

#BelBi2023 • Belgrade, Serbia

BOOK OF ABSTRACTS



4th Belgrade Bioinformatics Conference

HYBRID • 19 - 23 JUNE 2023

EDITORS

Dr. Ivana Morić

Dr. Valentina Đorđević

ISBN: 978-86-82679-14-1

belbi.bg.ac.rs

Title	4 th Belgrade Bioinformatics Conference BOOK OF ABSTRACTS
Publisher	Institute of Molecular Genetics and Genetic Engineering, University of Belgrade Vojvode Stepe 444a, Belgrade, Serbia https://www.imgge.bg.ac.rs/
Editors	dr. Ivana Morić dr. Valentina Đorđević
Technical editor	Dušan Radojević
ISBN	978-86-82679-14-1
Copyright	© 2023 Institute of Molecular Genetics and Genetic Engineering, University of Belgrade

BelBi2023 Committees

International Advisory Board

Alessandro Treves,

International School for Advanced Studies,
Trieste, Italy

Guanglan Zhang,

Health Informatics Lab, Metropolitan College,
Boston University, Boston, USA

Konstantin Severinov,

Waksman Institute for Microbiology,
Rutgers University, Piscataway, USA

Lou Chitkushev,

Health Informatics Lab, Metropolitan College,
Boston University, Boston, USA

Milena Stevanović,

Institute of Molecular Genetics and Genetic
Engineering, University of Belgrade,
Belgrade, Serbia

Oxana Galzitskaya,

Institute of Protein Research,
Russian Academy of Sciences, Moscow, Russia

Paul Sorba,

Laboratory of Theoretical Physics and CNRS,
Annecy, France

Peter Tompa,

VIB Structural Biology Research Center,
Flanders Institute for Biotechnology,
Brussels, Belgium

Predrag Radivojac,

Khoury College of Computer Sciences,
Northeastern University, Boston, SA

Sergey Volkov,

Bogolyubov Institute for Theoretical Physics,
National Academy of Sciences, Kiev, Ukraine

Vladimir Brusić,

Li Dak Sum Chair Professor of Computer Science,
University of Nottingham, Ningbo, China

Vladimir Uversky,

Department of Molecular Medicine,
University of South Florida, Tampa, USA

Yuriy Orlov,

I.M. Sechenov First Moscow State Medical
University, Moscow, Russia

Zoran Ognjanović,

Mathematical Institute, Serbian Academy
of Sciences and Arts, Belgrade, Serbia

International Program Committee

Alexandre de Brevern,

INSERM, Université Paris Cité, Université
de la Réunion, Paris, France

Branislava Gemović,

VINCA Institute of Nuclear Sciences,
University of Belgrade, Serbia

Branka Zukić,

Institute of Molecular Genetics and Genetic
Engineering, University of Belgrade, Serbia

Branko Dragovich,

Mathematical Institute, Serbian Academy of
Sciences and Arts, Belgrade, Serbia

Dragan Matić,

Faculty of Sciences, Department of Mathe-
matics and Informatics, University of Banja
Luka, Bosnia and Herzegovina

George Patrinos,

Department of Pharmacy, University of
Patras, Greece

Gordana Pavlović-Lazetić,

Faculty of Mathematics,
University of Belgrade, Serbia

Hong-Yu OU,

Shanghai Jiao Tong University, The Microbial
Bioinformatics Group, State Key Laboratory
of Microbial Metabolism, Shanghai, China

Ivana Morić,

Institute of Molecular Genetics and Genetic
Engineering, University of Belgrade, Serbia

Jelena Bojović,

Center for the Fourth Industrial Revolution,
Belgrade, Serbia

Marko Đorđević,
Faculty of Biology, University of Belgrade, Serbia

Mirjana Maljković,
Faculty of Mathematics,
University of Belgrade, Serbia

Nataša Kovacevic-Grujičić,
Institute of Molecular Genetics and Genetic
Engineering, University of Belgrade, Serbia

Nataša Pržulj,
Catalan Institution for Research and
Advanced Studies (ICREA), Spain;
Barcelona Supercomputing Center, Spain;
University College London, UK

Nenad Mitić,
Faculty of Mathematics,
University of Belgrade, Serbia

Nevena Veljković,
VINCA Institute of Nuclear Sciences,
University of Belgrade, Serbia

