XIII EDITION BIOINFORMATICS OPEN DAYS

ABSTRACT BOOK

UNIVERSITY OF MINHO
GUALTAR CAMPUS
MARCH 14, 15 AND 16TH, 2024

Bioinformatics Open Days 2024

/bioinformaticsopendays

Bioinformatics Open Days 2024

CONTENTS

Bioinformatics Open Days 2024

ORAL COMMUNICATIONS

Session 1 - Nº 1

Get to know MIMt! A new, smaller and curated 16S rRNA reference database with less redundancy and higher accuracyat species-level identification

M. PILAR CABEZAS [1,2], NUNO A. FONSECA [3,4] AND ANTONIO MUÑOZ-MÉRIDA [3,4]

1. Centre of Molecular and Environmental Biology (CBMA), Department of Biology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
2. Institute of Science and Innovation for Bio-Sustainability (IB-S), University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
3. CIBIO-InBIO, Research Center in Biodiversity and Genetic Resources, 4485-661 Vairão, Portugal
4. BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal

Microorganisms are one of the most diverse and abundant group of living organisms on Earth. However, an accurate determination and quantification of the taxonomic composition of microbial communities, especially at the species level, is one of the major issues in metagenomics. This is primarily due to the limitations of commonly used 16S rRNA reference databases, which either contain a lot of redundancy, or a high percentage of sequences with missing taxonomic information. The use of these incomplete or biased databases may lead to erroneous identifications and, thus, to erroneous conclusions regarding the ecological role and importance of those microorganisms in the ecosystem. The current study presents MIMt, a new 16S rRNA database for archaea and bacteria's identification, encompassing 39 940 sequences, all precisely identified at species level (https://mimt.bu.biopolis.pt). We evaluated MIMt against the most used reference databases, namely Greengenes, RDP, GTDB and SILVA, in terms of sequence distribution and accuracy of taxonomic assignments. Our results showed that MIMt contains less redundancy, and despite being five to 85 times smaller in size than existing databases, outperforms them in completeness and taxonomic accuracy, enabling more precise assignments at lower taxonomic ranks and thus, significantly improving species-level identification.

ORAL COMMUNICATIONS

Session 1 - Nº 2

Prediction of novel small RNAs from Pseudomonas aeruginosa

JOANA SILVA [1], CECÍLIA MARIA ARRAIANO [1] AND VÂNIA POBRE [1]

1. Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Portugal.

Pseudomonas aeruginosa is an opportunistic pathogen that infects immunocompromised patients, particularly cystic fibrosis patients. When untreated, the infections result in lung failure and ultimately death. P. aeruginosa is very difficult to eradicate due to its strong resistance to several classes of antibiotics, as well as its capability to develop into a multi-drug resistant (MDR) bacterium and form biofilms. Consequently, it is crucial to develop alternative strategies to combat antibiotic resistance, such as small non-coding RNA (sRNA)-based therapies. These molecules play a crucial role in gene expression and have been shown to modulate antibiotic resistance and sensitivity in P. aeruginosa. Nonetheless, there is limited knowledge regarding sRNAs in this Gram-negative bacterium. In this work we predicted novel sRNAs using RNA-seq data from wild type PAO1 planktonic cells and biofilms grown with sub-lethal concentrations of four commonly used antibiotic classes: aminoglycosides (kanamycin), beta-lactams (ceftazidime), quinolones (nalidixic acid), and polymyxins (polymyxin B). As a control we grown PAO1 without any antibiotics. Using two bioinformatic tools (Artemis and Rockhopper), we successfully identified regions exhibiting significant expression of potential sRNAs, resulting in the prediction of 1102 novel sRNAs in at least one condition. Of these 809 were classified as potential cis-encoded sRNA, 268 were not clearly defined and only 25 were easily classified as trans-encoded. Moreover, we found that there is a significant difference in expression of majority of these sRNAs when comparing planktonic with biofilm growth conditions. However, the differences in expression are less pronounced when comparing the different antibiotics used in this study. Overall, this work considerably expanded our knowledge of P. aeruginosa sRNAs especially in the context of antibiotic resistance mechanisms and biofilm formation.

