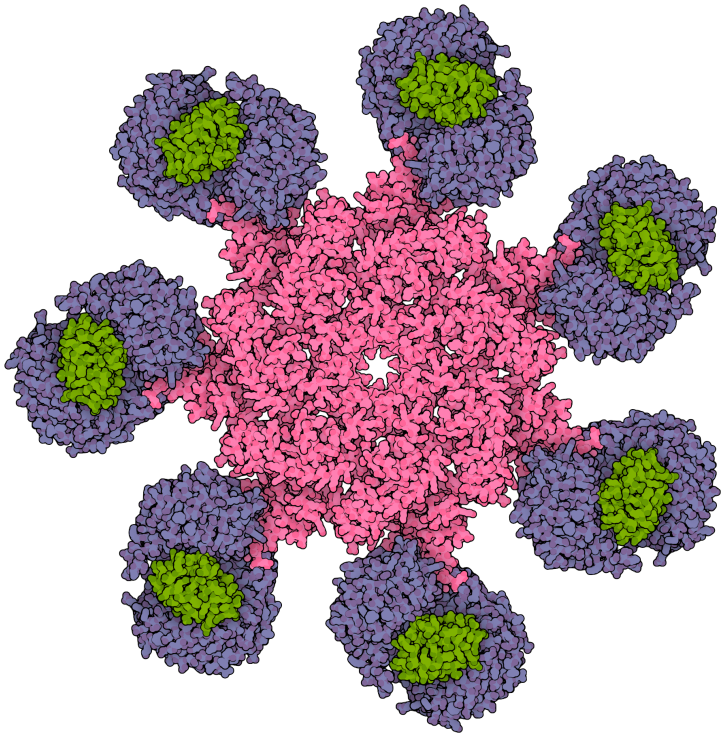
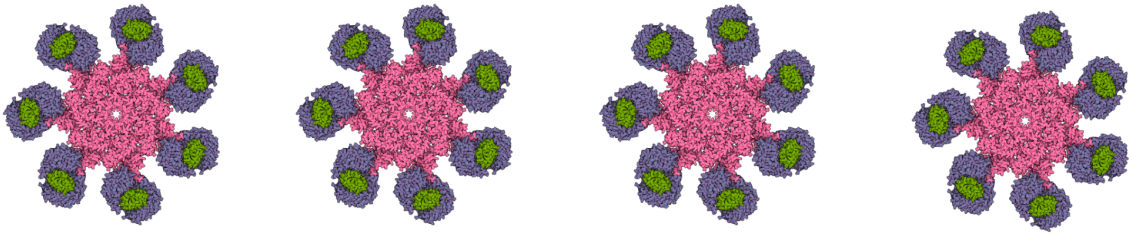


Meeting of the EMBO Young Investigator Network on computational methods in ecology and evolutionary biology of microbes

Book of Abstracts

7-9.07.2022

Chęciny, Poland



Organizers

Anna Karnkowska (University of Warsaw)

Rafał Mostowy (Jagiellonian University in Kraków)

Stanisław Dunin-Horkawicz (University of Warsaw)

Funding

EMBO

Program

Thursday, 7th of July

12:30 - 13:30 Lunch

13:30 - 13:40 Welcome

13:40 - 14:10 Adam Gudyś
K-mer-based analysis of microbial genomes

14:10 - 14:23 Metody Hollender
Workflow for genome recovery of protistan prokaryotic symbionts from single-cell data

14:23 - 14:36 Małgorzata Orłowska
Fucose in fungi

14:36 - 14:49 Paweł Hałakuc
How to investigate complex genomes in diverse taxa

14:49 - 15:02 Michał Karlicki
From huge metagenomes to single plastid genomes – preliminary results of searching for plastid genomes in freshwater metagenomic data

15:05 - 15:30 Coffee break

15:30 - 16:00 Tomasz Kościółek
Understanding and Shaping the Human Gut Microbiome for Health using Bioinformatics and Machine Learning

16:00 - 16:13 Rafał Madaj
In silico analysis of selected nerve agents' non-covalent binding affinity towards acetylcholinoesterase

16:13 - 16:26 Bogna Smug
Domain architecture shows extensive mosaicism of phage proteins engaged in host tropism

16:26 - 16:39 Krzysztof Szczepaniak
Evolutionarily conserved fragments: an attempt to better understand domain composition and mosaicism of phage proteins

