Siberian Branch of the Russian Academy of Sciences
Federal Agency for Scientific Organizations
Federal Research Center Institute of Cytology and Genetics

THE EIGHTH INTERNATIONAL YOUNG SCIENTISTS SCHOOL "SYSTEMS BIOLOGY AND BIOINFORMATICS" – SBB-2016

Abstracts

SBB-2016
Novosibirsk, Russia
22–25 August, 2016

Novosibirsk
2016

INTERNATIONAL PROGRAM COMMITTEE

Prof. Nikolay A. Kolchanov, Academician RAS, Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia (Chairman)
Prof. Ralf Hofestädt, University of Bielefeld, Germany
Prof. Inna Lavrik, Otto von Guericke University, Magdeburg, Germany
Prof. Evgeniy Rogayev, University of Massachusetts, USA
Prof. Elsa Khusnutdinova, Institute of Biochemistry and Genetics, Russian Academy of Sciences, Ufa, Russia
Prof. Guoliang Li, Huazhong Agricultural University, Wuhan, China
Dr. Maxim Zakhartsev, Hohenheim University, Institute of Plant Physiology and Biotechnology, German
Prof. Alexander Samsonov, Ioffe Physical Technical Institute, Saint Petersburg, Russia
Prof. Ming Chen, Zhejiang University, Hangzhou, China
Prof. Yuriy Orlov, Novosibirsk State University, Russia
Prof. Dmitriy Sherbakov, Limnological institute, Irkutsk, Russia
Prof. Ivo Grosse, Halle-Wittenberg University, Halle, Germany
Prof. Rene te Boekhorst, University of Hertfordshire, UK
Prof. Ancha Baranova, George Mason University, USA
Prof. Haihua Bai, Inner Mongolia University for the Nationalities, China
Dr. Vsevolod Makeev, VIGG RAS, MIPT, Moscow, Russia

LOCAL ORGANIZING COMMITTEE

Dmitriy Afonnikov, PhD, ICG SB RAS
Tatyana Chalkova, ICG SB RAS
Nadezhda Glebova, ICG SB RAS
Tatyana Karamysheva, PhD, ICG SB RAS
Galina Kiseleva, PhD, ICG SB RAS
Andrey Kharkevich, ICG SB RAS
Victoria Mironova, ICG SB RAS
Olga Petrovskaya, ICG SB RAS (chair)
Erlan Tokpanov, ICG SB RAS
Svetlana Zubova, ICG SB RAS

CONTACTS

The Federal Research Center
Institute of Cytology and Genetics
The Siberian Branch of the Russian Academy of Sciences

Russian Federation
630090 Novosibirsk, Lavrentyeva, 10

Tel: +7(383) 363-49-80
Fax: +7(383) 333-12-78

URL - ICG SB RAS: www.bionet.nsc.ru
URL - BGRS\SB-2016: http://conf.bionet.nsc.ru/bgrssb2016/
Organizing committee: sbb2016@icg.sbras.ru, popik-olga@bionet.nsc.ru

ISBN 978-5-91291-027-2

Organizers

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

Chair of Information Biology, Novosibirsk State University

Siberian Branch of the Russian Academy of Sciences

Company “Scientific service” Ltd.

The Vavilov Society of Geneticists and Breeders

The program “Genomics, proteomics and bioinformatics” of the SB RAS

Vavilov Journal of Genetics and Breeding

Novosibirsk State University

Sponsors

Russian Foundation for Basic Research

Siberian Branch of the Russian Academy of Sciences

Federal Agency for Scientific Organizations (FASO Russia)

SILVER SPONSORS

Khimexpert Ltd.

DIA-M Ltd.

Eppendorf Russia Ltd.

Roche Diagnostics Rus Ltd.

Kazan Federal Universiti

Agilent Technologies Russia

Bio-Rad

Qvadros-Bio

Albiogen

BRONZE SPONSORS

Genotek Ltd.

ИЦиГ СО РАН

Федеральное государственное бюджетное научное учреждение
«Федеральный исследовательский центр Институт цитологии и генетики
Сибирского отделения Российской академии наук» (ИЦиГ СО РАН)

директор – академик РАН Николай Александрович Колчанов

ИЦиГ СО РАН – это мультидисциплинарный, многопрофильный биологический институт с 60-летней историей, проводящий исследования по ключевым направлениям, связанным с феноменом наследственности (на уровне генов, геномов, клеток, организмов и популяций). ИЦиГ СО РАН по праву считается одним из ведущих биологических институтов в России, обладающим высокопрофессиональным кадровым составом.
В 2015 году образовался Федеральный исследовательский центр путем присоединения к Институту цитологии и генетики СО РАН в форме филиала Сибирского научно-исследовательского института растениеводства и селекции СО РАСХН (СибНИИРС).

Миссия ИЦиГ СО РАН: Решение приоритетных задач развития научно-технологического комплекса РФ в области генетики и селекции растений, генетики и селекции животных, генетики человека и биотехнологии на основе методов молекулярной генетики, клеточной биологии и биоинформатики.

Стратегическая задача: Проведение полных циклов исследований от генерации фундаментальных знаний до прикладных разработок в области генетики и селекции растений, генетики и селекции животных, генетики человека и биотехнологии на основе методов молекулярной генетики, клеточной биологии и биоинформатики, имеющих приоритетное значение для решения задач агропромышленного, биотехнологического, биологического и фармацевтического комплексов России.

Кадровый состав: в составе ИЦиГ СО РАН насчитывается 63 научных подразделения, в которых работает 990 человек, в том числе 417 научных работников, из которых 175 сотрудников в возрасте до 39 лет, 2 советника РАН, 4 академика РАН, 3 члена-корреспондента РАН, 58 докторов наук, 247 кандидатов наук (по данным на конец 2015 года).

Аспирантура: На конец 2015 года в аспирантуре ИЦиГ СО РАН обучалось 69 аспирантов.

Публикации: Институт активно печатается в российских и зарубежных журналах и является в российской биологии одним из признанных лидеров. В 2015 году опубликовано 202 статьи, зарегистрированных в международной базе Web of Knowledge (WoS), общее количество статей в рецензируемых журналах более 370. Статьи Института цитировались в WoS в 2015 г. более 4 000 раз.

Инфраструктура ИЦиГ СО РАН: УНУ «Центр генетических ресурсов лабораторных животных (ЦГР)» и 7 центров коллективного пользования (http://www.bionet.nsc.ru/uslugi/).

Имущественный комплекс: земельный участок площадью 35 тыс. га, закрепленный на праве постоянного пользования; 85 тыс. кв. м рабочих площадей, расположенных на территориях Советского района г. Новосибирска, Барышевского сельского совета НСО, в Искитимском и Черепановском районах НСО, в пос. Краснообск НСО.

Федеральный исследовательский центр ИЦиГ СО РАН приглашает к сотрудничеству научные и коммерческие организации!

Адрес: 630090, Россия, Новосибирск, проспект Академика Лаврентьева, 10, ИЦиГ СО РАН
Тел./факс: +7(383) 363-49-80 / +7(383) 333-12-78
URL: www.bionet.nsc.ru
E-mail: icg-adm@bionet.nsc.ru

CONTENTS

16S RRNA GENE SEQUENCE ANALYSIS OF EPIPHYTIC BACTERIA ISOLATED FROM ORGANIC RICE (*ORYZA SATIVA*)

S. Aphimeteetamrong[1], C. Kittiwongwattana[1]*
[1] *Department of Biology, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand*
* *Corresponding author: chokchai.ki@kmitl.ac.th*

Key words: *16S rRNA, bacteria diversity, epiphytic bacteria, rice*

Motivation and Aim: Epiphytic bacteria are able to colonize the surface of various plant organs including leaves, stems, roots, flowers, seeds and fruits. A number of epiphytic bacteria have been reported to promote plant growth by increasing the nutrient availability, producing phytohormones and inhibiting the growth of phytopathogens. Additionally, the use of plant-growth-promoting epiphytic bacteria in agriculture has been increasing in many parts of the world. In this preliminary study, we isolated and investigated the diversity of epiphytic bacteria on organic rice (*Oryza sativa* L.) plants for future applications in agriculture.

Methods and Algorithms: Epiphytic bacteria were isolated from rice plants collected from organic rice paddies in Thailand. Cell morphology and 16S rRNA sequencing analysis were used for identification of all isolates. Universal primers 41F and 1492R were used to amplify the 16S rRNA gene fragment. Pair-wise alignment analysis was performed using the EzTaxon database. Partial sequences of the 16S rRNA were compared with those of recognized bacterial species obtained from Genbank database for the multiple-alignment analysis. Phylogenetic tree was reconstructed using the neighbor-joining method.

Results: We isolated 120 epiphytic bacteria from rice plants. Currently, genomic DNA of thirty-six bacteria was obtained. The result showed that these isolates were members in phyla Bacteroidetes (Chitinophaga and Chryseobacterium), Firmicutes (Bacillus, Paenibacillus, Fictibacillus and Staphylococcus) and Proteobacteria (Pandoraea, Xanthomonas, Pseudomonas, Brevundimonas, Acinetobacter, Klebsiella and Citrobacter). The results from pairwise alignment analysis were consistent with that from phylogenetic analyses.

Conclusion: The analysis of the 16S rRNA gene showed that the bacterial community was consisted at least three different phyla. Four isolates were members of phylum Bacteroidetes, 20 isolates were under phylum Firmicutes and twelve isolates were in phylum Proteobacteria.

Availability: Not applicable.

HOW TO ACCOMPLISH A RAPID DEFENSE AGAINST FOREIGN DNA – RESTRICTION-MODIFICATION SYSTEMS AND IMPLICATIONS FOR SYNTHETIC GENE CIRCUITS

B. Blagojevic[1] *, M. Djordjevic[1], M. Djordjevic[2]
[1] *Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia*
[2] *Institute of Physiology and Biochemistry, Faculty of Biology, University of Belgrade, Belgrade, Serbia*
** Corresponding author: bojanab@ipb.ac.rs*

Key words: *R-M systems, transcript processing, synthetic gene circuits, CRISPR/Cas*

Motivation and Aim: Restriction-modification (RM) systems consist of restriction endonuclease (R), which cuts specific DNA sequences, and methyltransverase (M), which methylates and protects the same sequences from cleavage. It is considered that R must make a fast transition from "OFF" to "ON" state during RM establishment, so that the host cell becomes rapidly protected from foreign DNA. On the other hand, to prevent the host genome being cut by R, methylation must precede R expression. The relationship between these constraints on the enzyme expression dynamics, and the features controlling RM expression, is however unclear. To this end, we here develop a biophysical model of gene expression regulation in RM systems, to analyze dynamics of RM establishment in a naïve host. We then use this quantitative understanding to propose a synthetic gene circuits that can control how rapidly a potentially toxic molecule is expressed.

Methods and Algorithms: We develop a model of the enzyme synthesis in RM systems, based on statistical thermodynamics. We apply it to EcoRV, which is RM system with divergent CR and M promoters [1] that is under control of specialized control (C) protein. Overlapping RC and M promoters is the main feature of EcoRV, which we show is enough to ensure: i) the time-delayed expression of R with respect to M ii) the fast transition of the toxic molecule (R) from "OFF" to "ON" state iii) the increased stability of the steady-state of R [1]. Furthermore, we consider a novel synthetic gene circuit, which is capable of achieving a rapid cell defense against foreign DNA [2]. To this end, we combine transcription control inherent to RM systems, with the transcript processing inherent to CRISPR/Cas system [3].

Results and Conclusion: This, to our knowledge, represents the first quantitative model of expression regulation for a divergent RM system architecture. We show that EcoRV satisfies the proposed dynamical constraints, while any perturbation of system features makes these constraints less optimal. Combining transcription control of RM systems with transcript processing in CRISPR/Cas systems, allows a significantly faster transition from OFF to ON state.

References:

1. A. Rodic, B. Blagojevic, E. Zdobnov, M. Djordjevic, M. Djordjevic. (2016), submitted.
2. A. Rodic, B. Blagojevic, E. Zdobnov, M. Djordjevic, M. Djordjevic. (2016) to be submitted.
3. M. Djordjevic, M. Djordjevic, K. Severinov. (2012) CRISPR transcript processing: a mechanism for generating a large number of small interfering RNAs, Biology Direct, **7**:24.

COMPARATIVE STUDYING OF MULTICLUSTER STRUCTURE OF CHLOROPLAST GENOMES

E.I. Bondar[1]*, M.G. Sadovsky[2], Y.A. Putintseva[1, 3], M. Yu. Senashova[2]
[1] *Siberian Federal university, Krasnoyarsk, Russia*
[2] *Institute of computational modeling of SD RAS, Krasnoyarsk, Russia*
[3] *Institute of forest of SD RAS, Krasnoyarsk, Russia*
* *Corresponding author: bone-post@ya.ru*

Key words: *frequency; triplet; order; cluster; elastic map*

Motivation and Aim: Structures observed within a DNA sequence reveal an order and provide further understanding of functional roles of a sequence or its fragments. Previously, seven-cluster structure in various genomes has been reported [1]. It has been verified for the L. Sibirica chloroplast genome. Here we describe the different structure patterns observed over chloroplast genomes of various plant species.

Methods and Algorithms: 30 chloroplast genomes have been taken in Embl database. Firstly, each genome sequence was mapped into a set of equal length overlapping fragments. Numbers of different non-overlapping triplets in each fragment were counted. Then, a frequency dictionary, which contains the set of all the triplets counted within the fragments, was developed. The clusters in 63-dimensional space were identified with elastic map technique, where the objects to be clusterized are the fragments identified in the sequence. The 63-dimentional space is provided with frequencies of non-overlapping triplets found in the different fragments of the genome.

Results: The resulting structures were divided by numbers of clusters. In 10 out of 30 genomes 3 clusters were identified and in 12 out of 30 7 clusters were identified. Genomes with 7 clusters mostly include gymnosperms (Gnetum montanum, Juniperus bermudiana, Pinus taeda, Welwitschia mirabilis, Pseudotsuga sinensis, Larix decidua) and mosses (Physcomitrella patens, Anthoceros formosae, Aneura mirabilis), also diatoms (Thalassiosira pseudonana, Phaeodactylum tricornutum) and red alga (Cyanidioschyzon merolae). Genomes with 3 clusters mostly include angiosperms (Allium cepa, Brachypodium distachyon, Populus alba, Glycine max, Triticum aestivum, Vitis vinifera) and spore bearing plants (Angiopteris evecta, Equisetum arvense, Psilotum nudum) and one of gymnosperms (Cycas revoluta). The remaining eight species have different structures with undetermined numbers of clusters.

Conclusion: Unlike nuclear genome of bacteria, the chloroplast genome yields more complex structure. Further studies may bring new understanding of a fine structure details, or of relations between structure and taxonomy of plant species.

References:

Gorban A. N. et al. (2005) Four basic symmetry types in the universal 7-cluster structure of microbial genomic sequences. In Silico Biology 5: 25 – 37.

DRUG-INDUCED DYSKINESIA AND POLYMORPHISMS OF SGK1 GENE IN RUSSIAN SCHIZOPHRENIC PATIENTS

E.G. Boyarko*, O.Y. Fedorenko, A.V. Semke, S.A. Ivanova
Mental Health Research Institute, Tomsk, Russia
** Corresponding author: egboyarko@mail.ru*

Key words: *gene polymorphism, schizophrenia, dyskinesia*

Motivation and Aim: Extrapyramidal symptoms (including tardive dyskinesia) were observed in 20-30% of patients receiving conventional antipsychotics [1]. Antipsychotics change dopaminergic neurotransmission, which in turn may cause extrapyramidal disorders. An important influence of potassium channels on dopaminergic neurotransmission is well documented [2]. Studies show that serum and glucocorticoid-regulated kinase 1 (SGK1) modulate the activity of neuronal potassium channels [3]. The aim of our study was to investigate the association of polymorphisms of SGK1 (rs1743964, rs1057293, rs1009840) gene with drug-induced dyskinesia in Russian schizophrenic patients.

Methods: Blood samples were taken from 443 Russian Caucasian patients (61,4% male and 38,6% female) with a clinical diagnosis of schizophrenia (ICD-10: F20). The average age of patients 38 ± 14,5 years; duration of the disease at the time of the survey 23 ± 8,9 years. The drug-induced dyskinesia was assessed using standard international scale AIMS.

Results: When comparing the group of patients with and without dyskinesia we did not find any associations between polymorphic variants of SGK1 gene and tardive dyskinesia. Additional comparison were carried out in groups of men and women. In the group of men we identified the association of A allele ($\chi2 = 3,80$, $p = 0,049$) and AA genotype ($\chi2 = 6,17$, $p = 0,046$) of rs1057293 in SGK1 gene with drug-induced dyskinesia. In the group of women we identified the association of AA and GG genotypes of rs1743964 in SGK1 gene with dyskinesia ($\chi2 = 6,006$, $p = 0,049$). To assess clinical heterogeneity of tardive dyskinesia we subdivided group of patients with extrapyramidal disorders into groups of patients with orofaciolingual symptoms and patients with limb-truncal symptoms. It was found that orofaciolingual form was associated with AA and CC homozygous genotypes of rs1009840 in SGK1 ($\chi2 = 5,818$, $p = 0,049$) (in the group of women).

Conclusion: Thus, SGK1 gene are involved in the development of tardive dyskinesia induced by long-term therapy with neuroleptics and phenotypically different forms of tardive dyskinesia - orofaciolingual and limb-truncal - characterized by different genetic features.

The study was supported by a Russian Science Foundation № 14-35-00023.

References:

1. Loonen, A.J., van Praag, H.M. (2007) Measuring movement disorders in antipsychotic drug trials: the need to define a new standard. Journal of Clinical Psychopharmacology, 27(5): 423-430.
2. Zhang, H., Rodgers, E.W., Krenz, W.D., Clark, M.C., Baro, D.J. (2010) Cell specific dopamine modulation of the transient potassium current in the pyloric network by the canonical D1 receptor signal transduction cascade, J Neurophysiol, 104: 873–884.
3. Lang, F., Boehmer, C., Palmada, M., Seebohm, G., Strutz-Seebohm, N., Vallon, V. (2006) (Patho) physiological significance of the serum- and glucocorticoid-inducible kinase isoforms, Physiol Rev 86(4): 1151-78.

TARGETING CD95/FAS SIGNALING WITH PEPTIDES DERIVED FROM V-FLIP/NEMO CRYSTAL STRUCTURE

J.H. Buchbinder[1], N.V. Ivanisenko[2], M. Bollmann[1], M. Richter[1], V.A. Ivanisenko[2], I.N. Lavrik[1] *

[1] *Translational Inflammation Research, Otto-von-Guericke University, Magdeburg, Germany*

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

** Corresponding author: inna.lavrik@med.ovgu.de*

Key words: *CD95/Fas/APO-1, NF-κB, c-FLIP, Nemo, peptides*

Motivation and Aim: Stimulation of the death receptor CD95/Fas/APO-1 results in activation of apoptotic and anti-apoptotic pathways. In CD95 mediated signaling the DISC component c-FLIP is an important key player in NF-κB activation. Interestingly, the viral c-FLIP homolog v-FLIP can interact with Nemo/IKKγ, which is an important regulatory subunit in the canonical NF-κB pathway. Homology modeling of c-FLIP into the crystal structure of v-FLIP-Nemo suggests that c-FLIP can directly bind to Nemo. Based on structural modeling we designed a peptide derived from Nemo that is supposed to inhibit the c-FLIP-Nemo interaction. We used it to analyze the c-FLIP-Nemo interaction in the complex and check the influence of the Nemo-derived peptide on CD95-signaling.

Methods and Algorithms: The cervix carcinoma cell line HeLa overexpressing CD95 was selected for experiments and stimulated with chemically synthesized peptides and CD95L. Under these conditions, protein-protein interactions were analyzed by immunoprecipitation. Additionally, the activation of NF-κB and apoptosis was analyzed via western blotting and imaging flow cytometry. Noticeably, imaging flow cytometry allows monitoring the activation of these two signal pathways in single cells level in a quantitative manner.

Results: Our approach was able to reveal an interaction between Nemo and c-FLIP. The different Nemo-derived peptides were coupled to cell penetrating peptide and then able to enter HeLa cells. The peptides showed various effects on pro- and anti-apoptotic CD95 signaling. Finally, we obtained an interaction between Nemo-derived peptides and different c-FLIP isoforms.

Conclusion: c-FLIP can interact with Nemo similar to its viral homolog. Blocking this interaction with Nemo derived peptides and thereby CD95-induced NF-κB activation might be a promising therapeutic strategy of sensitizing tumor cells towards apoptosis.

Acknowledgements: The work is supported by Russian Science Foundation 14-44-00011

UGENE: A TOOLKIT FOR TEACHING STUDENTS

I.V. Bykova[1] *, O.I. Golosova[1], A.Y. Bakulina[2, 3], D.A. Afonnikov[2, 4], D.Y. Kandrov[1], A.Y. Palyanov[2, 5], G.A. Grekhov[1], Y.E. Danilova[1]

[1] Unipro Center of Information Technologies, Novosibirsk, Russia

[2] Novosibirsk State University, Novosibirsk, Russia

[3] State Research Center of Virology and Biotechnology VECTOR, Koltsovo, Novosibirsk region, Russia

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

[5] Institute of Informatics Systems SB RAS, Novosibirsk, Russia

** Corresponding author: ibykova@unipro.ru*

Motivation and Aim: Modern biological experiments in many cases require bioinformatics methods for planning and subsequent data analysis with application of complex computing algorithms.

Though there are serious commercial bioinformatics packages for these purposes, they sometimes are not available for the students. On the other hand free useful program tools and algorithms are uncoordinated in many cases and young biologists have to be experienced in programming to work with them successfully. Thus, young scientists encounter big obstacles in mastering bioinformatics hands-on.

Methods and Algorithms: UGENE is developed as an open-source free software aimed to assist a molecular biologist. It comprises a lot of analysis tools, including both the experiments design and data processing without any programming skills for a user.

UGENE provides an easy way to work with DNA, RNA and protein sequences. The functionality list is very wide: sequence annotation with access to remote databases, multiple alignment and phylogenetic trees, 3D protein structures, processing of Sanger and NGS sequencing data (genome assembling and variations, processing of RNA-Seq and ChIP-Seq data), etc. UGENE can be run on MS Windows, Linux and Mac OS X platforms.

Due to its accessibility and wide functionality UGENE can be used as an excellent tool to teach different biological methods.

Results and Conclusion: Currently UGENE is used in tutorials of several universities, for example, Roskilde University (Denmark). “Practical Bioinformatics”, interdisciplinary course utilizing UGENE, starts in September 2016 at NSU. Growth of educational programs using UGENE promotes its popularization within the international community of molecular biologists.

Availability: http://ugene.net/download.html

ANTIOXIDANT RESPONSE ELEMENT CONTROLS LYSOSOMAL BIOGENESIS MASTER-REGULATOR GENES

A.V. Chechushkov*, N.K. Zenkov, E.B. Menshchikova
Research Institute of Experimental and Clinical Medicine, Novosibirsk, Russia
** Corresponding author: avchechushkov@centercem.ru*

Key words: *Antioxidant Response Element, TFE3, TFEB, lysosomal biogenesis*

Motivation and aim: Antioxidant response element (ARE) is a highly evolutionary conserved transcription factor binding site that regulates expression of antioxidant defense genes and many other genes participating in cellular redox homeostasis [1]. Also, ARE was found in autophagy receptors genes and, vice versa, activity of transcription factor Nrf2, that binds ARE, has been shown to be regulated by autophagy [2]. TFEB and TFE3 are transcription factors that regulate lysosomal biogenesis and autophagy genes via CLEAR genomic motif, participating in stimulation of autophagosome formation and autophagy flux[3]. As TFEB and TFE3 participate in redox defense mechanisms, we were interested whether these transcription factors may be interrelated with ARE-binding transcription factors.

Methods and Algorithms: Using RSAT program package we created position frequency matrix for CLEAR and ARE motifs. Then genes of CLEAR-binding transcription factors (TFEB and TFE3) and ARE-binding transcription factors (Nrf1-3, BACH1-2) were searched for ARE or CLEAR motifs. Found motifs were screened for evolution conservation and ChIP-seq data for histone modifications using UCSC genome browser instruments. Real-time PCR was performed to investigate TFEB and TFE3 expression.

Results: No CLEAR motifs were found in ARE-binding transcription factors. Conversely, TFEB and TFE3 both have conservative ARE motifs with high score-similarity between mammals. ARE motif of TFEB gene is located in promoter region as indicated by H3K3me3 markers flanking the region. ARE of TFE3 is located in intron 2 of TFE3 gene and is flanked by H3K27Ac and H3K4m1 indicating intron enhancer region. Real-time PCR shows that ARE-inducing pharmacological agents tert-butylhydroquinone and its synthetic water-soluble analogue TS-13 increase TFEB and to a lesser extent TFE3 gene expression indicating conceivable functional role of found regulatory elements.

Conclusion: Our results may show possible evolutionary conservative control of master-regulator of lysosomal biogenesis genes by antioxidant response element. Such interactions may rely new pharmacological *basis for control of TFEB and TFE3 activity, by ARE-inducing pharmacological agents.*

References:

1. J.D. Hayes, A.T. Dinkova-Kostova, The Nrf2 regulatory network provides an interface between redox and intermediary metabolism, Trends Biochem. Sci. 39 (2014) 199–218. doi:10.1016/j.tibs.2014.02.002.
2. A. Jain et al., p62/SQSTM1 is a target gene for transcription factor NRF2 and creates a positive feedback loop by inducing antioxidant response element-driven gene transcription., J. Biol. Chem. 285 (2010) 22576–91. doi:10.1074/jbc.M110.118976.
3. M. Sardiello et al., A Gene Network Regulating Lysosomal Biogenesis and Function, Science. 325 (2009) 473–477. doi:10.1126/science.1174447.

SOFTWARE MODULE FOR INTEGRATION OF SBML-WRITTEN MATHEMATICAL MODELS OF MOLECULAR GENETIC SYSTEMS FOR THE HAPLOID EVOLUTIONARY CONSTRUCTOR 3D SOFTWARE PACKAGE

A.D. Chekantsev*, S.A. Lashin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: chekantsev@bionet.nsc.ru*

Motivation and Aim: Microbes forming large communities in nature regularly exchange genes via horizontal transfer. It gives microbial cells the ability to acquire novel metabolic functions [1] and consequently may lead to ecological changes in community as a whole. Nowadays, there are a lot of resources for warehousing mathematical models of metabolic [2-4] and gene regulatory [4] systems. It is a great challenge to integrate these data in a complex hierarchical model of evolving microbial community. The Haploid Evolutionary Constructor 3D (HEC 3D) framework allows constructing and simulating such communities consisting of cells of various strains/species/populations living in spatially heterogeneous habitats. Cells consume, utilize, synthesize and secrete metabolites according to genetic programs written modeled as gene networks [5]. Some of cells consume one metabolites and synthesize another, which respectively may be consumed by third cells, i.e. they form trophic cycles of exchanging metabolites.