ORAL COMMUNICATIONS

Session 1 - Nº 3

Finding novel genes coding for eco-friendly surfactants from hypersaline Iberian locations using metagenomic approaches

CÁTIA SANTOS-PEREIRA [1,2*], JOANA S. GOMES [1*], JOANA SOUSA [1], MARTA F. SIMÕES [3], RICARDO FRANCO-DUARTE [4,5], SUSANA R. CHAVES [4,5], SARA C. SILVÉRIO [1,2], ANDRÉ ANTUNES [3], LÍGIA R.RODRIGUES [1,2]

1. CEB - Centre of Biological Engineering, Universidade do Minho, Braga, Portugal.
2. LABBELS - Associate Laboratory, Braga/Guimarães, Portugal.
3. SKL Planets - State Key Laboratory of Lunar and Planetary Sciences, MUST, Macau SAR, China.
4. CBMA - Centre of Molecular and Environmental Biology, University of Minho, Braga, Portugal.
5. IB-S - Institute of Science and Innovation for Bio Sustainability, University of Minho, Braga, Portugal.

* equal contribution

Surfactants are tensioactive chemical compounds extensively used worldwide in a myriad of industrial sectors and in our daily lives, being present in numerous products including cosmetics, detergents, fabric softeners, toothpaste, among many others. Millions of tonnes of surfactants are thus manufactured every year. Most commercially available surfactants are non-renewable petroleum-based compounds that can have a profound environmental impact. This has prompted the search for new eco-friendly alternatives, including the so-called biosurfactants, which are surfactants produced by microorganisms that are sustainable alternatives to their chemical counterparts. Hypersaline environments are an attractive source of microbial communities that, due to their adaptation to extreme abiotic conditions, produce special secondary metabolites being hotspots for the discovery of new biosurfactants. Sampling campaigns were conducted at strategic hypersaline locations holding distinct features, namely Peña Hueca lagoon (hypersaline sulphated lagoon, Spain), and salinas of Aveiro (solar coastal salina, Portugal) and Rio Maior (terrestrial inland salina, Portugal). DNA was extracted from collected water samples and two different metagenomic approaches were carried out, sequence- and function-based, both aiming to identify new biosurfactant-producing genes. The latter used a robotic screening system to screen more than 500 clones for biosurfactant production. Physicochemical characterization of samples showed an interesting variability in terms of salinity, pH and ionic content. Analysis of shotgun metagenomic sequencing data revealed that the isolated metagenomes are enriched in genes involved in biosurfactant biosynthesis, and that the microbial community is shaped by the physicochemical features. The screening tests showed an interesting number of clones with biosurfactant activity. Therefore, the bioprospection of hypersaline locations of the Iberian Peninsula allowed the identification of biosurfactant-producing clones, which can have promising industrial applications and contribute to the quest for more sustainable alternatives to chemical surfactants.

ORAL COMMUNICATIONS

Session 1 - Nº 4

Individual-based modelling elucidates about the role of nanomaterials on methane production

JOÃO C. SEQUEIRA [1], BOWEN LI [2], M. MADALENA ALVES [2], MIGUEL ROCHA [1], TOM P.CURTIS [2], ANDREIA F. SALVADOR [1]

1. CEB-Centre of Biological Engineering, University of Minho, Braga, Portugal
2. School of Engineering, Newcastle University, Newcastle upon Tyne, United Kingdom
3. School of Computing, Newcastle University, Newcastle upon Tyne, United Kingdom

Modelling stands as a potent tool for the simulation of real-world systems, offering predictive insights into their dynamic behaviours. Individual-Based Models (IBMs), with their ability to capture intricate details at the microscale, emerge as particularly valuable tools. In the context of anaerobic digestion (AD), the application of IBMs gains significance. Some nanomaterials (NM) enhance methane production in anaerobic ecosystems, by yet unknown mechanisms. This study employed IBMs through the NUFEB software to investigate possible mechanisms of action by NM in AD. The models incorporated Monod-based growth dynamics for methanogenic-based growth and production of extracellular polymeric substances, a Lennard-Jones interaction between NM and microorganisms for simulating attachment, and Hooke forces between microorganisms and themselves and EPSs to simulate the attachment between biological entities and formation of biofilms. The simulations also explored boundary conditions (Dirichlet or Neumann), considering scenarios of both open systems and closed environments. Shear force was introduced as a variable to assess its impact on microbial behaviour, and different shapes of NM were simulated, differentiating between smaller and larger particles, and smoother and rougher surfaces. Results indicate that the rugosity of the material plays a pivotal role in providing increased surface area for microorganism attachment. In open systems, higher rugosity leads to the displacement of microorganisms, exposing them to the environment while concurrently attached to NM. These displaced microorganisms have higher access to their substrate and produce more methane. Conversely, in closed systems, a more rugous material serves as a protective shield, reducing the washout of organisms under the influence of shear force. With more microorganisms remaining in the system, more methane is produced. This research provided valuable insights into the role of NM morphology in shaping microbial interactions during AD. The findings enhance our understanding of the complex interplay between NM structures and microbial populations, providing essential knowledge for the optimization of AD processes in the presence of NM.