16:39 - 16:54 Kamil Kamiński
What does a neural actually learn? The case of reengineering protein-ligand binding affinity

16:55 - 17:30 Free time

17:30 - 19:00 Castle

19:00 Dinner

Friday, 8th of July

08:30 - 09:30 Breakfast

09:30 - 10:00 Jakub Barylski
Tell me who your enemy is, and I will still not tell you know who you are - the hard problem of phage-host prediction

10:00 - 10:13 Valentina Smacchia
Use of metabarcdoing for the identification of freshwater microbial communities along an eutrophication gradient

10:13 - 10:26 Marta Sałek
Finding interaction patterns between protists and eukaryotes based on single-cell microbiome approach: obtaining good quality data for computational analyses

10:26 - 10:39 Kamil Krakowski
Functional analysis of prokaryotic homologs of programmed cell death factors

10:40 - 11:10 Coffee break

11:10 - 11:40 Maria Górna
CLIPPERs mediate targeted protein degradation in Gram-negative bacteria

11:40 - 11:53 Wanangwa Ndovie
ANImm: A fast and accurate tool for calculating average nucleotide identity between pairs of viral genomes

11:53 - 12:06 Julia Gołębiowska
Predicting absorption wavelengths of rhodopsins using machine learning-based methods

12:07 - 12:20 Vyshakh R Panicker
Lytic Phage Depolymerases and Capsule Specificities: Insights from Literature and Sequence Analysis

12:30 - 13:30 Lunch

13:30 - 14:00 Anna Muszewska
Ancestral traits in the genomes of Mucoromycota fungi

14:00 - 14:13 Kacper Maciszewski
In the spotlight, losing IR region: challenging the importance of inverted repeats in plastid genomes

14:13 - 14:26 Stanisław Antonowicz
Workflow for inferring microbial interaction networks based on amplicon sequencing data

14:26 - 14:39 Jędrzej Kubica
Modeling of a putative programmed cell death receptor

14:39 - 14:45 Closing

14:45 - 15:15 Coffee break

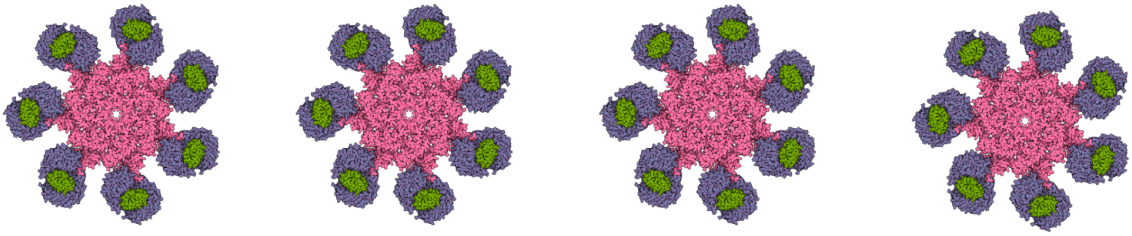
15:15 - 16:30 Free time

16:30 - 18:00 Walk to the nature reserve

19:00 Bonfire

Saturday, 9th of July

08:30 - 09:30 Breakfast



Abstracts

Understanding and Shaping the Human Gut Microbiome for Health using Bioinformatics and Machine Learning

Tomasz Kościółek[1]

[1] Małopolska Centre of Biotechnology, Jagiellonian University in Krakow, Poland

The microbiome harbors tens of trillions of microbes represented by more than 2 million unique genes. Through its malleability and links to health the microbiome is an attractive target for research and therapeutic interventions. Hindering this is our limited understanding of gene functions and metabolic potential encoded within the microbiome.

Currently, we can functionally annotate less than 50% of microbial genes. To address this, we devised a synergistic approach in which through large-scale grid computations we predict de novo 3D protein structures of microbial proteins from sequence. Then, using those structures and a deep learning we annotate gene function with higher accuracy and coverage. Those results are used in a custom metagenomic annotation pipeline for high-accuracy and high-coverage annotations of metagenomic data.

Thus, we build a protein sequence-structure-function link, which opens new avenues for understanding microbiome-phenotype relationships and, in future, microbiome-oriented therapies.