Predrag Radivojac,
Khoury College of Computer Sciences,
Northeastern University, Boston, USA

Saša Malkov,
Faculty of Mathematics,
University of Belgrade, Serbia

Sergei Kozyrev,
Department of Mathematical Physics,
Steklov Mathematical Institute RAS,
Moscow, Russia

Valentina Đorđević,
Institute of Molecular Genetics and Genetic
Engineering, University of Belgrade, Serbia

Vladimir Babenko,
Institute of Cytology and Genetics,
Novosibirsk, Russia

Local Organizing Committee

Anđela Rodić,
Faculty of Biology, University of Belgrade

Branislava Gemović,
VINCA Institute of Nuclear Sciences,
University of Belgrade

Branko Dragovich,
Mathematical Institute, Serbian Academy
of Sciences and Arts, Belgrade

Gordana Pavlović-Lazetić,
Faculty of Mathematics,
University of Belgrade

Ivana Morić,
Institute of Molecular Genetics and Genetic
Engineering, University of Belgrade

Jovana Kovačević,
Faculty of Mathematics,
University of Belgrade

Marko Đorđević,
Faculty of Biology, University of Belgrade

Marko Živanović,
Bioengineering Research and Development
Center

Nenad Filipović,
Faculty of Engineering, University of Kragujevac

Nenad Mitić,
Faculty of Mathematics,
University of Belgrade

Radoslav Davidović,
VINCA Institute of Nuclear Sciences,
University of Belgrade

Saša Malkov,
Faculty of Mathematics,
University of Belgrade

Valentina Đorđević,
Institute of Molecular Genetics and Genetic
Engineering, University of Belgrade

Željko D. Popović,
Faculty of Sciences, University of Novi Sad

FOREWORD

Dear colleagues and friends,

The 4th Belgrade Bioinformatics Conference - BelBi2023, where many high-quality scientific contributions were presented, has just ended. With great thanks to all participants, we now proudly present a book of abstracts that both reflects the scientific abundance and diversity of the conference and serves as a reminder of a memorable event.

Several research institutions, faculties, and scientific societies from Serbia joined forces in organizing this international conference, which covered numerous topics in computational biology, bioinformatics, and biomedical and health informatics. The main goal of BelBi2023 was to foster contact between scientists, both early stage career and senior researchers, allowing them to share experiences and latest advances in their fields. We sincerely hope that BelBi2023 has served as a platform for researchers from around the world to meet, initiate new collaborations, and expand professional contacts, and that all of you would become a part of the growing BelBi community.

We are grateful and proud to have welcomed more than 250 researchers from 21 countries. We have had 28 scientific sessions, consisting of more than 60 lectures (including eight Keynote talks), 47 presented posters, as well as three workshops and one satellite event – COST action. We have also organized seven industry lectures, including the NGS Challenge,

two Meet the Expert Sessions, and one Business Coffee Break where ten start-up companies took part. And finally, the future BIO4 campus was presented and first panel on Serbia's resources for storage and analyses of genetic data was organized.

We would like to thank all the members of the International Advisory Board and the International Program Committee for their efforts and help in making this event a success. We are very grateful to the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, SAIGE project, and UNDP-Serbia for their support. Finally, the Local Organizing Committee is very grateful to all the sponsors of the conference - BGI, Illumina & Elta'90MS, PacBio & East Diagnostics, ThermoFisher Scientific & Vivogen, Huawei, Labena, DSP Chromatography, RNIDS, Telekom Srbija, Alfa Genetics, Kefo and Superlab, hoping that they will stay with us for many years to come.

Looking forward to seeing you again at the 5th Belgrade Bioinformatics Conference.