ORAL COMMUNICATIONS

Session 2 - Nº 5

An Automated Structure-Based Platform for Gene Discovery and Enzyme Engineering

JOÃO CORREIA,[1] SOFIA FERREIRA,[1] ISABEL ROCHA,[1] DIANA LOUSA,[1] CAIO S. SOUZA,[1] CLÁUDIO M. SOARES[1]

1. ITQB NOVA, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras

Organisms and enzymes are currently used for catalyzing the production of value-added compounds, in diverse areas with economic impact. However, poor protein expression, low catalytic levels, low concentrations of co-factors or substrates, toxicity towards the final product and other factors, all contribute to the low performance of microorganisms and enzymes in these areas. Protein engineering is an effective strategy for improving metabolic pathways and overcoming these problems, by re-designing an enzyme's catalytic properties in favor of a certain reaction.
To address this problem, we developed an automated platform for gene discovery and enzyme engineering, which aims to enhance or change the enzymes that catalyze the limiting steps in pathways that can be exploited for the production of added value compounds. The objectives ranged from increasing an enzyme's efficiency to enabling it to catalyze a different transformation than the one catalyzed by the wild-type sequence. The steps of enzyme engineering are divided into three major modules: a random mutation generator, an atomistic homology modeler and a binding energy evaluator. This method may therefore build a set of mutant enzymes and filter which sequences are putative candidates for experimental expression and enzymatic testing.
Experimental validation of several case studies highlights the promising performance of the platform to identify suitable natural or mutant enzyme variants for catalyzing specific reactions, showcasing the potential of enzyme engineering approaches in overcoming challenges in the improvement of metabolic pathways.

ORAL COMMUNICATIONS

Session 2 - Nº 6

Computationally-designed miniproteins showing neutralization activity against SARS-CoV-2

PEDRO MOREIRA[1,2], MARIANA PARADA[1], BÁRBARA FERNANDES[1,3], CAROLINA C. BUGA[1], RITA I. TEIXEIRA[1], CHRISTINE OLIVEIRA[3], STEPHEN MICHAEL BUCKLEY[4], SANDRINE GEORGEON[4], MIGUEL A. R. B. CASTANHO[3], ANA S. VEIGA[3], ISABEL ABREU[1], JOÃO B. VICENTE[1], MIGUEL ROCHA[2], BRUNO E. CORREIA[4], CLÁUDIO M. SOARES[1], DIANA LOUSA[1]

1. INSTITUTO DE TECNOLOGIA QUÍMICA E BIOLÓGICA ANTÓNIO XAVIER, UNIVERSIDADE NOVA DE LISBOA, OEIRAS, PORTUGAL.
2. CENTRO DE ENGENHARIA BIOLÓGICA, ESCOLA DE ENGENHARIA DA UNIVERSIDADE DO MINHO, BRAGA, PORTUGAL.
3. INSTITUTO DE MEDICINA MOLECULAR JOÃO LOBO ANTUNES, FACULDADE DE MEDICINA, UNIVERSIDADE DE LISBOA, LISBOA, PORTUGAL.
4. SCHOOL OF LIFE SCIENCES, ECOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, LAUSANNE, SWITZERLAND.

As the COVID-19 pandemic demonstrated, the need for robust antiviral therapies against SARS-CoV-2 remains substantial. A promising strategy to meet this demand is through the development of proteins tailored for enhanced binding affinity to crucial viral targets, such as epitopes located on their fusion proteins. Among these targets is the Spike protein's receptor binding domain (RBD), which interacts with the human receptor ACE2 protein, initiating the viral entry process. Targeting the RBD epitope that binds to ACE2 is a promising strategy to fight COVID-19 and is used here as a model target to validate our approach.

In this work, we implemented a computational pipeline leveraging artificial intelligence-based computational design methodologies, RFDiffusion and ProteinMPNN, to de novo design miniproteins targeting the RBD epitope that interacts with ACE2. The most promising designs were selected based on a combination of relevant criteria, including metrics derived from the protein structure prediction tool AlphaFold2 and properties like surface hydrophobicity and shape complementarity to the target.

We expressed in vitro selected designs and evaluated their binding affinity to the RBD using Bio-layer Interferometry and Yeast Display assays. Proteins demonstrating favourable binding affinity to the RBD were subsequently subjected to neutralization assays, evaluating the proteins' ability to inhibit SARS-CoV-2 infection.