Domain architecture shows extensive mosaicism of phage proteins engaged in host tropism

Bogna J Smug[1], Krzysztof Szczepaniak[1], Stanislaw Dunin-Horkawicz[2], Rafal J. Mostowy[1]

[1] Małopolska Centre of Biotechnology, Jagiellonian University in Krakow, Poland
[2] Centre of New Technologies, University of Warsaw, Warsaw, Poland.

For milions of years bacteriophages have been coevolving with their hosts. We hypothesized such coevolution has shaped the domain architecture of phage proteins engaged in host tropism, like receptor-binding-proteins (RBPs). To address it, we downloaded all complete phage genomes from NCBI RefSeq and created deep HMM profiles for all proteins therein using UniClust30. We used remote homology detection (HHblits) for an all-by-all profile-profile comparison, protein domain detection using Pfam and putative function assignment to each profile using homology detection with three complementary approaches (PHROGs, GOs, PhANN).
We show that the function of a structural protein often cannot be uniquely determined and, that this is related to domain sharing between proteins of various functions. We then systematically compared domain architectures of structural proteins from various functional categories. Domains found within these proteins were shared between otherwise dissimilar proteins and co-occurred in multiple combinations. Such understood mosaicism was most frequent within RBPs.

Ancestral traits in the genomes of Mucoromycota fungi

Anna Muszewska[1], Małgorzata Orłowska[1], Blanka Sokołowska[1,2], Kamil Steczkiewicz[1] and Julia Pawłowska[2]

[1] Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5a, 02-106, Warsaw, Poland.
[2] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw, Zwirki i Wigury 101, 02-089 Warsaw, Poland.

Mucoromycota is a diverse group of fully terrestrial fungi often associated with plant-material. Their ecological significance is hard to be overestimated yet this lineage is relatively understudied.
In our studies, we try to fill this knowledge gap by screening publicly available and newly sequenced genomes. They harbour transposons of families rarely found in Dikarya They encode a profile of secondary metabolite clusters with the dominance of terpenoid and NRPS-like clusters. Unexpectedly, we revealed that they possess genes involved in cobalamin metabolism. Mucoromycotina is of particular interest for lipidome studies because of the high share of unsaturated fatty acids which have biotechnological and nutritional value. Yet some of the lineages lost ergosterol. These fungi have a high share of fucose in their chitin-chitosan cell walls and an array of fucosyltransferases working on diverse substrates. The fucose metabolic network may resemble in its complexity the one present in animals.
Most of these traits are likely inherited from early Opisthokonta ancestors. Our results extend the list of traits separating model Ascomycota from early-diverging fungal lineages.

ANImm : A fast and accurate tool for calculating average nucleotide identity between pairs of viral genomes

Wanangwa Ndovie[1], Jan Havranek[2], Leonid Chindelevitch[3], Rafal Mostowy[1]

[1] Małopolska Centre of Biotechnology, Jagiellonian University in Krakow
[2] The Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University in Krakow
[3] Department of Infectious Disease Epidemiology, School of Public Health, Imperial College London, UK

One of the most common metrics for the assessment of a relationship between pairs of microbial genomes is Average Nucleotide Identity (ANI). We present a new approach called ANImm that uses MMseqs2 to systematically calculate ANI and alignment fraction (AF) values for pairs of genomes. We assessed the accuracy and speed of ANImm using 200 NCBI RefSeq phage sequences and found that it reproduces the accuracy of PyANI (R2=0.999), a BLAST-based tool, at about 1/100 of running time. Moreover, with right parameterization, ANImm was able to accurately predict ANI values even below the PyANI prediction threshold of 60%. In addition, we were able to process the entire NCBI RefSeq (~4500) in 2 hours 56 minutes. Altogether, ANImm is a fast, accurate and scalable tool to investigate genetic relationships between pairs of viral genomes.

In the spotlight, losing IR region: challenging the importance of inverted repeats in plastid genomes

Kacper Maciszewski[1], Anna Karnkowska[1]

[1] Institute of Evolutionary Biology, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, Poland

Plastid organelles have vestigial genomes (ptDNA), whose quadripartite structure has been observed to be strictly conserved in most land plants and green algae. Despite the abundance of available genomic data from a variety of algal taxa carrying secondary plastids, the evolution of their ptDNA structure has been largely understudied. In order to fill this gap in knowledge, we investigated the structure of chloroplast genomes of euglenophytes – a group of plastid-bearing microbial eukaryotes, known to possess a variety of ptDNA organization types.