The aim of this study is development and implementation of a software module for HEC 3D framework in order to import mathematical models of molecular-genetic systems from the existing databases to the HEC.

Methods and Algorithms: We used models written in SBML format. The integration is provided via libSBML and SOSlib libraries. To resolve issues with different synonyms of the same metabolites, we used the REST API for ChEBI database.

Results and discussion: The module designed allows us to extract the parameters and formulas reactions from the model loaded from the repositories such as BioModels or SABIO-RK and to replace existing HEC 3D generalized synthesis strategies with real SBML models. ChEBI database of chemical names was integrated which solved the problem of metabolites wrong usage. Thus, the novel module allows users to use the "real-world" models in the HEC 3D and to investigate the behavior and the evolution of complex microbial communities.

Acknowledgement: The study has been partially supported by the RFBR grant №150703879 and Budget Project 0324-2015-0003.

References:

1. T.Y. Pang, S.S. Maslov (2011) A toolbox model of evolution of metabolic pathways on networks of arbitrary topology, PLOS Comp. Biol., 7: e1001137.
2. U. Witting et al. (2012) SABIO-RK – database for biochemical reaction kinetics, Nucleic Acid Res., 40: D790–D796.
3. S.A. Coskun et al. (2013) An online model composition tool for system biology models, BMC Syst. Biol., 7: 88.
4. F.V. Kazantsev et al. (2016) MAMMOTh: a new database for curated MAthematical Models of bioMOlecular sysTems, bioRxiv, 053652.
5. S.A. Lashin et al. (2014) HEC 2.0: improved simulation of the evolution of prokaryotic communities, In Silico. Biol. 11: 125-135.

LONG-TERM SPACEFLIGHT MEDIATED CHANGES IN PROMOTER LANDSCAPE IN ZEBRAFISH TISSUES

A.V. Cherkasov[1] *, K.V. Arshavsky[1], V.N. Sychev[2], M.A. Levinskikh[2], O.A. Gusev[1, 3, 4]
[1] *Institute of Fundamental Biology and Medicine, Kazan Federal University, Kazan, Russia*
[2] *Institute for Biomedical Problems, Russian Academy of Sciences, Moscow, Russia*
[3] *Division of Genomic Technologies, CLST, RIKEN, Yokohama, Japan*
[4] *Preventive Medicine and Diagnosis Innovation Program, CLST, RIKEN, Yokohama, Japan*
* *Corresponding author*: *cherkasovav@yandex.ru*

Key words: *CAGE, transcriptional activity, promoter landscape, spaceflight, zebrafish*

Motivation and Aim: Animal models are important to understanding influence of different factors of long-term spaceflights on living organisms and can be helpful for forecasting and prevention negative effects of spaceflights on humans. Some previous experiments, which performed on fish models using simulated microgravitation conditions and exposure aboard International Space Station (ISS), shows significant changes in whole genome gene expression. To define impact of spaceflight to tanscriptional activity on promoter level in Zebrafish tissues, we perform experiments using cap-analysis gene expression (CAGE) methodology.

Methods and Algorithms: Two groups of Zebrafish individuals was used as experimental and control group. The individuals from experimental group were maintained in Aquatic Habitat (AQH) in ISS and fixated in RNA stabilization reagent immediately after arriving and after 36 days of staying aboard. Part of experimental group animals were returned alive from ISS for RNA fixation in ground conditions in two time-points: 2 and 36 days after return. The animals from ground control group were fixated at the same time-points.

Results and Conclusion: First results of CAGE shows significant impact of spaceflight on transcriptional initiation landscape. More than 600 genes are changing their expression in eye samples, after arriving aboard ISS. Notably, the number of differentially expressed genes decreased to 154 after 36 days in space, it can be supposed successful adaptation to spaceflight conditions. From genes, that significantly activated in space, we found several important transcription factors: Fos, FosB and Jdp2, that regulate expression of many genes involved in growth and tissue development processes. Gene Ontology analysis of space-responsive genes shows significant over-representation of functional categories associated with circadian clock system, that confirm influence of microgravitation conditions to regulation of rhythmic processes in animals.

Acknowledgements: The work is performed according to the Russian Government Program of Competitive Growth of Kazan

MODELING OF THE BLOOD FLOW IN THE NARROWED VESSELS

S.G. Davydova[1] *, E.A. Biberdorf[1, 2]
[1] Novosibirsk State University, Novosibirsk, Russia
[2] Sobolev Institute of Mathematics SB RAS, Novosibirsk, Russia
** Corresponding author: svetamira_davydova@mail.ru*

Key words: *one-dimensional model of hemodynamics, mathematical modeling, pulse wave velocity, aorta*

Motivation and Aim: A complex mathematical model of the cardiovascular system realized on the BioUML platform was created as a result of the cooperation of Sobolev Institute of Mathematics and Design Technological Institute of Digital Techniques SBRAS. It proved to be very effective in many experiments [1]. This model assumes that all vessels have the cylindrical form. But at the same time a number of large vessels such as the aorta have a conical shape. A transition from the cylindrical vessels to the conic ones allows to significantly improve this model, especially in questions of modeling of a pulse wave. Thus the purpose of this work is to construct an one-dimensional hemodynamics model for the vessel's conical shape and research the distinctions between the vessels of conical and cylindrical forms.

Methods and Algorithms: Theoretical methods of mathematical physics, computational methods - a method of straight lines and orthogonal pro-race.

Results: The system of hemodynamics equations for the conical shaped vessels has been obtained and its program realization in the MATLAB system has been carried out. The velocity of blood flow and the distribution of the pulse wave, the influence of a filtration coefficient on the formation of the reflected wave and the reflection of the pulse wave from the place of the joint of two cylindrical vessels have been investigated. The calculations results lead to the conclusion that using of conic vessels and also the special filtration coefficient on the ends of the terminal vessels allows avoiding of the emergence of «excess» reflections of the pulse wave.

Conclusion: The received model for the vessels of conical shape can be recommended for the improvement of the existing model on the BioUML platform, in particular, for more accurate modeling of the pulse wave profile.

Availability: http://lib.nsu.ru:8081/xmlui/handle/nsu/10420

References:

1. I. N. Kiselev, E. A. Biberdorf, V. I. Baranov, T. G. Komlyagina, V. N. Melnikov,
2. I.Yu. Suvorova, S. G. Krivoshchekov, F. A. Kolpakov, «Personalization of Parameters and Validation of Model of the Human Cardiovascular System», Mat. Biolog. Bioinform., Vol. 10, № 2, 2015, P. 526—547.

EVIDENS OF VIOLETIONS OF GLUTAMATERGIC NEUROTRANSMISSION IN PATIENTS WITH SCHIZOPHRENIA USING PROTEOMICS METHODS

E.M. Dmitrieva[1] *, L.P. Smirnova[1], A.A. Seregin[1], A.A. Letova[3], A.V. Semke[1], S.A. Ivanova[1], V.G. Zgoda[2]

[1] *Mental Health Research Institute, Tomsk, Russia*

[2] *Institute of Biomedical Chemistry, Moscow, Russia*

[3] *Siberian State Medical University, Tomsk, Russia*

** Corresponding author: Lena-312tom@yandex.ru*

Key words: *Schizophrenia, glutamate, proteomics, biomarker*

Introduction: Schizophrenia is a disease with unknown pathogenesis, but today it is known that in schizophrenia there are disturbances of protein metabolism. Nevertheless, protein – marker, inherent in only this illness, has not still been detected. Analysis of the high molecular weight proteins is possible using mass spectrometry that allows identification of the potential markers of disease. Consequently, if we would know the content of protein markers in the blood serum so we could judge on the efficiency of the therapy, and relapse of the disease before the occurrence of clinical manifestations.

Materials and methods: We used the serum of 8 healthy persons and 20 patients with paranoid schizophrenia. Diagnostics was carried out in accordance with the current classification ICD-10. Preparation of samples included: purification from serum major proteins by affinity chromatography, separation of proteins by 1-D electrophoresis, the proteins in the gel trypsin digestion followed by extraction. Analysis of serum proteins was performed using an ion trap Thermo Scientific LTQ Velos. The quantitative mass spectrometric analysis, using synthetic peptide standards was carried out on QQQ TSQ Vantage (Thermo Scientific) equipped with a nano-electrospray ion source. Identification of proteins was carried out using program resources Mascot Ver. 2.1 (Matrix Science). For qualitative analysis method by Western blotting was used. Determination of glutamate concentration was performed using a set of «Glutamat Assay Kit» фирмы BioVision Research Products, Montain, USA. Statistical analysis was performed using Fisher's exact test with Yates' correction using the program STATISTICA 8.0.

Results: A result of mass spectrometry the protein of metabotropic glutamate receptor (mGlurR6) – 95376 Da was discovered in serum of blood of patients with schizophrenia. Quantitative analysis revealed a two-fold increase in protein concentration of mGluR6 of patients with paranoid schizophrenia. The presence of mGluR6 in the serum of patients also was confirmed by a qualitative analysis using Western blotting. The concentration of glutamate in the blood serum of patients with schizophrenia is twice higher than in the serum of healthy.

Thus, the protein of mGluR6, identified in our work, is directly associated with changes in the glutamatergic neurotransmission in patients with schizophrenia.

Support by Grant of RSF no.14-15-00480 «The search for biomarkers of socially significant endogenous mental disorders» 2014–2016.

PROGRAM COMPLEX FOR DEAFNESS' SOCIAL ASPECTS SIMULATION

I.S. Dyachenko*, S.A. Lashin, O.L. Posukh
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: i.s.djachenko@gmail.com*

Motivation and Aim: Both inborn and acquired deafness affects more than 10% of world population. Frequency of inborn deafness' forms varies from one part of the world to another and could be defined by population genetic parameters. It is assumed that abundance of the most frequent genetic form of mutations-caused hearing loss relates to our society's adaptation for deaf people's needs and closure of deaf societies, which members typically marry within themselves. Therefore, study of social and genetic characteristics of deaf people societies has important meaning for predicting the abundance of different forms of inborn deafness and understanding social factors' part in evolutionary processes, which take place in humans' populations [1].

Methods and Algorithms: We combined multilayer population model [2] with models of assortative mating [3] taking into account genetic and social levels of population. Our model represents genetic mutations' spreading across the population considering people's assertiveness.

Results: We developed prototype of program model based on agent-oriented simulation platform "Diploid evolutionary constructor" [2] for studying hearing loss inheritance. Base objects of this model are an individual (an agent) and population structure. Phenotype strategy computes agent's phenotype by his genotype, which is represented by two alleles. It allows changing phenotype expression mechanism without modifying whole model structure. Partner choice strategy defines marriage partner choice, taking into account both common for all humans and specific for current problem factors.

Conclusion: Developed program complex will allow examining trends of inborn hearing loss forms abundance considering obtained social, demographic, molecular genetic and popular genetic parameters.

References:

1. O.L. Posukh et al. Impact of socio-demographic structure of the deaf people communities in prevalence of hereditary hearing loss // Vavilov J. Genetics and Breeding. 2016;20(1):7-15
2. Lashin S.A., Matushkin Yu.G. Multi-layer computer models of gene network evolution in diploid populations // FEBS Journal, 2013, Vol. 280 (Suppl. 1), p. 563.
3. S. Karlin, Some mathematical models of population genetics // Amer. Math. Monthly, 79 (1972), pp. 699-739.

RECONSTRUCTION AND ANALYSIS OF A NETWORK OF PROTEIN-PROTEIN INTERACTIONS FUNCTIONING IN MATURE MAMMALIAN ADIPOCYTE

E.V. Ignatieva*
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: eignat@bionet.nsc.ru*

Key words: *Adipocyte, Protein-protein interaction network*

Motivation and Aim: Elevated body mass index (BMI) and lipid abnormalities are a substantial risk factorы for human disease emergence. To reveal genes which may be most important in control of biological processes in mature adipocytes, we reconstructed and analyzed a network of protein-protein interactions (PPIs) functioning in this cell type.

Methods and Algorithms: We used: (1) data on insulin signaling in adipocyte [1]; (2) data on triglyceride biosynthesis pathway from Reactome; (3) data on lipolytic enzymes from KEGG; (4) a compilation of human genes regulating body weight [2]; (5) gene expression patterns from Genotype-Tissue Expression project (GTEx). PPI network was reconstructed and analyzed using STRING, Cytoscape and MCODE.

Results: The set of 127 proteins, controlling metabolic and signaling pathways in mature mammalian adipocyte was formed. PPIs between proteins encoded by these 127 genes and additional 212 genes from [2], which, according to GTEx, are expressed in all tissues, were extracted from STRING and analyzed using Cytoscape. The reconstructed PPI network contained 183 nodes and 532 edges. We revealed that the maximal number of neighborhoods had proteins/genes involved in signal transduction (MAPK1, IRS1, AKT1, PTPN11, STAT3, etc.). Three clusters with score > 3 and number of nodes from 13 to 23, formed by PPIs between signal transduction molecules were revealed. The maximal score (which was equal to 5.6) had a cluster which included six proteins involved in BBSome and cilia formation. Genes with the decreased tolerance to functional genetic variation were revealed using RVIS values published in [3].

Conclusion: Analysis of a mammalian adipocyte PPI network revealed proteins with maximal number of neighborhoods or involved in clusters. We propose to keep in mind these proteins/genes as a potential drug targets in the treatment of elevated body weight.

Acknowledgements: This work was supported by the Russian Science Foundation (project no. 14-24-00123).

References:

1. T.N. Kuznetsova et al. (2008) The analysis of structure of insulin-dependent regulatory contours of mature adipocyte, Usp Fiziol Nauk. 39(1):3-22.
2. E.V. Ignatieva et al. (2014) Human genes controlling feeding behavior or body mass and their functional and genomic characteristics: a review, Vavilov Journal of Genetics and Breeding, 18: 867-875
3. S. Petrovski et al. (2013) Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes, PLoS Genet, 9(8): e1003709.

NEXT GENERATION SEQUENCING DATA: QUALITY ASSESSMENT FROM DIAGNOSTIC POINT OF VIEW

M.V. Ivanov[1, 2] *, A. Baranova[1, 2, 3, 4], V.A. Mileyko[1, 2]
[1] *Atlas Oncology Diagnostics, Moscow, Russia*
[2] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
[3] *Research Centre for Medical Genetics, Moscow, Russia*
[4] *School of Systems Biology, George Mason University, Fairfax, VA, USA*
* *Corresponding author: maksim.v.ivanov@phystech.edu*

Key words: *NGS, Next Generation Sequencing, BRCA, quality control, breast cancer*

Motivation and Aim: Next generation sequencing allows to develop diverse diagnostic solutions by varying the sets of target regions, coverage, techniques for target enrichment and sequencing platforms. Width of the spectrum of available protocols calls for the cross-validation of diagnostic assays. There is a necessity for assessment of the quality of NGS data that is required for successful clinical implementation of this promising technology.

Methods and Algorithms: Here we propose a technique for evaluation of analytical sensitivity and specificity for the detection of germline and somatic mutations by reverse variant calling under null hypothesis that each variant is present in sequencing reads. To calculate germline mutations presence/absence probabilities we used Bayesian model, while somatic mutations were modeled by Poisson distribution. Calculated probabilities are further summarized using coefficients related to the frequencies of respective mutations.

Results: Here we present the software that allows one to evaluate 'clinical quality' of sequencing data independently from particularities of sequencing platform and protocol employed. Suggested approach translates a variety of bioinformatics parameters, including coverage depth and uniformity, base call quality, amplicon drop-out and platform-specific sequencing errors into analytical sensitivity and specificity. In order to validate the proposed method, we tested this analytic technique on the detection of BRCA1/2 mutations. Mutation spectrum and frequencies were obtained from Breast Cancer Information database. Sequencing datasets were sourced from various platform and sequencing protocols were obtained from SRA and platform supplier repository. After coverage normalization by random read sampling, the germline mutation detection specificity varied from 90% to 99,29%, thus, highlighting the need for monitoring the quality of the data and the necessity of platform-to-platform comparisons. A study of the variance in lab-to-lab and run-to-run outputs revealed the CV of 1% demonstrating overall robustness of the analyzed protocols. Interestingly, when various sequencing panels were compared, the specificity for somatic mutation detection varied from 40% to 98%, thus demonstrating the need for the application-specific panel selection.

Conclusion: Here we propose an approach to the quality control of NGS data applicable to the cases when we expect certain mutation spectrum and frequencies (such as hereditary breast cancer syndrome and cystic fibrosis). Moreover, it can be also used for cross-comparing the results obtained using different protocols, labs and even runs. Proposed approach can be further optimized for genes with known somatic mutation hotspots like EGFR, ERBB2, RAS, cKIT and PDGFRA.

ASSOCIATION STUDY OF THE ENDOTHELIAL NITRIC OXIDE SYNTHASE GENE POLYMORPHISMS AND METABOLIC SYNDROME IN RUSSIAN PATIENTS WITH SCHIZOPHRENIA

N.S. Fattakhov[1, 2] *, L.P. Smirnova[2], D.A. Parshukova[2], D.A. Skuratovskaia[1], L.S. Litvinova[1], A.V. Semke[2], S.A. Ivanova[2]

[1] *Immanuel Kant Baltic Federal University, Kaliningrad, Russia*

[2] *Mental Health Research Institute, Tomsk, Russia*

** Corresponding author: NiFattakhov@kantiana.ru*

Key words: *schizophrenia, metabolic syndrome, endothelial nitric oxide synthase, single nucleotide polymorphism*

Motivation and Aim: Incidence rates of metabolic syndrome (MetS) are significantly higher in patients with schizophrenia compared to the general population [1]. Genetic variation within the endothelial nitric oxide synthase gene (NOS3) may result in impaired endogenous nitric oxide formation and has been associated with cardiovascular diseases [2]. There is growing evidence that polymorphisms in NOS3 influence the development of MetS, however, there is also a controversy regarding the association of polymorphisms in the gene encoding NOS3 and MetS in patients with schizophrenia [3]. In this study, we aimed to evaluate the effects of NOS3 polymorphisms on MetS risk in Russian patients with schizophrenia.

Methods and Algorithms: 70 Caucasian patients with schizophrenia and MetS and 127 schizophrenic patients with normal BMI were enrolled in the study and genotyped for T-786C (rs2070744), G894T (rs1799983) and C774T (rs1549758) in NOS3. MetS was diagnosed using International Diabetes Federation (IDF) criteria.

Results: The allelic and genotypic frequencies of rs2070744 (promoter region) polymorphism in schizophrenic patients with MetS were significantly different from those in schizophrenic patients with normal BMI. These patients had significantly higher frequencies of rs2070744 T allele ($\chi 2$=6.80; p=0.009, OR=0.59; 95%CI: 0.40-0.88), rs2070744 C allele ($\chi 2$=6.80; p=0.009, OR=1.69; 95%CI: 1.14-2.51) and rs2070744 TT genotype (p=0.006, OR=0.45; 95%CI: 0.25-0.82). Strong linkage disequilibrium between rs1799983 and rs1549758 was observed (D' > 0.9). No association was observed between NOS3 haplotypes and MetS risk in patients with schizophrenia.

Conclusion: Our results point to a role for NOS3 polymorphisms in MetS in Russian patients with schizophrenia. These findings indicate that rs2070744 polymorphism may serve as a prognostic biomarker for MetS among Russian schizophrenic subjects.

References:

1. S. Malan-Müller et al. (2015) A systematic review of genetic variants associated with metabolic syndrome in patients with schizophrenia, Schizophr Res, 170(1): 1-17.
2. A.Uthor, O.N.E.More. (2003) Title of the paper, Journal, 12: 13-16.
3. P.L. Huang. (2009) eNOS, metabolic syndrome and cardiovascular disease, Trends Endocrinol Metab, 20(6): 295-302.
4. K. Burghardt et al. (2014) Endothelial nitric oxide synthetase genetic variants, metabolic syndrome and endothelial function in schizophrenia, J Psychopharmacol, 28(4): 349-356.

PHYLOGENY DEVELOPED OVER THE TRIPLET COMPOSITION OF MITOCHONDRIAL GENOMES

V.S. Fedotova[1] *, M.G. Sadovsky[2], Yu.A. Putintseva[1, 3]
[1] *Siberian Federal University, Krasnoyarsk, Russia*
[2] *Institute of Computational Modeling, Krasnoyarsk, Russia*
[3] *Forest Institute, Krasnoyarsk, Russia*
** Corresponding author: vasilinyushechka@gmail.com*

Key words: *population genomics, cluster, k-means, coevolution, elastic map*

Motivation and Aim: Genomics is an up-to-date field of molecular biology studying the genome of the organism in its entirety. Nowadays, a new direction in Genomics that is Population Genomics is rapidly developing. One of the main problem in Population Genomics is to identify the features of genomes that manifest at the population level.

In this study, we present some preliminary results obtained in this new direction. Rather nontrivial genetic object, namely mitochondrial genomes of different species, have been chosen for a study. Previously, it was reported that there is fundamental connection between the genome structure of organells and taxonomy of the bearers. Here we explore the problem and expand the results obtained earlier.

Methods and Algorithms: In our study, a population diversity has been studied by genomic methods: we examine the mitochondrial genome clusterization of different organism groups, in the space of triplet frequencies. All mitochondrial genomes were taken from EMBL-bank and were used to build up the database (http://www.ebi.ac.uk/ena).

Source database of mitochondrial genomes contained variety of species in different genuses. It included 3726 entries. Variety numbers of species in genuses caused bias in results. Unbiased dataset was created with uniform distribution of species in genuses and it includes 2990 genomes of different organisms.

We used K-means to cluster the dataset. The clusterisation was conducted in ViDaExpert program. As usual, K-means classifies a part genomes stably (in a series of the runs of k-means), while others show unstable classification [1]. The construction of layered graph is based on K-means: the classes developed for specific K yield a layer.

Results: The high synchrony in the evolution of two genetic systems manifesting in nonrandom distribution of taxonomic categories in over the vertices of layered graph has been found. As K grows up, the species comprising the classes redistribute among them in highly nonrandom pattern. The stability of clusterisation has been found.

Conclusion: High synchrony of classic phylogeny and clusterisation developed over the triplet composition of mitochondrial genomes has been founded.

References:

1. Sadovsky M. G., Putintseva Yu. A., Fedotova V. S., Chernyshova A. I. (2015) Genome structure of organells strongly relates to taxonomy of bearers. LNCS vol. 9044, Part 2: 481-490

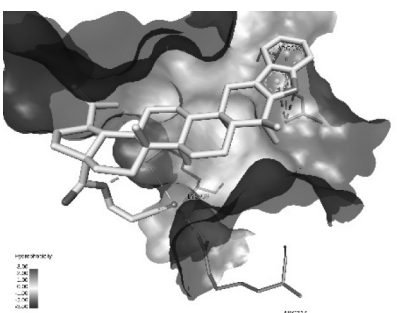
MOLECULAR MODELING OF THE INTERACTION BETWEEN INDOLE LUPANE DERIVATIVES AND C-MYC/MAX HETERODIMER

T.S. Frolova[1, 2, 3] *, D.S. Baev[2], A.V. Petrova[4], E.F. Khusnutdinova[4], O.I. Sinitsyna[1, 3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk Institute of Organic Chemistry SB RAS, Novosibirsk, Russia*
[3] *Novosibirsk State University, Novosibirsk, Russia*
[4] *Bashkir State University, Ufa, Russia*
** Corresponding author: frolova@bionet.nsc.ru*

Key words: *betulonic acid derivatives, selective cytotoxicity, molecular docking, c-Myc/Max interaction*

Motivation and aim: Pentacyclic triterpene acids are one of the most interesting compounds because of their various biological properties, for example, antioxidative, antiviral, antiallergenic, antiangiogenic and antispasmodic activity [1]. One of the most promising triterpenes in plant raw material are lupane acids, which are very popular for chemical modifications to enhance their biological activity, especially selective cytotoxicity against cancer cells. According to the literature, one the possible mechanisms of this effect is the inhibition of c-Myc/Max interaction [2].

Methods and Algorithms: Previously we synthesized four new lupane derivatives with indole functional group, all tested compounds have selective cytotoxic properties. Presently we assessed binding energy between c-Myc/Max heterodimer and indole lupane derivatives using molecular docking.



Interaction between c-Myc/Max heterodimer and F-118 derivative

Results: It was shown that all tested compounds probably could interact with binding site of c-Myc and prevent the association of c-Myc/Max heterodimer which results in cell death. Conclusion: Considering that c-Myc gene is overexpressed in many cancers, its directed inhibition may be used in antineoplastic therapy.

References:

1. A. Bishayee et al. (2011) Triterpenoids as potential agents for the chemoprevention and therapy of breast cancer, Front Biosci, **16:** 980–96.
2. H. Wang et al. (2015) Direct inhibition of c-Myc-Max heterodimers by celastrol and celastrol-inspired triterpenoids, Oncotarget, ***32*:**32380-95

RECONSTRUCTION OF TRANSCRIPTION CONTROL NETWORK IN GENOME-REDUCED BACTERIA BY HIGH-THROUGHPUT PROMOTERS IDENTIFICATION

I.A. Garanina*, G.U. Fisunov, D.V. Evsutina, V.M. Govorun
Scientific Research Institute of Physical-Chemical Medicine SRI PCM, Moscow, Russia
** Corresponding author: irinagaranina24@gmail.com*

Key words: *mollicutes, genome reduction, transcription, promoters*

Motivation and Aim: Class Mollicutes is a specialized clade of gram-positive bacteria that lack cell wall and feature significant genome reduction. Genome reduction resulted in the loss of most of the conserved transcriptional regulators. Attempts to elucidate transcriptional regulation in Mollicutes using high-throughput technologies resulted in a limited progress. In current work we aimed to investigate transcription regulation and reveal most important transcription factors for Mollicutes.

Methods and Algorithms: We utilized comparative genomics methods and algorithms to analyze conservation of regulators throughout Mollicutes phylogeny. For regulation network reconstruction we prepared and sequenced RNA 5`-ERS libraries of 4 samples of every bacteria analyzed. To identify transcription start sites we searched for a local maximum in the read coverage that was supported by at least 5 reads. Then, we modeled the coverage at each local maximum while considering 5 nt in each direction as background using a GLM with a quasi-binomial distribution.

