[2], A. Bugrov[1, 3], Yu. Ilinsky[1, 2]
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[3] *Institute of Systematics and Ecology of Animals SB RAS, Novosibirsk, Russia*
** e-mail:a.tikhomirova9@g.nsu.ru*

Key words: Acrididae, grasshopper, population, endosymbiont, *Wolbachia*, Siberia

Motivation and Aim: *Wolbachia* are maternally inherited symbionts that infect a wide range of arthropods. *Wolbachia* may influence on host biology by mutualistic or parasitic ways. For some insect taxa there are no data on *Wolbachia* diversity and its effects on the host. In this work we address to one of those taxa – family Acrididae. This family contain about 9000 species [1] but very few of them were studied on *Wolbachia*. We aim to characterize prevalence and genetic diversity of *Wolbachia* in Siberian populations of Acrididae.
Methods and Algorithms: Total collection includes 339 samples of 17 species from 10 genera: *Arcyptera*, *Bohemanella*, *Bryodema*, *Chorthippus*, *Gomphocerippus*, *Oedaleus*, *Ognevia*, *Omocestus*, *Podisma*, *Stauroderus*. Insects were collected from Novosibirsk and Irkutsk provinces, Altai Republic and Altai Krai. The DNA extraction was individually performed from legs and/or abdomens for each specimen. *Wolbachia* infection status was determined by PCR with primers specific to *coxA* and *wsp Wolbachia* genes. Genetic diversity of *Wolbachia* isolates was determined by PCR with primers to five *Wolbachia* loci according to the multilocus sequence typing (MLST) protocol [2].
Results: We found *Wolbachia* infection in 16 of 17 studied Acrididae species. The infection rates of some hosts were high. For instance, *Chorthippus biguttulus* population of Novosibirsk province was infected with the rate of 91.0±2.2 %. New *Wolbachia* haplotypes were found in grasshopper hosts.
Conclusion: This is the first report on *Wolbachia* infection in 15 species of Acrididae. We demonstrate high *Wolbachia* infection rate in Siberian grasshopper's populations and revealed new haplotypes of bacteria.
Acknowledgements: Supported by the RFBR (18-316-00099, 16-04-00980, 18-04-00192).

References

1. Richards O.W., Davies R.G. (2013) Imms' General Textbook of Entomology: Vol. 2: Classification and Biology. Springer Science Business Media.
2. Baldo L. et al. (2006) Multilocus sequence typing system for the endosymbiont Wolbachia pipientis. Applied Environmental Microbiology. 72(11):7098-7110.

DOI 10.18699/BioGenEvo-2018-43

The reasons for mtDNA structural instability: evolutionary physico-chemical retrospective

V.N. Timonina[1]*, D.A. Knorre[2], K.Yu. Popadin[1, 3], K.V. Gunbin[1, 4]
[1] *School of Life Science, Immanuel Kant Federal Baltic University, Kaliningrad, Russia*
[2] *A.N. Belozersky Institute of Physico-Chemical Biology, MSU, Moscow, Russia*
[3] *Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland*
[4] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** e-mail: valeratimonina@gmail.com*

Key words: mtDNA, non-B-DNA conformations, dinucleotide properties, molecular evolution

Motivation and Aim: It is well known that unevenness of physico-chemical DNA properties along the human mtDNA is associated with the unevenness in mtDNA deletion breakpoints and SNPs and, ultimately with longevity.
Methods and Algorithms: More than 1000 completely sequenced mtDNAs of Mammalia and more than 24000 human mtDNAs were analysed. To uncover the basis of mtDNA structural instability (SI) we analyzed both intraspecial (human) and interspecial (Mammalians) mtDNA variations with (1) the non-B-DNA conformations (cruciform, triplex, hairpins, G-quadruplex, Z-DNA, etc.) and (2) the dinucleotide properties of mtDNA enriched with variations. Our analysis was based on (1) various available software tools for non-B-DNA identification (various EMBOSS package programs; R packages triplex and pqsfinder; SIST and triplexator software) and (2) DiProDB dinucleotide properties, respectively. Random forest approach implemented in R randomForest and Boruta libraries and various sliding window lengths were used to dissect mtDNA dinucleotide properties unevenness; phylogenetic comparative methods implemented in R ape and caper libraries were used for linking Mammalian species longevity and variations in mtDNA properties or between ancestry of human mtDNAs (based on reconstructed maximum likelihood phylogeny) and their properties.
Results and conclusions: During the analysis, we identified various important regularities in intraspecial and interspecial evolution of mtDNA properties.
For example, interspecial analysis allowed us to find positive relations between quadruplex frequencies, CpG island frequencies and species longevity; the relations of DNA bend, free energy and enthalpy with species longevity. All these relations associated with mtDNA replication optimization. Various intriguing relations between di-/terta-nucleotide under- or overrepresentations were found, for example, we found a tendency for CG and TT dinucleotides loss in long-lived mammals that can be related with SI minimization due to high mutational pressure on such nucleotide patterns.
The results of intraspecial analysis demonstrates nonrandom physical causes of SI. For example, it was found that DNA properties associated with 5' and 3' breakpoint hotspots are quite different: the properties of 5' breakpoints as well as SNP hotspots relates with the rigidity of DNA and with protein-DNA interactions while 3' breakpoints associates with the number of imperfect repeats. These facts allowed us to conclude that mtDNA deletion breakpoints and SNP fixations is initiated by protein-DNA interactions while the termination of breakpoints is defined by complementary interaction between the single stranded DNA by imperfect repeats.
Acknowledgements: This study was supported by the 5 Top 100 Project at the I. Kant Baltic Federal University.

DOI 10.18699/BioGenEvo-2018-44

Specific anthrax bacteriophages as a factor for selection of subcultures with different phenotypic and genetic characteristics out of populations of *Bacillus anthracis* strains

O.I. Tsygankova, E.A. Koteneva*, A.V. Kalinin
Stavropol Plague Control Research Institute, Stavropol, Russia
** e-mail: postgenom_stv@mail.ru*

Key words: *Bacillus anthracis*, phage resistance, variability of phenotypic properties, genotypes