Our analyses have shown that the secondary plastids of euglenids do not share the previously described correlation between the structural features of the ptDNA, such as inverted repeat presence, and the rate of evolution of protein-coding genes. This indicates that primary and secondary plastid organelles may exhibit more different evolutionary dynamics than previously thought.

Evolutionarily conserved fragments: an attempt to better understand domain composition and mosaicism of phage proteins

Krzysztof Szczepaniak[1], Bogna J. Smug[1], Stanislaw Dunin-Horkawicz[2], Rafal J. Mostowy[1]

[1] Małopolska Centre of Biotechnology, Jagiellonian University in Krakow
[2] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw, Zwirki i Wigury 101, 02-089 Warsaw, Poland.

Some phage proteins are known to be complex and multi-domain. Since these proteins may be underrepresented in databases, we designed the following bioinformatic pipeline: First, we collected a representative subset of phage genomes from the NCBI RefSeq database and clustered all protein sequences. Then, from each cluster, a representative protein was selected and used to build an HMM profile. Finally, we performed all-vs-all profile-profile searches to identify evolutionarily conserved fragments (ECFs), i.e., alignable sequence regions shared between distantly related clusters. Analyses of the obtained ECFs showed that they frequently overlap with known Pfam domains. However, we also found cases where ECFs were localized in regions not annotated by any Pfam domain, e.g. in receptor-binding proteins, suggesting that these could represent novel domains. Overall, our approach highlights the need for a more systematic characterization of phage protein domains, particularly in more complex and multi-domain functional classes.

In silico analysis of selected nerve agents' non-covalent binding affinity towards acetylcholinoesterase

Rafał Madaj[1], Arkadiusz Chworoś[2], Bartlomiej Gostyński[3]

[1] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw, Zwirki i Wigury 101, 02-089 Warsaw, Poland.
[2] Division of Bioorganic Chemistry, Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Lodz, Poland
[3] Division of Structural Chemistry, Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Lodz, Poland

Within the last few years, a series of in silico studies were performed to assess the binding affinity and inhibition potential of some organophosphorus compounds towards the enzyme crucial for nerve system functioning – acetylcholinesterase (AChE). Due to the current world situation, a specific concern is raised towards a relatively unknown group of compounds - Novichoks.

In this study, a set of in silico quantum mechanical calculations, molecular docking, and molecular dynamics simulations were applied to assess quantitatively the differences between the binding energies of acetylcholine, the natural agonist of AChE, and its selected neurotoxic, human-made substituents. We confirmed that Novichoks' binding affinity is comparable to most toxic 'classical' nerve agents regardless stereoisomery and is more stable inside binding cavity than natural agonist. To our knowledge, it is the first quantitative in silico description of the AChE-Novichok non-covalent binding process that may facilitate the search for an efficient and effective treatment for Novichok intoxication and, hopefully, for other warfare nerve agents as well.

Fucose in fungi

Małgorzata Orłowska[1], Drishtee Barua[1], Sebastian Piłsyk[1], Anna Muszewska[1]

[1] Institute of Biochemistry and Biophysics PAS, Warsaw

Fucose is a deoxyhexose, found in the L-configuration in a diverse range of organisms, playing a variety of biological roles. Fucose metabolism in fungi has not been described so far. Current knowledge about fucose in fungi is mostly limited to describing its presence in the cell wall of several early diverging fungi (EDF). In our study, we are aiming to create a full picture of fucose metabolism in fungi. Using in-silico approaches, we were able to identify proteins putatively related to O-fucosylation among different classes of EDF. The presence of these predicted proteins is being validated with the help of whole transcriptome studies conducted in several species of EDF along with an experimental follow-up of the same. The results might be an indicator of protein O-fucosylation being an important process contributing to the ecological plasticity of fungi and their ability to interact with the external environment.