Belgrade, July 2023

*Dr. Valentina Đorđević
& Dr. Ivana Morić,*
On behalf of BelBi2023
Organizing Committee

ORGANIZER



Institute of Molecular Genetics
and Genetic Engineering,
University of Belgrade

MAIN CO-ORGANIZERS



Faculty of Biology,
University of Belgrade



Faculty of Mathematics,
University of Belgrade



Mathematical Institute
of SAsA,
Belgrade



Vinča Institute of
Nuclear Sciences,
University of Belgrade



Serbian Society for Bioinformatics
and Computational Biology

CO-ORGANIZERS



Faculty of Sciences,
University of Novi Sad



BioIRC - Bioengineering
Research and
Development Center



Faculty of Engineering,
University of Kragujevac



C4IR Serbia

Centre for the Fourth
Industrial Revolution
in Serbia



Serbian Society for
Molecular Biology



Biophysical Society
of Serbia

TABLE OF CONTENT

KEYNOTE LECTURES

Big Data in Biology: How EMBL delivers big data for biology, and some highlights of its application to human disease biology <i>Ewan Birney</i>	1
A tale of two stories: data-driven precision medicine and precision public health <i>Kristel Van Steen</i>	2
Machine intelligence and network science for complex systems big data analysis <i>Carlo Vittorio Cannistraci</i>	3
Targeting LLPS in disease: a new modality in drug development <i>Peter Tompa</i>	4
Translating Bioinformatics Back To Healthcare: Facilitating the use of Artificial Intelligence at UW Medicine <i>Sean D. Mooney</i>	5
Bioinformatics education course on gene networks reconstruction using online tools <i>Yuriy L. Orlov, Anastasia A. Anashkina</i>	6
A New Framework for the Use of Variant Interpretation Tools in Clinical Practice <i>Predrag Radivojac</i>	7
Privacy-preserving Systems Medicine <i>Jan Baumbach</i>	8

INVITED LECTURES

Using AI/ML to transform molecular biology databases <i>Alex Bateman</i>	9
Stereo-seq: Large Field of View-Spatially Resolved Transcriptomics at Nanoscale Resolution <i>Javier Batista Perez</i>	10
“What is life?”: Open quantum systems approach <i>Andrei Khrennikov</i>	11
Exploiting the linear organisation of omics network embedding spaces <i>Noël Malod-Dognin, Alexandros Xenos, Sergio Doria Belenguer, and Nataša Pržulj</i>	12
Inverting convolutional neural networks for super-resolution identification of regime changes in epidemiological time series <i>Jose M. G. Vilar</i>	13
Can we use biobanks to study infectious diseases? <i>Andrea Gelemanovič</i>	14

An integrated platform for genome assembly, comparative genomics and management of genomic variation databases <i>Jorge Duitama</i>	15
Bioinformatics and evolution of non-model organisms <i>Mikhail S. Gelfand</i>	16
Computational bioengineering for heart disease <i>Nenad Filipovic, Themis Exarchos, and Djordje Jakovljevic</i>	17
A Similarity-based Normative Framework for Bio-plausible Neural Nets <i>Anirvan Sengupta</i>	18
Complexity driven evolution of Alternative splicing <i>Vladimir Babenko and Timophej Boltunov</i>	19
Pangenomic Alignment: Strings plus Graphs <i>Travis Gagie</i>	20
Circular Codes in the Genetic Information <i>Elena Fimmel and Lutz Strüngmann</i>	21
Some Applications of Graph-Based Machine Learning Methods on Biological Data <i>Mladen Nikolić</i>	22
From multifunctionality to polypathogenicity with intrinsic disorder <i>Vladimir N. Uversky</i>	23
Exploring the impact of rare Copy Number Variants on miRNA genes in CAKUT: Insights from integrated bioinformatic analysis and experimental validation <i>Ivan Jovanović</i>	24
Persistence of plasmids targeted by CRISPR interference in bacterial populations <i>Konstantin Severinov</i>	25
Omics Data Fusion for Understanding Molecular Complexity Enabling Precision Medicine <i>Nataša Pržulj</i>	26
An agnostic analysis of the human AlphaFold2 proteome using local protein conformations <i>Alexandre G. de Brevern</i>	27
Uncovering resistance to microtubule targeting drugs <i>Mattia Pavani, Elena Chiroti, Paolo Bonaiuti, and Andrea Ciliberto</i>	28
Computational tools and repositories for precision therapeutics in the post-genomic era <i>George P. Patrinos</i>	29
Development of hybrid and optimized deep learning classifiers for speech recognition in tracheostomy patients: a case study <i>Themis Exarchos</i>	30