Motivation and Aim: The aim of the work was to study population composition of two virulent strains of the causative agent of anthrax by the factor of phage resistance to specific bacteriophages, to select phage resistant subcultures, and to study their properties comprehensively.
Methods and Algorithms: We used virulent strains *B. anthracis* 1 (CO) and 81/1 which were isolated from pathological material. Concentrations of phage corpuscles in the preparations of bacteriophages Gamma A-26, BA-9, K-VIEV were, correspondingly, 8×10^9, 4×10^8 and 2×10^8 per 1 ml. Population composition of *B. anthracis* strains by their resistance to bacteriophages was studied by the method [1]. Subcultures of *B. anthracis* 1 (CO) strain were studied according to the basic identification tests and additional methods to study *B. anthracis* cultures. Six chromosomal loci were used to characterize strains by VNTR-loci [2].
Results: Not a single colony grew on plates of both strains treated with bacteriophage Gamma A-26. Plates treated with bacteriophage K-VIEV showed a 2.9 and a 4.8 % growth of colonies of the strains *B. anthracis* 1 (CO) and 81/1 respectively. Plates treated with phage "BA-9" showed a 10.9 % growth of colonies of the strain *B. anthracis* 1 (CO) and a 17.3 % growth of colonies of the strain *B. anthracis* 81/1, correspondingly. Distribution of the 22 variants of each strain into groups differing by their sensitivity to various bacteriophages was as follows. The retest showed that in both strains 16.7 % of variants separated on the basis of their resistance to bacteriophage BA-9 were sensitive to all the three bacteriophages. In variants of the strain *B. anthracis* 81/1 which were selected from the plates treated with bacteriophage BA-9 such variants made up 20 %, and in *B. anthracis* 1 (CO) – 10 %. Variants, resistant to the action of bacteriophage Gamma A-26, were found among cultures selected on the basis of their resistance to the other bacteriophages. Among 20 phage resistant variants of the strain *B. anthracis* 1 (CO) we found variants which were atypical in capsule formation, toxin production, nutritional requirements, protease, lecithinase, and hemolysin activities. Genetic studies revealed three variants of the plasmid structure, and four MLVA-genotypes.
Conclusion: A considerable variety of phenotypic and genetic properties among phage resistant subcultures of the strain *B. anthracis* 1 (CO) can testify to complex change of some of them which makes specific anthrax bacteriophages an effective factor for selection of atypical in many properties cultural variants out of populations of strains.

References

1. Tsygankova O.I. et al. A method of quantitative definition of resistance of population composition of *Bacillus anthracis* strains to anthrax bacteriophage. Russian Patent No. 2266963. 27.12.2007.
2. Keim P., Price L.B., Klevytska A.M. et al. (2000) Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. 182:2928-2936.

DOI 10.18699/BioGenEvo-2018-45

Fractal analysis of otolith microrelief as a method for determines relationship of species

M.L. Tyagun[1]*, A.A. Golovko[2]**
[1] *Limnological Institute SB RAS, Irkutsk, Russia*
[2] *Institute of Solar-Terrestrial Physics SB RAS, Irkutsk, Russia*
** mary@lin.irk.ru, ** golovko@iszf.irk.ru*

Key words: crystalline surface, otolith, fractal analysis, multifractal spectrum

The micromorphology of fish otolith (texture diagnostics) in recent years has attracted interest from specialists in various fields. The mechanisms of the organization of the microstructural tissue and its functional features are not clear. On the otolith, in an important functional part of the acoustic apparatus of fish, a hierarchy of crystallized elements is formed in the zone of the sulcus acousticus. A specialized neural tissue, the auditory macula, adjoins a macular spot containing similar elements. The macula consists of sensory cells interacting with the surface of the microstructure. Acoustic vibrations of the otolith cause mechanical action and polarization of the cells. Structural organization of the macula is usually presented in the form of a simple scheme consisting of regions that symbolize groups of cells of different lengths, with different polarization vectors. It is considered that a polarized macular pattern is identical to the otolith macular pattern including areas with structures of different types as well. Each area is visualized as hierarchic levels of multidimensional geometrical objects. Therefore, to concept its organization, a scheme similar to the sensory macula is not enough. Conception the principle of structural organization of tissue of this kind is possible with the help of fractal analysis, capable of describing the nature of the geometry of a complex object. Microcanonical analysis of the multifractal spectra confirmed the fractal origin of the hierarchic organization of the crystallized otolith surface. Multifractal pattern of this biomineral surface is consistent with the ideas on these structures formation in nature (insentient substance). Instantaneous temporal analysis of the complex self-organizing processes in nature, such as lithospheric orogeny or fluid flow through porous media showed their fractal geometry. It seems that regulation of the otolith microrelief formation by the organism is facilitated by sustaining non-isotropy of its tissue that is required for fixation of the arriving signal by the acoustic organs and space orientation. The similarity of multifractal spectral patterns of stone sculpins (*P. knerii*), described for different areas of crystallized surface using different scales, was found. Thus, we are convinced that the scale variations during analysis do not affect the quality of the data obtained. The concave curves of the multiscale images of the same area of stone sculpins are located within the same range of values and reach the same level. The multifractal spectra of the crystalline patterns of four species selected for analysis is differ. This supports a hypothesis on the unique species organization of the crystallized surface of the biomineral. Based upon the species spectral curves integrated into pairs by their relationship distance, we conclude that the closer is the relationship, the greater is the number of common features of the external microreliefs (stone and sand sculpins, *P. knerii*, *L. kesslerii*) and the larger is the taxonomical divergence (grayling and dace, *T. baicalensis*, *L. leuciscus*), the greater are the differences in the multifractal spectra. The data obtained allow us to state that the fractal analysis provides a possibility to describe the organization of crystalline elements of the biomineral and use its fractal parameters, multifractal spectrum in particular, for taxonomic analysis.

DOI 10.18699/BioGenEvo-2018-46

Genomic characterization of *DEP1* gene in the Triticinae species with compact, compactoid and normal spike shape

V. Vavilova*, I. Konopatskaia, A. Blinov
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** e-mail: valeriya-vavilova@bionet.nsc.ru*

Key words: *DEP1* gene, Triticinae, *Triticum*, *Aegilops*, spike shape, plant architecture

Motivation and Aim: Spike shape of Triticinae species is one of the most important taxonomically characteristic of this tribe. There are four main variants of spike shape (spelt, compact, compactoid and normal), that widely distributed among wheat and *Aegilops* species. Several genes and loci associated with spike shape trait have been identified previously [1]. Nevertheless, there are practically no data about nucleotide sequences of this genes and loci. *DENSE AND ERECT PANICLE 1* (*DEP1*) gene is related to several traits in rice (erect panicle, number of grains per panicle and panicle dense) [2]. In case of *T. aestivum* the experiments with the transgenic line showed that downregulation of *DEP1* homologue affects the length of the ear, ear density and number of spikelets [3].The aim of this study was investigation and genomic characterization of *DEP1* gene in wheat and *Aegilops* species with compact, compactoid and normal spike shape.
Methods and Algorithms: A combination of bioinformatical tools and standard molecular biology methods was used.
Results: We determined the spike shape of several accessions of *Triticum* and *Aegilops* species (*T. antiquorum*, *T. macha*, *T. sphaerococcum* and *Ae. tauschii*) by calculation of Flaksberger's formula. The full-length sequences of *DEP1* gene were obtained for all accessions studied. Nucleotide sequences comparing revealed *DEP1* gene regions which distinguish analyzed wheat and *Aegilops* species with compact, compactoid and normal spike phenotypes. Phylogenetic analysis allowed to determined the origin of *DEP1* alleles and different Triticinae species.
Conclusion: *DEP1* gene variability could contribute to the spike shape formation in wheat and *Aegilops* species.
Acknowledgements: The study was supported by the Russian Science Foundation (grant number: 16-16-10021).