Functional analysis of prokaryotic homologs of programmed cell death factors

Kamil Krakowski[1], Karolina Łabędzka-Dmoch[2], Jakub Piątkowski[2], Paweł Golik[2] and Stanisław Dunin-Horkawicz[1,3]
[1] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw
[2] Institute of Genetics and Biotechnology, Faculty of Biology, University of Warsaw
[3] Department of Protein Evolution, Max Planck Institute for Biology, Tübingen, Germany

Programmed cell death (PCD) is typically associated with multicellular eukaryotes. However, to some extent, this phenomenon can be also observed in simpler organisms, including bacteria. Previously we have identified eukaryotic-like bacterial protein networks (ELB) encompassing many homologs of eukaryotic PCD systems. The ELB networks are especially extensive in prokaryotes characterized by a multicellular stage in their life; however, their function remains unknown.
The central component of PCD systems in eukaryotes is the large, homooligomeric complex (Apaf-1, CED-4, and DARK in humans, worms, and flies, respectively) that is the main component of the apoptosome. The aim of this study is to investigate the ability of eukaryotic and bacterial Apaf-1 homologs to induce cell death in yeast, which natively lacks the Apaf-1 homolog.
Bioinformatic analysis on early-branching Eukaryota resulted in the discovery of distant Apaf-1 homologs. Further search, using hidden Markov models, revealed an abundance of other homologs in prokaryotes (components of the ELB networks). Laboratory work is underway towards the introduction of selected Apaf-1 homologs into yeast.

Workflow for inferring microbial interaction networks based on amplicon sequencing data

Stanisław Antonowicz[1], Michał Karlicki[1], Anna Karnkowska[1]

[1] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw

The goal of my master's thesis is to construct an easy-to-use workflow that performs several common tasks needed for inferring protist interaction networks. These tasks include taxonomic classification of OTU/ASV sequences using BLAST (Camacho et al., 2009) and PR2 database (Guillou et al., 2013), transforming abundance tables to a uniform format and running several inference methods. At the end the results are compared and visualized. The workflow is tested on the data collected during the Tara Oceans expedition (Pesant et al., 2015). The results are compared with an already published Tara interactome (Lima-Mendez et al., 2015) and PIDA (Bjorbækmo et al., 2019) – a database of known protist interactions.

Use of metabarcdoing for the identification of freshwater microbial communities along an eutrophication gradient

Valentina Smacchia[1], Małgorzata Chwalińska[1], Anna Karnkowska[1]

[1] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw

Freshwater ecosystems, even covering only 0.01% of the water on Earth, have a fundamental ecological role and provide important economical services. Such environments are menaced by eutrophication, i.e. the increasing of nutrients caused by human activity (tourism, runoff of fertilizers, etc.). Microorganisms, protists and bacteria, play an important role in the regulation of the biogeochemical cycles of aquatic environments, and they are strongly influenced by physicochemical changes in their environment. Moreover, there is evidence that the increase in eutrophication status is related to the occurrence of pathogenic bacteria.
To investigate such a correlation, we will analyze the microbial communities in two important niches of freshwater systems, epiphytic biofilms and water of 5 lakes with different eutrophication states of the Great Masurian Lakes District (GMLS). Their characterization will be realized through metabarcoding analyses by short and long-read techniques (Illumina and Nanopore technology), for better detection of pathogenic strains.

From huge metagenomes to single plastid genomes – preliminary results of searching for plastid genomes in freshwater metagenomic data

Michał Karlicki[1], Katarzyna Piwosz[2], Jason Woodhouse[3], Hans-Peter Grossart[3], Anna Karnkowska[1]

[1] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw
[2] Department of Fisheries Oceanography and Marine Ecology, National Marine Fisheries Research Institute
[3] Plankton and Microbial Ecology, The Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB)

Microbial eukaryotes (protists) are a diverse, widespread group. They play an important role in biogeochemical cycles on earth. One of the most crucial processes which is conducted by protists is photosynthesis indicated by possession of plastids – organelles with its own genomes. Recently, the increase of availability of metagenomic sequencing allowed the development of a new field - eukaryotic metagenomic, however analysis of nuclear genomes due to its large size and complexity is still challenging. Thus, we believe the analysis of organellar genomes that not only bears a phylogenetic potential but also points out an ability to conduct photosynthesis, it's a good starting point.
Here we designed a pipeline to reconstruct plastid genomes from a unique dataset of 1096 metagenomic samples from four freshwater lakes. A preliminary result of analysis of two out of eight co-assemblies showed a huge diversity of obtained plastid genomes (100 genomes), representing major eukaryotic groups.

Finding interaction patterns between protists and eukaryotes based on single-cell microbiome approach: obtaining good quality data for computational analyses

Marta Sałek[1], Metody Hollender[1], Anna Karnkowska[1]

[1] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw

Examples of interactions between microorganisms are a multitude and consist of many kinds of symbiosis, such as endosymbiosis between microeukaryotic (protist) host and prokaryote endosymbiont. Here we developed a workflow for finding interaction patterns between microorganisms based on a single cell microbiome.