Advancing Genomics with OrthoDB, BUSCO, and the LEM Framework <i>EV Kriventseva, M Manni, M Seppay, F Tegenfeldt, M Berkeley, D Kuznetsov, EM Zdobnov.....</i>	31
Multomics Integration by Non-Negative Tri-Matrix Factorization Reveals New Target Genes in Parkinson's Disease <i>Alexander Skupin.....</i>	32
Prediction of cell types using single-cell mRNA profiles <i>Vladimir Brusic.....</i>	33
The complete solution and interpretation algorithms for large field-of-view and high-resolution spatial transcriptomics <i>Shuangfang Fang.....</i>	34
To be folded, to be unfolded or to be aggregated with important functions: application of the directed coaggregation mechanism to combat bacterial communities <i>O.V. Galzitskaya, S.Yu. Grishin, A.V. Glyakina, M.V. Slizen, A.V. Panfilov, P.A. Domnin, A.P. Kochetov, A.A. Surin, S.V. Kravchenko, A.K. Surin, S.A. Ermolaeva.....</i>	35

ORAL PRESENTATION

CADDIE - An online knowledge base for network-based mechanism exploration and drug repurposing in oncology <i>Michael Hartung, Elisa Anastasi, Zeinab M. Mamdouh, Cristian Nogales, Harald HHW Schmidt, Jan Baumbach, Olga Zolotareva, and Markus List.....</i>	36
Mapping of Disease Names to Disease Codes based on Natural Language Processing Techniques <i>Anđelka Zečević, Jovana Kovačević, and Radoslav Davidović.....</i>	37
Zero- and Few-Shot Machine Learning for Named Entity Recognition in Biomedical Texts <i>Miloš Košprdič, Nikola Prodanović, Adela Ljajić, Bojana Bašaragin, and Nikola Milošević.....</i>	38
Clustering and classification of SARS-COV-2 isolates using RSCU <i>S. Malkov, M. Beljanski, G. Pavlović Lažetić, B. Stojanović, M. Maljković, A. Veljković, S. Kapunac, and N. Mitić.....</i>	39
The use of Active Machine Learning for Protospacer-Adjacent Motif recovery in Class 2 CRISPR-Cas systems <i>Bogdan Kirillov, Aleksandra Vasileva, Oleg Fedorov, Maxim Panov, and Konstantin Severinov.....</i>	40
Application of classification algorithms for hip implant surface topographies <i>Aleksandra Vulović, Tijana Geroski, and Nenad Filipović.....</i>	41
Computational Modelling of Drug Effects on Cardiomyopathy and Analysis of Myocardial Work <i>Smiljana Tomasevic, Miljan Milosevic, Bogdan Milicevic, Vladimir Simic, Momcilo Prodanovic, Srbojjub M. Mijailovich, Nenad Filipovic.....</i>	42

Echocardiography-based Left Ventricle Cardiac Hypertrophy Simulations <i>Bogdan Miličević, Miljan Milošević, Vladimir Simić, Danijela Trifunović, Goran Stanković, Nenad Filipović, and Miloš Kojić.....</i>	43
Decoding Cystic Fibrosis Phenotype <i>Aleksandra Divac Rankov, Dušan Ušjak, Martina Mia Mitić, and Jelena Kusic Tisma.....</i>	44
Single cell 3' transcriptome profiling <i>Nevena Milivojević, Uršula Prosenec Zmrzljak, Biljana Ljujić, Valentina Đorđević, Marina Gazdić Janković, Marko Živanović, Feđa Puač, Miloš Ivanović, and Nenad Filipović.....</i>	45
Modulating Horizontal Gene Transfer through Bistability in the Dynamics of Bacterial Restriction-Modification Systems <i>Marko Djordjevic, Lidija Zivkovic, and Magdalena Djordjevic.....</i>	46
Cell-type-specific mechanistic drivers of progressive multiple sclerosis lesions <i>Elkjaer ML, Hartebrødt A, Oubounyt M, Weber A, Vitved L, Reynolds R, Thomassen M, Rottger R, Baumbach J, and Illes Z.....</i>	47
AI-powered framework to predict the toxicity of microplastics <i>Junli Xu.....</i>	48
Newest Advances on the FeatureCloud Platform for Federated Learning in Biomedicine <i>Niklas Probul, Mohammad Bakhtiari, Mohammad Kazemi Majdabadi, Balázs Orbán, Sándor Fejér, Supratim Das, Julian Klemm, Christina C Saak, Nina K Wenke, and Jan Baumbach.....</i>	49
Deciphering key regulatory networks and drug repurposing candidates through scRNAseq data analysis using SCANet <i>Mhaned Oubounyt, Jan Baumbach, and Maria L. Elkjaer.....</i>	50
From protein-protein to isoform-isoform interactions: the toolkit to map alternative splicing to interactome <i>Olga Tsoy, Zakaria Louadi, Chit Tong Lio, Jan Baumbach, Olga Kalinina, Alexander Gress, Tim Kacprowski, and Markus List.....</i>	51
Drugst.One - A plug-and-play solution for online systems medicine and network-based drug repurposing <i>Andreas Maier, Michael Hartung, The Drugst.One Initiative, and Jan Baumbach.....</i>	52
Fatty Acid Data Analysis Unravels Skeletal Site and Age-Specific Features of Human Bone Marrow Adiposity <i>Drenka Trivanović, Jovana Kovačević, Aleksandra Arsić, Marko Vujačić, Nikola Bogosavljević, Ivana Oklič Djordjević, Milena Živanović, Slavko Mojsilović, Mirjana Maljković, and Aleksandra Jauković....</i>	54
Exploration of Pharmacogenomic Biomarkers in Chronic Immune Diseases Using Single-Cell RNA Sequencing <i>Mario Gorenjak, Larisa Goričan, Boris Gole, Uršula Prosenec, Erik Melén, Michael Kabesch, Anke H Maitland-van der Zee, Susanne Reinartz, Susanne J H Vijverberg, Uroš Potočnik and the PERMEABLE consortium.....</i>	55