We observed that two of the selected designs can bind to the RBD and neutralize viral infection, thus successfully demonstrating that this computational framework is able to design proteins that are tailor-made to interact with specific epitopes. This paves the way for the next round of design, where the most promising candidates are being optimized using strategies that consider the oligomerization tendency observed in some of the designs tested.

ORAL COMMUNICATIONS

Session 2 - Nº 7

Deep Learning on Chaos Game Representation for Resistome

Javier Montoya [1,2], Francisco Fernandes [2], Ana Teresa Freitas [1,2]

1. Instituto Superior Técnico - Universidade de Lisboa,
2. Instituto de Engenharia de Sistemas e Computadores: Investigação e Desenvolvimento em Lisboa (INESC-ID)

Emerging evidence has established a connection between Colorectal Cancer (CRC) risk and antibiotic consumption. Nevertheless, it remains uncertain whether CRC is indeed linked to antibiotic resistance within the gut microbiota. In the meantime, Metagenomic studies keep producing vast amounts of sequencing data enabling a deep look within the genetic information of different microbial communities present in several micro-environments inside the human body. This work studies the use of Deep Learning techniques in understanding the microbiome of patients with CRC.
Chaos Game Representations (CGR) are fractal-like graphical representations of DNA sequences that can work as signatures for specific genes or genomes. We apply Deep Learning algorithms to analyse and interpret the Resistome data from CRC patients and healthy individuals, specifically using these CGR images as a data preprocessing and feature extraction technique. The use of Convolutional Neural Networks to analyse CGR images offers a powerful approach for recognizing patterns, relationships, and potential Antibiotic Resistance Genes (ARG) within the Resistome data. The Resistome data used comes from The Comprehensive Antibiotic Resistance Database (CARD). This is a biological database that collects and organizes referenced information on antimicrobial resistance genes, proteins and phenotypes.
Preliminary data analyses have tested different resolution on the images, image sizes, different window sizes of the sequences as well as different structures of CNN and have showed that CGR is a capable representation for distinguishing ARGs from other types of genes such as 16S genetic sequences. Further work has been done to confront ARGs against Metagenomic Assembled Genome fragments from faecal samples of patients suffering from CRC, these sequences of DNA have sizes going from 200 to 8,000 and an average length of around 1,500. The generated pilot models display an accuracy of around 80% and a precision of the class of interest of 85%. These promising results open lines for future work related to exploring unknown ARGs and classifying CRC afflicted patients in early stages from non-intrusive faecal samples data.

ORAL COMMUNICATIONS

Session 2 - Nº 8

Modeling deep neural network architectures to improve schizophrenia predictability using genotype data

Daniel Martins [1,2], **Maryam Abassi** [1,3,4], **Conceição Egas** [2,5], **Joel P. Arrais** [1]

1. Centre for Informatics and Systems of the University of Coimbra (CISUC), Department of Informatics Engineering, University of Coimbra, Coimbra, Portugal.
2. Centre for Innovative Biomedicine and Biotechnology (CIBB), University of Coimbra, Coimbra, Portugal.
3. Polytechnic Institute of Coimbra, Applied Research Institute, Coimbra, Portugal.
4. Research Centre for Natural Resources Environment and Society (CERNAS), Polytechnic Institute of Coimbra, Coimbra, Portugal.
5. Biocant - Transfer Technology Association, Cantanhede, Portugal.

Schizophrenia (SCZ) is a a psychiatric complex disorder with a polygenic architecture. The lack of a consistent set of causal genes hinders its diagnosis. Deep Learning (DL) models have been suggested to enhance the predictability of the disease. However, the most promising results require expression data for its training, which is not readily available for brain tissue.
This study introduces a novel DL model designed to enhance the prediction of SCZ phenotype based solely on genotype data. Two sequential components, involving a locally directed variant-gene network and an auto-encoder, were pre-trained to identify a set of genes with a baseline influence on the phenotype. The learned information was transferred within the encoder to a final model, which incorporated a parallel network branch to capture complementary genes and a final post-concatenation Fully-Connected Neural Network to learn the relevant genetic interplay leading to SCZ.
The first component captured an enrichment on neurodevelopment and synaptogenesis pathways. The final model demonstrated enhanced SCZ classification performance (AUC = 0.83) compared to existing literature, using solely genotype data. Additionally, an enrichment analysis complemented the prioritized terms for the first component, unraveling sub-cellular processes associated to the microtubule cytoskeleton or protein localization.
The final model was trained for diversified groups of samples within the training set. As the model was capable of prioritizing different sets of genes for each subset, this design overcomes challenges in SCZ research, as the genetic heterogeneity of the condition and the inherent subjectivity on non-genetic-based diagnosis. In terms of clinical implications, this model provides valuable insights into the genetic architecture of SCZ and offers potential applications in personalized medicine.