We studied 83 single cells of protists, DNA from single cells was isolated and the whole genome amplification was performed, followed by the 18S rDNA amplification with a variety of general and specific primers to identify the eukaryotic host. Obtained products were sequenced using the Sanger method. Then we amplified 16S rDNA to identify symbionts, to be further validated with 16S amplicon and metagenomic sequencing. During the workflow data from 31-65% of cells was lost (percentage of lost data varies depending on studied taxa). Currently, we validate the workflow on protists from two lakes. The main challenge is to assess studied interactions based solely on sequencing.

How to investigate complex genomes in diverse taxa

Paweł Hałakuc[1], Rafał Milanowski[1], Anna Karnkowska[1]

[1] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw

Among many eukaryotic groups we can observe large spectrum of nuclear, mitochondrial and plastid genomes. Euglenids are one of them, with only intron-rich plastid genomes relatively well studied. Their repetitive nuclear and mitochondrial genomes remain difficult to sequence, a case not unique to this particular group

We used combination of non-standard assembler settings (metaSPAdes) and manual follow up (Bandage, Geneious) to acquire more complete and less fragmented mitochondrial genomes. We successfully identified mitochondrial genomes for 10 of 15 phototrophic genera. Euglenid mitochondrial genomes encode conserved set of genes and appear to have non-trivial structure.

Those results provide a glimpse into the evolution of mitochondrial genomes of Euglenids and show that unconventional approaches may yield useful conclusions from seemingly unusable data.

Lytic Phage Depolymerases and Capsule Specificities: Insights from Literature and Sequence Analysis

Vyshakh Rajachandra Panicker[1], Rafał Mostowy[1]
[1] Małopolska Centre of Biotechnology, Jagiellonian University in Kraków, Poland

Receptor Binding Proteins (RBPs) are important structural proteins carried by bacteriophages infecting Klebsiella pneumoniae. RBPs have a modular architecture, characterized by N-terminal domain, a depolymerase domain and a C-terminal domain. Depolymerases degrade the capsular polysaccharides, serving as the first stage of infection. To explore the tripartite relationship between depolymerase specificity, capsular serotype and phage host range, we generated a dataset comprising 74 lytic depolymerases with validated/predicted capsule specificities via systematic literature survey. A preliminary sequence level analysis of depolymerases showed that the vast majority of sequence overlap happens predominantly in the N and centre domains. Network analysis of depolymerases showed very limited sharing of domains even between depolymerases specific to same capsule types, suggestive of sequence variability of depolymerases. In the next phase of our analysis, we will be using AlphaFold 2 to determine the sharing of domains between different depolymerases specific to same capsule type at the structural level.

Predicting absorption wavelengths of rhodopsins using machine learning-based methods

Julia Gołębiowska[1], Kamil Kamiński[1], Anna Karnkowska[1], Stanisław Dunin-Horkawicz[1]
[1] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw

Rhodopsins are photosensitive proteins that contain retinol. They can absorb different wavelengths depending on their function and ecological niche. It is known that rhodopsin's absorption spectrum depends on particular amino acids. The study aimed to build a machine learning model for the prediction of absorption maximum. We made a group model using: linear regression with lasso, random forest, support vector regression and neural network. We implemented the alignment-free approach for sequence representation - word embeddings. This solution enables to omit long step of aligning sequences for prediction and ensures reproducibility. Finally, we obtain a good model with the following error metrics: RMSD – 18.99, MAE - 13.06 and MAPE – 0.02. Additionally, we perform an analysis of a cluster of sequences with a broad absorbance spectrum. It suggested that the most important amino acids for light absorption are near the retinal and have different chemical properties compared to the conserved amino acids.

CLIPPERs mediate targeted protein degradation in Gram-negative bacteria

Matylda Izert[1], Maria Klimecka[1], Anna Antosiewicz[1], Patrycja Szybowska[1], Maria Górna[1]
[1] Biological and Chemical Research Centre, Department of Chemistry, University of Warsaw

Targeted Protein Degradation is a new exciting approach to drug discovery that relies on degradation of disease-causing proteins, typically via the ubiquitin-proteasome pathway. The urgent need for next generation antibiotics could be answered by this type of drugs, but a different degradation pathway must be used in bacteria since they lack the proteasome. We present our invention, the Clp-Interacting Peptidic Protein Erasers (CLIPPERs) which are chimeric peptides that recruit the ClpXP protease to knock down target endogenous proteins in Escherichia coli. CLIPPERs show antimicrobial activity when directed against essential proteins and can be used as tools to study protein function and validate antimicrobial drug targets.