Dehydrins in the service of protecting the DNA helix from the aspect of molecular dynamics (MD) <i>Milan Senčanski, Ivana Prodić, Ana Pantelić, and Marija Vidović.....</i>	57
Machine learning approach in inferring main population-level COVID-19 risk factors <i>Sofija Marković, Anđela Rodić, Ognjen Milićević, Igor Salom, Magdalena Đorđević, and Marko Đorđević.....</i>	58
Using whole exome sequencing to explore genetic basis of unicuspid aortic valve disease <i>Martina Mia Mitić, Dušan Ušjak, Maja Milošević, Marija Cumbo, Sofija Dunjić Manevski, Branko Tomić, Ivana Petrović, Petar Otašević, Slobodan Micović, Milovan Bojić, and Valentina Đorđević.....</i>	59
Online <i>in silico</i> validation of disease and gene sets, clusterings or subnetworks with DIGEST <i>Klaudia Adamowicz, Andreas Maier, Jan Baumbach, and David B. Blumenthal.....</i>	60
Alternative splicing impacts microRNA regulation within coding regions <i>Lena Maria Hackl, Amit Fenn, Zakaria Louadi, Jan Baumbach, Tim Kacprowski, Markus List, and Olga Tsoy</i>	61
Using AI to design antibodies <i>Goran Rakočević.....</i>	62
Semantic unification and search of bioinformatics databases <i>A. Veljković, and N. Mitić.....</i>	63
Beyond the Global Health Security Index: A Machine Learning Approach to Analyzing the Official COVID-19 Deaths and Excess Deaths Data <i>Andjela Rodic, Sofija Markovic, Igor Salom, and Marko Djordjevic.....</i>	64
Integration of differential transcriptomic and proteomic data in hydrated and desiccated leaves of <i>Ramonda serbica</i> Panc. <i>Marija Vidović, Ilaria Battisti, Ana Pantelić, Dejana Milić, Giorgio Arrigoni, Antonio Masi, and Sonja Veljović Jovanović.....</i>	65

POSTER PRESENTATION

Possible role of estrogen metabolism and aldo-keto reductase activity in chemoresistance of ovarian cancer <i>Nika Marolt, Andrew Walakira, Tadeja Režen, Damjana Rozman and Tea Lanišnik Rižner.....</i>	66
Seven miRNAs potentially included in the chilling response of maize plants in early stages of development <i>Manja Božić, Dragana Ignjatović-Mičić, Nenad Delić, Marko Mladenović, Jelena Vančetović, Bojana Banović Đeri, Ana Nikolić.....</i>	67
Two contrasting late embryogenesis abundant protein family groups of <i>Ramonda serbica</i> Panc. <i>Ana Pantelić, Strahinja Stevanović, Sonja Milić Komić, Nataša Kilbarda and Marija Vidović.....</i>	68