References

1. Konopatskaia I. et al. (2016) Spike morphology genes in wheat species (*Triticum* L.). Proc. Latv. Acad. Sci. Sect. B. 70:345-355.
2. Huang X. et al. (2009) Natural variation at the *DEP1* locus enhances grain yield in rice. Nat. Genet. 41:494-497.
3. Xu H. et al. (2016) The DENSE AND ERECT PANICLE 1 (*DEP1*) gene offering the potential in the breeding of high-yielding rice. Breed Sci. 66:659-667.

DOI 10.18699/BioGenEvo-2018-47

The review of *Trichoplax adhaerens* genome: comparative analyses

E. Vodiasova*, E. Chelebieva
A.O. Kovalevsky Institute of Marine Biological Research of RAS, Sevastopol, Russia
** e-mail: eavodiasova@gmail.com*

Key words: *Trichoplax adhaerens,* mitochondrial genome, Placozoa

Motivation and Aim: *Trichoplax adhaerens* Schulze 1883 is the basal multicellular animal, which structure is very simple: the body consists from three layers, it doesn't have any tissue differentiation, neurons, muscles, synapsis [1]. At that time, this interesting organism is able to change the body form in response to environmental conditions, has different cycles of moving, incomprehensible behavior, the life cycle stages aren't studied completely (for example planktonic forms or sexual reproduction). The Trichoplax has 19 haplotypes with the high genetic divergence. The mitochondrial genome was sequenced in 2006 [2]. The mitochondrial genome is bigger than the other types of animals and has 43079 nucleotides. And two years later the whole genome was sequenced, the size – 98 Mb, 11514 genes was found. The presence of different transcription factors (LIM-homeobox and POU– homeobox), genes which involved to syntheses of dopamine, adrenaline and noradrenaline; genes coding different intracellular matrix proteins [3]. But all these genes are characteristic for more complex Animals which have neural systems or intracellular matrix. All of these make the *Trichoplax adhaerens* as an interesting object for investigation of different evolution process and comparative genome analyses among others animal phyla.
Methods and Algorithms: We analyzed the whole genome (NCBI: GCA_000150275.1) and mitochondrial genome (NCBI: DQ112541.1) of *Trichoplax adhaerens*, made statistic of presence protein families and comparison with the representatives of different animal phyla.
Conclusion: Despite of simplicity of body structure, the genome of the Trichoplax is very complicated and comparable with other animals. The whole complex of genes, which regulated biological processes absent in the Trichoplax, apparently aren't expressed in the cells. This requires genomic investigation of transcriptome on the cell level in the nearest future.
Acknowledgements: Supported by the Ministry of Education and Science of the Russian Federation grant No. 14.W03.31.001.

References

1. Smith C.L., Varoqueaux F., Kittelmann M., Azzam R.N., Cooper B., Winters C.A., ... Reese T.S. (2014) Novel cell types, neurosecretory cells, and body plan of the early-diverging metazoan *Trichoplax adhaerens*. Current Biology. 24(14):1565-1572.
2. Dellaporta S.L., Xu A., Sagasser S., Jakob W., Moreno M.A., Buss L.W., … Schierwater B. (2006). Mitochondrial genome of Trichoplax adhaerens supports Placozoa as the basal lower metazoan phylum. Proceedings National Academy Sciences. 103(23):8751-8756.
3. Srivastava M., Begovic E., Chapman J., Putnam N.H., Hellsten U., Kawashima T., ... Signorovitch A.Y. (2008) The Trichoplax genome and the nature of placozoans. Nature. 454(7207):955.

DOI 10.18699/BioGenEvo-2018-48

Mirtrons as a possible inherent source of silencing variability

P.S. Vorozheykin[1], I.I. Titov[1, 2]*
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
* *e-mail: titov@bionet.nsc.ru*

Key words: miRNA, introns, secondary structure, mirtrons

Background: MicroRNAs proceeds through the different canonical and non-canonical pathways; the most frequent of the non-canonical ones is the splicing-dependent biogenesis of mirtrons. We compare the mirtrons and non-mirtrons of human and mouse to explore how their maturation appears in the precursor structure around the miRNA [1].
Results: A typical structure of the annotated mirtron pre-miRNAs differs from the canonical pre-miRNA structure and possesses the 1- and 2 nt hanging ends at the hairpin base. The mirtron Dicer cleavage site shows the excessive variability (partially due to guanine at its ends inherited from splicing) than the canonical one. In contrast with the canonical miRNAs the mirtrons have higher SNP densities and their pre-miRNAs are inversely associated with diseases. Therefore we supported the view that mirtrons are under positive selection while canonical miRNAs are under negative one and we suggested that mirtrons are an intrinsic source of silencing variability which produces the disease-promoting variants. Finally, we considered the interference of the pre-miRNA structure and the U2snRNA:pre-mRNA basepairing. We analyzed the location of the branchpoints and found that mirtron structure tends to expose the branchpoint site what suggests that the mirtrons can readily evolve from occasional hairpins in the immediate neighbourhood of the 3′ splice site.
Conclusion: The miRNA biogenesis manifests itself in the footprints of the secondary structure. Close inspection of these structural properties can help to uncover new pathways of miRNA biogenesis and to refine the known miRNA data, in particular, new non-canonical miRNAs may be predicted or the known miRNAs can be re-classified.

References

1. Titov I.I., Vorozheykin P.S. (2018) Comparing miRNA structure of mirtrons and non-mirtrons. BMC Genomics. 19(Suppl.3):114.

DOI 10.18699/BioGenEvo-2018-49

Genetic diversity and phylogeny of *Wolbachia* in lepidopteran hosts

M. Yudina[1, 2]*, V. Dubatolov[3], R. Bykov[1], I. Mazunin[4], Yu. Ilinsky[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Institute of Systematics and Ecology of Animals SB RAS, Novosibirsk, Russia*
[4] *Immanuel Kant Baltic Federal University, Kaliningrad, Russia*
** e-mail: Judina@bionet.nsc.ru*

Key words: *Wolbachia*, Lepidoptera, multilocus sequence typing, forest pests

Motivation and Aim: *Wolbachia* are maternally inherited intracellular symbionts of arthropods and some nematodes. The genetic diversity of *Wolbachia* is subdivided into 16 supergroups (phyletic lines). Significant number of insect families appears to be unexplored for *Wolbachia* symbionts. Seventeen families of Lepidoptera are known to be infected with *Wolbachia*, while this order includes up to 200 families. Here we try to find *Wolbachia* infection in hosts of poorly investigated lepidopteran families and further analyse genetic data of *Wolbachia* isolates.
Methods and Algorithms: The insect collection consisted of 519 samples and included 255 species of 23 lepidopteran families. Some of these species are dangerous forest pests. The DNA extraction was individually performed from each specimen. Infection status was determined by PCR with *Wolbachia* specific primers targeted to the *wsp* and *16SrRNA* loci. Genetic diversity of *Wolbachia* was determined by the multilocus sequence typing (MLST) protocol [1], which included five loci. Phylogenetic analysis was conducted using MEGA 6 [2].
Results: *Wolbachia* symbionts were found in 36 species of 12 lepidopteran families. Some of these species were forest pests. We found both new haplotypes and previously described *Wolbachia* variants. Most of symbiont haplotypes belonged to supergroup B, and remaining haplotypes belonged to supergroup A.
Conclusion: We present comprehensive results of *Wolbachia* incidence and *Wolbachia* genetic diversity in Lepidoptera. The results of this study are necessary to develop *Wolbachia*-based biotechnological methods of pest control.
Acknowledgements: Supported by the RFBR (16-04-00980, 18-316-00099).