Results: We described all transcription factors and DNA-binding proteins of 47 Mollicutes and found that there is only one common function requiring transcriptional control in all species: protein folding maintenance via chaperones. We carried out whole-genome mapping of transcription start sites of Acholeplasma laidlawii and Spiroplasma melliferum and compared promoter structure between three bacteria. A. laidlawii demonstrate the most organized promoter structure, whether M. gallisepticum show some extent of promoter degeneration. Based on comparative genomics approach we predicted binding sites of transcription factors and probable targets of these regulators. Most of the transcription factors with identified binding sites in the three species are involved in the metabolism control. Progressive loss of the conserved TFs from A. laidlawii to S. melliferum and M. gallisepticum is associated with drastic loss of metabolic pathways.

Conclusion: We suggests that only one regulatory component is conserved between analyzed bacteria. This fact reflects high plasticity and adaptive potential of Mollicutes. Our study thus provides insight into evolution and organization of transcriptional regulation of genome-reduced bacteria.

MUTATIONAL LANDSCAPE OF PROSTATE TUMORS BASED ON WHOLE EXOME SEQUENCING

I.R. Gilyazova[1, 2] *, M.A. Yankina[1], G.B. Kunsbaeva[2], A.A. Izmaylov[3], A.T. Mustafin[3], V.N. Pavlov[3], E.K. Khusnutdinova[1, 2]

[1] I nstitute of biochemistry and genetics Ufa Scientific Centre RAS, Ufa, Russia

[2] Bashkir State University, Ufa, Russia

[3] Bashkir State Medical University, Ufa, Russia

** Corresponding author: gilyasova_irina@mail.ru*

Key words: *prostate cancer, candidate genes, tumor tissue, variants, functional significance, rare alleles*

Motivation and Aim: Prostate cancer (PC) is one of the most common malignancies of the male population worldwide. Extremely rapid increase in the incidence of prostate cancer and the high prevalence of the disease in the world testify to the need to study the mechanisms underlying the development of prostate cancer.

Methods and Algorithms: New possibilities in molecular genetics investigations using NGS technology can expand the possibilities of tumor heterogeneity investigation, which is especially important for the genetic heterogeneity of pathologies with a large number of candidate genes. To find new genes involved in the pathogenesis of prostate cancer, we conducted a whole exome sequencing in samples of normal and tumor tissue for 8 patients with prostate cancer. The variants revealed were annotated using ANNOVAR software tool [1]. To identify the events associated with the tumor, the variants with frequencies more than 0.03 in any of the databases 1000 Genomes Project (European, East Asian, All), ExAC, ESP were excluded. The functional significance of the observed changes was carried out using following databases: SIFT, PolyPhen (using annotations for rare alleles), MutationTaster, MutationAssessor [2], FATHMM, CADD.

Results: We detected 41542 variants in normal tissue sample, 45948 - in tumor, in average. All the samples contained mutations in the ATM, and TP53 genes. After all stages of bioinformatic analysis 35 candidate genes, involved in cell cycle control, apoptosis signaling, androgen processes of cell growth and differentiation, transcription repair were selected - MUC16, MUC6, MTCH2, ZNF844, PRSS3, SSTR1, PDE11A, L2HGDH, etc.

Conclusion: The study revealed a number of genes which role in prostate cancer has not been described previously. The analysis is to be continued to determine the involvement of the identified genes in the pathogenesis of prostate cancer.

References:

1. H.Yang, K.Wang. (2015) Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR, Nature Protocols 10,1556–1566.
2. F.Gnad, A.Baucom, K.Mukhyala, G.Manning, Z.Zhang. (2013) Assessment of computational methods for predicting the effects of missense mutations in human cancers, Gnad et al. BMC Genomics, 14 (Suppl 3):S7.

DO MONOCOT AND DICOT PLANTS HAVE SIMILAR EVOLUTIONARY PATTERNS OF THE UNIVERSAL METABOLIC GENETIC NETWORKS?

A. Glagoleva[1, 2] *, O. Shoeva[1], E. Khlestkina[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
* *Corresponding author: ardemira@mail.ru*

Key words: *flavonoid biosynthesis genes, plant species, selective pressure, structural genes*

Motivation and Aim: We hypothesized that in universal metabolic genetic networks, such as flavonoid biosynthesis network, the specific features occurring in different taxa may result in different evolutionary patterns of the structural genes. The aim of the current study was to compare monocot and dicot species by synonymous and nonsynonymous substitution rates in the flavonoid biosynthesis genes.

Methods and Algorithms: Homologous sequences search was performed using BLAST algorithm in 3 databases (www.ncbi.nlm.nih.gov/Database/, https://urgi.versailles.inra.fr/blast/blast.php, http://webblast.ipk-gatersleben.de/barley/). Multiple sequence alignment was done with MULTALIN 5.4.1. Cluster analysis and Ka/Ks calculations were performed with MEGA v6.06 software using Neighbor-Joining algorithm.

Results: The nucleotide sequences of the Chs (chalcone synthase), Chi (chalcone-flavanone isomerase), F3h (flavanone 3-hydroxylase), F3'h (flavonoid 3'-hydroxylase), Dfr (dihydroflavonol 4-reductase), Ans (anthocyanindin synthase) genes in four monocot (Triticum aestivum, Hordeum vulgare, Oryza sativa, Zea mays) and four dicot (Arabidopsis thaliana, Vitis vinifera, Petunia hybrida, Malus domestica) species were identified. Validation of the orthologs was performed using phylogenetic analysis. To avoid sequences with rare SNP variants the alleles identified were compared with homologous ESTs. Then, C- and N-terminus signal peptides were eliminated to avoid their affect on assessment of enzyme-encoding parts evolutionary rates. The species were compared pairwise for Ka/Ks ratios of each gene. In monocot, negative correlation between structural genes Ka/Ks ratios and numbers of flavonoid classes synthesized by the enzymes encoded was shown. The Chi gene doesn't meet this rule (due to possible spontaneous conversion of chalcones to flavanones the Chi gene may possess relatively low selective pressure in both monocots and dicots). In dicots, there are no correlation between Ka/Ks ratios and numbers of flavonoid classes synthesized by the enzymes encoded. Intraspecies comparisons (between wheat genomes A and D) revealed unexpectedly high selective pressure on the F3h gene, whose activation is crucial for the flavonoid pigments anthocyanins production.

Conclusion: (1) Monocot and dicot plants differ by evolutionary patterns of the universal flavonoid biosynthesis genetic network. (2) Unexpectedly high selective pressure on certain gene in some species is suggested to be an indirect evidence of importance of this target gene activation as a key regulatory point of the biosynthetic pathway in this species.

PREDICTING OF THERMODYNAMIC DATA OF MORPHOLINO ANALOGOUS OF NA BY COMPUTER APPROACH AND COMPARING WITH EXPERIMENTS

V.M. Golyshev*, A.A. Lomzov
Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia
** Corresponding author: golyshevvictor@gmail.com*

Key words: *molecular dynamics (MD), novel analogous of NA, thermodynamics of binding and cooperativity*

Motivation and aim: Development of approaches for reliable calculations of physic-chemical properties of novel analogous of nucleic acids (NA) is a big challenge. Recently, the ability to calculate enthalpy of binding ΔH0 for native oligonucleotides (ON) with DNA was shown. The applicability of proposed method for NA derivatives is under dissuasion. A new morholino derivative [1] is an attractive object for verification of the predictive capacity. In this work the hybridization enthalpy of native and morpholino derivative oligonucleotide complexes were calculated using MD simulation and determined experimentally.

Methods and Algorithms: The structures of the two isomers of morpolino adenine was minimized using QM calculation at HF/6-311+G(d,p) in Gaussian'09 and particular atoms charges were calculated using RESP method. Based on this data the library files for MD simulation were prepares. MD simulations of native and the morpholino derivative (two isomer and one isomer mixed chain) oligonucleotide complexes were performed using AMBER12 software. The 1 μs trajectories were obtained in the explicit solvent in periodic condition (TIP3P water model, 12Å cuboid box) and NPT ensemble (1 bar, 300K). MD data we analyzed using MMGBSA calculation.

Thermal stability of pentamers of adenine and morhpholino derivative of adenine with oligothymidines (15, 20, 25 and ~300 base) was studied by thermal denaturation with optical registration of signal method. The data analyses were performed with originally developed method which includes simultaneous fitting of three different complex denaturation curves.

Results: The values of enthalpy changes of binding (ΔH0 in kcal/mol) and cooperative interaction (ΔH0k in kcal/mol) obtained by MD were ΔH0=28.3±0.2, ΔH0k=11.1±0.4 for native and ΔH0=25.3±0.2, ΔH0k=14.0±0.2 for morpholino complexes . Corresponding values, obtained by experimental techniques were ΔH0=37.7±2.1, ΔH0k=27.2±3.2 for deoxyriboadenosine and ΔH0=22.6±1.1, ΔH0k=15.5±0.7 morpholino derivative complexes.

Conclusion: We have shown that thermodynamics parameters of binding and oligonucleotide cooperative interaction in tandem duplexes can be determined by proposed using experimental and computer simulation techniques. Thermodynamics data obtained with MD and thermal denaturation approaches are perfectly matched. Thus the MD simulations are compatible for prediction of hybridization enthalpy of analogous of NA.

References:

1. Abramova, T. V., et al. (2012). Synthesis and properties of nucleic acid methylene carboxamide mimics derived from morpholine nucleosides. Russian Journal of Bioorganic Chemistry. 38(4) 400-411.

IN SILICO SEARCHING STRATEGY OF NEW ANTIMICROBIAL PROTEINS AND PEPTIDES: THE CASE OF TRANSCRIPTOME STUDY OF THE MEDICINAL LEECH

E.N. Grafskaia[1, 2] *, V.V. Babenko[2]
[1] *Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region, Russia*
[2] *Federal Research and Clinical Center of Physical-Chemical Medicine of Federal Medical Biological Agency, Moscow, Russia*
** Corresponding author: grafskayacath@gmail.com*

Key words: *antimicrobial peptides and proteins (AMPs), transcriptome, medicinal leech*

Motivation and Aim: Search of new antibacterial agents is an important and nontrivial task due to the rapidly growing antibiotic resistance of pathogenic agents to existing drugs. Means for studying biological multicomponent mixtures at the moment are very limited and consist principally of mass spectrometric and chromatographic methods of analysis. The aim of this work is to develop in silico transcriptome analysis to AMPs' search using transcriptome of three types of medicinal leeches: Hirudo medicinalis, H. verbana and H. orientalis.

Methods and Algorithms: To screen the medicinal leech transcriptome, we have been designed a joint database containing sequences of antimicrobial proteins, peptides and toxins out of 5 different databases: APD2, ADAM, CAMPR3, UniProt and ATDB2.0. Using InterProScan 5.15-54.0 software package we conducted the functional analysis of sequences and compared transcriptomic data against obtained database using BLAST. Furthermore, in silico analysis includes sequence physicochemical and structural properties investigation using EMBOSS PEPSTATS, TANGO and AGGRESCAN.

Results: Based on in silico analysis 19 candidate polypeptides were selected. Proteins of interest exhibit a different sequence homology to well-known AMPs. Extensive homology shows to neuromacin, lumbricin and histones; considerable homology extends to scolopendin, lectins. Several proteins display weak similarity to bactericidal permeability-increasing proteins, lactotransferrin and lactoperoxidase. However, selected proteins exert domain folding resembling bactericidal permeability-increasing proteins, chitin-binding proteins, lectins and different proteinases such as adamalysins, cystatins and zinc metalloproteinases.

Conclusion: In silico searching technique of new antimicrobial proteins and peptides was developed.

Acknowledgements: The authors thank the laboratory of genetic engineering, the laboratory of postgenomic research in biology and proteome analysis laboratory of FRCC PCM of FMBA for providing the experimental data and invaluable contribution to the work. Also, the authors are grateful to the Russian Science Foundation (agreement № 14-14-00696) for financial support.

TRANSCRIPTION BY ALTERNATIVE SIGMA FACTORS: REVISING THE RIGIDNESS PARADIGM

J. Guzina*, M. Djordjevic
Faculty of Biology, University of Belgrade, Belgrade, Serbia
** Corresponding author: jelenag@bio.bg.ac.rs*

Key words: *ECF sigma, bacterial promoters, transcription initiation, σ^{70} family*

Motivation and Aim: In distinction to the housekeeping σ factors, which transcribe a majority of bacterial genes, alternative σ factors have a more specialized regulon, necessary for coping with stress or development. Among alternative σ factors, ECF (ExtraCytoplasmic Function) σs are the most numerous and diverse group, which is, on the other hand, insufficiently studied [1]. In particular, the current ECF paradigm, which assumes recognition of rigid promoters with well conserved elements, is supported by a very limited data originating from a few (canonical) representatives. This is contrary to the mix-and-matching paradigm, which is well established for the housekeeping σ70 factors, implying flexibility in the promoter recognition. With the goal of gaining a better insight in the transcription initiation by ECF σ factors, we did a comprehensive study of ECF σ protein and DNA recognition motifs, and analyzed flexibility (mix-and-matching) in ECF σ promoter recognition.

Methods: By combining protein multiple global and local sequence alignment, domain search and DNA regulatory elements detection, we extensively computationally analyze all the available bacterial ECF σ subgroups [3]. We also use the ECF group outliers obtained by recently sequenced bacteriophages, as a source of independent (self-contained) data, which we analyze through a novel procedure for detecting phage-promoters [2]. Additionally, we systematically (quantitatively) analyze the canonical bacterial ECFs, through a biophysics based procedure, to investigate contribution of their promoter elements to transcription activity.

Results: We found an extreme qualitative example of mix-and-matching for phage ECFs, where a long -10 element extension - interacting with an extension of σ2 domain - complements the absence of the major (-35) promoter element. We also report examples of the putative novel interactions between ECFs and their promoters, exhibited by the conserved promoter spacer elements and the σ-motifs outside of the main σ2 and σ4 DNA-binding domains [3]. Finally, we provide quantitative evidence of substantial promoter element complementation (mix-and-matching) in ECFs [4].

Conclusion: We reveal a much larger flexibility in ECF σ functioning than previously recognized, which suggests that mix-and-matching may provide a common kinetic framework for promoter recognition in the entire σ70 family. Novel protein and DNA recognition motifs that we discovered, will guide future experiments on alternative σs.

Acknowledgement: This work is funded by the Swiss National Science Foundation (IZ73Z0_152297), European Commission (GA-2010-276996) and Serbian Ministry of Science and Technology (ON173052).

References:

1. Staroń A et al., 2009. Molecular microbiology 74:557-581.
2. Guzina J, Djordjevic M., BMC Evolutionary Biology 15: S(1) (2015).
3. Guzina J, Djordjevic M., J Bacteriol, in press (2016), doi: 10.1128/JB.00244-16
4. Guzina J, Djordjevic M., to be submitted.

SCORING OF PROTEIN DOCKING BY GENE ONTOLOGY

A. Hadarovich[1, 2] *, I.Anishchenko[1], A.V. Tuzikov[2], P.J. Kundrotas[1], I.A. Vakser[1]
[1] *Center for Computational Biology and Department of Molecular Biosciences, The University of Kansas, Lawrence, Kansas, USA*
[2] *United Institute of Informatics Problems, National Academy of Sciences, Minsk, Belarus*
** Corresponding author: ahadarovich@gmail.com*

Key words: *template-based predictions, modeling of protein complexes, molecular recognition*

Motivation and Aim: Structural characterization of protein-protein interactions (PPI) is important for understanding life processes at the molecular level. Experimental techniques, due to their inherent limitations, can determine structures only for a fraction of known PPI. Thus, most structures of protein complexes have to be modeled by docking techniques. In the template-based (comparative) docking, suitable templates are detected by sequence and/or structure similarity between the target and the template. When, structures of individual proteins are not available, comparative docking has to deal with modeled structures of the interactors, which are typically less accurate than the ones determined by experimental techniques [1], thus reducing target/template structure similarity score. This requires additional scoring of the target/template match, which would compensate for such drop in structure similarity.

Results: We present a functional score, based on target/template similarity of the Gene Ontology (GO) annotations [2] (GO-score), which is complementary to the structurebased scoring (TM-score) [3]. A scoring function that combines TM- and GO-scores was tested on a non-redundant set of 165 protein-protein complexes. The set includes six models for each structure, generated with predefined C-alpha root mean square deviations from the native structure [4]. For templates, we used the set of 4,950 template complexes [5] from the DOCKGROUND resource [6].

Conclusion: The results show that the new combined score improves the template detection and can be successfully applied to template-based docking of modeled proteins.

References:

1. I. Anishchenko et al. (2014) Protein models: The Grand Challenge of protein docking, Proteins, 82: 278-87.
2. Gene Ontology Consortium (2012) Gene Ontology annotations and resources, Nucleic Acids Research, 41: D530-D535.
3. Y. Zhang et al. (2005) TM-align: A protein structure alignment algorithm based on the TM-score, Nucleic Acids Research, 33: 2302-2309.
4. I. Anishchenko et al. (2015) Protein models docking benchmark 2. Proteins, 83: 891- 897.
5. I. Anishchenko et al. (2015) Structural templates for comparative protein docking, Proteins, 83: 1563-1570.
6. D. Douguet et al. (2006) DOCKGROUND resource for studying protein-protein interfaces, Bioinformatics, 22: 2612-2618.

FUNCTIONAL ANALYSIS OF RNA-SEQ TRANSCRIPTOMES FROM OESOPHAGEAL CANCER SPECIMENS OF KAZAKHSTANI PATIENTS

U. Kairov[1] *, A. Molkenov[1], S. Rakhimova[1], A. Abilmazhinova[1], M. Zhalbinova[1], D. Yerezhepov[1], A. Akhmetova[1], Y. Zhukov[2], M. Omarov[2], M. Popova[3], A. Zinovyev[4], A. Akilzhanova[1], Zh. Zhumadilov[1]

[1] *Center for Life Sciences, NLA, Nazarbayev University, Astana, Kazakhstan*
[2] *Oncology Center, Astana, Kazakhstan*
[3] *Department of Pathology, Astana Medical University, Astana, Kazakhstan*
[4] *Institute Curie, Paris, France*
** Corresponding author: ulykbek.kairov@nu.edu.kz*

Key words: *oesophageal cancer, transcriptome, RNA-seq, next-generation sequencing*

Motivation and Aim: Oesophageal cancer is the eighth most common cancer in the world and the highest in Eastern Asia. The incidence rate in Kazakhstan is 10.1: 100 000. Oesophageal squamous cell carcinoma (ESCC) is the dominant histological type (>90%) of esophageal cancer cases. Transcriptomic profiling of cancer specimens with next-generation sequencing technologies has provided a comprehensive opportunity for in-depth investigation of gene expression and affected molecular pathways. In our study we aimed to perform whole-transcriptome sequencing to identify affected molecular pathways and extract meaningful molecular signals from oesophageal cancer specimens of Kazakhstani patients.

Methods and Algorithms: Twenty three patients with esophageal cancer that underwent surgery at Oncology center (Astana, Kazakhstan). Fresh frozen cancer and its adjacent normal tissue specimens were obtained from each patient (in total 23 tumor center samples and 23 normal tissue samples). Whole-transcriptome sequencing was performed on Illumina HiSeq2000 platform at the Center for Life Sciences, Nazarbayev University. mRNA libraries were prepared using TruSeq RNA library prep kit according to standard protocol. Raw *.bcl files were converted and demultiplexed using bcltofastq. STAR and HTSeq have been used for alignment and mapping of sequencing reads. Differentially expressed genes have been identified using DeSeq. KEGG and Reactome databases were processed for analysis of signaling networks.

Results: Grouped analysis of cancer and normal samples has identified 1072 down-regulated and 1963 up-regulated genes. Functional analysis of up-regulated genes revealed the most significant enrichment for genes encoding products in the category of 'cell cycle' (p-value = 2.3×10-6), 'DNA replication' (p-value = 1.8×10-4) and 'lysosome' (p-value = 2.31×10-5), whereas down-regulated genes in the category 'metabolism of lipids and lipoproteins' (p-value = 8.62×10-4), 'valine, leucine and isoleucine degradation' (p-value=1×10-6) and 'propanoate metabolism' (p-value = 6.4×10-6).

Conclusion: Here, we report functional analysis of transcriptomic profiles from oesophageal cancer and matched adjacent normal specimens from twenty-three Kazakhstani patients.

DIVERGENCE OF PARALOGOUS GROWTH HORMONE GENES IN SALMONIDS

D.N. Kamenskaya[1] *, M.V. Pankova[2], D.M. Atopkin[2, 3], V.A. Brykov[1, 2]
[1] *A.V. Z hirmunsky Institute of Marine Biology FEB RAS, Vladivostok, Russia*
[2] *Far Eastern Federal University, School of Natural Sciences, Vladivostok, Russia*
[3] *Institute of Biology and Soil Science FEB RAS, Vladivostok, Russia*
** Corresponding author: kamenskaya.daria@mail.ru*

Key words: *growth hormone genes, nucleotide diversity, gene duplication, Salvelinus, Salmonidae*

Motivation and Aim: Duplication of DNA sequences, especially genes, is one of the main evolutionary factors. The stage of whole-genome duplication is believed to take place in the history of all organisms, and many duplicated genes could originate from it. Subsequently, full or partial diploidization stages occurred in various phylogenetic branches of eukaryotes, and these duplications can be manifested in the form of multiple gene families [1, 2]. Salmonids are a unique group, which developed in the past, after the last autotetraploidization event. Many genes in this taxonomic group were multiple, including the growth hormone (GH) gene. [3] The GH gene is represented by two unrelated paralogous genes, gh1 and gh2, in salmonids' genome. Both genes exist throughout the time of divergence of species in this group. Therefore, salmonids are a suitable model system for investigating the origin, evolution, and functions of duplicated genes. The aim of the current research is to compare paralogous GH genes of three genera in the family Salmonidae—Salvelinus, Salmo and Oncorhynchus—in order to determine their potential differences and possible functions.

Methods and Algorithms: To obtain nucleotide sequences of the GH genes, the following conventional techniques of molecular genetics were used: PCR, electrophoresis, molecular cloning, and sequencing. Nucleotide sequences were analyzed using the MEGA-6.0 and DnaSP-5.10.01 software packages.

Results: A comparison of the complete paralogous GH genes gh1 and gh2 of salmonids has shown that the conserved regions are associated with exons, and the peaks of variable regions correspond to intron sequences. It should be noted that not all intron sequences are variable; conserved regions can also be found. The presumable regulatory elements, localized in some introns (Pit-1 motifs, CRE, ERE), are also conserved. We found that gh1 gene in charrs is more conserved than gh2 gene; the amount of variability in gh2 gene is 2–3 times as large as that in gh1 gene, but it is not so obvious when comparing genes of the all investigated species.

Conclusion: A high conservation of coding sequences (exons) in paralogous GH genes of salmonid fish can be determined by the fact that both genes are functional or probably subfunctional. Both exons and regulatory regions are under the influence of negative selection. But the different rate of changes can be explained by the different selection intensity in paralogous genes.

References:

1. J.Zhang. (2003) Evolution by gene duplication: an update, Trends Ecol. Evol., 18: 292–298.
2. J.B.Walsh. (1995) How often do duplicated genes evolve new functions?, Genetics, 139: 421–428.
3. F.W.Allendorf, G.H.Thorgaard. (1984) Tetraploidy and the evolution of salmonid fishes, In: Evolutionary Biology of Fishes, B.J.Turner (Eds.), 1–53 (N.Y.: Plenum Press).

DIOXIN-MEDIATED REGULATION OF KEY GENES INVOLVED IN THE FUNCTIONAL CHARACTERISTICS OF MACROPHAGES

E.V. Kashina[1] *, D.Y. Oshchepkov[1] *, D.P. Furman[1, 2], V.A. Mordvinov[1, 3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
[3] *Institute of Molecular Biology and Biophysics SB RAS, Novosibirsk, Russia*
** Corresponding authors: alena_kashina@bionet.nsc.ru; diman@bionet.nsc.ru*

Key words: *macrophage, dioxin, AhR*

Motivation and Aim: The environmental pollutant 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD, dioxin) is the most toxic among the dioxin xenobiotics and induces a broad spectrum of biological responses including immunotoxicity and cancer [1]. Macrophages are key regulators of the innate immune response, however their response to TCDD is still poor understood. TCDD acts as a modulator of gene expression via AhR receptor, the important regulator of cell physiology and organ homeostasis [2]. To study the effects of TCDD on macrophages functional characteristics and measure expression profiles we performed RNA-Seq experiments using U937 cell culture model.

Methods: For the study of the TCDD-mediated alteration of the gene expression levels total RNA was isolated from U937 macrophages, treated with 10 nM (or 0.1% (v/v) DMSO as a control) for 1 and 6 hours. cDNA library preparation was performed by NEBNext mRNA Library Prep Master Mix set for Illumina according to the manufacturer's instruction. The RNA-Seq experiments were performed on Illumina HiSeq platform. Over 20 million cDNA reads were included in the RNA-seq analysis performed by ZAO Evrogen (http://evrogen.com/index.shtml). Categories of gene ontology for differentially expressed genes were determined using Reactome Pathway Database (http://www.reactome.org). Expression levels of some key genes were verified by Real Time-PCR with SY

BR Green I.

Results: For the first time we demonstrated that TCDD significantly enhanced expression of genes that are important for cellular responses to stress (HIST3H2BB, HIST3H2BB), eicosanoids and fatty acids metabolism (CYP4F22, CYP3A43), IGF1R signaling cascade (IGF1, CAMK2A), innate immune system signaling (PDE1A, MAPK10).

Conclusion: Obtained data shed a light on a wide range of dioxin effects including immunotoxicity and carcinogenicity. It is important to note that these effects can be induced via AhR signaling pathway.

References:

1. Mandal P. (2005) J Comp Physiol B. 175: 221-30.
2. Mulero-Navarro S., Fernandez-Salguero P.M. (2016) Front Cell Dev Biol. May 11;4:45.

PHAGE INFECTION SLOWS DOWN SPECIATION CAUSED BY GENE LOSS AND HORIZONTAL GENE TRANSFER OF METABOLIC GENES IN MODELS OF SPATIALLY DISTRIBUTED BACTERIAL COMMUNITIES

A.I. Klimenko*, Yu.G. Matushkin, N.A. Kolchanov, S.A. Lashin
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: klimenko@bionet.nsc.ru*

Motivation and Aim: Bacteriophages are known to be one of the driving forces of bacterial evolution. Besides promoting horizontal gene transfer between cells, they may induce directional selection (for instance, according to more or less resistance of cells to phage infection). However, the impact of phages on metabolic evolution and formation of community trophic structure remains obscure. Spatial organization of the environment is another factor that is crucial for a bulk of processes in the corresponding microbial community including its evolution and infection patterns. We have simulated and analyzed a series of computer models of microbial communities evolving in spatially distributed habitats under the pressure of phage infection.