De novo Genome Assembly of Sweet Chestnut (<i>Castanea sativa</i> Mill.) Insights into the Molecular Basis of its Nutritional Properties <i>M. Aydin Akbudak and Ali Tevfik Uncu</i>	69
Numerical and Biological Modeling Approach in the Analysis of the Cancer Viability and Apoptosis <i>Katarina Virijević, Marko Živanović, Marina Gazdić Janković, Amra Ramović Hamzagić, Nevena Milivojević, Katarina Pecić, Dragana Šeklić, Milena Jovanović, Nikolina Kastratović, Ana Mirić, Tijana Đukić, Ivica Petrović, Vladimir Jurišić, Biljana Ljujić, Nenad Filipović</i>	70
Root colonization ability of herbicide-resistant PGP bacteria evaluated by 16S rRNA metabarcoding <i>Cristina Bez, Ivana Galic, Iris Bertani, Nada Stankovic, Vittorio Venturi</i>	71
Genetic Complexity and Synteny Analysis of Castanea Genomes: Unveiling the Significance of Chestnut Species in Ecological and Genomic Perspectives <i>Ali Tefvik Uncu and M. Aydin Akbudak</i>	72
Elongation factor P (-like) protein and polyproline motifs <i>Marina Parr, Alina Sieber, Prof. Dr. Dmitrij Frishman and Dr. Jürgen Lassak</i>	73
Comparative study of in silico protein design techniques <i>Ivan Tanasijević and Branka Rakić</i>	74
Energy and information exchange between “donor” and “molecular bridge” structures: non adiabatic polaron model <i>Dalibor Chevzovich, Vasilije Matic, and Zeljko Przulj</i>	75
Profiling Pre-Replication Complex Mutations in Cancer <i>Jelena Kusic Tisma, Marija Orlic Milacic, Quang Trinh, Rhea Ahluwalia, Lincoln D. Stein</i>	76
Combined experimental and theoretical study of Type-II toxin-antitoxin system response to antibiotics <i>Bojana Ilic, Marko Đorđević, Hong-Yu Ou</i>	77
Methodology, performance and retrainability survey of intrinsic disorder predictors <i>Nevena Ćirić and Jovana Kovačević</i>	78
Evaluating ND1 and Cytb mitochondrial genes as markers for diversity analysis of protected White-tailed eagle species from Serbia <i>Slobodan Davidovic, Milica Stanković, Pavle Erić, Katarina Erić, Aleksandra Patenković and Marija Tanasković</i>	79
Analysis of nucleotide sequence repeats in coronaviruses <i>S. Kapunac, S. Malkov, M. Beljanski, G. Pavlović Lažetić, B. Stojanović, M. Maljković, A. Veljković, N. Mitić</i>	80
Genomic Surveillance and Phylogenetic Analysis of SARS-CoV-2 Variants in Serbia: Insights into Evolutionary Dynamics and Genetic Diversity <i>Mirjana Novkovic, Bojana Banovic Djeri, Sasa Todorovic, and Valentina Djordjevic</i>	81