References

1. Baldo L. et al. (2006) Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Applied Environmental Microbiology. 72(11):7098-7110.
2. Tamura K. et al. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology Evolution. 30:2725-2729.

DOI 10.18699/BioGenEvo-2018-50

Diversity of genomic variants and population genetics of ethnic and regional groups across Russia

D.V. Zhernakova[1, 2], V. Brukhin[1], S. Malov[1, 3], T.K. Oleksyk[1, 4] K.P. Koepfli[1, 5], A. Zhuk[1], P. Dobrynin[1], S. Kliver[1], N. Cherkasov[1], G. Tamazian[1], M. Rotkevich[1], K. Krasheninnikova[1], I. Evsyukov[1], S. Sidorov[1], A. Gorbunova[1, 6], E. Chernyaeva[1], A. Shevchenko[1], S. Kolchanova[1, 4], A. Komissarov[1], S. Simonov[1], A. Antonik[1], A. Logachev[1], D.E. Polev[7], A.S. Glotov[7], V. Ulantsev[8], E. Noskova[8, 9], T.K. Davydova[10], T.M. Sivtseva[10], S. Limborska[11], O. Balanovsky[12, 13, 14], V. Osakovsky[10], A. Novozhilov [15], V. Puzyrev[16], N. Kropachev[17], S.J. O'Brien[1]*

[1] *Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russia*
[2] *University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands*
[3] *Department of Mathematics, St. Petersburg Electrotechnical University, St. Petersburg, Russia*
[4] *Biology Department, University of Puerto Rico at Mayaguez, Mayaguez, Puerto Rico*
[5] *Smithsonian Conservation Biology Institute, National Zoological Park, Washington, USA*
[6] *I.I. Mechnikov North-Western State Medical University, St. Petersburg, Russia*
[7] *Research Resource Center for Molecular and Cell Technologies, Research Park, St. Petersburg State University, St. Petersburg, Russia*
[8] *Computer Technology Department, St. Petersburg Electrotechnical University, St. Petersburg, Russia*
[9] *Jet Brains Research, St. Petersburg, Russia*
[10] *Institute of Health, North-Eastern Federal University, Yakutsk, Russia*
[11] *Department of Molecular Bases of Human Genetics, Institute of Molecular Genetics, RAS, Moscow, Russia*
[12] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*
[13] *Research Centre for Medical Genetics, Moscow, Russia*
[14] *Biobank of North Eurasia, Moscow, Russia*
[15] *Department of Ethnography and Anthropology, St. Petersburg State University, St. Petersburg, Russia*
[16] *Institute of Medical Genetics, Tomsk Research Center RAS, Tomsk, Russia*
[17] *Office of the Rector, St. Petersburg State University, St. Petersburg, Russia*
** e-mail: lgdchief@gmail.com*

Key words: genome-wide variation, Russian Federation, disease-causing mutations, gene flow, geological barriers

The Russian Federation spans 11 time zones and is the home of ~146,000,000 people: 80 % are the ethnic Russians and the remainder identify themselves as one of ~200 indigenous ethnic minorities. Despite the large population size and high ethnic diversity, no centralized reference database of functional and endemic genetic variation has been established to date. Such data are crucial for medical genetic purposes and would be essential for studying population history. The Genome Russia Project aims at filling this gap by performing high coverage whole genome sequencing and analysis of peoples of the Russian Federation. Here we report general methodology and inferences of genome-wide variation (SNPs, indels, and copy number variation) from 264 healthy adults, including 60 newly sequenced samples consisting of family trios from three geographic regions: Pskov, Novgorod and Yakutia, now contributed to the 1000 Genomes database. People of Russia are shown to carry known and novel genetic variants of adaptive, clinical and functional consequence. We identified 31 SNPs associated with disease-causing mutations from Human Gene Mutation Database; 758 loss-of-function SNPs and more than 20 SNPs associates with diseases, drug response and other phenotypes showing appreciable occurrence or allele frequency divergence from the neighboring Eurasian populations. Principal component and phylogenetic analyses of overall variation revealed strong geographic partitions among indigenous ethnicities corresponding to the geographic locales where they have lived. Allele frequency spectra identified strong constraints to gene flow that correspond to the geological barriers (e. g. the Ural Mountains and Verkhoyansk mountain range). The first haplotype-based genetic maps have been generated for western ethnic Russian and Yakut populations that can be used for identification of functional gene variants associated with diseases, and as a tool for the population genetic analyses. This study presents a genomic characterization of population-specific variation in Russia with results important for medical genetics as well as for population natural history studies.

DOI 10.18699/BioGenEvo-2018-51

Allelic diversity of the *GJB2* gene in deaf patients and ethnically matched populations from South Siberia

M. Zytsar[1, 2]*, M. Bady-Khoo[3], E. Maslova[1, 2], V. Danilchenko[1, 2], N. Barashkov[4, 5], I. Morozov[2, 6], A. Bondar[6], O. Posukh[1, 2]

[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Scientific Research Institute of Medical-Social Problems and Management of the Republic of Tuva, Kyzyl, Russia*
[4] *Laboratory of Molecular Biology, MK Ammosov North-Eastern Federal University, Yakutsk, Russia*
[5] *Laboratory of Molecular Genetics, Yakut Scientific Centre of Complex Medical Problems, Yakutsk, Russia*
[6] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
** e-mail: zytzar@bionet.nsc.ru*

Key words: hereditary deafness, *GJB2,* allelic diversity, populations of Siberia