Methods and Algorithms: We used a multilayer simulation tool HEC 3D [1] taking into account genetic, metabolic, cellular, population, and ecological levels of community organization. It simulates both high-level (cellular chemotaxis and diffusion, substrates flow and diffusion) and low-level (mutations, horizontal gene transfer, gene regulation and metabolism) processes allowing combining various mathematical modeling approaches (agent-based modeling, differential equations, automata etc.) in one model.

Results: We modeled evolving microbial communities living in spatially distributed aquatic habitats characterized by a nutrient gradient. We varied time and location of initial phage infestation as well as switched chemotaxis on and off. Simulations have shown that phage infection decreases the speciation rate by more than one order as far as intensified selection blocks the origin of novel viable populations/species, which could carve out potential ecological niches. The dependence of speciation rate on the invasion node location varied on the invasion time corresponding to different stages of community formation.

Conclusion: Our study has shown that phage infection affects evolution of microbial community slowing down speciation caused by gene loss and horizontal gene transfer of metabolic genes and stabilizing the system as a whole [1]. This influence varied in its magnitude depending on spatially-ecological factors as well as community state at the moment of phage invasion.

Acknowledgements: The study has been partially funded by the RFBR grant 150703879 and Budget Project 0324-2015-0003.

References:

1. Klimenko A.I., Matushkin Yu.G., Kolchanov N.A., Lashin S.A. Bacteriophages affect evolution of bacterial communities in spatially distributed habitats: a simulation study // BMC Microbiology, 2016, Vol. 16 (Suppl. 1), No.1, p. 31

WHEATDB2: PLANT TRAIT DATABASE AND INFORMATION SYSTEM BASED ON CROP ONTOLOGY TERMS

E.G. Komyshev[1] *, M.A. Genaev[1], A.V. Akushkina[2], D.A. Afonnikov[1, 3]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State Agrarian University, Novosibirsk, Russia*
[3] *Novosibirsk National Research State University, Novosibirsk, Russia*
** Corresponding author: komyshev@bionet.nsc.ru*

Key words: *extreme climatic conditions, database, plant selection, crop, ontology*

Motivation and Aim: Extreme weather conditions such as hot summer or cold winter frequent in many regions of Russia and have a negative impact on crop yields. Increasing grain yields can be achieved by identifying those plant varieties which are most suitable for cultivation in certain target areas. Thereby a systematic collection of data on the genotypes and phenotypes of plants in different climatic conditions is required. This requires developing integrated databases that collect large scale data on phenotype, genotype and environment. An effective approach to solve this problem is to use ontology terms together with a detailed documentation of an experimental protocol to fix important information such as a condition and location of the experiment, used measurement tools. We present a database of specific plant traits based on Crop Ontology ontology terms [1] with a reference to genetic data: varieties/lines, genetic markers, sequences.

Methods and Algorithms: The basic structural unit of the database structure is the "experiment" that combines information on the protocol of the experiment, its conditions, measurement tools, methods and participants. Experiment data types can be bound to the terms of ontology Crop Ontology. As a database management system was selected the MongoDB [2] that allow to represent a variety of flexible data structure in single database.

Results: A user can save the data in structures in a way which is convenient for him. The data can be selected from a number of various experiments and used for comparative analysis to identify the most suitable crop for cultivation in certain climatic conditions.

Conclusion: A proposed database structure is suitable for storing information such as a condition and location of the plant breeding experiment, used measurement tools.

Availability: Wheatdb2 system is available at http://pixie.bionet.nsc.ru/wheatdb2

Acknowledgements: The work supported by RFBR grants №16-37-00304, №16-34-00688 and budget project № 0324-2015-0003.

References:

1. Shrestha R. et al. (2010) Multifunctional crop trait ontology for breeders' data: field book, annotation, data discovery and semantic enrichment of the literature, AoB plants, **2010:** plq008.
2. Chodorow K. (2013) MongoDB: the definitive guide, In: O'Reilly Media, Inc.

MOLECULAR EVOLUTION ANALYSIS OF GENES RELATED TO PLANT ROOT HAIR AND TRICHOME DEVELOPMENT

D.K. Konstantinov[1], A.V. Doroshkov[2] *
[1] *Novosibirsk State University, Novosibirsk, Russia*
[2] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
* *Corresponding author: ad@bionet.nsc.ru*

Key words: *molecular evolution, epidermal cells, trichomes, root hairs*

Motivation and Aim: Specialized cell formation is an fruitful model system for analyzing the molecular mechanisms of plant cell differentiation, including cell fate choices, cell cycle control, and cell morphogenesis (Yang, C., & Ye, Z. 2013). In plants, epidermal cells are easily accessible and allow in vivo study. Arabidopsis thaliana trichomes were used as a model system for identification of the activator–inhibitor and the activator–depletion pattern formation models, studying the interplay between cell cycle and cell differentiation and numerous of genes involved in these processes were found. However, the evolution of specialized epidermal cell formation genetic network remains unclear. In this study, we analyze the phylogenetic relationships of genes associated with the formation of trichomes and root hairs from various species of flowering plants (monocotyledonous and dicotyledonous).

Methods and Algorithms: Extracting sets of homological sequences presets from databases was carried out using the reciprocal BLAST search. Multiple sequence alignment was conducted with MAFFT algorithm. The PhyML maximum likelihood algorithm was used to reconstruct the phylogeny and bootstrap resampling technique was used for testing the topology. Genetic networks containing target genes were reconstructed using Cytoscape and Pathway Studio software.

Results and conclusion: Our results argue that there is a large fraction of genes involved in the formation of trichomes and root hairs and do not reveal a direct correspondence between monocotyledonous and dicotyledonous plants. This fact is in the agreement with the latest data confirming that the closest homolog of Arabidopsis thaliana trichome and root hair-related gene in rice do not affect the rice pubescence (Zheng et.al., 2016). Also, this facts justify that part of the cellular morphogenesis mechanisms evolved independently in dicots and monocots. At the same time, we observe a good correspondence between studied genes of cell morphogenesis inside dicotyledonous as well as monocotyledonous clades. This allowed us to find orthologous genes in wheat genomic sequences and predict its chromosome localization to compare with known leaf hairiness QTLs.

Acknowledgements: This study was funded by Russian Science Foundation grant №14-14-00734.

Reference:

1. Yang, C., & Ye, Z. (2013). Trichomes as models for studying plant cell differentiation. Cellular and molecular life sciences, 70(11), 1937-1948.
2. Zheng, K., Tian, H., Hu, Q., Guo, H., Yang, L., Cai, L & Wang, S. (2016). Ectopic expression of R3 MYB transcription factor gene OsTCL1 in Arabidopsis, but not rice, affects trichome and root hair formation. Scientific reports, 6, 19254.

MATHEMATICAL MODELING A RECIPROCAL INTERACTIONS BETWEEN AUXIN AND ITS PIN TRANSPORTERS IN THE ROOT TIP OF *A. THALINA* L.

V. V. Kovriznykh[1, 2] *, F.V. Kazantsev[1], N.A. Omelyanchuk[1, 2], V.V. Mironova[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: vasilinakovr@gmail.com*

Key words: *auxin, Arabidopsis thaliana, PIN-FORMED (PIN), mathematical modeling*

Motivation and Aim: Plant hormone auxin is the main regulator of plant growth and development. In the root, auxin is distributed nonuniformly, with concentration gradients and a maximum in the quiescent center (QC), which are necessary for the meristem maintaining. The PIN-FORMED family proteins (PIN1-PIN4, PIN7) asymmetrically locate on the cell membrane and by this form directed auxin flows in the root tissue. Auxin regulates the speed of its efflux from the cell by mediating PINs expression. Comparing the experimental data on auxin response distribution and PIN proteins localizations, we hypothesized that different auxin concentrations might be optimal for expression of various PINs. This assumption was verified by mathematical modeling.

Methods and Algorithms: 2D cellular automaton was developed on the basis of the model [1] to study auxin distribution along longitudinal direction of the A. thaliana root tip. The cell layout consists of inner and outer layers, which differ by the set of expressed PINs. Auxin flows into the upper cell raw of the inner layers, then it is distributed across the cell layout by diffusion and PIN-mediated polar transport. In accordance with the assumption, PINs expression is established by auxin concentration in a cell in a dose-response manner.

Results: We started the model calculation from zero initial condition, which at the beginning of calculation led to generation of PIN2 expression in the outer layers with shootward polarization. As auxin entered to the cell layout, a PIN3, PIN4, PIN7 and then PIN1 began to be expressed with rootward polarization. All these events resulted in formation of auxin flow towards the root tip, which reached the root end and accumulated there. Auxin accumulation in columella cells mediated the changes in PINs expression there from rootward to non-polar. Gradually the auxin concentration maximum shifted to fifth cells from the root end. Finally, auxin concentration gradients were formed at both sides of the concentration maximum.

Conclusion: As a result of the numerical simulation, from the uniform auxin distribution in a tissue we got a steady-state solution, in which the auxin distribution organized itself simultaneously with the pattern of PIN proteins. The results of in silico experiments fitted well the in vivo data on auxin and PIN proteins pattern in wild-type. Consequently, our assumptions confirmed that auxin provides positional information to root tip cells.

Acknowledgements: The work was supported by RSF grant 14-14-00734.

Reference:

1. Mironova V.V. et.al. (2012). Annals of Botany, 110(2), 349-360.

TRAJECTORIES OF THE DNA KINKS IN THE SEQUENCES CONTAINING CDS REGIONS

L.A. Krasnobaeva[1, 2] *, L.V. Yakushevich[3]
[1] *Tomsk State University, Tomsk, Russia*
[2] *Siberian State Medical University, Tomsk, Russia*
[3] *Institute of Cell Biophysics RAS, Pushchino, Moscow region, Russia*
* *Corresponding author: kla1983@mail.ru*

Key words: *DNA dynamics, kink trajectories, gene energy profile, CDS regions, interferon alpha 17*

Motivation and Aim: In our previous works [1-2] to model the DNA kinks, we used a simple model based on the sine-Gordon equation with parameters that were averaged over all length of the gene sequence. However, this approach does not allow us to take into account the effect of the internal structure of the DNA sequence, and, in particular, the presence of the CDS regions, on the DNA kinks dynamics. In this paper, we just consider this problem and solve it for the gene encoding interferon alpha 17, the sequence of which consists of three regions: the coding (CDS) region (50..619) and the two regions (1..49 and 620..980) with unknown functional significance.

Methods and Algorithms: To solve the problem, we use several methods: the method of McLaughlin and Scott, the average field approximation and the block method where the parameters of the model equation are averaged separately for each of the three regions. To analyze the DNA kinks dynamics, we use the physical approach which includes the calculation of the energy profile of the sequence and the construction of the trajectory of the movement of the DNA kink in the potential with this profile.

Results: We have obtained the energy profile of the sequence of the gene coding interferon-alpha 17. It was shown that the CDS region corresponds to the region of the energy barrier. The minimum value of the kink initial velocity required to overcome the barrier was estimated. The trajectories with different initial kink velocities were constructed. The trajectories were calculated both with and without dissipative effects. It was shown that with the increasing of the initial kink velocity the trajectories became more independent on the inhomogeneity of the sequence. We suggest that the proposed approach can be applied to analyze the movement of transcription bubbles through the CDS regions of the DNA sequences.

References:

1. L.A. Krasnobaeva, L.V. Yakushevich (2015) Rotational dynamics of bases in the gene coding interferon alpha 17 (IFNA17), Journal of Bioinformatics and Computational Biology, **13:** №1, 1540002 (13 pages).
2. L.V. Yakushevich, L.A. Krasnobaeva (2016) Forced oscillations of DNA bases, Biophysics, **61:** № 2, 286-296.

GENELO – PROGRAM FOR STATISTICAL ANALYSIS OF GENES LOCATION RELATIVE TO CHROMOSOME CONTACTS REVEALED BY CHIA-PET AND HI-C

E.V. Kulakova*, A.M. Spitsina
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: kylakovaekaterina@gmail.com*

Key words: *sequencing, ChIA-PET, Hi-C, chromosome contacts, genome, CTCF sites*

Motivation and Aim: Several technologies based on chromatin immunoprecipitation (ChIP) have been developed to study the binding of transcription factors (TF) to genomic DNA including microarray (ChIP-chip), ChIP-PET and ChIP-Seq [1]. The challenge is to define whether such distal binding sites are functional, i.e. physically proximal to target gene promoters via chromosome loops attracting RNA polymerase II complex for gene transcription [2]. Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET) method fits these demands still requiring development of specialized high-throughput software for data integration [2]. The aim of the work was to develop a computer program for statistical data analysis and test it on CTCF (CCCTC-binding factor) binding sites, genes and spatial topological domains.

Methods and Algorithms: We used data on the location of CTCF binding sites clusters obtained by ChIA-PET as well as obtained experimentally by methods Hi-C, ChIA-PET [2]. Gene annotation was obtained from UCSC Genome Browser (http://genome.ucsc.edu).

Results: In result has been developed computer software for statistical analysis and visualization of results for experimental data obtained by ChIA-PET and Hi-C. The program was developed in Java language that calls modules based on R and Matlab environment using library such as Rserve and MatlabControl. The program has graphical user interface. This tool has function such as identification gene location near to domains boundary; near to binding sites of transcription factor; visualization of heatmap based on pairs of CTCF binding sites; distributions of human genes relative CTCF binding sites and a randomly generated list of such sites.

Conclusion: We considered a model the location of genes relative chromosome loops and binding sites. Genes of RefSeq are located inside the loop between the sites accounted for half of the total. It was revealed that most of the genes in the chromosomal loops are arranged individually

Availability: Software is available from the author upon request.

Acknowledgements: The work is supported by ICG budget project 0324-2015-0003.

References:

1. E.V. Ignatieva et al. (2015) Regulatory Genomics: Combined Experimental and Computational Approaches, Russian Journal of Genetics, V. 51, **4:** 334–352.
2. G. Li et al. (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation, Cell V.148. **1-2:** 84-98.

GENOTYPE DISTRIBUTION IN PATIENTS WITH CHRONIC HEPATITIS C ANALYSIS USING MULTIFACTOR DIMENSIONALITY REDUCTION METHOD

A.D. Liaudanski[1] *, M.S. Rodzkin[1], V.S. Pankratov[1], D.E. Danilau[2], I.A. Karpov[2], O.G. Davydenko[1]

[1] *Institute of Genetics and Cytology of the National Academy of Sciences of Belarus, Minsk, Belarus*

[2] *Belarusian State Medical University, Minsk, Belarus*

** Corresponding author: 666555@tut.by*

Key words: *chronic hepatitis C, gene polymorphism, MDR method*

Motivation and aim: hepatitis C virus is one of the major causes of chronic liver pathologies. Near 2% of world population suffer from chronic hepatitis C, which determine high relevance of the studying of spontaneous elimination and therapy efficiency influencing factors [1]. The prevalent type of chronic hepatitis C therapy includes pegylated interferon in combination with ribavirin, and IL28b gene polymorphism has been announced to be the most informative predictor of such therapy efficiency [2,3]. Taking in account the accuracy of the therapy efficiency prediction based on IL28b gene polymorphism (up to 50% in some population) it is still actual to find other genetic markers to enhance the precision of prognostication.

Methods and Algorithms: 100 patients with chronic hepatitis C, who had taken pegylated interferon and ribavirin therapy with different (67 unsuccessful and 33 successful) outcomes were analyzed. Polymorphism of IL28b, TNFα, CCR5 and CCL5 genes was defined using PCR or PCR-RFLP methods. Multifactor dimensionality reduction method was used to find the best model for therapy efficiency prediction (MDR ver.3.0.2 (build 2)).

Results: IL28b was found to be the best single marker, as expected (accuracy=0,61, CV consistency 10/10, p=0,0003). But the best accuracy was demonstrated by two-factor model, including IL28b and CCL5 genotypes (accuracy=0,77, CV consistency 10/10, p<0,0001). CCL5 genotypes distribution didn't vary significantly in patients with different therapy outcome, but including this in the prognosis model improve the prediction significantly. Entropy distribution demonstrated more than 1,8-fold value increase when comparing IL28b and CCL5 together with IL28b alone: IL28b entropy was 11.57%, CCL5 – 1.75%, IL28b and CCL5 interaction – 7,56% (20,88% in total).

Three- and four-factor models were found to be much less informative.

Conclusion: multifactor dimensionality reduction proved to be an effective method for genotype distribution in patients with chronic hepatitis C analysis. Two-factor model including IL28b and CCL5 genotypes was found to be the most accurate model for prediction of the chronic hepatitis C therapy efficiency. Taking in account CCL5 genotype of the patient together with IL28b could help to improve the prognosis precision of such therapy outcome.

References:

1. C.W. Shepard et al. (2005) Global epidemiology of hepatitis C virus infection, Lancet Infect. Dis., 5: 558-567.
2. D. Ge et al. (2009) Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance, Nature, 17: 399-401
3. C.N. Hayes et al. (2012) Genetics of IL28B and HCV--response to infection and treatment, Nat Rev Gastroenterol. Hepatol., 9:406-417

PREDICTION OF FUNCTIONAL EFFECTS OF REGULATORY SEQUENCE VARIATIONS

M. Malkowska[1] *, J. Zubek[2, 3], D. Plewczynski[3], L.S. Wyrwicz[1]
[1] *Maria Sklodowska-Curie Memorial Cancer Centre and Institute of Oncology, Warsaw, Poland*
[2] *Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland*
[3] *Center of New Technologies, University of Warsaw, Warsaw, Poland*
** Corresponding author: bioinfo@coi.waw.pl*

Key words: *regulatory variation, function prediction, machine learning*

Motivation and Aim: In the era of huge amounts of data from high-throughput DNA sequencing is necessary to search for more effective methods of analysis of functional regions of genomes, in particular regulatory regions, that may be crucial in the search for single-nucleotide polymorphisms potentially responsible for the development of the diseases.

Our research aims to create and evaluate an integrative machine-learning model for regulatory variants identification within promoter regions of genes. We intend to utilize various gene annotations and DNA shape features, which have recently become available and determine the significance of these new components.

Methods and Algorithms: To improve the accuracy of predicting the functional consequences of regulatory variants we are implementing machine learning algorithms using newly available features. Our positive examples dataset was constructed using regulatory mutations from the Human Gene Mutation Database1. Negative examples were derived using single nuclear variations from the 1000 Genomes Project2. The annotation data include the local DNA 3D shape, phylogenetic conservation, transcriptomic and epigenetic measurements. The model prediction performance will be evaluated across human genome to determine potential, not reported, true functional variants and compared to existing methods.

Results: Currently, we are in process of model selection and optimization. There are strong indications that chosen features will improve the accuracy and precision of functional effects of regulatory sequence variations predictions.

Conclusion: The results of the project will improve our understanding of the molecular mechanisms of disease pathogenesis and complex traits.

Availability: The machine-learning model and variant predictions will be available through GitHub repository (project in progress).

References:

1. P.D. Stenson at el. (2014) The human gene mutation database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine, Human Genetics, **133**:1–9.
2. 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes, Nature, **491(7422):**56-65.

MOLECULAR EVOLUTION AND SYSTEMATICS OF FLAT LEECHES (HIRUDINEA: GLOSSIPHONIIDAE)

N.B. Mandzyak[1] *, I.A. Kaygorodova[1, 2]
[1] *Limnological Institute SB RAS, Irkutsk, Russia*
[2] *Irkutsk State University, Irkutsk, Russia*
* *Corresponding author: mandzyak@lin.irk.ru*

Key words: *species divergence, species delimitation, genetic distances*

Motivation and Aim: Species as a basic unit of evolution have to be identifiable. Meanwhile, the presence of cryptic species poses significant challenges for the accurate assessment of biodiversity. Increasingly, molecular data are being used to evaluate species boundaries and to assist in the identification of groups such as hirudinids in which morphological characters often do not distinguish species reliably [1]. However, the development of species-specific DNA barcodes for leeches has lagged behind that of other groups.

Methods and Algorithms: To study species-specific DNA barcodes, leech samples were collected during multiple expeditions in 2003-2015. Sequences of universal for Metazoan barcoding fragment (CoxI) of 73 leech specimens from different sites of Northern Eurasia were newly generated for the present study. Bioinformatic analyses were conducted in MEGA V6.06 [2]. The evolutionary distances were computed separately for both intraspecific and congeneric comparisons using the Kimura two-parameter (K2P) model in accordance with DNA barcode techniques [3].

Results: The molecular phylogenetic analysis of barcoding fragment of the Glossiphonia representatives allowed to reveal four evolutionary branches among glossifonias inhabiting Siberian and Far Eastern freshwater bodies. On the resulting phylogenetic tree, each group of nucleotide sequences belonging to Glossiphonia sp.1, Glossiphonia sp.2, Glossiphonia sp.3 and Glossiphonia sp.4 was clustered apart of each other and separately from homologous sequences of the rest world fauna representatives. Genetic distances between Siberian leeches and phylogenetically close lineages exceed significantly the threshold of 3% [3] that could indicate their independent taxonomic position and assume that four potentially new species exist. In addition, intraspecific genetic polymorphism varies within the range of 0-0.73% and confirms this supposition. Moreover, due to impressive barcoding gap (over 5%), a long branch attraction phenomenon observed in this group of parasitic organisms [4] becomes clear.

Conclusion: Thus, the use of barcoding thresholds bode well for delineating closely related species and taxonomically understudied groups within Hirudinea.

This work was supported by the Russian Foundation for Basic Research (14-04-00345).

References:

1. A.E. Bely, D.A. Weiseblat (2006) Evolution & Development, **8**: 491-501.
2. K. Tamura et al. (2013) Molecular Biology and Evolution, **30**: 2725-2729.
3. P. Hebert et al. (2003) Proc. Royal Soc. B: Biological Sciences, **270**:313–321.
4. P. Martin et al. (2000) Mol. Phyl. Evol., **15**: 355-368.

PREDICTION OF AMYLOID AGREGATION INHIBITORS AS A THERAPEUTIC ALTERNATIVE FOR TREATMENT OF ALZHEIMER'S DISEASE. A MOLECULAR DYNAMICS STUDY

A.B. Martínez[1, 3] *, Y.M. Álvarez Ginarte[2], M.S. Carrazana[1], E.M. Frías[3], Ch.R. Tanty[1]
[1] *Cuban Neuroscience Center, Havana, Cuba*
[2] *Chemistry Faculty, Havana University, Havana, Cuba*
[3] *Center for Molecular Immunology, Havana, Cuba*
* *Corresponding author: albertobm@cim.sld.cu*

Motivation and Aim. Alzheimer's disease (AD) is a neurodegenerative disease characterized by the accumulation of amyloid plaques in the brain. AD is the leading cause of dementia in the aging population and an effective treatment is still unavailable. Increasing evidence suggests that the aggregation of the small peptide β-amyloid (Aβ) plays an important role in the development of AD. Understanding the interactions of Aβ with aggregation inhibitors on an atomic level is essential for the rational development of diagnostic and therapeutic tools. Several naphthalene derivative compounds (NDCs) have been previously identified as inhibitors of the Aβ aggregation and some of them were studied by QSAR.

Methods and Algorithms. In our laboratory novel NDCs were synthetized. Docking techniques were used to find and describe potential sites of interaction of NDCs with Aβ.

Results. According to the calculations, these compounds interact preferentially with S8-G9 E11-H13 Q15-L17 F19 amino acids. This region has been reported as crucial for the fibril formation. A virtual screening was done and the interaction energies between NDCs and Aβ were in the range of -7.2 and -5.7 Kcal/mol. Molecular dynamics calculations were done to assess the stability of the complex NDCs-Aβ. The simulations showed that hydrophobic and some polar interactions stabilize the complex formation. Finally, the ΔG energies were calculated to evaluate the affinity of the predicted compounds to Aβ. All these studies suggest that NDCs could be used as potential inhibitors of amyloid aggregation.

LATER ANTIBODIES WITH OXIDOREDUCTASE ACTIVITY OF PATIENTS WITH SCHIZOPHRENIA AND MULTIPLE SCLEROSIS

I.A. Mednova[1] *, L.P. Smirnova[1], L.E. Sinianskii[2], E.A. Ermakov[1], A.V. Semke[1], S.A. Ivanova[1]

[1] *Mental Health Research Institute, Tomsk, Russia*
[2] *Siberian State Medical University, Tomsk, Russia*
* *Corresponding author: i.mednova@yandex.ru*

Key words: *abzymes, oxidoreductase activity, IgG, multiple sclerosis, schizophrenia*

Motivation and Aim: According to generally accepted notions primary function of immunoglobulins is binding of foreign antigens , but the new one was opened relatively recently – catalytic function of antibodies. These catalytically active antibodies were named abzymes. The abzymes with proteolytic, DNA and RNA hydrolysis and other activities were found. The free radical mechanism of injury is relevant to the pathogenesis of multiple sclerosis (MS) and schizophrenia. This stipulates an actuality of determining of oxidoreductase activity of IgG along these pathologies.

Methods and Algorithms: The study included 20 patients with definite diagnosis of MS relapsing-remitting subtype (McDonald, 2010), 20 patients with a diagnosis of schizophrenia (ICD-10: Paranoid schizophrenia), and 22 healthy persons. The allocation of IgG was carried out with help of affinity chromatography on columns Protein-G-Sepharose. The homogeneity of the substances was proved by 1D-SDS PADGE. Determination of catalase (CAT), superoxidedismutase (SOD), peroxidase (PER) activity of IgG was performed by spectrophotometry. Statistical treatment was applied in the program «Statistica 8.0». Marginal level of significance in verification of statistical hypothesis was assumed to be 0.05.

Results: IgG in MS and schizophrenia is shown the first time to have a CAT, SOD, PER activity. On the analysis of the affinity of sorbent, the homogeneity of the selected IgG and gel filtration under conditions of pH shock the studied activities are proved to be a private property AB. MS patients were found to have IgG with maximum PER activity that is 4 times superior to such activity in healthy individuals. Patients with schizophrenia have shown the double increase of the peroxidase activity in comparison with healthy people. MS patients showed SOD activity 4.5 times, and schizophrenia ones 2.5 times higher than healthy individuals. 50mM triethylenetetramine ferment inhibitor of superoxide dismutase inhibited this activity in 100% cases of IgG patients and healthy people. Identified CAT activity in schizophrenic patients is 7 times, and in patients with RRMS is 3.5 times higher than the activity in the healthy group. Enzyme catalase inhibitor 3-amino-1,2,4-triazole also inhibited catalase activity IgG.