Acknowledgements: FCT - UIDB/00326/2020; UIDP/00326/2020; UIDB/04539/2020; UIDP/04539/2020; LA/P/0058/2020; CENTRO-01-0145-FEDER-029266; CEECINST/00077/2021; SFRH/BD/146094/2019. dbGaP - phs000473.v2.p2.

ORAL COMMUNICATIONS

Session 3 - Nº 9

A network science approach to interpret multidimensional associations between human transcriptome, tissues and traits

Darmit Kumar [1,2], Luísa Pereira [1,3] and Pedro G. Ferreira [2,4]

1. Instituto de Investigação e Inovação em Saúde (i3S), Universidade do Porto, Porto, Portugal
2. Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Ipatimup), Porto, Portugal
3. Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), Porto, Portugal
4. Faculty of Sciences of the University of Porto (FCUP), Porto, Portugal

Complex trait aetiology and pathology remains a great challenge of modern medicine. The advent of large omics consortia, such as the Genotype-Tissue Expression (GTEx, https://gtexportal.org/) project, pave the way to address this thematic by providing high-resolution data. However, the vast, multidimensional, and sometimes heterogeneous omics data inferred from dynamic human biology is a major challenge for interpretability. In this work, we explore the expression patterns of 19,000 genes that are differentially expressed (DE) in 21 phenotypes across 46 human tissues and organs (estimated by Garcia-Pérez et al. 2022 from GTEx). Using network science, we built a map of DE-tissue-trait. We implemented a bipartite multigraph where a gene is connected to clinical or demographic phenotypes in which it is differentially expressed.

The overall network was large and dense but highlighted interesting biological features. Protein-coding and long intergenic non-coding RNA (lincRNA) genes displayed topological differences, with the first ones being more central and the latter peripheral. Several sex chromosome genes stood out with high edge weights and tissue-wide relevance, testifying that gender-related gene expression is a main driver of several traits. In a second step, we explored the functional information that could be retrieved by conducting the analysis on a specific tissue specifically linked to a disease, exploring the case of the thyroid and Hashimoto thyroiditis (HT). Genes exclusively connected to HT were enriched for several landmarks of auto immune disease, such as immune response activation, antigen processing/presentation and T cell processes. Genes shared between HT and Sex emphasised B cell processes. Genes shared between HT and Ancestry indicated upregulation of cell cycle, DNA damage and metabolism.
With this work we demonstrate that bipartite multigraphs can highlight the relation of gene expression changes to key biological mechanisms behind diseases and their associated demographic phenotypes at a tissue level.

ORAL COMMUNICATIONS

Session 3 - Nº 10

The Portuguese Variome from over 12,000 Exomes: Streamlining Clinical Diagnostics and Translational Research

Mariana Ribeiro [1,2], Susana Valente [1,2], Filipe Alves [1], Jorge Sequeiros [1,3], João Parente Freixo [1], Jorge Oliveira [1*], Paulo Silva [1*]

1. Centro de Genética Preditiva e Preventiva (CGPP), Instituto de Biologia Molecular e Celular (IBMC), Instituto de Investigação e Inovação em Saúde (i3S), Universidade do Porto, Portugal;
2. Departamento de Ciências Médicas, Universidade de Aveiro, Aveiro, Portugal;
3. ICBAS – Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal.

*- equally contributing

In the rapidly expanding field of biomedicine, genomics holds not only great promise but has already entered clinical practice. As sequencing technology evolved and routine bioinformatics approaches stabilised, clinical genetic testing has progressed from targeted sequencing (e.g. multigene panels), to more comprehensive approaches such as whole-exome sequencing (WES) and whole-genome sequencing (WGS). This has exponentially increased the number of variants identified in each patient, making data interpretation progressively more challenging. International initiatives like the 1000 Genomes Project, the Exome Sequencing Project (ESP), the Exome Aggregation Consortium (ExAC), and the Genome Aggregation Database (gnomAD) have played crucial roles in distinguishing between polymorphisms and candidate variants, through the public release of information about variants' frequencies. Yet, up to this day, no existing genetic variant database displays the unique genetic composition of the Portuguese population.