Motivation and Aim: Pathogenic variants in gene *GJB2* (MIM 121011, 13q11-q12) account for a significant portion of hereditary hearing loss (HL). Spectrum of *GJB2* variations (pathogenic, benign and yet unclassified) and their prevalence are highly population-specific. We aimed to analyze allelic diversity of *GJB2*, obviously being under selection, in deaf patients from indigenous Siberian peoples (Tuvinians and Altaians), from Russian patients living in Siberia and in ethnically matched controls and to compare these data with appropriate data for worldwide populations.
Methods: Sanger sequencing was applied for analysis of non-coding (exon 1), coding (exon 2) and flanking intronic regions of the gene *GJB2*. Pedigree analysis and molecular cloning were applied for verification of *cis*-configuration of variants p.V27I and p.E114G. Genetic differentiation of populations was analyzed by standard population structure analysis softwares.
Results: Different spectrum of the *GJB2* pathogenic variants (c.-23+1G>A, c.35delG, p.V37I, p.R75Q, c.235delC, c.299_300delAT, c.313_326del14, p.W172C) and their variable contribution to HL of patients were previously found (15.1 % and 17.5 % in Altaians and Tuvinians, respectively, vs 55.9 % in Russians). We also evaluated the frequencies of known and novel benign *GJB2* variants: rs117685390, rs9552101, NC_000013.11:g.20193014C>T, c.-23+27G>A (rs899667206), p.V27I (rs2274084), p.E114G (rs2274083), p.V153I (rs111033186), p.F191L (rs397516878), p.I203T (rs76838169), rs3751385, rs5030700 in all examined samples. Ambiguous association of combination of variants p.V27I and p.E114G with HL is widely discussed in literature. Our important results: the proved *cis*-configuration of p.V27I and p.E114G and the absence of association of allele p.[V27I;E114G] with HL based on higher frequency of p.[V27I;E114G] in Tuvinian and Altaian controls compared with the patient's samples. All obtained data were used for comparative analysis of genetic differentiation of studied samples and populations from global human genomic data (1000 Genomes, HapMap Projects etc).
Conclusion: Despite the limited number of SNPs found in analyzed the *GJB2* gene region, strong ethno-specific genetic differentiation was revealed between the samples of deaf patients and controls from indigenous Siberian populations vs corresponding samples of Russians.
Acknowledgements: Supported by the project 0324-2018-0016 and by the RFBR (18-34-00166_mol-a, 17-29-06016_ofi-m).

Author index

Компания Huawei является ведущим мировым поставщиком ИКТ-решений. Благодаря установлению взаимовыгодных отношений с нашими партнерами и заказчиками компании Huawei удалось добиться существенных преимуществ в сфере операторских сетей, корпоративного и потребительского бизнеса, а также в сфере облачных технологий. Мы стремимся создавать максимальные преимущества для операторов связи, предприятий и потребителей путем разработки конкурентных ИКТ-решений и услуг. Оборудование и решения Huawei используются в более чем 170 странах мира. Компания обслуживает более трети населения земного шара.

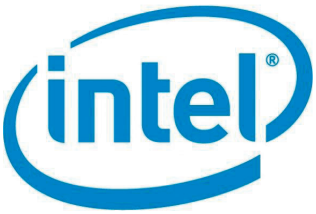
Имея богатый опыт и технические знания в области НИОКР, Huawei придерживается стратегии тесного сотрудничества и интеграции с корпоративными заказчиками и предоставляет им широкий спектр высокоэффективных клиентоориентированных ИКТ-решений и услуг, на базе глубокого понимания их потребностей. Согласно этой стратегии Huawei предлагает широкий выбор передовых ИКТ-решений в сфере государственного управления, общественного сектора, финансов, транспорта, электроэнергетики, крупных предприятий, а также малых и средних предприятий (SME). Эти решения охватывают корпоративные сети, универсальные системы связи и взаимодействия (UC&C), системы облачных вычислений и центры данных, системы корпоративной беспроводной связи, сетевого электропитания, а также инфраструктурные услуги.

ООО «Техкомпания Хуавэй»
Филиал в СФО:
630112, Новосибирск, ул. Фрунзе, 242,
11-й этаж
Тел.: +7(383) 328 00 70
Факс: +7(383) 328 00 71
E-mail: Kroshin.Fyodor@huawei.com
URL: e.huawei.com/ru

Huawei is a leading global ICT solutions provider. Through our dedication to customer-centric innovation and strong partnerships, we have established end-to-end capabilities and strengths across the carrier networks, enterprise, consumer, and cloud computing fields. We are committed to creating maximum value for telecom carriers, enterprises and consumers by providing competitive ICT solutions and services. Our products and solutions have been deployed in over 170 countries, serving more than one third of the world's population.

By leveraging our strong R&D capabilities and comprehensive technical expertise, Huawei's strategy in the enterprise domain focuses on close cooperation and integration with partners to deliver a wide range of highly efficient customer-centric ICT solutions and services that are based on a deep understanding of customer needs. In line with our strategy, we offer a broad portfolio of innovative ICT solutions that cater to global vertical industry and enterprise customers across government and public sector, finance, transportation, energy, large enterprises and small and midsize enterprises (SMEs). Our portfolio covers enterprise networking, unified communications & collaboration (UC&C), cloud computing & data center, enterprise wireless, network energy and infrastructure services.

HUAWEI Technologies Co Ltd., Russia
Siberia office:
630112, Russia, Novosibirsk, Frunze Str., 242
Business Center "New Height"
Tel.: +7(383) 328 00 70,
Fax: +7(383) 328 00 71
Email: Kroshin.fyodor@huawei.com
URL: e.huawei.com/ru



Корпорация Intel

Корпорация Intel была основана в 1968 году Робертом Нойсом и Гордоном Муром. На протяжении 50 лет Intel создает инновационные технологии, открывающие новые возможности для людей.

Корпорация Intel является мировым лидером в области микроэлектроники и информационных технологий. Intel создает технологии для умного мира эпохи больших данных. Основное внимание корпорация уделяет созданию интеллектуальных решений для умного мира, от устройств Интернета вещей и пользовательских ПК до коммуникационной инфраструктуры, технологий для центров обработки данных и суперкомпьютеров.

Штаб-квартира корпорации расположена в г. Санта-Клара, шт. Калифорния. Общий штат Intel насчитывает более 100 тыс. сотрудников в более, чем 60 странах по всему миру. Главным исполнительным директором корпорации является Роберт Свон (Robert Swan).

Intel в России

Первое представительство Intel в России было открыто в 1991 году в Москве. Сегодня в российских офисах Intel в Москве и Нижнем Новгороде работают более 800 человек.

В московском офисе компании представлены отделы маркетинга и развития бизнеса, группы по разработке программного обеспечения, юридический отдел.

В НИОКР центре Intel в Нижнем Новгороде создаются новые и инновационные продукты для разработки ПО. Сегодня он является одним из крупнейших центров исследований и разработок Intel в Европе. Более **700 специалистов и инженеров** разрабатывают программные инструменты и приложения для архитектур Intel. В Нижнем Новгороде также размещаются различные группы поддержки бизнеса (например, административно-хозяйственная часть, финансовый отдел, отдел ИТ, отдел кадров).

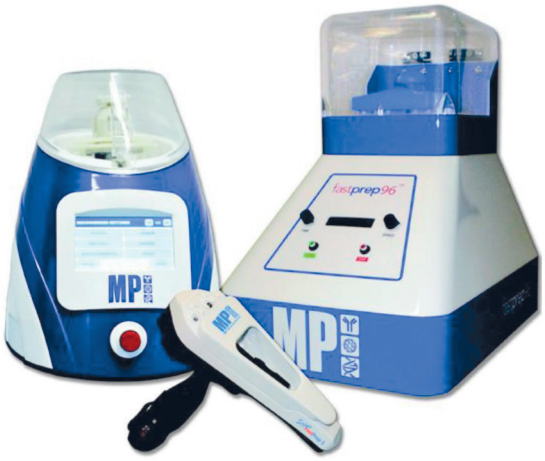
Центр исследований и разработок Intel в Нижнем Новгороде

Нижегородский офис Intel был является центром экспертизы корпорации в области высокопроизводительных вычислений, разработки программного обеспечения в области численных методов и беспроводной связи.