Conclusion: We can assume that in the conditions of oppression AOC activity in MS patients , abzymes partially take over the function of protecting the body from patients with generalized oxidative stress. Catalytic antibodies with oxidoreductase activity involved in the removal of AKM in the intercellular space and blood .Specific inhibitors of SOD and catalase enzymes similar to inhibit the catalytic activity of IgG, which suggests a similar mechanism of catalysis at the abzyme.

SEARCH FOR THE CONTEXT FEATURES IN THE REGULATORY SEQUENCES OF *ARABIDOPSIS THALIANA* GENES ASSOCIATED WITH THE HOURGLASS PATTERN

A.M. Mukhin*
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: acamukhin@gmail.com*

Motivation: We can classify genes by their emergence time in evolution comparing the amino acid sequences of paralogs in different species. For example, young genes are species or genus specific, but old genes are present in kingdom or phylum. Indexes TAI and TDI were created to estimate the evolution age of the transcriptome. Index TAI is based on the evolution age of each gene into its phylostratum (if we know, which vertex of taxon tree are associated with our gene, we can substitute depth of this vertex in formula). TDI is based on epy ratio of asynonymous and synonymous substitutions between closely related species, for example, Arabidopsis thaliana and Arabidopsis lyrata. It has been shown that Arabidopsis thaliana on early (before quadrant) and late (bent cotyledon and mature embryo) stages of embryogenesis expresses young genes, but between these stages older genes are expressed more. That profile of indexes was linked to so called hourglass pattern. As molecular-genetics mechanisms that shapes hourglass pattern in development are unknown, we suggested that it can be specific transcription regulation for young and old genes. So that the footprints of this regulation might be found in the promoter regions as conservative motifs or other features recognizing by RNA polymerase transcription machinery selectively in young and old genes. The aim of this work was the search for new qualitative and quantitative characteristics of genes which have an hourglass pattern in embryo development.

Materials and methods: We used genome assembly Arabidopsis thaliana from database TAIR, its genome in FASTA format from Ensemble Plants and microarray data for Arabidopsis thaliana embryo on different developmental stages from [1]. We proposed TgcI index to calculate the contribution of GC-context in an associated sequence (see below), analogous to TAI and TDI. Also we developed a program in Java to calculate the index. Program gets specific sequences with considering (upstream regions of different lengths, 3' and 5' UTR, exons, introns and CDS). Then it calculates GC-context in all associated with the gene sequences, find for the expression level of the gene and calculates TgcI. On the last stage, the program draws a graph for the index against the developmental stages.

Results and Conclusion: We developed the program and the method for evaluation the context features in the regulatory sequences associated with a specific pattern of gene expression changes in development of Arabidopsis thaliana embryo. The method was tested on the example GC context feature, which can play a role in the gene activity. The index TgcI values for GC context in regulatory and coding regions of the genes appeared to be the same for all embryo stages. That is GC context of the sequences do not play a role in formation of hourglass pattern. The same approach will be implemented for evaluation of contribution for position-weight matrix for different transcription factors of plants.

References:

1. http://figshare.com/articles/Supplementary_table_S3/1244948

ORTHOSCAPE: A CYTOSCAPE PLUGIN FOR EVOLUTIONARY ANALYSIS OF GENE NETWORKS

Z.S. Mustafin[1] *, D.A. Afonnikov[1, 2], K.V. Gunbin[1], Yu.G. Matushkin[1, 2], S.A. Lashin[1, 2]
1 Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
2 Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: mustafinzs@bionet.nsc.ru*

Key words: *Cytoscape plugin; ortholog; paralog; metabolic pathway; gene regulatory network; evolution; phylostratigraphy; evolution*

Motivation and Aim: There are a number of software intended to visualization and analysis of biological networks. Among them, Cytoscape (http://cytoscape.org/) is one of the most comprehensive tools. There are a lot of plugins extending the base functionality of Cytoscape providing analysis of protein-protein interaction networks, gene regulatory and gene co-expression networks, metabolic, signaling, neuronal networks as well as ecological-type networks including food webs, communities network etc. Nevertheless, there is still lack of evolution-oriented plugins: just three plugins tagged 'network evolution' in Cytoscape official app store and in literature. We have developed a new Cytoscape plugin Orthoscape aimed to perform evolutionary analysis of gene networks and visualize its results.

Methods and Algorithms: We used KEGG (http://www.kegg.jp/) Pathway database to get gene networks, KEGG Orthology to get lists or homologs with identity and SW Score values, KEGG Genes to get protein domains, also nucleotide and amino acid sequences and KEGG Organism to get taxonomic information. The domain composition and sequence similarity approaches used to discriminate between paralogs and orthologs. Last common taxonomic level overlapped with inferred orthology status were used as a phylostratigraphic data about (sub)network origin/divergence. We also allow users to create new networks from gene sets using GeneMANIA or CyPath2 (for BioPax format support) plugins. In this case, we used KEGG sequence similarity search to associate proteins from new networks with data annotated in KEGG. Divergence index is calculated by KaKs calculator using pairwise sequence comparisons for the taxa under analysis. The average value of this index allow us to discriminate between diversifying and stabilizing selection in orthologous groups.

Results: Cytoscape plugin Orthoscape has been developed. The plugin allows users to analyse evolutionary information in the gene sets and networks: (1) the orthology relationships between genes; (2) the evolutionary origin of gene network components; (3) the evolutionary regime (diversifying or stabilizing, negative or positive selection) of orthologous groups in general and/or branch-oriented mode. The distinctive feature of Orthoscape is the ability to control all data analysis steps via user-friendly interface.

Conclusion: Orthoscape allows user to analyze gene networks or separated gene sets to know the evolutionary origin of genes (sub)networks and selective pressure. It also provides convenient visualization and data manipulation abilities on each data analysis step.

Availability: Upon the requests to the authors.

Acknowledgements: The study is supported by the RSF 14-24-00123 grant.

IDENTIFICATION OF STURGEON SPECIES WITH MTDNA AND MICROSATELLITE MARKERS IN BELARUS

A.Yu. Nosova, O.I. Zaitseva, M.V. Bogdanova
Institute of Genetics and Cytology NAS of Belarus, Minsk, Belarus
** Corresponding author: A.Nosova@igc.by*

Key words: *Genotyping, sturgeon, species identification, hybrids*

Motivation and Aim: Sturgeons are a unique ancient group of fishes and are commercially important. Since the natural populations have dramatically decreased in size, sturgeon domestication and artificial rearing become widespread. Usually, there is no sufficient work on the preservation of the purity of livestock in aquaculture. However, the cost of products is significantly different depending on the species. Identification of the species is complicated by the morphological similarity of sturgeons of commercial size, or impossible for fish products and caviar. This study was conducted to determine the efficiency of using mtDNA and microsatellite markers for the identification of sturgeon species in Belarus.

Methods and materials: We obtained fins fragments of 32 individuals including 7 Siberian sturgeons, 5 Russian sturgeons, 5 starlets, 5 belugas and 10 interspecific hybrids from local fish farm. For species identification variations of the mtDNA control region (D-loop) were studied by PCR analysis [1]. Sturgeons were genotyped by fragment analysis of a set of five microsatellite loci (Afug41, Afug51, An20, AoxD161, AoxD165) [2].

Results: The analysis of mtDNA variations and microsatellite alleles confirmed that 22 individuals of pure species and 10 of interspecific hybrids were provided. Hybrid individuals were represented by 4 Besters (beluga×starlet), 4 Steroses (starlet×Siberian sturgeon). Presumably, two samples differing from the declared were reciprocal hybrids starlet×beluga and Siberian sturgeon×starlet.

Conclusion: The methods of DNA identification can be conventionally divided into two groups addressing nuclear genetic markers and mitochondrial DNA (mtDNA). However, the maternal inheritance of mtDNA limits the use of this methodology for the hybrid individuals. The differences in STR allele frequencies among different species enable to identify individuals of hybrid origin. However, it should be taken into account that STR loci may vary from population to population. In our case, not all species-specific loci according to Barmintseva [2] were presented in studied animals.

Availability: Thus, the use of PCR for identification is appropriate for maintaining the genetic purity of livestock and producers in aquaculture, and for detection of falsified caviar and other sturgeon products. An analysis of the polymorphism of the five microsatellite loci should be used in the genetic pasportisation of aquaculture surgeon stocks and the species verification of the sturgeon products.

References:

1. N. S. Mugue et al. (2008) Polymorphism of the Mitochondrial DNA Control Region in Eight Sturgeon Species and Development of a System for DNA-Based Species Identification, Russian Journal of Genetics, 44: 793–798.
2. A. E. Barmintseva, N. S. Mugue. (2013) The Use of Microsatellite Loci for Identification of Sturgeon Species (Acipenseridae) and Hybrid Forms, Russian Journal of Genetics, 49: 950–961.

MEMBRANE-ASSOCIATED KINASE REGULATORS OF MAKR FAMILY GENES IN *ARABIDOPSIS THALIANA* L.

D.D. Novikova*, N.A. Omelyanchuk, V.V. Mironova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: novikovadd@bionet.nsc.ru*

Key words: *kinase receptor regulators, A. thaliana, kinase receptors, hormone signaling*

Motivation and Aim. Tyrosine kinase receptors and membrane-associated kinase regulators are involved in most signaling pathways, including those for plant hormone signal transduction. Recently, a family of membrane-associated kinase regulators (MAKR) was described in Arabidopsis thaliana [1]. MAKR family proteins have distant homology with the basic domain of BRI1 kinase inhibitor 1 (BKI1) which controls activity of tyrosine kinase receptor BRI1 in brassinosteroid signaling pathway. Here, we study phylogeny and systematize the publically available information about protein structure, expression and regulation of this poorly annotated gene family.

Methods and Algorithms. First, we analyzed the MAKRs protein sequences to reveal their features and signatures, which might imply their potential functions by PhosphoSVM and CBS Prediction servers. Further to investigate phylogenetic background and evolvement of MAKRs we have searched for their homologs throughout the plant kingdom using BLAST and PLAZA on-line services and constructed phylogenetic trees in MEGA 6.06 program. We analyzed data on their mRNA patterns, which is presented in publicly available microarray experiments (eFP browser). To study the hormonal regulation of MAKRs in depth we analyzed their upstream [-1000; +1] regions for the presence of potential cis-regulatory sites using online tools AtCOECIS, TRANSFAC programs and CIS-BP database.

Results and conclusions. MAKR family genes probably evolved together with land plants and as a result of duplication series all members of the family appeared. These proteins are widely spread among terrestrial plants. The latter uncovers their potentially important role in plant kingdom in general and in Arabidopsis in particular. In addition, some conservation was discovered among orthologs of all MAKRs, but none among paralogs, except C-terminus. According to protein modification prediction, all of MAKRs have plenty of modification sites. MAKRs are expressed tissue specifically and response to different hormones, they possess hormone responsive motifs corresponding to the sensitivity to the one. These findings suggest that MAKR family genes might play an important role in plant development regulation via signaling pathways. Obtained results may provide new ideas in searching for new regulators of growth and herbicides.

References.

1. Jaillais Y. et al. Tyrosine phosphorylation controls brassinosteroid receptor activation by triggering membrane release of its kinase inhibitor //Genes & development. – 2011. – Т. 25. – №. 3. – С. 232-237.

CLUSTER ANALYSIS OF PHYSICAL CHARACTERISTICS IN *E. COLI* PROMOTERS

M.A. Orlov*, A.A. Ryasik, E.A. Temlyakova, A.A. Sorokin
Institute of Cell Biophysics RAS, Pushchino, Russia
** Corresponding author: orlovmikhailanat@gmail.com*

Key words: *DNA physical properties, machine learning, bacterial promoters, molecular recognition, prokaryote transcription*

Motivation and Aim: Development of high-throughput sequencing methods provides large-scale genomic data that cannot be processed manually. Thus, a need for automated analysis and annotation methods has become evident. Here we propose using the method of non-supervised learning algorithms in the attempt for prediction of the promoters location by analyzing profiles of physical and chemical properties of DNA.

Methods and Algorithms: We suggest dividing all physical properties into three main classes: global smooth (i.e. is determined by a large sequence's part and manifests on a large sequence as well), global burst (i.e. is determined globally but manifests locally), and locally determined ones. According to this electrostatic potential (EP) profile, stress-induced duplex destabilization (SIDD), and dynamic properties of DNA open states (calculated using the coarse-grained model of DNA open states) were used in the analysis to represent 3 classes listed above. Corresponding profiles were calculated for a complete set of experimentally found promoters of E. coli (primary structures were obtained from RegulonDB). Following clusterization of the profiles was performed with Ward method and its results consistency were assessed using consensus clustering technique.

Results: By dividing SIDD, EP, and dynamic properties profiles sets into 4, 4 and 3 clusters accordingly we have obtained the most stable partitions. The dissimilarities for the clusters were confirmed by comparison of their partitions concordance and adjusted Rand indices, while the corresponding resemblance of the dendrograms was measured using cophenetic coefficients values. All the properties were shown to be independent and poorly correlated with each other.

Conclusion: It was shown that correlation between the studied physical and chemical properties of DNA is insignificant. Therefore they can be used cooperatively (along with textual characteristics) for bacterial promoters recognition. We also suggest that this approach could enable researchers to predict promoters location in de novo sequenced prokaryotic genomes.

Acknowledgements: This work was supported by RFBR grants mol_a 16-37-00303 and r_centr_a 14-44-03679.

IN SILICO MOUSE CHROMOCENTERS CONTENT

D.I. Ostromyshenskii*, A.S. Komissarov, I.S Kuznetsova, O.I. Podgornaya
Institute of Cytology RAS, St. Petersburg, Russia
** Corresponding author: necroforus@gmail.com*

Motivation and Aim: Chromocenters are interphase nuclei landmark structures of constitutive heterochromatin. Heterochromatin enriched in tandem repeat (TR). There is progress in recent years in revealing chromocenters protein content, though it is not clear what DNA underlay constitutive heterochromatin apart from TR. The aim of the current work was to find out what DNA sequences involved in chromocenters formation.

Methods and Algorithms: Chromocenters was biochemically isolated and chromocenters' DNA library prepared by Nextera kit. The library were sequenced using Illumina MiSeq with paired-end reads of 35 bp.

Results: Bioinformatics comparison of chromocenters MiSeq (chcMiSeq) with whole genome sequencing on Illumina HiSeq (gnHiSeq) revealed NNN content:

Family\Source	chcMiSeq	gnHiSeq
Tandem repeats (TR)	70%	10%
LINE	7%	4%
ERV	1.5%	1%

Among chcMiSeq TR the most abundant is MaSat (61%) and MiSat (4%). The rest of TR (5%) represents the TRs families previously described [1]. The rest 20% of chcMiSeq dataset is mostly unannotated sequences, but some of them have been identified when part of chcMi-Seq dataset have been assembled into contigs by IDBA_UD program. In the contigs assembled there are many fragments of heterochromatic Y chromosome, rRNA and six other pseudo-genes and ncRNA gene. Full scale gene sfi1 is found in contigs and it is localized to the chromosome 11 pericentromeric region. The ERV based fragments from chcMiSeq assembled contigs went to all the possible locations being mapped to different ERV consensuses from Repbase. This indicate that the whole ERV could be built in TR arrays. In contrast, there is very few full-length LINEs in chcMiSeq or in its' part assembled. Most of the LINE fragments collected in the same 2 kb region at the end of the 2nd ORF and its' flanking region. The same region of LINE is the origin for the L1-based TR [1]. Full-length LINEs enrich facultative heterochromatin, but it is nearly absent in constitutive heterochromatin [2, 3].

Conclusion: The sequencing of chromocenters' DNA (chcMiSeq) reveal full length ERVs and precise LINE' fragment of 2 kb as the substantial mouse constitutive heterochromatin components together with TR of different families.

Acknowledgements: This work was supported by the granting program 'Molecular and cell biology' of the Russian Academy of Sciences.

References:

1. A.S. Komissarov, E.V. Gavrilova, S. Demin et al. (2011). Tandemly repeated DNA families in the mouse genome. BMC genomics, 12(1):531-552.
2. A.T. Chinwalla et al (2002). Initial sequencing and comparative analysis of the mouse genome. Nature, 420(6915): 520-562.
3. I. Solovei, M. Kreysing, C. Lanctot et al. (2009). Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. Cell 137: 356-368.

THE GENOME WIDE ANALYSIS OF THE LARGE TANDEM REPEATS IN THE CLOSELY RELATED GENOMES

D.I. Ostromyshenskii*, O.I. Podgornaya
Institute of Cytology RAS, St. Petersburg, Russia
** Corresponding author: necroforus@gmail.com*

Motivation and Aim: Large tandemly repeated sequences (TR, or satellite DNA) are necessary part of higher eukaryotes genomes and can comprise up to tens percent of the genomes. Much of TRs' functional nature in any genome remains enigmatic because there are only few tools available for dissecting and elucidating the TR functions.

Material and methods: The modified pipeline of the one used previously in our Lab [1] applied to the several databases. Four mammalian genera was used: (1) mice Mus: *M. musculus*, *M. caroli* (unassembled genome); (2) guinea pigs Cavia: *C. porcellus*, C. apperea; (3) bats Myotis: *M. brandii*, *M. davidii*, *M. lucifugus*; (4) cows Bos: *B. taurus*, *B. mutus*, *B. indicus*.

Results: We tried to find all the 62 *M. musculus* TR families [1] in raw reads of *M. caroli* genome (Caroli Genome Project, PRJEB2188). There are only few TR of *M. musculus* in *M. caroli* genome. *M. musculus* major satellite (MaSat or GSAT-MM) occupied nearly 0,7% of *M. caroli* genome, while in *M. musculus* genome - ~ 11 %. In *M. caroli* genome we found 5 other *M. musculus*'s TR's families.

Genus *Cavia*. *C. porcellus* genome possesses 25 TR and *C. apperea* – only 10 TR. 9 out of 10 C. apperea TR's family exist also in *C. porcellus* genome except the major TR for this species – Capp-1518. In *C. porcellus* genome there are two major TR – Cpor-783 is absent in the 2nd genome and Cpor-123 exists in *C. apperea* genome as the minor one. Genus *Myotis*. There is no any TR of Myotis in Repbase, but 133 TR's families are found in *M. brandtii* genome, 105 - in *M. davidii* genome and 26 - in *M. lucifugus* genome. Only 5 TR families exist in three genome but most of TR families are species-specific. Major TR for *M. davidii* and *M. lucifugus* is common in sequence though differ in monomer length, but the same TR is minor one in *M. brandtii*. The major for *M. brandtii* is not identified in both other genomes at all.

Genus *Bos*. There are three TR known for Bos in Repbase and all of them are found in all Bos assemblies. Still the major TR in all Bos assemblies differ: in *B. taurus* genome BTSAT4/BTSAT5 is a major TR while BTSAT6 major TR family in *B. indicus* genome. It is visible that most of the top TR families in genus Bos exist only in two genomes or even in one, i.e. is species-specific.

Conclusion: The most exhausting analysis of major TR (one for each species) of ~300 animals and plants display no readily apparent conserved characteristics [2]. We compared the TR sets. Our data evidenced that there are species-specific top TR, which are absent in genome of closely related species. In all genera examined major TRs are species-specific and hardly exist in other species of genera even as a minor ones.

Acknowledgements: This work was supported by the granting program 'Molecular and cell biology' of the Russian Academy of Sciences.

References:

1. A.S. Komissarov, E.V. Gavrilova, S. Demin et al. (2011). Tandemly repeated DNA families in the mouse genome. BMC genomics, 12(1):531-552.
2. D.P. Melters, K.R. Bradnam, H.A. Young, et al. (2013). Comparative analysis of tandem repeats from hundreds of species

SEARCH FOR FUNCTIONAL NF-KB BINDING SITES VIA META-ANALYSIS OF NGS EXPERIMENTS IN HUMAN CELL LINES

N. Panyushev[1,2], E. Lomert[1], D. Tentler[1], A. Predeus[2*]
[1] *Institute of Cytology RAS, St. Petersburg, Russia*
[2] *Bioinformatics Institute, St. Petersburg, Russia*
* *Corresponding author: predeus@gmail.com*

Key words: *NF-kB, ChIP-seq, Binding sites, TNF-a, Histone modifications*

Motivation and Aim: The NF-kB family of transcription factors plays the critical role in inflammation, immunity, cell proliferation and metastasis. NF-kB dimers recognize 9-11 nucleotide sequences called kB sites. There are about 13600 human genes with kB sites in promoter region; however not all of them are transcriptionally active. Thus search for functional kB sites could assist in understanding the basic principles that underlie NF-kB regulation.

Methods and Algorithms: In order to indentify the regulatory mechanisms and NF-kB cofactors on a genomic level, we have gathered the following three sets of data: 1) RNA-seq transcriptional profiling experiments of cells before and after NF-kB activation; 2) p65 ChIP-seq experiments before and after stimulation; 3) ChIP-seq profiles of histone marks that allow comprehensive chromatin role annotation. The first two sets of experiments allowed us to identify enhancer and promoter elements, as well as genes that are activated upon TNF-alpha. The latter combination of the histone marks has allowed us to identify latent regulatory elements that were not marked in the unstimulated cells but only became activated upon NF-kB induction.

Results: The analysis revealed a consistent pattern of the regulation of NF-kB-dependent transcription, the correlations of histone modifications and localization and consensus sequence of kB-sites. Regulatory elements (promoters and enhancers) were identified genes via the chromatin context information. Obtained data can be used for experimental validation of NF-kb -dependent regulation mechanisms by the binding kB-sites in the regulatory regions.

This work is supported by the Russian Science Fund grant (14-50- 00068).

IDENTIFICATION OF RECOMBINATION SITES IN THE GENOMES OF THE EUROPEAN SUBTYPE OF TICK BORNE ENCEPHALITIS VIRUS

A.I. Paramonov[1] *, Yu.P. Dzhioev[1, 2], I.V. Kozlova[1]
[1] *Cientific Center ofFamily Health and Human Reproduction Problems, Irkutsk, Russia*
[2] *Irkutsk State Medical University, Institute of Biomedical Technology, Irkutsk, Russia*
** Corresponding author: paramonov_a.i@mail.ru*

Key words: *recombination, tick borne encephalitis virus, TBEV*

Motivation and Aim: Tick borne encephalitis virus (TBEV) of the Flaviviridae family is the causative agent in human neuroinfections which often cause disability and death. There are three main subtypes - Far Eastern, Siberian and European (Western). Each subtype has its own habitat; European subtype TBEV has extensive undivided area in Western Europe and mosaic in Asia. Representatives of this serotype circulate in ecosystems significantly different by composition of biocenoses, vectors and hosts. Special attention should be paid to the genetic variation of this subtype, and genetic recombination is one of its leading factors [1, 2, 3]. The aim of this study was to detect potential recombination sites in the genomic sequences of the isolates of European subtype TBEV.

Methods and Algorithms: Genomes of 26 strains of European subtype TBEV available in the GenBank data base, as well as 8 strains we sequenced, were used in this work. The phylogenetic test for the presence of recombination was obtained using Splits Tree v4.1, by Neighbor-net method. Statistical test was carried out using the Phi Test for Recombinations method of Splits Tree software system [4]. Positioning of recombination sites was performed using software methods implemented in the programs package RDP v.4.46 [5].

Results: Phylogenetic network constructed by Neighbor-net has multiple splits indicating the possibility of reticulated evolution and accordingly recombination events at least in some strains. Phi Test for Recombinations showed the presence of recombination in this sequence set with p = 0,008. RDP v4.46 software package found recombination points in strains Joutseno and Absettarov. Interestingly, Joutseno strain contained two independent recombination points.

Acknowledgements: The work was supported by Russian Science Foundation grant №14-15-00615.

References:

1. Gritsun TS, Nuttall PA, Gould EA. Tick-borne flaviviruses //Adv Virus Res. 2003;61:317
2. I.V. Kozlova et al. Genetic and Biological Properties of Original TBEV Strains Group Circulating in Eastern Siberia //Monograph.«Encephalitis» Edited by S.Tkachev, 283 pages, Publisher: InTech, Published: January 09, 2013. P. 95-112.
3. Růžek D, Dobler G, Donoso Mantke O. Tick-borne encephalitis: pathogenesis and clinical implications // Travel Med Infect Dis. 2010 Jul;8(4):223-32.
4. Huson DH, Bryant D: Application of phylogenetic networks in evolutionary studies // Mol BiolEvol. 2006, 23.-P. 254-267.
5. Martin DP, Murrell B, Golden M, Khoosal A, & Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evolution 1: vev003 doi: 10.1093/ve/vev003

PROTEOLYTIC ACTIVITY OF IMMUNOGLOBULINS G OF PATIENTS WITH SCHIZOPHRENIA

D.A. Parshukova[1] *, L.P. Smirnova[1], V.N. Buneva[2], A.V. Semke[1], S.A. Ivanova[1]
[1] *Mental Health Research Institute, Tomsk, Russia*
[2] *Institute of Chemical Biology and Fundamental Medicine SB RAS, Novosibirsk, Russia*
** Corresponding author: Susl2008@yandex.ru*

Key words: *Schizophrenia, abzymes, myelin basic protein*

Motivation and Aim: Involvement of the immune system in the pathogenesis of schizophrenia was shown in many articles. There is the observed dysregulation between the nervous and immune systems, the causes of which may be changes in brain structure and dysfunction of immune cells. Abzymes are found under pathological conditions, although their pathogenic or beneficial role remains unclear.

Materials and methods: Here we present evidence demonstrating that highly purified IgGs from patient with schizophrenia catalyze specifically hydrolysis of human myelin basic protein (hMBP). IgG fraction were obtained by affinity chromatography of serum proteins on protein G-Sepharose under conditions that remove nonspecifically bound proteins. Number of strict criteria was testing to assign the detected catalytic activity to the antibodies: electrophoretic homogeneity of Abs, gel exclusion chromatography of Abs at conditions of dissociation of immune complexes (pH shock analysis).