Deciphering the reward-related impulsivity domains in rats: The big data study of historical control <i>Jovana Arandelović, Kristina Mirković, Jana Kojić, Miroslav Savić</i>	82
Computer analysis of glioma gene network structure <i>Iarema P.O., Turkina V.A., Mayorova A.A., Orlov Y.L.</i>	83
Genome-wide association analysis for severe COVID-19 in Serbian population <i>Marko Zecevic, Nikola Kotur, Bojan Ristivojevic, Vladimir Gasic, Branka Zukic, Sonja Pavlovic and Biljana Stankovic</i>	84
Impact of different mapping tools on detection of small RNAs in bacterial outer membrane vesicles <i>Bojana Banović Đeri, Sofija Nešić, Ana Pantelić, Jelena Samardžić, Dragana Nikolić</i>	85
<i>In silico</i> pre-selection of β -glucosidase gene for heterologous recombinant expression <i>Marija Atanaskovic, Ivana Moric, Milos Rokic, Lidija Senerovic</i>	86
Supervised Machine Learning Approach for Prediction of Occult Lymph Node Metastasis in T1-T2 Papillary Thyroid Carcinoma <i>Marina Popović Krneta, Nemanja Krajinović, Zoran Bukumirić, and Miljana Tanić</i>	87
Determinants of CRISPR array non-canonical adaptation mechanism <i>Marko Tumbas and Marko Đorđević</i>	88
Data mining for long-non coding RNAs deregulated in colon cancer through analysis of Gene Expression Omnibus database <i>Iva Pruner and Aleksandra Nikolic</i>	89
Efficient bioinformatics workflow for <i>de novo</i> transcriptome assembly of <i>Pelargonium zonale</i> <i>Dejana Milić, Ana Pantelić, Jelena Samardžić, Bojana Banović Đeri, Marija Vidović</i>	90
Application of principal component analysis (PCA) and analytical hierarchy process (AHP) in analysis of articulatory characteristics of phonemes of children with 22q11.2 Deletion Syndrome <i>Danijela Drakulic, Marijana Rakonjac, Goran Cuturilo, Natasa Kovacevic-Grujicic, Jelena Kusic-Tisma, Ivana Moric, Branka Zukic, and Milena Stevanovic</i>	91
Integrated relational database of human protein-protein interactions <i>Bojana Jošić, Jovana Kovačević, Vladimir Perović, Nevena Veljković</i>	92
Mining for the data about glycosylation in the bovines- the analysis of the recently published studies <i>Anđelo Beletić, Ivana Duvnjak Orešković, Tea Pribić, and Gordan Lauc</i>	93
Different approaches in microRNA analysis <i>Barbara Jenko Bizjan, Bine Stančić, Iva Sabolić, Maja Štalekar and Uršula Prosenč Zmrzljak</i>	94

<i>In silico</i> analysis and prediction of novel pharmacogenomic markers of pediatric ALL treatment <i>Vladimir Gašić, Nikola Kotur, Biljana Stanković, Đorđe Pavlović, Marina Jelovac, Jelena Perić, Bojan Ristivojević, Sonja Pavlović, and Branka Zukić</i>	95
Exploring Changes in Diagnoses during the COVID-19 Era: Comparative Analysis <i>Despina Misheva, Marija Stojcheva, Hana Hasanicaaj, Ana Mladenovska, Jovana Dobрева, Mary Lucas, Irena Vodenska, Lou Chitkushev, Dimitar Trajanov</i>	96
Shotgun metagenomics reveals gut microbiota features associated with the efficacy of myeloid derived suppressor cells in the prevention of neuroinflammation <i>Marina Bekić, Nataša Ilić, Jelena Đokić, Dušan Radojević, Dragana Vučević, Saša Vasilev, and Sergej Tomić</i>	97
Seeking an optimal variant calling pipeline for medical genetics <i>Yury A. Barbitoff, Alexandra Panteleeva, Alexander V. Predeus</i>	98
Groundwater and soil as a reservoir for polyurethane-degrading bacteria <i>Milica Ciric, Brana Pantelic, Vladimir Šaraba, and Jasmina Nikodinovic-Runic</i>	99
Developing bioinformatics pipeline for processing environmental DNA metabarcoding sequencing data <i>Iva Sabolić, Lucija Markulin, Teja Petra Muha, Barbara Jenko, Uršula Prosenc Zmrzljak</i>	100
Evaluation of variant calling tools for detection of SNVs in BRCA1 and BRCA2 genes in patients from the Institute of Oncology and Radiology of Serbia <i>Isidora Pantović, Katarina Živić, Ivana Boljević, Milica Nedeljković, Radmila Janković, Miljana Tanić</i>	101
Transcriptome analysis of <i>Pseudomonas aeruginosa</i> after MhqO dioxygenase treatment <i>Andjela Djokic, Ivana Moric, Lidija Senerovic and Lidija Djokic</i>	102
PACSIN2 modifies miRNAs in extracellular vesicles, modulating thiopurine response <i>Alessia Norbedo, Marianna Lucafò, Carlotta Bidoli, Marco Gerdol, Metka Lenassi, Giuliana Decorti, Gabriele Stocco</i>	103
Pathway analysis of CD8 ⁺ T cell transcriptome in glioblastoma patients reveals multiple sclerosis signaling pathway as the top rated upregulated disease pathway in tumor infiltrating cells <i>Milan Stefanović, Ivan Jovanović, Aleksandra Stanković and Maja Živković</i>	104
Transcriptomic profiling of white blood cells reveals new insights into the molecular mechanisms of thalidomide in children with inflammatory bowel disease <i>Marianna Lucafò, Letizia Pugnetti, Debora Curci, Carlotta Bidoli, Marco Gerdol, Fulvio Celsi, Sara Renzo, Monica Paci, Sara Lega, Paolo Lionetti, Alberto Pallavicini, Giuliana Decorti, Gabriele Stocco, Matteo Bramuzzo</i>	105
The past, the present, and the future of RNA secondary structure prediction <i>Lazar Vasović</i>	106