Workflow for genome recovery of protistan prokaryotic symbionts from single-cell data

Metody Hollender[1], Marta Sałek[1], Michał Karlicki[1], Anna Karnkowska[1]

[1] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw

Prokaryotic symbioses with protists are highly diverse and complex, but understudied. One of the promising approaches here is the analysis of data from sequencing DNA acquired by single cell whole genome amplification. However, it faces numerous challenges, including reliable assembly of high quality genomes, as well as contamination identification. Here we present the new workflow of bioinformatic analyses which allows for obtaining de novo high-quality prokaryotic genome drafts. Such result is achieved by a novel combination of contig classification and binning tools. The workflow does not require knowledge about present symbionts and classifies obtained genomes, thus fuelling discovery of unknown interactions. Additionally, no manual analyses are needed. We show that together with simple annotation tools the workflow is able for providing valuable insights into symbiont physiology, as demonstrated on the selected cases of symbioses.

What does a neural actually learn? The case of reengineering protein-ligand binding affinity

Kamil Kamiński[1], Stanisław Dunin-Horkawicz[1]

[1] Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw

In my talk, I will describe some properties derived from the graph neural network model applied to protein structures. I will show that representations extracted from the different depths of the network allow measuring local and global similarities between protein structures. This, in turn, enables identifying residues crucial for protein-ligand affinity and designing new protein variants with altered specificity.

Modeling of a putative programmed cell death receptor

Jędrzej Kubica[1,2], Dominik Gront[2], and Stanisław Dunin-Horkawicz[1,3]

[1] Laboratory of Structural Bioinformatics, Institute of Evolutionary Biology, University of Warsaw, Poland
[2] Laboratory of Theory of Biopolimers, Faculty of Chemistry, University of Warsaw, Poland
[3] Department of Protein Evolution, Max Planck Institute for Biology, Tübingen, Germany

Multicellularity is a process strictly coupled to the mechanisms of programmed cell death (PCD). A homology between well-studied eukaryotic PCD proteins (human Apaf-1, Caenorhabditis elegans Ced-4) and yet not fully characterized prokaryotic PCD-like proteins has been reported. Wrap1 is a transmembrane protein, which is a component of a putative PCD apparatus in a multicellular cyanobacterium Nostoc punctiforme. It is equipped with a highly-repetitive interaction-mediating β-propeller domain, which suggests its possible antibody-like role. Upon activation, Apaf-1 and Ced-4 to form apotosomes, which trigger signaling pathways in apoptosis and innate immunity. Due to homology, Wrap1 is also expected to oligomerize. This study has been focused on the possible state of NTPase domains of two proteins and Wrap1 using Rosetta and AlphaFold2. Additionally, control modeling for Apaf-1 and Ced-4 (for which the oligomerization states are known) has been performed. The obtained results might serve as indication of the most probable state for the Wrap1 protein and its potential interaction partner.

Tell me who your enemy is, and I will still not tell you know who you are - the hard problem of phage-host prediction

Jakub Barylski[1], Andrzej Zieleziński[2]

[1] Department of Molecular Virology, Adam Mickiewicz University in Poznań, Poland
[2] Department of Computational Biology, Adam Mickiewicz University in Poznań, Poland

Relationship between a phage and its host is probably the main determinant of the virus biology. It is also a crucial information guiding the application of phages in biotechnology and medicine. Unfortunately, this information is usually unavailable for phages detected in metagenomic data, and experimental attempts to bridge this gap are painfully slow. Therefore recent years brought numerous computational methods that predict hosts based on viral sequences.
Here we discuss these methods (including a few designed by our team) and underlying theoretical assumptions.
The main conclusion of our review is that popular approaches may not generalize outside the "comfort zone" of known, well-studied viruses. Thus, we call for a rigorous benchmarking test of prediction methods that will expose the strengths and limitations of each tool. This will allow bacteriophage researchers to rationally choose programs to use and create a roadmap for further development of more effective algorithms.