Results: It was shown that Abs of schizophrenic patients specifically hydrolyzed protein substrate such as: human serum albumin, collagen, myelin basic protein (MBP) and its 21- and 25-mer oligopeptides (OP21, OP25). IgG of schizophrenic patient demonstrate a high rate of proteolytic activity (73,07%) towards hMBP. It is 5 times higher as compared to indices of proteolytic activity of healthy person's IgG. Inhibition of proteolysis and products of hydrolysis 21- and 25-mer MBP oligopeptides (OP21, OP25) were analyzed by thin layer chromatography. Specific inhibitor of serine (PMSF) and metal-dependent (EDTA) proteases significantly inhibit activity of proteolytic abzymes. High rate of proteolitic activity of Abs patient with schizophrenia towards BMP and its oligopeptides is probably associated with elimination of defective BMP fragment in serum. The estimated role of Abs with catalytic activities is lowering of autointoxication by products of degradation of BMP.

Supported by Grant of RSF № 14-15-00480 «The search for biomarkers of socially significant endogenous mental disorders» 2014-2016.

PQ: A NEW PROGRAM FOR PHYLOGENETIC RECONSTRUCTION

D. Penzar[1] *, M.S. Krivozubov[2], S.A. Spirin[3]
[1] *Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia*
[2] *Gamaleya Center of Epidemiology and Microbiology, Moscow, Russia*
[3] *Belozersky Institute of Physico-Chemical Biology of Moscow State University, and Institute of System Studies, Moscow, Russia*
* *Corresponding author: dmitry_penzar_1996@mail.ru*

Key words: *Molecular phylogeny, algorithm, software, web interface*

Motivation and Aim. Programs for protein phylogenetic reconstruction are widely used by evolutionists and molecular biologists. A recent study [1] show that distance methods of phylogenetic reconstruction (such as minimum evolution) outperforms symbol-oriented ones (such as maximum likelihood). The aim of the work is to elaborate and implement a new symbol-oriented method that could compete with distance methods.

Methods and Algorithms: With a given protein multiple sequence alignment and a given tree on sequences of the alignment, the PQ (position-quartets) score is calculated by the following formulas: $W = \Sigma p \Sigma q \, Wpq$, where p runs over all positions of the alignment and q over all quartets (forths) of sequences of the alignment; $Wpq = \max(S(aip, ajp) - M, 0) + \max(S(alp,akp) - M, 0)$, where $M = \max(S(aip, akp), S(aip, alp), S(ajp, akp), S(ajp, alp))$, aip is the letter of i-th sequence in p-th position of the alignment, S(a, b) is a scoring matrix (e.g., BLOSUM62), and i, j, k, l are sequences of the quartet q such that {i, j} is separated from {k, l} by at least one branch of the tree. The search for the tree with maximal W is performed with standard protocols: tree growing, nearest-neighbor interchange, and subtree pruning and regrafting.

Results: The program PQ is implemented and tested on a number of sets of alignments of protein evolutionary domains. PQ outperforms maximum likelihood and maximum parsimony programs for all tested sets. For some sets PQ outperforms distance-oriented programs, too. We also investigate effectiveness of different search strategies.

Conclusion: The new program PQ can be a good alternative to phylogenetic programs based on maximum likelihood and distance methods, especially for small (<30 sequences) protein alignments.

Availability: The program code in plain C is available on request from the authors. The web interface is available without registration at http://mouse.belozersky.msu.ru/tools/pq.

Acknowledgements: the work is supported by the Russian Scientific Foundation, grant No. 16-14-10319.

References:

1. G. Gonnet (2012) Surprising results on phylogenetic tree building methods based on molecular sequences, BMC Computational Biology, 13: 148.

SEARCH OF GENETIC SEQUENCES OF POTATOES IN DATABASES

A.I. Perfileva
Siberian Institute of Plant Physiology and Biochemistry SB RAS, Irkutsk, Russia
** Corresponding author: alla.light@mail.ru*

Key words: *databases, potatoes, NCBI, Spud DB*

Motivation and Aim: The potato genome is interesting object of research because potatoes is impotent for agriculture. However, now there is an insufficient number of the works devoted to research of a genome of potatoes. The problem consists in complexity of definition of the necessary gene of potatoes because of a large number of databases, and also ambiguity of information provided in them, sequence considerably differ. The most widespread databases, containing gene sequences of potatoes are: GenBank (http://www.ncbi.nlm.nih.gov/genbank/), UniProt (http://www.uniprot.org/), Spud DB (http://solanaceae.plantbiology.msu.edu/). Most of scientists-biologists actively uses the BLAST NCBI base, however, it contains information on a wide range of biological objects of various kingdoms therefore not all genes of potatoes are present at her. Some works refer to the UniProt base, however, of information in her less, than in BLAST NCBI.

Methods and Algorithms: Most of researchers in the works refer to the Spud DB base (http://solanaceae.plantbiology.msu.edu/). The real base is characterized by what contains data not only on potatoes genes, and also on tomato genes. All base consists of these two research groups which have carried out a genome sequencing. One group of researchers has carried out a sequencing of a genome of a tomato and potatoes (designation of a gene begins "S. tuberosum group Phureja DM1-3-ITAG"), and other group potatoes sequencing (designation of a gene begins "S. tuberosum group Phureja DM1-3-PGSC"). The base contains information on a gene, his transcript, proteinaceous sequence, and also allows to carry out search of homologous genes.

Results: Potatoes belong to family Solanacea therefore for search of genes of potatoes it is possible to use homologous genes of a tomato, and also other organisms. More truly to pick up the necessary sequence of a gene it is possible to bring sequence of the interesting gene in a search box of the Spud DB base and to find a potato homolog (with the indication of a protsentn of identity). Further the found sequence can be checked in Blast NCBI GenBank (http://www.ncbi.nlm.nih.gov/genbank/), at her existence in this base.

Conclusion: Distinctions in the found genetic povtornost can be connected with high-quality features, various techniques of research groups. However the main part of a gene at various sequencings is identical that allows to carry out work with these sequences.

Acknowledgement: The work has been conducted with the support of Russian Foundation for Basic Research, Grant mol_a No. 16-34-00806.

MIRNA BINDING SITES IN THE MRNA OF HUMAN TITIN GENE

I.V. Pinsky[1] *, A.T. Ivashchenko[1], S.B. Labeit[2]
[1] *Al-Farabi Kazakh National University, Almaty, Kazakhstan*
[2] *Institute of Integrative Pathophysiplogy, Mannheim, Germany*
* *Corresponding author: ilya.pinskyi@mail.ru*

Key words: *titin, sarcomeres, isoform, miRNAs, mRNAs, CDS*

Motivation and Aim: Titin, the human muscle protein, is the largest in the nature (the longest isoform IC contains 35991 amino acid residues) and plays an important role in providing the elasticity and structural integrity of sarcomeres. Interruption of its synthesis leads to the development of a number of serious cardiovascular diseases such as heart failure, cardiomyopathy, ischemic heart disease, and myocardial infarction. Titin gene expression is controlled by miRNAs (microRNAs) that bind with the mRNAs of the gene and block their translation. Therefore, it is important to determine which miRNAs most strongly regulate the synthesis of human titin and what exons of the gene contain the binding sites because different exons of the gene are expressed in different types of muscle tissue at different stages of the human body development.

Methods and Algorithms: The binding of 2563 human microRNAs with mRNA of human titin IC isoform, including all 363 exons of the human titin gene was determined using program miRTarget. The human miRNA sequences were taken from miRBase site (www.mirbase.org/), and the mRNA sequence of the titin gene was taken from Genbank (www.ncbi.nlm.nih.gov/genbank). The degree of binding (ΔG/ΔGm, %) was estimated according to the value of the ΔG/ΔGm ratio, where ΔG was equal to the free energy of miRNA-mRNA binding and ΔGm was equal to the energy of miRNA binding with its perfect complementary nucleotide sequence.

Results: As a result of this research, 15 miRNA binding sites with scores not less than 90% were found and marked in exons of titin mRNA. miR-6861-5p has the largest number of binding sites. This miRNA bound with the mRNA of titin at positions 37324, 38077, and 38830 nt at the boundaries of exons 178-179, 187-188, and 196-197, respectively. Other miRNAs had only one binding site each. miR-494-5p bound with titin mRNA at position 1301 nt in the seventh exon. miR-578 bound with titin mRNA in the eleventh exon at position 1960 nt. The 58th exon contained overlapping binding sites of two miRNAs (miR-374b-3p and miR-374c-3p) in positions 17239 nt and 17241 nt, respectively. Exon 59 was a target for miR-3714, which interacted with the mRNA at position 17450 nt. Exon 75 was the target of miR-34a-3p, which interacted with the mRNA at position 22116 nt. The 85th exon of the titin gene was the target for miR-1278, which interacted with titin mRNA at the position of 24928 nt. The 89th exon had a binding site for miR-544b at position 26044 nt. The 326th exon contained binding sites for miR-4738-3p and miR-136-3p. miR-4738-3p interacted with titin mRNA at position 74955 nt and miR-136-3p bound with mRNA of titin at position 71469 nt. Exon 339 also contained binding sites for miR-4693-5p and miR-4495, which bound with the mRNA for titin at positions 92464 nt and 93909 nt, respectively. The results of the computer analysis provide a theoretical basis for further experiments to validate the miRNA binding sites found in the mRNA of titin and to determine the miRNA concentrations in the blood and other cells and tissues of humans and mice.

ANALYSIS OF MIRNA-MRNA INTERACTOME IN HUMAN: GENERAL CHARACTERISTICS AND PREDICTIONS EVALUATIONS

O.M. Plotnikova[1] *, M.Y. Skoblov[1, 2]
[1] *Moscow Institute of Physics and Technology, Moscow, Russia*
[2] *Research Centre for Medical Genetics, Moscow, Russia*
** Corresponding author: plotnikova@phystech.edu*

Key words: *miRNA-mRNA interactions, CLASH, miRNA target predictions*

Motivation and Aim: MicroRNAs are a short RNA molecules (~ 22nt long), which plays a key role in the regulation of many biological networks. Nowadays it is known more than 2'500 human's miRNAs, but the question about the mechanism of interaction between mRNA and miRNA is still open. It is considered, that miRNA has a special region - 'seed', which is crucial for a miRNA-mRNA binding. There are five widely used predictive tools: TargetScan, Pictar2, PITA, RNA22 and miRanda. They have different approaches to make prediction. It was impossible to understand, which one is better and can we trust these predictions.

Methods and Algorithms: However recently the new high-throughput experimental method «CLASH» was developed to identify all miRNA-mRNA interactions. It allowed us to find out all miRNA- mRNA cases in HEK293 cell line. We created an algorithm to analyse and compare data from "CLASH" experiment with predicted miRNA sites by all five algorithms. Expression data for mRNA and miRNA were used from FANTOM5 and GEO DataSets.

Results: For the comparative analysis, we obtained 16'190 miRNA-mRNA interactions (according to CLASH data) and 19'398 predicted miRNA-mRNA interactions. We estimated working of miRNA-mRNA prediction programs by the following criteria: sensitivity, positive predicted value, predictions in different mRNA regions (3'UTR, CDS, 5'UTR), predictions for different types of interactions (5 classes), predictions of "canonical" and "nocanonical" interactions, and testing by using random data for miRNA binding sites. Expression analyze of miRNA revealed several interesting groups: highly expressed miRNAs without any interactions, highly expressed miRNAs with small number of interactions and lowly expressed miRNAs, which take a part in a lot of interactions with mRNAs. Deep expression analysis found out mRNA formed the most interactions.

Conclusion: For all five miRNA prediction softwares we demonstrated a low sensitivity and positive predictive value that doesn't allow to use them for experimental researching. Also we found group of interesting miRNAs and mRNAs, which could play a key role in cell regulation processes.

DNA DAMAGE AND GENERATION OF REACTIVE OXYGEN SPECIES BY PLATINUM DRUGS: EXPERIMENTS ON BACTERIA

E.V. Prazdnova*, V.A. Chistyakov, M.S. Mazanko, M.N. Churilov, V.K. Chmyhalo
Academy of Biology and Biotechnologies of Southern Federal University, Rostov-on-Don, Russia
** Corresponding author: prazdnova@sfedu.ru*

Key words: *cisplatin, DNA damage, oxidative stress, mutagenesis, antibiotic resistance*

Motivation and Aim: Cisplatin is anticancer drug that provides the cytotoxic effect by inducing apoptosis through the formation of intrachain DNA–adducts [1]. The use of cisplatin is limited by a wide range of side effects, strong mutagenic and genotoxic effects [2]. We suppose that the drug may cause mutations in the patient's microflora metagenome. That leads to secondary infections, caused by microorganisms resistant to antimicrobial agents. The effects of cisplatin may also be mediated by the generation of reactive oxygen species [3]. Therefore, antioxidants may reduce the mutagenic potential of cisplatin. We performed a study to investigate whether the mutagenic effect of cisplatin to bacteria is also due to oxidative stress.

Methods and Algorithms: We used a range of bacterial biosensors reacting to oxidative stress and DNA damage, based on E.coli strains MG1655 pKatG-lux (registers formation of hydrogen peroxide in the cell), pSoxS-lux (reacts to the increased superoxide-anion-radical level), and pColD-lux (registers DNA damage). To estimate mutation rate, we applied the standard serial dilution method.

Results: The biosensor assay demonstrated high genotoxic activity of cisplatin, and a slight induction of superoxide anion radical, with no generation of hydrogen peroxide. It was shown that ascorbate reduces the genotoxic effect of cisplatin by 41% in this model system. Non-lethal doses of cisplatin induced 3-7-fold increase in the frequency of mutant resistant to rifampicin and ciprofloxacin in E. coli MG 1655. It was found that ascorbate reduces mutagenesis induced by cisplatin by 65%. However, it decreased the toxicity of drug too.

Conclusion: The mechanism of drug action on bacteria appears to be associated with the generation of superoxide anion-radical. To reduce the risk of secondary infections, complicated by antibiotic resistance, it seems reasonable to use antioxidants, but it should be taken into account that they can reduce the general cytotoxicity of drug.

Acknowledgements: This work was supported by the Ministry of Education and Science of the Russian Federation (project No. 6.1202.2014/K)

References:

1. M.J.Silva et al. (2005) Comparative analysis of the mutagenic activity of oxaliplatin and cisplatin in the Hprt gene of CHO cells, Environmental and molecular mutagenesis, 46.2: 104-115.
2. G.Brozovic et al. (2008) Evaluation of DNA damage in vivo induced by combined application of cisplatin and sevoflurane, European journal of anesthesiology, 25.08: 642-647.
3. A. Laurent et al. (2005) Controlling tumor growth by modulating endogenous production of reactive oxygen species, Cancer research 65.3: 948-956

MODELING RESTRICTION-MODIFICATION SYSTEMS: EXPRESSING TOXIC MOLECULES WITHIN A CELL

A. Rodic*, M. Djordjevic
Faculty of Biology, University of Belgrade, Serbia
** Corresponding author: andjela.rodic@bio.bg.ac.rs*

Key words: *restriction-modification systems, modeling transcription regulation, single cell experiments, expression dynamics*

Motivation and Aim: Bacterial restriction-modification (RM) systems encode a restriction enzyme (R), which cuts specific DNA sequences, and a methyltransferase (M), which methylates and protects the same sequences. Expression of these enzymes during the system establishment in a naïve bacterial host has to be tightly regulated to prevent the host DNA being damaged by the toxic molecule (R), with this regulation often being accomplished by a control (C) protein. We aim understanding which features of RM systems allow this tight regulation. To that end, we provided modeling for recent state-of-the-art single cell measurements of RM dynamics, and also systematically in silico abolished the main RM features.

Methods and Algorithms: We developed a model for in vivo dynamics of R and M expression in a cell, which we compared with the first single-cell measurements done for Esp1396I RM system [1]. We used statistical thermodynamics to model the system transcription regulation, which was then used as an input for the dynamical modeling, that we subsequently implemented in a numerical procedure allowing direct comparison with the measured data. We used a similar model to perturb (in silico mutate) the main system features in another RM system with convergent architecture (AhdI). These features include an extremely high binding cooperativity, the differential translation efficiency, and the high dimerization constant. We then systematically analyzed the effect of these perturbations on the system dynamics [2].

Results: Our model successfully reproduces the main experimentally measured qualitative features of the expression dynamics – the significant delay of R with respect to M expression, including a high pic in M expression for the early times [1]. Regarding the in silico mutations, the perturbations generally abolish the three main dynamical features of the system: a delay in R expression, a fast transition from "OFF" to "ON" state, and the steady state stability [2].

Conclusion: The developed theoretical model can satisfactory explain the first direct measurements of the enzyme expression in RM systems. Our results suggest that the prominent RM features are likely optimized to satisfy few dynamical constraints. The inferred principles provide guidelines for constructing synthetic gene circuits capable of efficiently expressing toxic molecules [3].

Acknowledgement: This work is funded by the Swiss National Science Foundation (IZ73Z0_152297), and Serbian Ministry of Science and Technology (ON173052).

References:

1. N. Morozova, A. Sabantsev, E. Bogdanova, Y. Fedorova, A. Maikova, A. Vedyaykin, A. Rodic, M. Djordjevic, M. Khodorkovskii, K. Severinov (2015), Nucleic Acids Research, **44:** 790-800.
2. A. Rodic, et al. (2016) submitted.
3. A. Rodic, et al. (2016) in preparation.

GENETIC FITNESS OF DEAF PEOPLE IN THE SAKHA REPUBLIC

G.P. Romanov[1] *, N.A. Barashkov[1, 2], F.M. Teryutin[1, 3], L.A. Klarov[3], A.V. Solovyev[1], N.N. Gotovtsev[1], V.G. Pshennikova[1, 2], N.N. Sazonov[1], I.V. Morozov[4, 5], A.A. Bondar[4], L.U. Dzhemileva[6], E.K. Khusnutdinova[6, 7], O.L. Posukh[5, 8], S.A. Fedorova[1, 2]

[1] *Institute of Natural Sciences, M.K. Ammosov North-Eastern Federal University, Yakutsk, Russia*
[2] *Yakut Scientific Center of Complex Medical Problems, Yakutsk, Russia*
[3] *Republican Hospital # 2 – Center of Emergency Medicine, Yakutsk, Russia*
[4] *Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia*
[5] *Novosibirsk State University, Novosibirsk, Russia*
[6] *Institute of Biochemistry and Genetics, Ufa Scientific Centre, Ufa, Russia*
[7] *Bashkir State University, Ufa, Russia*
[8] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** Corresponding author: gpromanov@gmail.com*

Key words: *GJB2, deafness; genetic fitness; Eastern Siberia, Russia*

Motivation and Aim: Introduction of a sign language in schools for deaf people led to growth of their genetic fitness, which has doubled the GJB2 gene, associated deafness in the USA over the past 200 years. High prevalence of the GJB2-deafness and relatively recent (~ 60 years ago) introduction of sign language among deaf people were recorded among indigenous Yakut population (Eastern Siberia, Russia).

Methods and Algorithms: We have performed study of fertility of deaf people compared to their hearing siblings in Eastern Siberia. Fertility was determined as mean number of children born to specific group. Genetic fitness of deaf people was calculated as the ratio of their fertility to fertility of their hearing siblings [1].

Results: Data on 83 deaf people and 185 hearing siblings, aged 35-69 years was collected. 143 children accounted for 83 deaf individuals, whereas 185 hearing siblings had 422 children. Fertility of deaf people was estimated as 1.72 vs 2.28 of their hearing siblings. Overall the genetic fitness for deaf individuals is 0.75. There was no difference between genders. Our results are comparable with fitness of deaf women in Sweden – 0.76 [2], and lower than in USA – 0.88 [1] and higher than in Mongolia – 0.62 [3].

Conclusion: Thus, genetic fitness of deaf in Eastern Siberia is slightly reduced, which is still could possibly increase frequency of GJB2-deafness.

Acknowledgements: The study was supported by the RFBR grants #15-04-04860_a, #16-34-00564_mol_a, the State project #6.656.2014/K.

References:

1. Blanton SH, Nance WE, Norris VW, Welch KO, Burt A, Pandya A, Arnos KS Fitness among individuals with early childhood deafness: Studies in alumni families from Gallaudet University Ann Hum Genet. 2010 Jan;74(1):27-33.
2. Carlsson PI, Danermark B, Borg E. Marital status and birthrate of deaf people in two Swedish counties: the impact of social environment in terms of deaf community. Am Ann Deaf. 2004-2005 Winter;149(5):415-20.
3. Tekin M, Xia XJ, Erdenetungalag R, Cengiz FB, White TW, Radnaabazar J, Dangaasuren B, Tastan H, Nance WE, Pandya A. GJB2 mutations in Mongolia: complex alleles, low frequency, and reduced fitness of the deaf Ann Hum Genet. 2010 Mar;74(2):155-64.

MATHEMATICAL MODELING OF ACTIVE SUBSTANSES AND FACTORS INFLUENCE ON FUNCTIONING OF PLANT ROOT MERISTEM

M.S. Savina[1, 2] *, F.V. Kazantsev[1, 2], V.V. Mironova[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: captains-daughter@yandex.ru*

Key words: *mathematical model, auxin, salicylic acid, Arabidopsis thaliana*

Motivation and Aim: Plant hormone auxin is the most important regulator of plant growth and development. Auxin maintains of meristems, involves in gravity response and organogenesis. The influence of external factors and active substances on plant root development is usually mediated through variations in auxin synthesis or transportation, which subsequently affects auxin distribution pattern. With auxin regulating division, growth and differentiation of cells in a dose-dependent manner, the variations in auxin distribution would lead to morphological changes in plants. The problem is that auxin distribution is still elusive as no direct methods exist to study cellular concentrations of the low-molecule substance in a tissue. Mathematical modeling of the auxin transportation and synthesis allows to predict auxin distribution in a tissue and the range of auxin-dependent morphological changes. In order to study the influence of various environmental factors on plant root development, we proposed an approach based on the mathematical modeling of auxin distribution in the tissue.

Methods and Algorithms: Here we extended the mathematical model [1] to describe in more details auxin synthesis and transportation. Rectangular cell layout MxN corresponded to the root tip of Arabidopsis thaliana. Concentration changes of auxin and four of its transporters (PIN1, PIN2, PIN3, PIN7) in every cell of cell layout were described by ordinary differential equations. The model took into account synthesis and degradation of all substances, auxin diffusion and active transport. Experimental data on the changes of PINs expression in control and after salicylic acid (SA) treatments were used to fit the model, as an example of the approach application (Pasternak et al., unpublished).

Results: The model was extended to describe the effects of SA treatments on synthesis rates of PIN transporters. At the first step, the model parameters for control were adjusted. At the second step, the model parameters for PINs expression were adjusted to the experimental data. Auxin and PINs distribution were analysed in the steady state solution. The model predicted auxin accumulation in the root tip tissues after SA treatments. The sites of auxin accumulations coincide with the phenotypic changes induced by the treatments.

Conclusion: Developed approach can be applied for studying the effects of various environmental factors on the plant root growth.

Acknowledgements: The work was supported by RSF 14-14-00734. We thank Nadya Omelyanchuk for valuable advice and Taras Pasternak for experimental data.

References:

1. Mironova V.V. et al. (2012) Combined *in silico/in vivo* analysis of mechanisms providing for root apical meristem self-organization and maintenance, Annals of botany, **110(2)**: 349-360.

PHYLOGENETIC ANALYS OF DAHPS II TYPE AMINOACID SEQUENCES

A.I. Semashko*, E.G. Veremeenko, N.P. Maksimova
Belarusian State University, Minsk, Belarus
** Corresponding author: tassigo@gmail.com*

Key words: *3-deoxy-D-arabino-heptulosonate 7-phosphate synthase, aromatic amino acids, phenazine antibiotics, phylogeny*

Motivation and Aim: DAHP (3-deoxy-D-arabino-heptulosonate 7-phosphate) synthase catalyses the condensation reaction between phosphoenolpyruvate and D-erythrose 4-phosphate as the first committed step in the biosynthesis of aromatic compounds in plants, fungi and bacteria. It is known about two types of DAHP synthases[1]. The first type of this enzyme contributes formation of aromatic amino acids, siderophores and quinones. The second type promotes formation of phenazine antibiotics and was discovered in plants, fungi and bacteria in contrast to the first type which was find only in bacteria[2]. That is why the main aim of this research was detection of DAHP synthase II type evolution pathway.

Methods and Algorithms: Sequencing of genes which encodes DAHP synthase II type in Pseudomonas bacteria was carried out. Sequences for plants, fungi, agrobacteria, cyanobacteria and mammals were found in GeneBank database. Alignment and plotting of a phylogenetic tree were made with help of program MEGA (version 6.0) using sequences of unique binding sites of DAHP synthase II type with substrate[3].

Results: The UPGMA-tree shows unexpected results. The first major clade comprises all members of plants: both with cloroplastic and nuclear DAHP-genes. The second clade unites Basidiomycota fungi. The third clade is complex of two following groups: cyanobacteria Mastigocladus and Ascomycota fungi such as Penicillium and Aspergillus. The forth clade unites bacteria such as Agrobacterium and Rhizobium which can interract with plants. The last clade combines cyanobacteria Scytonema, Pseudomonas bacteria and Pantholops. Thus such unusual allocation of different organisms can be explained by the old origin of DAHP-genes and absence of reverse mutation detection. For these reasons it can be observed the join of organisms from various kingdoms in one group.

Conclusion: According to received cladogram it can be assumed that gene which encodes DAHP synthase II type primary appeared in ancestral cyanobacteria. After this part of them gave rise for modern bacteria, cyanobacteria, fungi and plants' chloroplasts. Plants' nuclear genes of DAHP synthase descended from cloroplastic genes by acquisition of introns. Appearance of such genes in grass-feeding mammals genome can be connected with endosymbiotic bacteria which decompose cellulose.

References:

1. Diversity and evolution of the phenazine biosynthesis pathway / D.V. Mavrodi, T.L. Peever, O.V. Mavrodi, J.A. Parejko, J.M. Raaijmakers, P. Lemanceau, S. Mazurier, L. Heide, W. Blankenfeldt, D.M. Weller, L.S. Thomashow // Appl. Environ. Microbiol. – 2010. – Vol. 76, № 3. – P. 866-879.
2. Light, S.H. The diversity of allosteric controls at the gateway to aromatic amino acid biosynthesis / S.H. Light, W.F. Anderson // The Protein Society. – 2013. – P. 395-404.
3. In Silico sequence analysis and molecular modeling of the three-dimensional structure of DAHP synthase from Pseudomonas fragi / S. Tapas, G.K. Patel, S. Dhindwal, Sh. Tomar // J. Mol. Model. – 2011. – Vol. 17. – P. 621-623

COMPUTATIONAL MODELLING OF QUIESCENT PLATELET ENERGETIC METABOLISM

T.O. Shepelyuk[1, 2, 3] *, A.N. Sveshnikova[1, 2]
[1] *Center for Theoretical Problems of Physicochemical Pharmacology RAS, Moscow, Russia*
[2] *Federal Research and Clinical Center of Pediatric Hematology, Oncology and Immunology, Moscow, Russia*
[3] *Faculty of Medicine, Lomonosov Moscow State University, Moscow, Russia*
** Corresponding author: shepelyuktaisiya@gmail.com*

Key words: *platelet, FBA, metabolism*

Motivation and Aim: Platelets are anucleate blood cells circulating in the bloodstream. Upon vessel wall injury platelets become activated, thus forming a thrombus and preventing the blood loss. For these processes they can use oxidative phosphorylation and glycolysis, utilizing blood glucose, stored glycogen or endogenous and exogenous fatty acids as fuel. There is no agreement in experimental data on platelet functioning in quiescent and activated states.