✉ MP Biomedicals
(ООО «МПБА диагностика»)
Адрес: 109147, г. Москва, ул. Марксистская, д. 3,
стр. 2, оф. 2.1.20/2
☎ Тел./факс:
+7(495)604-13-44
E-mail rus@mpbio.com

WEB mpbio.com; mpbio.ru

Компания ООО «МПБА диагностика» является дочерней компанией MP Biomedicals, ранее известной как ICN Biomedicals, основанной в 1959 году, признанного лидера в области производства широкого спектра химических реактивов, оборудования для пробоподготовки (система для гомогенизации FastPrep) и наборов реагентов. Каталог продукции компании MP Biomedicals включает более 55000 наименований высококачественных продуктов для проведения биохимических исследований, фармацевтического и биотехнологического производства, для различных отраслей иммунологии и генетики.



ООО «Рош Диагностика Рус» – официальный импортер продукции Roche в России и лицензиат компании F.Hoffmann–La Roche Ltd.

Roche Sequencing Solutions, подразделение Roche, ориентированное на решения для NGS, а в частности на пробоподготовку к NGS, предлагает:

-**Наборы KAPA Biosystems для приготовления библиотек ДНК** (включают баркодированные адаптеры, частицы для очистки, наборы для оценки концентраций ДНК и библиотек методом ПЦР в реальном времени).

-**Наборы для направленного отбора генов перед NGS:**

NimbleGen SeqCap EZ –гибридизационное обогащение панелей генов, экзомов, транскриптомов и метиломов;

HEAT-seq - амплификационное обогащение панелей генов, в том числе и панелей онкогенов;

AVENIO –гибридизационное обогащение панелей онкогенов из внеклеточной опухолевой ДНК и анализ данных.

ООО «Рош Диагностика Рус» предлагает комплексные решения, включающие в себя не только оборудование и реагенты, но и технический сервис, обучение персонала и постоянную методическую поддержку.

Unlock the Potential of Every Sample

ООО «Рош Диагностика Рус
115114, Россия, Москва
Ул. Летниковская, д. 2, стр. 2
Тел:+7 (495) 229-69-99
Бизнес-центр «Вивальди Плаза»
ras.russia@roche.com
sequencing.roche.com

BioLine
группа компаний

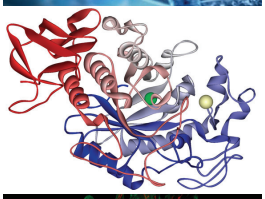
PerkinElmer For the Better
Leica MICROSYSTEMS
BioTek
BD

ООО «БиоЛайн» является официальным дистрибьютором компаний **BD Biosciences, Leica Microsystems, PerkinElmer, BioTek,** ведущих мировых производителей приборов и реагентов для биомедицинских исследований.

Благодаря уникальному портфолио продукции и опыту наших специалистов мы выполняем поставки и внедрение комплексных решений для разнообразных задач в области молекулярной и клеточной биологии.

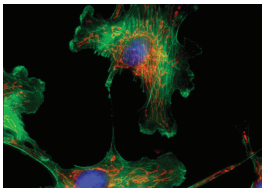
Молекулярно-генетические исследования

- Системы для выделения и молекулярного анализа одиночных клеток Becton Dickinson
- Станции для выделения ДНК, оборудование PerkinElmer для подготовки и контроля библиотек для NGS
- Наборы Nextflex для подготовки библиотек NGS PerkinElmer: полногеномное и таргетное секвенирование, транскриптомика, эпигенетика, метагеномика



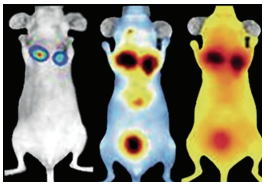
Протеомные исследования

- Передовые оптические технологии компании BioTek Instruments для биохимических исследований, идентификации и количественной оценки аналитов, исследования взаимодействия биомолекул
- Реагенты и расходные материалы PerkinElmer для протеомных исследований



Клеточные исследования

- Системы для проточной цитометрии и сортировки клеток компании BD Biosciences
- Оптическая визуализация клеток для моделирования процессов в клеточных культурах и на 3D сфероидах: решения PerkinElmer и BioTek Instruments
- Системы для конфокальной микроскопии Leica Microsystems



Исследования на животных

- Приборы для оптической визуализации ***in vivo*** Spectrum и Lumina, системы для КТ и ПЭТ компании PerkinElmer
- Оборудование для исследований на животных Leica Biosystems

Официальные дистрибьюторы BD Biosciences, Leica Microsystems, PerkinElmer, BioTek в России — компания «БиоЛайн»

ООО «БиоЛайн»
Россия, 197101, Санкт-Петербург
Пинский пер., д. 3, Лит. А
тел.: +7 (812) 320 49 49
факс: +7 (812) 320 49 40
e-mail: main@bioline.ru
www.bioline.ru

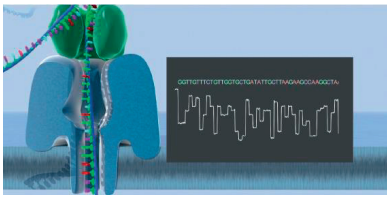
Москва, тел.: +7 (800) 555 49 40
Новосибирск, тел.: +7 (383) 227 09 63
Екатеринбург, тел.: +7 (343) 287 32 49
Владивосток, тел.: +7 (423) 201 18 08
Нижний Новгород, тел.: +7 (831) 278 61 47
Ростов-на-Дону, тел.: +7 (863) 268 99 32
Казань, тел.: +7 (843) 570 66 88
Самара: +7 (846) 246-06-54

Единый бесплатный номер сервисной службы для всех регионов России:

8 800 333 00 49

Компания Диаэм – крупнейший поставщик современного лабораторного оборудования на Российском рынке. Каталог компании насчитывает более 500 000 наименований приборов, реагентов и расходных материалов для медицинских и научно-исследовательских лабораторий.
В каталоге компании представлена продукция ведущих мировых производителей, как: Abcam, Applied Biosystems, Binder, Bio-Rad, Corning, Eppendorf, Illumina, Ion Torrent, Lexogen, Oxford Nanopore Technologies, Panasonic (Sanyo), Sage Sciences, Sigma-Aldrich, Thermo Fisher Scientific, Qiagen:

- Наборы для подготовки библиотек, для высокопроизводительного секвенирования NGS, для исследовательских работ и, в онкологии, репродуктивной медицине, в изучении наследственных заболеваний, реагенты и наборы для капиллярного секвенирования.
- Секвенаторы капиллярные и высокопроизводительные NGS, оборудование для анализа качества НК для NGS, роботизированные станции для подготовки библиотек и секвенирования.
- Все для ПЦР, реагенты,наборы, пластик, амплификаторы.
- Нанопоровые секвенаторы Oxford Nanopore Technologies, наборы для секвенирования ДНК и РНК.