The use of tryptic food protein digests data in public proteomic repositories to assess the effects of chemical and post-translational modifications on digestion outcomes <i>Ivana Prodić, Teodora Đukić, Vesna Jovanović, Katarina Smiljanić</i>	107
Machine learning-based data correlation between scanning electron microscopy images and energy-dispersive X-ray spectroscopy profiles <i>Ahmed Musa, Baekkyoung Sung, Leon Abelmann</i>	108
Protein structural differences in Cytochrome c oxidase subunit 1 of two <i>Heterogynis</i> species as a new approach for species delimitation <i>Marija Vidović, Vladislava Galović</i>	109
Potentially relevant variants of unknown significance in NGS-tested patients with suspected skeletal dysplasia <i>Marija Mijović, Goran Cuturilo, Jelena Ruml Stojanović, Aleksandra Miletić, Brankica Bosankić, Hristina Petrović, Bojana Vasić, and Nadja Vukasinović</i>	110
Analysis of Long COVID Phenotypes and their Impact on Mental Health and Daily Functioning: Insights from Twitter <i>Marko Marković, Jovana Dobrova, Mary Lucas, Irena Vodenska, Lou Chitkushev, Dimitar Trajanov</i>	111
Metagenomic Analysis of Bacterial Community and Isolation of Representative Strains from Vranjska Banja Hot Spring, Serbia <i>Jovana Curčić, Danka Matijasević, Nemanja Stanisavljević, Srđan Tasić, Milan Kojić, and Milka Malešević</i>	112

Clustering and classification of SARS-COV-2 isolates using RSCU

S. Malkov*¹, M. Beljanski¹, G. Pavlović Lažetić¹, B. Stojanović², M. Maljković¹, A. Veljković¹, S. Kapunac¹, and N. Mitić¹

¹Faculty of Mathematics, University of Belgrade, Studentski trg 16, 11000 Belgrade, Serbia

²Mathematical Institute SASA, Knez Mihaila 36, 11000 Belgrade, Serbia
sasa.malkov@matf.bg.ac.rs

The existence of a large number of sequenced SARS-COV-2 isolates provides an opportunity to observe genomic variability in a massive sample. The goal of our research was to use data mining techniques to study possible correlation between codon usage and classification by WHO-labels in a certain period of time.

The material includes 745,533 isolates with 12,236,672 coding sequences (proteins) from NCBI (10.08.2022.). RSCU was used as a measure of codon usage. Samples are associated with WHO-labels (based on Pango_Id) and time intervals. Inconsistency of WHO-labels with periods in which the respective strains were actually present was observed. The isolates with the observed discrepancy were excluded from the sample. Isolates without assigned WHO-labels were also excluded. In addition, individual coding sequences containing ambiguous nucleotide codes were eliminated.

Clustering was performed for each of the 12 common types of coding sequences (proteins), with multiple methods and a different number of clusters. Neural clustering gave the best results. For different protein types, different degrees of RSCU variability are observed. In the case of proteins with a small variation in nucleotide contents, over 95% of the material belongs to a single cluster, while the other clusters are of negligible size. In the case of proteins with more variations, a higher number of pure clusters (by WHO-labels) is obtained, with a small number of heterogeneous clusters (about 10% of the material). In those heterogeneous clusters, there are isolates with different WHO-labels that were present in parallel at some point, as a kind of transitional forms between two strains.

Different classification models were created on the same sample. Models based on protein types with higher diversity between coding sequences are highly accurate (96-100%). Using the classification models, the corresponding WHO-labels were associated with isolates without previously assigned WHO-labels.

Keywords: SARS-COV-2, RSCU, clustering, classification



ISBN: 978-86-82679-14-1