By leveraging a dataset of 12,167 patient exomes processed at the Center for Predictive and Preventive Genetics (CGPP) from the Institute for Molecular and Cell Biology (IBMC), a key provider of molecular genetic testing in Portugal, we created a representative reference set of the Portuguese population, by municipality, resampling our cohort to 3,972 individuals. From this subset, we focused on the variants present in the list of 81 secondary-finding genes elaborated by the American College of Medical Genetics and Genomics (ACMG), which are known for their direct medical actionability.

The open-source codebase of ExAC was adapted to be run locally and display this dataset, showing the allele counts and frequencies of each variant, in the 81 ACMG actionable genes, grouped by municipality, as well as the visualization of the distribution of each variant throughout the country in interactive maps.

Concluding, in this work, we established a reference set of common, rare, and ultra-rare genetic variants present in a representative sample of the Portuguese population, providing a comprehensive view of the genetic variations present in the list of 81 ACMG genes, which can ultimately provide valuable information for patients, practitioners, and stakeholders involved in genomic medicine in order to improve healthcare in Portugal.

ORAL COMMUNICATIONS

Session 3 - Nº 11

Runs of homozygosity: bioinformatic approaches for diagnostic purposes and population analysis in 12,000 exomes

Susana Valente [1,2], Mariana Ribeiro [1,2], Jennifer Schnur [3], Filipe Alves [1], Nuno Moniz [3], Jorge Sequeiros [3,4], João Parente Freixo [1], Paulo Silva [1 *], Jorge Oliveira [1 *]

1. Centro de Genética Preditiva e Preventiva (CGPP), Instituto de Biologia Molecular e Celular (IBMC), Instituto de Investigação e Inovação em Saúde (i3S), Universidade do Porto, Portugal
2. Departamento de Ciências Médicas, Universidade de Aveiro
3. University of Notre Dame, Indiana, US
4. ICBAS – Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal.

*- equally contributing

The present work offers a comprehensive exploration of runs of homozygosity (ROH) analysis within a dataset comprising over 12,000 whole-exome sequencing (WES) samples from the Center of Predictive and Preventive Genetics (CGPP) from the Institute for Molecular and Cell Biology (IBMC) focusing on genetic diagnosis and population analysis. ROH represents homozygous segments for genetic markers of the genome. These are particularly relevant in the context of consanguinity, given that their number and size depend on the degree of shared parental ancestry. Homozygosity mapping serves as a robust tool for ROH identification and gene discovery in human genetics. Next-generation sequencing (NGS) introduction allows both homozygosity mapping and variant detection. In this work, an individual analysis of ROH was performed.

To enhance applicability for diagnostic purposes, we aimed to streamline the creation of personalized multigene panels based on ROH. The approach combined two software based on distinct algorithms: ROHMMCLI (Hidden Markov Model) and HomozygosityMapper (Sliding-Window). We extended the functionality of multigene panel creation with the possibility of adding phenotypic information using Human Phenotype Ontology (HPO) terms. We resorted to a Django Web application to facilitate the creation of these panels. Further to the generation of standardized Excel files for ROH profiles, the impact of Copy Number Variants on ROH detection was evaluated. Collaborative efforts with the University of Notre Dame, US, attempted the development of a new clustering model predicting patients' consanguinity based on ROH features.

The analysis of ROH at a population level involved creating a representative sample of the Portuguese population and achieving the first Portuguese ROH characterization at a genomic scale. To address these points, a database was created as well as interactive maps for better visualization. The initial dataset of 12,167 exomes was reduced to 3,972 individuals. This sample was used to calculate the mean genome-wide autozygosity measure from ROHs per municipality.

In conclusion, this study presents significant advancements in ROH analysis using WES data for both diagnostic applications and population genetics. Our results contributed to the development of new bioinformatic tools for diagnostics and enables future research in the field of genomics and gene-discovery using ROH data.

ORAL COMMUNICATIONS

Session 3 - N° 12

Sequence to graph alignment based copy number calling using a flow network formulation

Hugo Magalhães [1], Timofey Prodanov [1], Jonas Weber [2], Gunnar W. Klau [3], Tobias Marschall [1]

1. Institute for Medical Biometry and Bioinformatics, Medical Faculty, and Center for Digital Medicine, Heinrich Heine University, Düsseldorf, Germany;
2. Institute of Medical Microbiology and Hospital Hygiene, Heinrich Heine University, Düsseldorf, Germany;
3. Algorithmic Bioinformatics, Heinrich Heine University, Düsseldorf, Germany;