Секвенирование теперь доступно каждому!

Диаэм сегодня представляет продукцию Oxford Nanopore Technologies – это секвенаторы третьего поколения – MinION, GridION, PromethION.
Технология секвенирования Oxford Nanopore Technologies позволяет делать прямое прочтение цепей ДНК или РНК в режиме онлайн, длина рида ограничена только длиной фрагмента, а портативность оборудования и быстрая подготовка библиотек дает возможность секвенировать даже в полевых условиях с минимальными требованиями к генетической лаборатории. С Oxford Nanopore Technologies секвенировать теперь может каждый, даже тот, кто ранее и не задумывался о секвенировании - это просто и доступно.
Секвенирование третьего поколения не заменяет и не отменяет применение капиллярных секвенаторов по Сэнгеру или платформ NGS второго поколения, наоборот, сочетание трех поколений генетического анализа открывает новые возможности получения ранее неизвестных данных.
Специалисты Диаэм прошли обучение в Oxford Nanopore Technologies, осуществляют профессиональное консультирование и техническую поддержку, помогут спланировать эксперимент и подобрать необходимые наборы реагентов для решения конкретной задачи независимо от бюджета лаборатории.
Обращайтесь в любой из наших офисов или пишите на sales@dia-m.ru

ООО «Диаэм» www.dia-m.ru

Москва	Новосибирск	Казань	Санкт-Петербург	Ростов-на-Дону	Пермь	Воронеж
ул. Магаданская, 7/3	пр. Акад. Лаврентьева, 6/1	ул. Парижской Коммуны, д. 6	ул. Профессора Попова, 23	пер. Семашко, 114	Представитель в УФО	тел./факс: (473) 232-4412
тел./факс: (495) 747-0508	тел./факс: (383) 328-0048	тел/факс: (843) 210-2080	тел./факс: (812) 372-6040	тел/факс: (863) 250-0006	тел./факс: (342) 202-2239	voronezh@dia-m.ru
sales@dia-m.ru	nsk@dia-m.ru	kazan@dia-m.ru	spb@dia-m.ru	rnd@dia-m.ru	perm@dia-m.ru	

Компания АЛЬБИОГЕН — официальный дистрибьютор illumina и Lucigen

Компания ООО «АЛЬБИОГЕН» с 2015 года является эксклюзивным (единственным) официальным торговым представителем и дистрибьютором компании illumina на территории Российской Федерации, Республики Беларусь, Республики Казахстан и Республики Узбекистан.

Нашей задачей является обеспечение полного доступа клиентов к передовым технологиям и сервисам illumina, включая современные системы NGS и анализа ДНК-биочипов, программное обеспечение для биоинформатики и весь спектр реактивов.

ООО «АЛЬБИОГЕН» предоставляет полный комплекс услуг, связанных с продажей, технической поддержкой и сервисным (гарантийным и постгарантийным) обслуживанием продукции компании Illumina, а также обучением пользователей работе на данном оборудовании.

Инновационная и стремительно развивающаяся компания illumina Inc., являющаяся мировым лидером в области геномных технологий, заключила соглашение с компанией АЛЬБИОГЕН, специализирующейся на поставках оборудования и расходных материалов для секвенирования нового поколения (NGS) и анализа на ДНК-биочипах.

Новейшие продукты компании illumina, создаваемые совместно с ведущими мировыми учеными, позволяют изучать геном на очень глубоком уровне и дают возможность для новаторских достижений в науке, медицине, сельском хозяйстве и потребительской геномике. Более 90% научных статей, связанных с технологиями секвенирования нового поколения, сделаны при помощи оборудования Illumina.

Сотрудничество с компанией АЛЬБИОГЕН направлено на то, чтобы сделать технологии NGS и анализа ДНК-биочипов более доступными на территории Российской Федерации и в странах СНГ.

Компания АЛЬБИОГЕН использует свой обширный опыт в области продаж и продвижения продукции, знания передовых технологий и сеть региональных представителей для обеспечения быстрой, эффективной и бесперебойной работы лабораторий клиентов illumina.

Компания АЛЬБИОГЕН также является официальным дистрибьютором компании Lucigen, основными продуктами которой являются ферменты и реагенты для секвенирования нового пколения и молекулярной диагностики.

Компания СкайДжин предлагает к поставке со склада в Москве и под заказ наборы реагентов, оборудование, расходные материалы, реактивы, а также специализируется на сервисном обслуживании и поверке дозаторов, лабораторных весов различных производителей. Мы предлагаем гибкие условия работы и очень большой ассортимент продукции.

Поставляемая нашей компанией продукция широко используется в научно-исследовательских лабораториях и R&D центрах, лабораториях секвенирования, при решении практически любых молекулярно-биологических задач.

Большая часть производителей в нашем портфолио - это прямые, эксклюзивные поставки. Мы являемся первым звеном в поставках для таких компаний как New England Biolabs, Agilent Technologies, Oxford Nanopore Technologies, QIAGEN, 10x Genomics, NIMAGEN, Integrated DNA Technologies, Thermo Fisher Scientific, SIGMA-ALDRICH, BioSan, Gilson.

К флагманским продуктам наших линеек относятся:

- Набор для пробоподготовки образцов от New England Biolabs ULTRA II FS с интегрированной системой фрагментации и другие наборы серии ULTRA для образцов ДНК, РНК и микроРНК;
- Digital NGS: готовые панели и наборы для обогащения на основе ПЦР от QIAGEN с мономолекулярным баркодированием;
- Специализированные наборы для работы с микроРНК и анализа экспрессии от QIAGEN-Exiqon;
- Нанопоровые секвенаторы третьего поколения: портативный секвенатор MinION, высокопроизводительный секвенатор GridION;
- Уникальная система Chromium производства 10x Genomics для автоматической пробоподготовки геномов и транскриптомов единичных клеток.

За дополнительной информацией о производителях, товарах, ценах и условиях поставки обращайтесь к нашим квалифицированным специалистам.

Будем рады ответить на Ваши вопросы и помочь выбрать качественное и недорогое решение для Ваших задач!

ООО «СкайДжин»
Адрес: 115093, Москва,
ул. Люсиновская, д. 36, стр. 1
Тел: 8 (495) 215 02 22
info@skygen.com
www.skygen.com

ХИМЭКСПЕРТ

Агентство Химэксперт

Информация о компании:

Компания Химэксперт существует 16 лет и давно зарекомендовала себя, как надежный поставщик приборов, реактивов и расходных материалов для молекулярной биологии. Мы собрали для своих клиентов самые интересные и перспективные бренды, большинство из которых в России можно приобрести только у нас.