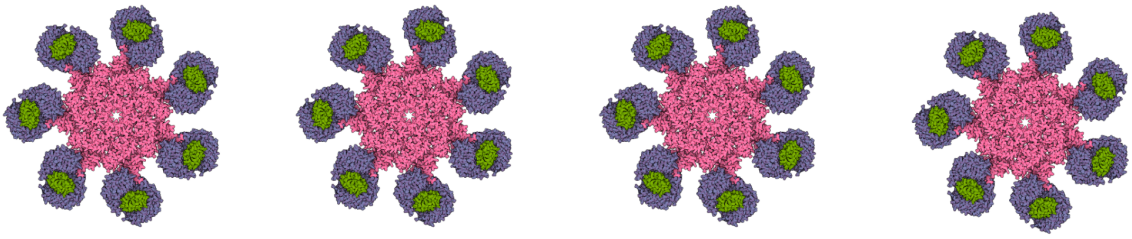
K-mer-based analysis of microbial genomes

Sebastian Deorowicz[1], Adam Gudyś[1], Maciej Długosz[1], Marek Kokot[1]

[1] Silesian University of Technology, Gliwice

Large volumes of data produced during sequencing thousands of prokaryotic organisms require fast analysis methods. Short substrings of nucleotide sequences, called k-mers, are commonly used in this area as they can be extracted from genomes or sequencing reads. As a result, tasks like phylogeny reconstruction or metagenomic classification can be done in alignment- or even assembly-free setting.

The existing solutions for k-mer-based evolutionary reconstruction (e.g., Mash) are time and memory consuming. This imposes the usage of small k-mer subsets, known as sketches. Our tool, Kmer-db, is free of this limitation. The estimation of similarities of 40 715 microbial genomes on the basis of the full 20-mer spectrum required 2h30, less than Mash for 500 times smaller representation (10 000 sketch). Moreover, the ability to index all k-mers from the investigated genomes greatly extends potential applications of our algorithm.



Participants

Invited speakers

Jakub Barylski

- Department of Molecular Virology, Faculty of Biology,
  Adam Mickiewicz University

Maria Górna

- Biological and Chemical Research Centre, University of Warsaw

Adam Gudyś

- Silesian University of Technology, Gliwice

Anna Muszewska

- Institute of Biochemistry and Biophysics,
  Polish Academy of Sciences

Tomasz Kościółek

- Małopolska Centre of Biotechnology, Jagiellonian University

Group leaders

Anna Karnkowska

- Institute of Evolutionary Biology, Faculty of Biology,
  University of Warsaw

Rafał Mostowy

- Małopolska Centre of Biotechnology, Jagiellonian University

Stanisław Dunin-Horkawicz

- Institute of Evolutionary Biology, Faculty of Biology,
  University of Warsaw

Speakers

Stanisław Antonowicz
- Institute of Evolutionary Biology, Faculty of Biology, UW

Julia Gołębiowska
- Institute of Evolutionary Biology, Faculty of Biology, UW

Paweł Hałakuc
- Institute of Evolutionary Biology, Faculty of Biology, UW

Metody Hollender
- Institute of Evolutionary Biology, Faculty of Biology, UW

Kamil Kamiński
- Institute of Evolutionary Biology, Faculty of Biology, UW

Michał Karlicki
- Institute of Evolutionary Biology, Faculty of Biology, UW

Kamil Krakowski
- Institute of Evolutionary Biology, Faculty of Biology, UW

Jędrzej Kubica
- Institute of Evolutionary Biology, Faculty of Biology, UW

Kacper Maciszewski
- Institute of Evolutionary Biology, Faculty of Biology, UW

Rafał Madaj
- Institute of Evolutionary Biology, Faculty of Biology, UW

Wanangwa Ndovie
- Małopolska Centre of Biotechnology, Jagiellonian University

Małgorzata Orłowska
- Institute of Biochemistry and Biophysics, PAS

Vyshakh Rajachandra Panicker
- Małopolska Centre of Biotechnology, Jagiellonian University

Marta Sałek
- Institute of Evolutionary Biology, Faculty of Biology, UW

Valentina Smacchia
- Institute of Evolutionary Biology, Faculty of Biology, UW

Bogna Smug
- Małopolska Centre of Biotechnology, Jagiellonian University

Krzysztof Szczepaniak
- Małopolska Centre of Biotechnology, Jagiellonian University