Methods and Algorithms: This study is a systematic analysis of the energetic abilities of quiescent and activated platelets through mathematical modelling of their energetic metabolism by Flux Balance Analysis (FBA) using the GNU Linear Programming Kit. The model mainly includes glycolysis, glycogenolysis, pentose phosphate pathway, β-oxidation, oxidative phosphorylation and ATP generalized consuming processes. Flux constraints were calculated from platelet proteomics data. The ATP production was considered as an objective function.

Results: Theoretical limits on glucose, oxygen and fatty acids consumption rates are 0,128, 0,145 and 0,0055 mM/s, respectively. The corresponding ATP production rate is 0,980 mM/s without glycogenolysis and 1,180 mM/s with glycogenolysis. These fluxes are constrained by key enzymes activities (hexokinase, pyruvate dehydrogenase complex, cytochrome c oxidase, specific acyl-CoA dehydrogenases, etc.). Analysis of experimentally measured fluxes for platelet ATP production (0,200 – 0,600 mM/s), glucose (0,055 – 0,120 mM/s) reveals that these fluxes are close to theoretical maximum. Analysis of the same fluxes during platelet activation has shown doubling of ATP production.

Conclusion: As a result of the FBA analysis that glycolysis is a principal source to platelet energy production, while the importance of oxidative metabolism is controversial. We also demonstrated that key platelet enzymes function close to their theoretical limits, which is unusual for human cells.

IDENTIFICATION OF NUCLEAR GENES CONTROLLING CHLOROPHYLL SYNTHESIS IN BARLEY BY RNA-SEQ

N.A. Shmakov[1, 2] *, G.V. Vasiliev[1], N.V. Shatskaya[1], A.V. Doroshkov[1], D.A. Afonnikov[1, 2], E.K. Khlestkina[1, 2]

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

[2] *Novosibirsk State University, Novosibirsk, Russia*

** Corresponding author: shmakov@bionet.nsc.ru*

Key words: *barley, near-isogenic lines, chlorophyll synthesis, albinolemma, nuclear genes, gene network, RNAseq, differential expression, IonTorrent sequencing platform*

Motivation and Aim: Albinism in plants is characterized by lack of chlorophyll and results in photosynthesis impairment, abnormal plant development and premature death. These abnormalities are frequently encountered in interspecific crosses and tissue culture experiments. Analysis of albino mutant phenotypes with full or partial chlorophyll deficiency can shed light on genetic determinants and molecular mechanisms of albinism.

Methods and Algorithms: Poly-A RNA was extracted from spikelets of barley of Bowman line and i:BwAlm line with tissue-specific albinism and sequenced using IonTorrent platform. Resulting short read libraries were mapped to Hordeum vulgare genome using cufflinks pipeline and STAR mapper. Differential expression search was conducted with cufflinks pipeline and edgeR package. Differentially expressed genes list was examined for enriched gene ontology terms from AgriGO database and significantly represented pathways from PlantCyc database. For a selected list of genes differential expression was checked with quantitative real-time PCR. Phentypic characterization and chlorophyll distribution patterns were examined using chlorophyll fluorescence microscopy. De novo transcriptome assembly was performed using Trinity tool.

Results: Microscopic analysis revealed segregation of cells in spikelets to chloroplast-containing and chloroplast-deficient. Our results demonstrated that alm mutant has decreased expression level of plastid genes. Statistically significant differential expression was observed for several plastid operons containing protein coding genes, rRNA and tRNA-coding genes. We identified nuclear genes with differential expression in two barley lines. Functional differentiation between genes with higher and lower levels of expression in i:BwAlm line was detected. As was demonstrated with gene ontology analysis, genes with lower level of expression in i:BwAlm line are mostly associated with photosynthesis and chlorophyll synthesis, while genes with higher expression level are functionally associated with vesicle transport. Differentially expressed genes are shown to be involved in several metabolic pathways, most represented being Calvin-Benson-Bassham cycle. Finally, de novo assembly of transcriptome contains several transcripts, not annotated in current H. vulgare genome version.

Conclusion: Our results provide the new information about genes which could be involved in formation of albino lemma and pericarp phenotype. They demonstrate the interplay between nuclear and chloroplast genomes in this physiological process.

Acknowledgements: This work was supported by the RFBR (№ 16-34-00924, bioinformatics data analysis), RSF (№ 14-14-00734, microscopic images preparation and analysis)

GENETIC CONTROL OF CIRCADIAN RHYTHMS: AN IMPACT OF MOLECULAR CLOCK EXPRESSION PROFILE CHANGES IN LONGEVITY

I.A. Solovyev[1, 2] *, E.V. Dobrovolskaya[1], A.A. Moskalev[1, 2, 3]
[1] *Institute of Biology of Komi Scientifice Center of Ural Branch of RAS, Syktyvkar, Russia*
[2] *Syktyvkar State University, Syktyvkar, Russia*
[3] *Moscow Institute of Physics and Technology (State University), Moscow, Russia*
** Corresponding author: soloviev@ib.komisc.ru*

Key words: *circadian rhythms, aging, longevity, transcriptome*

Motivation and Aim: Genes of circadian rhythms change their expression during aging of different organisms. We analyzed available transcriptomic data from different online bases and compared circadian genes' expression profile changes in animals. These findings have led us to an idea of normalizing expression profiles of circadian oscillator elements to compensate potential aging-associated changes during all lifespan on Drosophila model. The aim of the present research is to investigate the role of molecular oscillator elements (cry, per, tim, clk, cyc) in aging and longevity mechanisms [1].

Methods and Algorithms: We used standard methods of Drosophila cultivation, demographic methods to investigate the lifespan changes, RU486-inducible UAS-GAL4 system inserted before genes of interest was used as a tool which up-regulates expression.

Results: We overexpressed clk, per, cry, cyc and tim using neuron-specific RU486-inducible system, this resulted in the increase of median life expectancy (10%) in tim- and cry12-overexpessing females. Median lifespan of female fruit flies overexpressing per10 was 5.4% longer than in control group. Noteworthy, overexpression of clk shortened (-10%) only female's lifespan. 4% augmentation of median life expectancy was observed for males overexpressing per24 and cyc.

Conclusion: Thus, our data has shown that compensation of circadian clock genes' potential age-dependent expression decrease (cry, per) in the nervous system during all imago's life as well as hypercompensation of increased levels of other genes (tim, cyc) extends lifespan. The analysis of the literature shows that clock genes modulate the activity of various determinants of aging, which, probably, cause life extension [1].

Acknowledgements: The research is supported by RFBR grant N16-34-00734 and by Grant of Presidium of Ural Branch of RAS N15-4-4-23.

References:

1. I. A. Solovyev, E. V. Dobrovolskaya, A. A. Moskalev (2016) Genetic control of circadian rhythms and aging, Russian Journal of Genetics, **52(4)**: 393-412.

INVESTIGATION A ROLE OF DRUG EFFLUX PUMP IN AMINOGLYCOSIDE-RESISTANT *MYCOBACTERIUM TUBERCULOSIS* CLINICAL STRAIN

A. Sowajassatakul[1], T. Prammananan[2, 3], A. Chaiprasert[3, 4], S. Phunpruch[1, 5] *

[1] *Department of Biology, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Chalongkrung Road, Bangkok, Thailand*

[2] *Tuberculosis Research Laboratory, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Thailand Science Park, Pathum Thani, Thailand*

[3] *Drug Resistance Tuberculosis Research Fund, Siriraj Foundation, Bangkok, Thailand*

[4] *Department of Microbiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand*

[5] *Bioenergy Research Unit, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Chalongkrung Road, Bangkok, Thailand*

* *Corresponding author: kpsarany@kmail.ac.th*

Key words: *Mycobacterium tuberculosis, aminoglycoside, resistance, efflux pump*

Motivation and Aim: Aminoglycosides are one of effective drugs for treatment of multidrug-resistant tuberculosis (MDR-TB). Understanding the molecular mechanisms of drug resistance would help to develop rapid tests for accurate diagnosis, particularly for detection of drug-resistant strains. However, some resistant strains did not contain any known resistance mechanisms, indicating that there are other unknown mechanisms involved in the resistant phenotype. This study aims to investigate the putative aminoglycoside efflux pumps in the amikacin- and kanamycin-resistant M. tuberculosis clinical strain isolated in Thailand.

Methods and Algorithms: In this study, we investigated the amikacin- and kanamycin-resistant Mycobacterium tuberculosis strain MT433 (GenBank accession no. LGAX00000000). This strain contains none of any mutations at known resistance genes (rrs, eis promoter, and whiB7). We hypothesized that drug efflux mechanism might be involved in the resistant phenotype. Expression of 16 putative efflux pump genes and one regulator gene was determined in the presence and absence of kanamycin using Real-Time Quantitative Reverse Transcription PCR (Real-Time qRT-PCR) technique.

Results: Results revealed that after exposure to kanamycin, 3 out of 16 efflux pump genes (Rv1819c, Rv1877, and Rv2846c) showed significantly higher expression levels than those of other genes (>400-flod upregulation).

Conclusion: Our finding demonstrated the up-regulation of three drug efflux pump genes, namely Rv1819c, Rv1877, and Rv2846c, in response to kanamycin and was firstly found in M. tuberculosis clinical strain. This indicates that the drug efflux pump might play an important role in amikacin and kanamycin resistance in this strain and it needs to be further investigated.

Availability: Not applicable.

Acknowledgements: This study was financially supported by the Faculty of Science, King Mongkut's Institute of Technology Ladkrabrabang (KMITL) and Drug-Resistant Tuberculosis Research Fund, Siriraj Foundation. A. Sowajassatakul is also thankful for a scholarship for the Ph.D. Program that was provided by the Thailand Graduate Institute of Science and Technology (TGIST), National Science and Technology Development Agency (NSTDA).

EXPGENE – SOFTWARE FOR ANALYSIS AND PROCESSING OF GENE EXPRESSION DATA

A.M. Spitsina*
Novosibirsk State University, Novosibirsk, Russia
** Corresponding author: aspitsina@bionet.nsc.ru*

Key words: DNA microarray, RNA-Seq, bioinformatics software

Motivation and Aim: Bioinformatics analysis of gene expression molecular mechanisms has a great fundamental importance in various fields of science, particularly in medicine and statistics. Currently, there is a rapid development of genomic and biological technologies, which are leads to the accumulation of large experimental gene expression data in publicly available databases (most popular and free is BioGPS [1], GEO NCBI). Processing of such data requires development of new computer analysis methods, what will allow solve practical biomedical problems - reveal genes, associated with various diseases (cancer, neurodegenerative diseases and other). The purpose of this work is the creation of software for the analysis and visualization of transcriptomic and microarray data, which will be easy to use and multifunctional.

Methods and Algorithms: The Affymetrix GeneChip U133A data on the human genome and genomes of model organisms (mice and rats) were used as test data. The program is written in Python language using JSON modules, and also popular libraries for processing and visualization text and numerical data (pandas, numpy, scipy, matplotlib).

Results: Software package ExpGene has been developed. It includes a set of options to work with a large array of microarray data - preprocessing, statistical analysis of gene expression correlations and visualization. This tool is also designed to work with gene ontologies. It allows search genes in the chromosome loops formed by pairs of CTCF transcription factor binding sites. It is versatile for any type of text databases (allows the user to pre-select processed data). The program has a user-interaction interface (menu) and is easy to handle even by an inexperienced user.

Conclusion: Using this program we performed a comparative analysis of different samples of genes [1], such as genes from gene networks annotated in the ICG SB RAS regulating cholesterol levels and circadian rhythm. We also studied genes, which are responsible for aggressive behavior of laboratory animals (rats). Gene expression correlation matrices for gene lists were reconstructed as basis for qualitative analysis of the gene network studied.

Availability: Software is available from the author upon request.

References:

1. A.M. Spitsina et al. (2015) Supercomputer analysis of genomics and transcriptomics data revealed by high-throughput DNA sequencing, Program systems: theory and applications, 6:1(23): 157–174. (In Russian)

PREDICTING SMALL RNAS FROM BACTERIAL GENOME

T. Stankovic*, J. Guzina, M. Nikolic, M. Djordjevic
Faculty of Biology, University of Belgrade, Serbia
** Corresponding author: tamara.stankovic@bio.bg.ac.rs*

Key words: *CRISPR/Cas, small RNAs, transcription start site predictions, terminator predictions*

Motivation and Aim: Small RNAs perform important regulatory roles in bacteria. Their discovery is however complicated by still limited availability of dRNA-seq data in bacteria, and by the fact that they are often expressed under non-standard and poorly characterized conditions, which may not correspond to those of the experiment. Furthermore, small RNAs are often poorly conserved even between related bacterial strains, which complicate their computational discovery. As an alternative, small RNAs can be detected directly from the genome sequence, which requires accurately predicting transcription start site (TSS) and terminator signals. However, two main limiting factors in this approach are generally insufficient accuracy with which TSS are predicted in genome, and the fact that the terminator prediction parameters are trained on *E. coli* data, which may lead to suboptimal predictions in other bacteria. The aim of this work is improving TSS predictions, and assessing if the parameters for terminator predictions can be retrained to allow accurately predicting a specific group of small RNAs.

Methods and Algorithms: For TSS predictions, we start from accurate alignments of the promoter elements for σ70 and σE promoters, which we implement within a weight matrix search. As different TSS detection methods use a diverse description of the bacterial promoter structure, it is unclear which promoter futures should be included in TSS recognition, and how their accuracy impacts the search detection. We addressed this question for σ70 and σE (an alternative σ factor) promoters in *E. coli* [1]. For the terminator predictions, we start from the standard algorithm, where the parameters are trained on *E. coli* data. We then investigate if these parameters can be retrained to accurately detect small RNAs associated with CRISPR/Cas [2].

Results: We find that σ70 -35 element, which is considered exchangeable, contributes equally (for σ70), or more (for σE), to the search accuracy than the ubiquitous -10 element. Surprisingly, the sequence of the spacer between -35 and -10 promoter elements, significantly decreases the search accuracy for σ70 promoters, while improving the search accuracy for σE promoters. Overall, there is as much as ~50% false positive reduction for optimally implemented promoter features in σ70, underlying necessity for accurate promoter element alignments, for the purpose of small RNA predictions [1]. We also find that the terminator prediction algorithm can be successfully retrained to allow accurate predictions of small RNAs related with CRISPR/Cas [2].

Conclusion: Both standard TSS and terminator predictions can be significantly improved to allow accurate predictions for a specific group of small RNAs, across different bacterial genomes.

References:

1. M. Nikolic, T. Stankovic, M. Djordjevic (2016), to be submitted
2. J. Guzina, T. Stankovic, M. Djordjevic, M. Djordjevic (2016), in preparation

MYC GENE FAMILY IN CEREALS: TRANSFORMATION IN THE COURCE OF THE *TRITICUM* AND *AEGILOPS* GENERA EVOLUTION

K.V. Strygina[1] *, E.K. Khlestkina[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: pushpandzhali@bionet.nsc.ru*

Key words: *Aegilops, bHLH, gene divergence, gene duplication, flavonoids biosynthesis, MYC, Triticum, transcription factor*

Motivation and Aim: MYC family transcription factors are of an essential part of the regulatory complex «MYB + MYC + WD40», which is necessary for gene activation in flavonoid biosynthesis. In bread wheat (*Triticum aestivum* L., BBAADD, 2*n* = 6*x* = 42) the gene TaMyc1 controlling the synthesis of flavonoid pigments in wheat seed was earlier isolated and characterized [1]. The aim of the current study was to identify, characterize and compare full-length sequences of duplicated homoeologous and paralogous copies of the TaMyc1 gene.

Methods and Algorithms: The search of homologous sequences was made in databases for not annotated wheat sequences using BLAST. The cluster analysis using the MEGA software was based on UPGMA algorithm. The nucleotide substitutions rate (k) for Myc was calculated by the formula Ks/2T. The obtained value (k=10.04×109) was used for the calculation of divergence time of duplicated copies.

Results and discussion: We identified 10 copies of the gene TaMyc1 in a common wheat genome and 22 Myc–like genes in the genomes of related species (*T. durum; T. urartu, T. monococcum, Aegilops speltoides, Ae. sharonensis, Ae. tauschii*). Analysis of genetic similarity showed that the first duplication of Myc gene was in the diploid common ancestor of the tribe Triticeae. The duplication has undergone from two to four further acts of duplication in *Triticum* and *Aegilops* genomes. Time of occurrence of each new copy is calculated and presented in the report. Maintaining functional duplicated genes is likely due to their specialization. It is assumed that TaMyc1 copies may be involved in the synthesis of various flavonoid compounds in different parts of the plant.

Conclusion: Polyploid genome of bread wheat carries at least 11 copies of the TaMyc gene involved in flavonoid biosynthesis regulation. The duplications of this gene occurred several times in the course of evolution of diploid wheat progenitors. The exon-intron organization of these genes is the same to the TaMyc1 structure. All available sequences have a conserved bHLH domain. None significant change in the motive, which could lead to changes in gene function, has been identified among the annotated Myc genes of *T. aestivum* and related species.

Acknowledgements: This study was partially supported by the RAS MCB Programme (0324-2015-0016) and by RFBR (16-04-00912).

References:

1. Shoeva, O. Y., Gordeeva, E. I., Khlestkina, E. K. 2014. The regulation of anthocyanin synthesis in the wheat pericarp. Molecules, 19(12), 20266-20279.

DIVERSITY OF *MARINER*-LIKE DNA TRANSPOSONS IN THE GENOME OF *LOCUSTA MIGRATORIA*

I. Sukhih[1], K. Ustyantsev[1], A. Kononov[1], I. Konopatskaia[1, 2] *
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
* *Corresponding author: sormacheva@bionet.nsc.ru*

Key words: *DNA transposon, Tc1/mariner, insects, Locusta migratoria, genome, HMMER search, diversity*

Motivation and Aim: The genome of *Locusta migratoria* (Acrididae, Orthoptera) is the largest among sequenced animal genomes for today. Genomic studies revealed that the distribution of mobile elements is the major cause of such a large genome size in locusts. Thus, approximately 60% of *L. migratoria* genome consists of repetitive elements, about 24% of which are DNA transposons. Mobile elements of the migratory locust are poorly investigated despite their significant role in the genome evolution. In the present work, we study the diversity of mariner-like DNA transposons in the genome of *L. migratoria*.

Methods and Algorithms: To search and analyze the distribution of mariner-like elements from the *L. migratoria* genome the specific algorithm was designed. The algorithm includes: initial search for Tc1/Mariner transposase sequences using specific HMM profile in the HMMER3 suite; identification of the sequences of terminal inverted repeats near the HMM signals using the custom Python script designed by authors; selection of the Tc1/Mariner transposon sequences with the complete/(potentially intact) transposase open reading frame; Preliminary classification (grouping) of identified Tc1/Mariner transposons into subfamilies of Tc1/Mariners (mellifera, cecropia, etc.) using custom HMM profiles specific to their transposase sequences with hmmerscan tool from the HMMER3 suite; Rough clusterization within split subfamilies to identify typical representatives using CD-HIT Suite; Phylogenetic analysis using PhyML program.

Results and Conclusion: Analysis of mariner-like DNA transposons available from GenBank database showed that only nine subfamilies are distributed in the genomes of Insecta species (lineata, capitata, cecropia, irritans, mariner, mauritana, mellifera, DD37D_new_subf1 and vertumana). For the initial search the HMM profile was build based on the transposase sequences of the Tc1/mariner DNA transposons from various Metazoa species. HMM search revealed 127274 sequences of Tc1/mariner DNA transposons in the genome of *L. migratoria*, 27139 of which possess the complete (potentially intact) transposase sequence. Using hmmerscan tool from the HMMER3 and CD-HIT Suite 6074 out 27139 sequences were defined as mariner-like elements and clustered. Typical representatives for these groups were extracted and used for further phylogenetic analysis. Phylogenetic tree constructed based on 211 sequences revealed 28 clusters, 8 of which were known mariner-like subfamilies and 13 represent potentially new subfamilies. We conduct the first comprehensive search and analysis of mariner-like DNA transposons in *L. migratoria* genome. Diversity of mariner-like elements of *L. migratoria* is represented by capitata, cecropia, irritans, mariner, mauritana, mellifera, DD37D_new_subf1, vertumana and 20 potentially new subfamilies.

Acknowledgements: The investigation was financially supported by the Russian Foundation for Basic Research (Grant No. 16-34-00893).

COMPREHENSIVE ANALYSIS OF DRAFT GENOMES OF TWO CLOSELY RELATED *PSEUDOMONAS SYRINGAE* PHYLOGROUP 2B STRAINS INFECTING MONO-AND DICOTYLEDON HOST PLANTS

R.I. Sultanov[1, 2] *, G.P. Arapidi[1, 2], S.V. Vinogradova[3], V.M. Govorun[1, 2, 4], D.G. Luster[5], A.N. Ignatov[6]

[1] *Moscow Institute of Physics and Technology (State University), Moscow, Russia*
[2] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, the Russian Academy of Sciences, Moscow, Russia*
[3] *Research Center of Biotechnology, Moscow, Russia*
[4] *SRCC of Physical-Chemical Medicine, Moscow, Russia*
[5] *USDA-ARS Foreign Disease - Weed Science Research Unit, Ft. Detrick, USA*
[6] *R&D Center "Phytoengineering" LLS, Moscow region, Russia*
** Corresponding author: rhenium112@gmail.com*

Key words: *Pseudomonas syringae, dicots, monocots, pan-genome, core genome, T3SS, virulence factors*

Motivation and Aim: In recent years, the damage caused by bacterial pathogens to major crops has been increasing worldwide. *Pseudomonas syringae* is a widespread bacterial species that infects almost all major crops. Different *P. syringae* strains use a wide range of biochemical mechanisms, including phytotoxins and effectors of the type III and type IV secretion systems, which determine the specific nature of the pathogen virulence.

Methods and Algorithms: The genomes of Pseudomonas syringae strains 2507 (wheat) and 1845 (sunflower) isolated on the territory of the Russian Federation were determined by pyrosequencing and compared with previously published sequences of 18 genomes of the strains belonging to the same phylogroup and affecting dicots and monocots. We analyzed seven informative genes used in MLST genotyping of *P. syringae,* calculated the average nucleotide identity (ANI), and examined the compositions of the type III secretion system (T3SS) effectors and of the elements of insertion sequences (IS).

Results: We found that strains 2507 and 1845 and strains SM and B64 form a subgroup that is stable among the other strains of phylogroup 2b. The analysis of the genome of strain 1845 indicated the recent loss of several genetic elements (the cluster of genes responsible for the synthesis of syringolin and the prophage cluster) that are present in strains 2507, B64, and SM. We found three genes (YP_234264.1, YP_234265.1, and YP_237386.1), the acquisition of which by strain 1845 could lead to the change in its host class.

Conclusion: The results obtained by comparing the strain 1845 genome with the genomes of bacteria infecting monocots can help to identify the genes that define specific nature of the virulence of *P. syringae* strains.

Availability: The genomes of P. syringae strains 1845 and 2507 will be available soon at NCBI.

IDENTIFICATION OF PATHWAYS ASSOCIATED WITH CELL DEATH IN THE CORTEX OF OXYS RATS AS THE SIGNS OF ALZHEIMER'S DISEASE DEVELOP

G.K. Suvorov*, D.V. Telegina, E.A. Rudnitskaya, N.A. Stefanova, N.G. Kolosova
Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
** Corresponding author: augustocash@gmail.com*

Key words: *Alzheimer's disease, autophagy, apoptosis, cell death, OXYS rats, RNA-seq*

Motivation and aim: Alzheimer's disease (AD) is a progressive, age-dependent neurodegenerative disorder, featuring progressive impairments in memory, cognition, and ultimately leads to death. In spite of numerous studies on AD pathogenesis, the information about the molecular genetic preconditions of events leading to the death of neurons, as well as about the pathways of death, is extremely limited. Deregulation of autophagy and apoptosis plays a pivotal role in the etiology and/or progress of many of these diseases, including AD. Autophagy and apoptosis are basic physiologic processes contributing to the maintenance of cellular homeostasis. Autophagy is a major intracellular degeneration pathway involved in the elimination and recycling of damaged organelles and long-lived proteins by lysosomes. Apoptotic processes remove old and damaged cells to maintain tissue homeostasis without harming adjacent cells. To determine the role of cell death in the pathogenesis of AD, suitable animal models are needed. In this study, using nontransgenic OXYS rats that simulate key aspects of sporadic AD, we aimed to compare the gene expression profiles of the prefrontal cortex from OXYS rats and Wistar rats (as control) to identify the molecular mechanisms and the factors underlying of neuronal cell death in disease development. The transcriptome analysis was conducted at three stages of the disease (pre-symptomatic, 20 days; symptomatic, 5 month; and progressive stage, 18 month) in OXYS rats, using RNA-Seq technique.

Results and conclusion: Our results show that the development of the signs of AD (between ages 20 days and 5 months) in OXYS rats takes place during changes in mRNA expression of the 7 genes that are mostly related to processes of autophagy, such as Atg12, Atg7, and Atg8 (regulators of elongation). In addition, changes in mRNA expression of the 21 genes were related to apoptosis (proapoptotic genes and inhibitors) in the prefrontal cortex of OXYS rats between ages 20 days and 5 months. In OXYS rats, with progression of disease, 24 genes related to apoptosis and 7 genes related to autophagy change their expression. Importantly, Wistar rats show changes in expression of 21 genes related to apoptosis only between ages 20 days and 5 months. We also indicated the upregulation of 5 proapoptotic genes in 20-day-old OXYS rats compared Wistar rats. At the age of 5 and 18 months in OXYS rats, the balance between pro- and anti-apoptotic genes were changed (compared to Wistar rats). Accordingly, we demonstrated that the development of AD-like pathology in OXYS rats is related to the alterations in processes of autophagy and apoptosis.