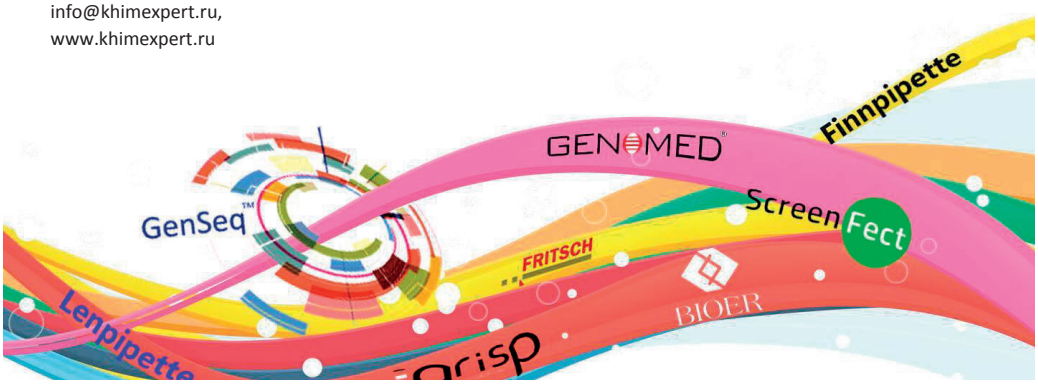
Химэксперт предлагает оборудование для анализа ДНК и РНК, в том числе и методами NGS, фундаментальных протеомных и цитологических исследований, фармацевтики и биотехнологий, прикладного тестирования, включая идентификацию личности и установление родства в криминалистике и судебно-медицинской экспертизе.

Наши клиенты выбирают Химэксперт потому что:

- Химэксперт всегда находит самые прогрессивные решения в области Life Sciences. Наша компания постоянно расширяет свое портфолио и в курсе последних веяний в области молекулярной биологии
- Химэксперт осуществляет полную техническую и методическую поддержку наших клиентов: обратившись к нам, вы получаете помощь квалифицированных сотрудников в подборе оборудования и реагентов под поставленные задачи и их последующем использовании
- Химэксперт стремится идти навстречу заказчикам и осуществлять быстрые поставки, так как скорость и четкость исполнения заказов очень важна.

Обратившись к нам, вы можете быть уверены в будущем своего эксперимента.
Начните сотрудничество с компанией Химэксперт и убедитесь в этом на своем опыте!

ООО «Агентство Химэксперт» 125009, г. Москва, Страстной б-р, д. 4, оф. 101
Тел: +7 (495) 629 28 69, 650 36 66
info@khimexpert.ru,
www.khimexpert.ru



www.genexplain.com

The geneXplain GmbH is glad to welcome you at the BGRS/SB'2018 conference and is proud to introduce you the following software and database solutions for the needs of bioinformatics, systems biology and systems medicine:

geneXplain platform – is a high-performance tool for multi-omics data analysis, which allows identification of new therapeutic targets and biomarkers. A unique feature of the geneXplain platform is its Upstream Analysis. You can register and immediately receive access to a free account.

TRANSFAC database – is a unique collection of transcription factors, their experimentally validated binding sites (TFBS) and a widely known library of positional weight matrices (PWMs). The database has its own integrated methods for TFBS search. It can also be used as an integral part of the geneXplain platform. TRANSFAC is available online or can be downloaded as a set of flat files.

TRANSPATH database – is one of the biggest and most famous collections of signaling and metabolic pathways, which counts over 489000 reactions. The database can be applied for master-regulators search within the geneXplain platform. TRANSPATH is also available online in one package with HumanPSD database or can be downloaded as a set of flat files.

HumanPSD database – is a collection of genes, proteins and micro-RNAs, which includes information about disease biomarkers and clinical trials for various diseases. Besides the detailed biomarkers data, the database contains information about drugs.

BRENDA database – is a comprehensive enzyme and enzyme-ligand information system. Its manually derived core contains over 3 million data points about 77,000 enzymes annotated from 135,000 literature references.

PASS – is a software tool for evaluating the general biological potential of organic compounds based on their structural formula. This program predicts main and side pharmacological effects, molecular mechanisms of action, specific toxicities, and antitargets, actions associated with the metabolism and transport of pharmaceutical substances and their influence on gene expression.

PharmaExpert – is a software tool for analysis of the biological activity spectra of substances predicted by PASS and selecting compounds with the desirable set of biological activity, for analyzing the relationships between biological activities, drug-drug interactions and for multiple targeting of chemical compounds.

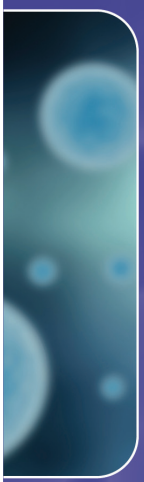
GUSAR – is a software tool for analysis of quantitative structure-activity/structure-property relationships (QSAR/QSPR) based on the structural formulas of the compounds and data on their activity/property, and for prediction of activity/property for new compounds. GUSAR can be easily applied to different routine QSAR/QSPR tasks, for building multiple models, and for prediction of the different quantitative values simultaneously.

If you got interested in any of the products, provided by GeneXplain, or you have any questions, please contact us by e-mail info@genexplain.com. We will be glad to help you!

ISSN 1386-6338

In Silico Biology

Journal of Biological Systems Modeling and Multi-Scale Simulation



```
function [ t, x ] = firstReactionMethod( ...
    stoich_matrix, propensity_fcn, tspan, x0,...
    rate_params, output_fcn, max_out)

if ~exist('rate_params', 'var')
    rate_params = [];
end

num_rxns = size(stoich_matrix, 1);
num_species = size(stoich_matrix, 2);

%Simulation loop
while t(rxnCount) <= max(span)
    % Step 1: calculate propensities
    a  = propensity_fcn(X(rxnCount,:), rate_params);
    % Step 2: identify the reaction that will occur
    r = rand(1,num_rxns);
    taus = -log(r)./a;
    [tau, mu] = min(taus);
    % Update time and execute reaction mu
    rxnCount = rxnCount + 1;
    T(rxnCount)    = T(rxnCount-1)    + tau;
    X(rxnCount,:) = X(rxnCount-1,:) + stoich_matrix(mu,:);

    if rxnCount > max_out
        warning('SSA:ExceededCapacity','');
        return;
    end
end

% Simulation completed
t = T(1:rxnCount-1);
x = X(1:rxnCount-1,:);
end
```

IOS Press

Научное издание

BIODIVERSITY:
GENOMICS AND EVOLUTION (BioGenEvo-2018)

Symposium

Abstracts

Printed without editing

БИОРАЗНООБРАЗИЕ:
ГЕНОМИКА И ЭВОЛЮЦИЯ (BioGenEvo-2018)

Симпозиум

Тезисы докладов

Публикуется в авторской редакции

Выпуск подготовлен информационно-издательским отделом ИЦиГ СО РАН

Подписано к печати 30.07.2018. Формат 70 × 108 $^{1}/_{16}$. Усл. печ. л. 6,3.
Тираж 100 экз. Заказ № 186

Федеральный исследовательский центр
«Институт цитологии и генетики Сибирского отделения Российской академии наук»
630090, Новосибирск, проспект Академика Лаврентьева, 10

Отпечатано в типографии ФГУП «Издательство СО РАН»
630090, Новосибирск, Морской проспект, 2