The majority of bioinformatics applications still rely on a linear reference genome, which cannot fully represent genomic variation within and between different individuals. Pangenomics addresses this issue, for example by using a graph representation of multiple reference genomes. However, the shift from linear sequences to graph structures raises a need for new algorithms and data structures.
Variation of copy number (CN) between individuals can be associated with phenotypical differences. Consequently, CN calling is an important step for disease association and identification, as well as in genome assembly. Traditionally, CN calling is done by mapping sequencing reads to a linear reference genome and estimating the CN from the observed read depth. This approach, however, can lead to inconsistent CN assignments and is particularly hampered by sequences not represented in a linear reference genome. To address this issue, we propose a method for CN calling with respect to a graph genome using a flow network formulation.
As input, we are given a bidirected genome graph, along with a set of read alignments to the same graph. We calculate raw CN probabilities for every graph node based on the Negative Binomial distribution and the base pair coverage across the node. Finally, we use integer linear programming to find a maximum likelihood flow through the entire graph, resulting in CN predictions for each node. This way, the method achieves consistent copy number assignments across the graph and is able to characterize complex loci.
Preliminary results: The proposed method is able to process a wide variety of input graphs and read mappings from different sequencing technologies. We processed reads aligned to a Verkko assembly graph for HG02492 (HGSVC) using high coverage mixed HiFi and ONT-UL reads in one thread under 2 hours with <2Gb peak memory. For 18% of the nodes, corresponding to 14% of the total sequence in the graph, our method estimated different CN values than those expected from read depth alone, showcasing how the graph topology informs CN assignments.
Further applications of our method include CN assignment as part of diploid and polyploid (pan)genome assembly workflows.

ORAL COMMUNICATIONS

Session 3 - Nº 13

Plugging the holes: an intuitive software tool for reproducibility in Molecular Biology

Patrícia Ataíde[1,2], Inês Abreu[1], Faezeh Ghasemi[1,2], Claúdia Barata-Antunes[1,2], Sandra Paiva[1,2] and Björn Johansson[1]

1. CBMA - Center of Molecular and Environmental Biology and Department of Biology, School of Sciences, University of Minho, Campus de Gualtar, Braga, 4710-057, Portugal
2. Institute of Science and Innovation for Bio-sustainability, University of Minho, Braga, Portugal

Life science research often depends on constructing and analyzing plasmids and other recombinant DNA molecules, where their precise sequence is vital for the results. Biosciences are currently suffering from a reproducibility crisis that is exacerbated by the lack of systematic, complete and verifiable documentation. Most cloning strategies are deterministic, so in principle possible to describe completely and unambiguously. Typically, publications include ad-hoc guides to the cloning strategies used. Following these descriptions manually is painstaking, contributing to irreproducible results slipping through peer review. The Python package “Pydna” simplifies the expression of reproducible cloning strategies, but requires the user to code, which is not the best option for all potential users. We have developed a tool to express cloning strategies as a collection of interlinked text files, using very little markup and a flexible syntax. Each text file is self-contained and essentially describes a molecular biology unit operation, such as PCR or CRISPr digestion. Each of these unit operations has input and output DNA sequences. These strategies can be automatically validated by processing with a command line application. This can be combined with automated verification through integration services like GitHub Actions, resulting in an altogether no-code solution.

This work was supported by the FCT project MetaFungal PTDC/BIA-MIC/5246/2020 (DOI 10.54499/PTDC/BIA-MIC/5246/2020).
F.G. received FCT Ph.D fellowship (2023.03135.BD).

POSTER COMMUNICATIONS

Session 1 - Nº 14

Exploring bioinformatics tools to characterize a new regulator of Candida glabrata biofilm matrix

Sónia Silva [1] ; Mariana Henriques [1] ; Bruna Gonçalves [1]