Acknowledgements: This work was supported by the grant from the Russian Foundation for Basic Research (project # 15-04-06066).

MULTIPLE NON-LINKED NUCLEAR LOCI ANALYSIS REVEALS A DIRECT CONNECTION BETWEEN HYBRIDIZATION AND OBSERVED NON-MONOPHYLY IN THE SUBGENUS *STROBUS* (GENUS *PINUS,* FAM. PINACEAE)

V. Vavilova[1] *, K. Ustyantsev[1], G. V. Vasilyeva[2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Institute of Monitoring of Climatic and Ecological Systems SB RAS, Tomsk, Russia*
* *Corresponding author: valeriya-vavilova@bionet.nsc.ru*

Key words: *Pinus, non-monophyly, hybrids, introgression, hybridization, phylogeny*

Motivation and Aim: The phylogeny of genus *Pinus* are characterized by frequent discordant, because of presence of gene variability within a single species. It is indicate a possibility of frequent non monophyly in this genus. Interspecific hybridization was suggested as a one of the possible explanations of the observed phylogenetic discrepancies. However, there was no direct evidence to support any of the proposed scenarios. Natural hybrids between *Pinus sibirica* Du Tour (subgenus Strobus) and *P. pumila* (Pall.) Regel, as well as their parental species can be used as a model to reproduce the scenario of non monophyly in the subgenus Strobus.

Methods and Algorithms: We have analyzed genetic diversity of *P. sibirica, P. pumila* and their hybrids by dint of molecular biological and bioinformatical methods.

Results: We analyzed 37 specimens of *P. sibirica*, *P. pumila* and their hybrids using PCR with primers specific for three non-linked nuclear DNA markers: 4CL, AGP6 and LEA genes. Comparative and phylogenetic sequence analyses revealed two clusters of species-specific haplotypes for each of the markers, characteristic for two studied species. No clusters of hybrid-specific haplotypes were found. We found no hybrid specimens with a genotype characteristic to only one of the parental species for all the three loci. On average, the hybrids were characterized by an equal ratio of haplotypes from the *P. sibirica* and *P. pumila* clusters. Introgression was also detected for the individual loci of both parental species, suggesting multiple backcrossing events of the hybrids among themselves and/or with the parental species. Despite the introgression, the haplotype ratio was always shifted towards haplotypes from the lineage-related clusters in morphologically "pure" species. This allows separation of the species on the genetic basis.

Conclusion: Based on the received data, we speculate that interspecific hybridization should be the preferred explanation of the complicated patterns of non-monophyly often detected in phylogenetic studies of pine species.

Availability: The obtained nucleotide sequences of 4CL, AGP6 and LEA genes of *P. sibirica, P. pumila* and their hybrids were placed to ENA (European Nucleotide Archive) and GenBank.

CHARACTERISTICS OF *ACDS*-GENE OF BACTERIA *PSEUDOMONAS PUTIDA* B-37 RESPONSIBLE FOR ACC-DEAMINASE SYNTHESIS

D.S. Volkava*, S.I. Leanovich, A.A. Melnikava, E.A. Khramtsova
Belarusian State University, Minsk, Republic of Belarus
** Corresponding author: 95volkovads@gmail.com*

Key words: *ACC-deaminase, acdS-gene, Pseudomonas putida, ethylene, transgenic plants*

Motivation and aim: Nowadays one of the main problems in agriculture to be solved is plants' resistance to the numerous environmental factors expanding every year due to the active anthropogenic intervention. Plant's natural reaction on stress is the production of stress hormone ethylene that inhibits growth and development of plant organism during unfavorable periods of time. This process attends to decrease of the biomass production which is very unprofitable for agriculture. One of the most perspective ways to decline the level of stress ethylene is creating transgenic plants with acdS-gene coding for 1-aminocyclopropane-1-carboxylate (ACC-deaminase). ACC-deaminase is required for increasing the concentration of ethylene's precursor, ACC. This process promotes roots elongation, tuber forming, and biomass accumulation. The aim of current work was analysis of the primary nucleotide sequence of acdS-gene of Pseudomonas putida strain B-37 for further development of the recombinant plant cells Nicotiana benthamiana and Nicotiana tabacum.

Methods and algorithms: Primary nucleotide sequence of ACC-deaminase gene from P. putida B-37 was analyzed using programs available on on-line (NCBI, ExPASy, NPS@) and off-line resources. Search for homologs was conducted using BLASTn, and analysis of conserved domains – in Conserved Domains, available on-line on NCBI resource. To estimate ACC-deaminase gene its amino acid sequence was constructed using Translator on ExPASy resource (CDS size – 1017 b.p.). Molecular weight, amino acid composition, estimated half-life, theoretical pI were valued using ProtParam on ExPASy resource. Secondary structures were forecasted by the consensus prediction from the multiple alignments using SOPMA on NPS@. Sequences coding acdS-genes of other species and strains of Pseudomonas genus were downloaded from UniProt and aligned using the ClustalW method. Phylogenetic tree was constructed using Neighbor-joining method implemented in MEGA 6.0.

Results: It was shown that analyzing sequence has high homology with ACC-deaminase genes from different species of Pseudomonas genus. In the protein coding for open reading frame of this gene was detected Aminocyclopropane-1-carboxylate deaminase (ACCD) domain refers to tryptophan synthase beta superfamily (fold type II). Molecular weight is above 36718.9 D, mean theoretical pI = 5.6, estimated half-life in Escherichia coli is above 10 hours. Among secondary structures were mostly predicted alpha helixes (31.95%) and random coils (34.91%). Assumptive 3D structure of analyzing protein was modeled using SWISS-MODEL on ExPASy. To estimate divergence and homology of analyzing protein and ACC-deaminase proteins of different species of Pseudomonas phylogenetic tree was constructed; the highest homology with analyzing protein was obtained for ACC-deaminases from Pseudomonas putida strains.

OPTIMIZATION OF THE PIGGYBAC TRANSPOSON SYSTEM FOR CULTURED DROSOPHILA CELLS

L.A. Yarinich[1, 2] *, M.O. Lebedev[1, 2], A.V. Pindyurin[1, 2]
[1] *Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
* *Corresponding author: l.yarinich@mcb.nsc.ru*

Key words: *gene expression, chromatin position effect, DNA barcode, multiplex analysis, Drosophila cultured cells*

Motivation and Aim: The TRIP (Thousands of Reporters Integrated in Parallel) approach allows studying the influence of the local chromatin context on the gene activity simultaneously at thousands of genomic loci in cultured mouse cells. The approach is based on piggyBac-mediated transposition of the DNA-barcoded reporter constructs into the genome with the subsequent identification of their insertion sites and analysis of their transcriptional activity using high-throughput sequencing. The aim of this study is the adaptation of the TRIP approach to cultured Drosophila cells, which are a convenient model system for understanding the regulation of gene expression.

Methods and Algorithms: Drosophila Kc167 and S2 cultured cells of embryonic origin were used. Both electroporation and chemical methods were used to transfect cells. The level of transcription of the gene encoding piggyBac transposase was assessed by RT-qPCR. Efficiency of transposition of the reporter constructs into the genome was evaluated by qPCR as well as FACS. The insertion sites of the reporter constructs were identified by inverse PCR.

Results and Conclusion: We generated plasmid constructs encoding the piggyBac transposase under the control of four different promoters (constitutively active and inducible ones) and optimized the conditions for the transient activity of the transposase to avoid the 'rehopping' of the reporter constructs after their integration into the genome. For this purpose the following two approaches were used: [i] the chimeric version of the piggyBac transposase (PB-L3-ERT2), whose activity is regulated by tamoxifen, and [ii] the constructs encoding the transposase and the reporter gene as one transcriptional unit allowing to turn off the piggyBac transposase expression immediately after the transposition of the reporter construct from the plasmid into the genome. The system with the transposase gene and the transposon present in two separate plasmids was used to manipulate the ratio of the transposon to the transposase. We observed that 5:1 ratio of the transposon to the transposase allows to obtain the maximum copy number of the transposon insertions in the genome of Drosophila cells. The protocols for transfection of Drosophila cultured cells by electroporation as well as chemical methods were optimized to get as much as possible insertions of reporter constructs. We found that the level of transcription of the gene encoding piggyBac transposase and the frequency of reporter construct integration into the genome depend not only on the selected promoter for the transposase expression, but also on the cell line, the method of transfection and the ratio of the transposon to the transposase used. The developed plasmid constructs and protocols can be used for effective transgenesis of Drosophila cultured cells.

Acknowledgements: We thank Dr. S.V. Kulemzin (IMCB SB RAS) for FACS analysis of cells. This study was supported by the grant from the Russian Science Foundation no. 16-14-10288.

ALTORFEV: A NOVEL TOOL FOR PREDCITION OF ALTERNATIVE ORFS BASED ON THE LINEAR SCANNING MODEL

B.S. Zuraev[1] *, A.V. Kochetov[1, 2], A.I. Klimenko, S.A. Lashin[1, 2]
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
[2] *Novosibirsk State University, Novosibirsk, Russia*
** Corresponding author: bulatzuraev@gmail.com*

Key words: *Cytoscape plugin; ortholog; paralog; metabolic pathway; gene regulatory network; evolution; phylostratigraphy; evolution*

Motivation and Aim: The ribo-seq and proteomics techniques have revealed a large number of alternative ORFs (altORFs) within eukaryotic mRNAs. Some bioinformatics resources were developed to explore the available ribo-seq data to locate altORFs within mRNAs of interest (e.g., Ribotools, RiboGalaxy, GWIPS-viz). Indeed, knowledge on the full set of polypeptides encoded by a eukaryotic gene under study is essential for detailed investigation of its functions. However, published ribo-seq data are still very limited and conventional nucleotide sequence databanks do not provide information on the altORFs. In addition, the individual genetic variants may cause changes in mRNA coding potential: if a nucleotide sequence of mRNA under study is non-identical to the available ribo-seq-checked reference sequence, the positions of altORF(s) and their relative translation rates may differ. Thus, development of new tools for altORFs prediction remains quite actual. However, an accurate prediction of altORFs is very complicated because of a large number of various parameters influencing their recognition and translation efficiency.

Methods and Algorithms: The altORFev is based on the linear scanning model of translation. It also considers the leaky scanning and reinitiation mechanisms. In brief, 40S ribosomal subunits bind to 5'-end of mRNA and move linearly along mRNA until start AUG codon is found. The probability of AUG recognition depends on its nucleotide context: start codon in the optimal context is recognized by the majority of 40S ribosomal subunits. Thus, if AUG codon is located in the optimal context and its ORF is larger than 30 codons, this ORF is defined as "terminal" since the majority of incoming 40S ribosomal subunits can't move beyond it. If AUG codon is located in a suboptimal context, some 40S ribosomal subunits will recognize it and initiate translation, whereas others skip it and may initiate translation downstream (leaky scanning). Finally, if AUG is located in the optimal context but the ORF size is small (lesser than 30 codons), the reinitiation is possible: in this case, some 40S ribosomal subunits after termination of translation of small ORF remain connected to mRNA and may continue movement in 3'-direction. During scanning they restore their initiation competence by acquiring the lacked eIFs and met-tRNAi and may initiate translation further downstream.

Results: We have implemented two versions of the altORFev: (1) web application (Java 1.8, Vaadin); (2) desktop application (Java 1.8, Swing).

Conclusion: The altORFev may be used to get additional information on eukaryotic genes taking into consideration alternative coding abilities of their mRNAs.

Availability: web-version: http://www-bionet.sscc.ru:7780/AUGWeb/, desktop version: upon the requests to the authors.

Acknowledgements: The study is supported by the RSF 14-24-00123 grant.

ХИМЭКСПЕРТ

Агентство Химэксперт – официальный дистрибьютор компании Thermo Fisher Scientific (Life Sciences), являющейся основным мировым поставщиком приборов, реактивов и расходных материалов для молекулярной биологии. В компанию входят такие известные бренды, как: Applied Biosystems, Invitrogen, Ion Torrent, Ambion, Gibco, Molecular Probe, Novex, TaqMan, Pierce, Fermentas, Nunc, Finnpipette, Ленпипет и другие.

Мы предлагаем оборудование для анализа ДНК и РНК, фундаментальных протеомных и цитологических исследований, фармацевтики и биотехнологий, прикладного тестирования, включая идентификацию личности и установление родства в криминалистике и судебно-медицинской экспертизе, тестирование пищевых продуктов и многое другое.

Наши клиенты выбирают Химэксперт потому что:

– Мы предоставляем комплексные решения для оснащения молекулярно-генетических и медико-генетических лабораторий: от дозаторов до секвенаторов.

– Мы осуществляем полную техническую и методическую поддержку наших клиентов. Всегда готовы помочь заказчикам с подбором реагентов под поставленные задачи. Регулярно проводим профильные обучающие школы и семинары.

– Мы всегда осуществляем своевременные поставки заказов.

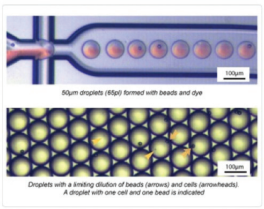
Компания Химэксперт существует 15 лет и давно зарекомендовала себя, как надежный поставщик оборудования и реактивов для широкого спектра исследований.

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

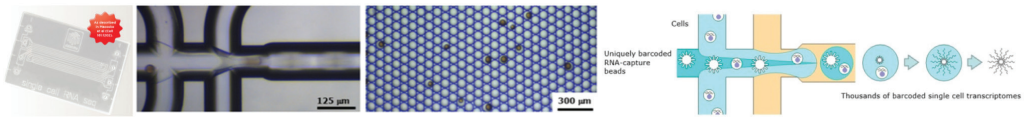
Использование микрофлюидики Dolomite для секвенирования транскриптомов отдельных клеток

Система для инкапсуляции клеток или нуклеиновых кислот в капли µEncapsulator 1

Отличная пробоподготовка для изучения экспрессии генов, ПЦР, сортировки и др.; инертные, биосовместимые материалы; возможность поддержания жизнеспособности клеток; пропускная способность: 300 000 клеток в 3 млн капель за 15 минут; инкапсуляция 100 мкл образца и реагента, или двух последовательных образцов по 100 мкл.



Система для создания библиотек единичных клеток для последующего секвенирования



Транскриптомика одиночных клеток; высокая точность, воспроизводимость, надежность, инертность материала чипа; скорость инкапсуляции, капель/сек – 4000 - выше чем у чипов из PDMS1.

Микрофлюидика Dolomite — технология, позволяющая работать с очень малыми объемами жидкостей, газов, с кристаллическими и полимерными частицами, клетками животного, растительного и бактериального происхождения, пузырьками и каплями с возможностью наблюдать за ними, манипулировать ими и контролировать процессы, протекающие с ними. Это дает возможность проводить «традиционные» исследования в миниатюрном формате, а также проводить исследования, которые ранее были невозможны.

Особенности и возможности микрофлюидики Dolomite:
работа с микрообъектами (капли, клетки, частицы, пузырьки); работа с микро- и нанообъемами (диаметр канала от 10 нм); высокая воспроизводимость: точность дозирования — порядка пиколитра; точный контроль параметров процесса: температуры, скорости потоков, давления, смешивания; большая библиотека «стандартных» чипов; чипы произвольной конфигурации и геометрии: многослойные и составные чипы с разными свойствами поверхности каналов, интеграция на одном чипе различных стадий процессов для ускорения и автоматизации методик исследований; интеграция с приборами, детекторами, системами пробоподготовки и сенсорами (хроматографами, масс-спектрометрами, лазерами, спектрофотометрами, микроскопами и т.д.); автоматизация процессов: удобство, высокий выход, воспроизводимость, точность; объединение разных стадий методик в одном приборе; уменьшение размеров приборов; появление новых методов и приборов.

Технология микрофлюидики Dolomite находит применения в таких областях как:
химический синтез, аналитическая химия, физико-химические исследования; разработка лекарственных препаратов, определение эффективности и цитотоксичности; биология, диагностика и медицина; экология, производство, приборостроение.

Биология, диагностика и медицина:
качественный и количественный анализ фрагментов НК на чипе капиллярного электрофореза; чипы для секвенирования НК; цифровая капельная ПЦР для количественной ПЦР-диагностики с высокой точностью; анализы крови (биохимические, ИФА, на глюкозу и т.д.); изоляция ДНК из цельной крови; наблюдение за иммобилизованными эмбрионами и клетками.

Компания Eppendorf занимается разработкой, производством и продажей продукции премиум-класса и услуг для лабораторий по всему миру и является одним из лидеров на рынке высокотехнологичного оборудования. Продукция Eppendorf используется в лабораториях самых различных профилей: академических, отраслевых научно-исследовательских, клинико-диагностических, экологических, криминалистических, а также в фармацевтической, биотехнологической, химической и пищевой промышленности и в лабораториях на промышленных предприятиях, которым необходим контроль качества и анализ производственного процесса.
Головной офис компании находится в Гамбурге, Германия.

Продукция Eppendorf представлена в трех направлениях:

LIQUID HANDLING:
- Механические и электронные дозаторы;
- Наконечники для дозаторов;
- Механические и электронные диспенсеры;
- Бутылочные дозаторы, цифровые бюретки;
- Станции автоматического дозирования.

SAMPLE HANDLING:
- Миксеры, термомиксеры и термостаты;
- Пробирки и планшеты;
- Центрифуги и роторы;
- Амплификаторы и расходные материалы для ПЦР;
- Низкотемпературные морозильные камеры.

CELL HANDLING:
- Фотометры, спектрофотометры;
- Микроманипуляторы, микроинъекторы и вспомогательное оборудование;
- Расходные материалы для культивирования клеток и микроскопии;
- CO2-инкубаторы, шейкеры;
- Биореакторы, ферментеры.

ООО «Эппендорф Раша» является представительством немецкой компании Eppendorf AG на территории Российской Федерации. Головной офис компании находится в Москве. Компания осуществляет всестороннюю поддержку пользователей, занимается продвижением продукции на территории РФ, участвует в тематических выставках, конференциях, проводит демонстрацию продукции Eppendorf в шоу-руме в московском офисе и семинары в Москве и других городах России для всех интересующихся продукцией Eppendorf.

ООО «ЭППЕНДОРФ РАША»
115114, Москва,
Дербеневская набережная, д. 11, оф. Б301
Т.: (495) 743-51-23, Ф.: (495) 743-51-22
URL: www.eppendorf.ru,
E-mail: info@eppendorf.ru

Компания предлагает оборудование и реагенты для ПЦР в реальном времени и NGS:

- Реагенты для секвенирования нового поколения 454 (до сентября 2016 года)
- Реагенты для обогащения целевых фрагментов перед NGS – NimbleGen SeqCap (экзомы, метиломы, транксриптомы, панели генов)
- Системы для ПЦР в реальном времени – LightCycler 96 и LightCycler 480
- Системы экстракции нуклеиновых кислот – MagNA Pure LC 2.0, MagNA Pure Compact
- Система для гомогенизации тканей и других твердых образцов – MagNA Lyser

Компания предлагает покупателю не просто свою продукцию, а единую систему, включающую технический сервис, обучение персонала, постоянную методическую поддержку, своевременную доставку реактивов и расходных материалов.

Контакты:
ООО «РОШ ДИАГНОСТИКА РУС»
115114, Москва, ул. Летниковская, д. 2 стр. 2
(БЦ «Вивальди Плаза»)
Т.: (495) 229-69-99
Ф.: (495) 229-62-64
URL: www.roche-applied-science.ru
E-mail: russia.ras@roche.com

Agilent Technologies

Компания Аджилент Текнолоджиз является мировым лидером в разработке и производстве контрольно-аналитического оборудования. На рынке аналитического оборудования Agilent Technologies представляет приборы для молекулярной спектроскопии, элементного анализа, системы для газовой и жидкостной хроматографии, хромато-масс-спектрометрические системы, ПЦР, ПЦР в реальном времени, системы анализа микрочипов, автоматизированные системы дозирования, системы капиллярного электрофореза. Также Agilent Technologies представляет широкий круг продукции для молекулярной биологии, биотехнологии и геномного анализа. Клиентами компании Agilent Technologies являются лаборатории контроля качества и научно-исследовательские лаборатории. Компания Agilent Technologies в России предоставляет консультации по вопросам подбора оптимальной комплектации оборудования, проводит запуск оборудования и обучение работе на приборах, оказывает услуги гарантийного, постгарантийного обслуживания, услуги по квалификации и валидации систем.
Более подробную информацию о компании Аджилент Вы можете найти на сайте компании:

ООО «АДЖИЛЕНТ ТЕКНОЛОДЖИЗ»
115054, Москва,
Космодамианская набережная 52, строение 1 Б
Т.: (495) 664-73-00,
Ф.: (495) 664-73-01
URL: www.agilent.com,
www.your-analytical-solution.com

Компания Qvadros-Bio предлагает современные комплексные решения для исследовательских и медицинских лабораторий, в том числе:

- оборудование для организации биобанков любого масштаба – от ручных систем хранения, сканеров и принтеров штрих-кодов для небольших лабораторий до надежных автоматизированных систем хранения LiCONiC;
- оборудование для организации криобанков – криохранилища, криорезервуары и криогенные трубопопроводы CryoTherm, система хранения образцов C+CRYO, криозамораживатели и 2D криопробирки;
- многозадачные роботизированные платформы для автоматизации различных лабораторных операций: выделение ДНК, постановка ПЦР, пробоподготовка для NGS, масс-спектрометрии и хроматографии, проведение ИФА и аликвотирование;
- современное оборудование для наблюдения за клеточными процессами от пролиферации до ангио- и нейрогенеза, формирования 3D культур.
- Высококвалифицированные специалисты компании QvadroS-Bio окажут максимальную поддержку и проконсультируют по всем вопросам. Мы подберем оптимальное решение под Ваши задачи, составим спецификацию и предложим необходимые расходные материалы и дополнительное оборудование с учетом Ваших требований и пожеланий.

Сертифицированные сервисные инженеры компании QvadroS-Bio, прошедшие обучение на заводах производителей, проводят установку оборудования, обучение персонала, и обеспечивают качественную техническую поддержку, гарантийное и постгарантийное сервисное обслуживание оборудования.

Контакты:
ООО «Квадрос-Био»
127287, г. Москва,
Петровско-Разумовский пр., д.29, стр.4
Тел: +7 (495) 981-80-35
e-mail: info@qvadrosbio.ru

АЛЬБИОГЕН

ООО «АЛЬБИОГЕН» – молодая, активно развивающаяся компания, специализирующаяся на предоставлении полного комплекса услуг, связанного с продажей, всесторонней поддержкой пользователей и сервисным обслуживаем продукции компании Illumina.

- Наша команда состоит из специалистов самого высокого уровня с большим опытом работы как с продукцией компании Illumina, так и в области продаж, продвижения и поддержки.
- Мы с гордостью можем предложить Вам высочайший уровень логистики, который позволяет бережно и в кратчайшие сроки доставлять нашу продукцию до конечного пользователя.

Компания Illumina Inc. (Сан-Диего, США) инновационная, стремительно развивающаяся компания, являющаяся мировым лидером в области секвенирования нового поколения (NGS). Секвенаторы Illumina позволяют осуществлять генетические исследования для науки, медицины, сельского хозяйства, ветеринарии и криминалистики. Более 90% научных статей, связанных с технологиями секвенирования нового поколения, сделаны при помощи оборудования Illumina.

Компания АЛЬБИОГЕН является единственным официальным представителем Illumina в Российской Федерации, Республике Беларусь и Казахстане. Нашей задачей является обеспечение полного доступа клиентов к передовым технологиям и сервисам Illumina, включая наиболее современные системы NGS, биочиповые технологии и весь спектр реактивов.

ООО «АЛЬБИОГЕН»
тел. +7 (499) 550 15 25
info@albiogen.ru, www.albiogen.ru
Адрес: 129085, г. Москва, ул. Звездный бульвар дом 21, строение 3, пом. I, ком.5

Genotek

Компания Genotek оказывает услуги по высокопроизводительному секвенированию и генотипированию различных организмов, начиная с 2010 года и является ведущим поставщиком на рынке генетических услуг.

Наши клиенты – ведущие научно-исследовательские институты и фармкомпании России и стран СНГ, такие как Институт Общей генетики им. Вавилова, Институт цитологии и генетики РАН, Московская медицинская академия им. И.М. Сеченова и др.

Широкий спектр возможностей

Только в Genotek Вы можете выбрать любую из платформ для секвенирования в зависимости от поставленных задач – Illumina HiSeq2000, Illumina MiSeq, SOLID, Ion Proton.

Высокое качество результата

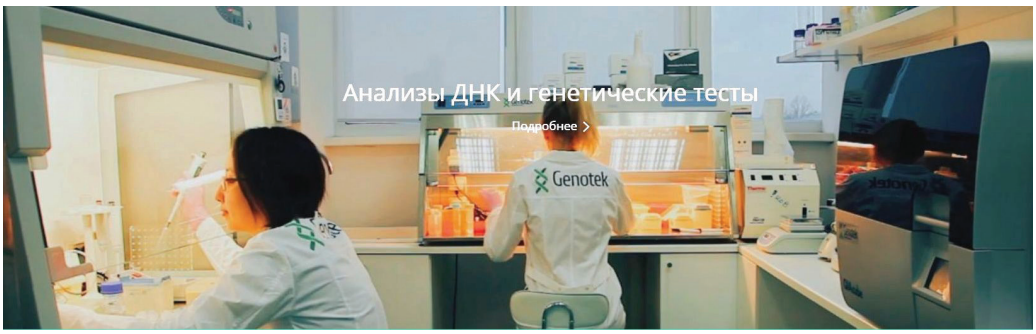
Строгое соблюдение протоколов производителей, использование только оригинальных реактивов и применение многоуровневой системы контроля качества позволяют нам гарантировать высокий уровень достоверности полученных данных.

Оптимальная стоимость

Значительный объем заказов, получаемых компанией, позволяет нам оптимально использовать производственные мощности оборудования, чтобы всегда предлагать Вам самые низкие цены.

Бесплатное консультирование

Не важно – заказываете Вы секвенирование одного ПЦР-продукта или полногеномное исследование сотен человек – Вы можете быть уверены, что получите ответы на любые возникающие вопросы совершенно бесплатно. Просто напишите или позвоните нам!



ООО «ГЕНОТЕК»
105120, г. Москва, Наставнический пер., д. 17,
стр. 1, пом. I, эт. 2, ком. 9.
Т.: (495) 215-15-14
URL: www.genotek.ru
E-mail: info@genotek.ru