1. Centre of Biological Engineering, University of Minho, 4700-057, Braga, Portugal

Candida glabrata is a clinically relevant human pathogen with ability to form high recalcitrant biofilms, which produce an extracellular matrix with suggestive virulent and protective roles. Thus, the elucidation of the matrix composition and regulation is crucial to disclose the matrix role in C. glabrata pathogenesis. As such, this study aimed to characterize, with various bioinformatics tools, a new C. glabrata biofilm matrix regulator identified by us, the transcription factor Zap1. For that, genes and matrix proteins targeted by Zap1 were firstly assessed through microarrays and LC-MS/MS analyses, respectively, using C. glabrata mutant strains. Then, Zap1 targets were analyzed with various bioinformatics tools and databases including: a) functional distribution using FungiFun and FunCat; b) predicted phenotype and Gene Ontology (GO) using Candida Genome Database; c) molecular interaction using STRING and Cytoscape; d) orthology using PathoYeastract; e) predictive secretory nature using Fungal Secretome Database and Fungal Secretome KnowledgeBase. The bioinformatics analyses suggested that Zap1 is a complex regulator of C. glabrata biofilm matrix, inducing and repressing various genes/matrix proteins involved in glucan metabolism and transport functions, including transferases and hydrolases with potential role in the delivery and organization of matrix components. Additionally, the bioinformatics analyses also suggested that Zap1 may be involved in relevant roles such as energy generation, adhesion, virulence, antifungal resistance, host immunity evasion and modulation of extracellular vesicles. Overall, this study, using a variety of bioinformatics tools, revealed that Zap1 is a relevant regulator of C. glabrata biofilm matrix and suggests that it may be an interesting target for the development of novel therapeutics to fight the complicated infections caused by C. glabrata biofilms.

A preliminary metabarcoding study of the diversity of Eukaryotic and Prokaryotic communities in the socio-ecological system from Rias Baixas (NW Spain)

Ríos-Castro, R; Ramilo, A; Rodríguez, H; Pascual, S; Abollo, E

Parasitome Lab-Ecobiomar, Instituto de Investigaciones Marinas, Consejo Superior de Investigaciones Científicas, IIM-CSIC, 36208 Vigo, Spain

Marine ecosystems play a crucial role in sustaining life on Earth, providing diverse ecosystem services that are fundamental to human well-being. With the growing impact of anthropogenic activities, monitoring strategies based on One Health approach in marine ecosystems and their services, are imperative to safeguarding both aquatic life and human populations. Therefore, understanding the community composition in anthropogenic influenced areas is essential for developing management strategies to preserve ecosystem services.

DNA metabarcoding was performed to describe the eukaryote and prokaryote communities (< 200 µm) in water samples collected in 2023 from 16 areas in Rías Baixas (NW, Spain) that provide different socio-ecological services and which are subjected to varying degrees of anthropogenic influence. Bioinformatic analysis of raw sequencing data were mainly performed using sequential pipelines in Qiime2. Quality assessment, trimming and merging of raw sequence data, followed by taxonomic assignment using specific reference databases and alpha and beta diversity analysis were performed.

While our analysis is ongoing, we present initial findings on 2 marine ecosystem services: an estuarine area near to a discharge point of a wastewater treatment plant (WWTP) and another marine area close to a shellfish depuration plant (SDP). Rarefaction plots showed that the sequencing effort was enough to achieve all community diversity. The highest bacteria and fungi diversity was obtained close to the WWTP, whereas the highest eukaryote diversity was detected in the samples near to the SDP. Different microbial communities were associated with each sampling area and only 6 fungi and 8 eukaryote ASVs were common between both locations. Dinoflagelata (Protist), Crustacea (Metazoa), Rhodobacteriales (Bacteria), and Pleosporales (Fungi) were the most representative ASVs in the SDP area; while Cercozoa (Protist), Corynebacteriales (Bacteria), Campylobacterales (Bacteria) and Trichosporonales (Fungi) were mainly associated with the WWTP area. Potential human and animal pathogens were also found in both samplings; however more exhaustive analysis would be necessary to confirm their identity.

Our preliminary results underscore the importance of ongoing research in DNA metabarcoding for monitoring and maintaining sustainable marine ecosystem services. Detecting microbial threats with this tool not only aids in preserving ecological integrity
but also safeguards human and animal health.

Computational design of antiviral biologics targeting Zika virus envelope protein

Maria Benedita Pereira[1], Pedro Moreira[1], Rita I. Teixeira[1,2], Cláudio M. Soares[1], Diana Lousa[1]

1. Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal
2. Centro de Engenharia Biológica, Escola de Engenharia da Universidade do Minho, Braga, Portugal

In the past two decades, the world has struggled with recurrent viral outbreaks, with viruses from diverse families demonstrating pandemic and epidemic potential. One of those viruses is the Zika virus. Despite disease cases having declined globally after 2017, Zika virus transmission persists at low levels in regions like the Americas and a total of 89 countries and territories have reported evidence of Zika virus infection. Despite active research, treatments for Zika virus infection are lacking, and vaccine development remains ongoing.

The field of protein design has arisen as a transformative discipline in molecular engineering, allowing precise tailoring of protein properties such as their stability and ability to bind to specific partners. Antiviral biologics, such as small proteins that can bind to and inhibit viral targets, appear as a promising therapeutic option. In the Zika virus, the envelope protein (E) has a pivotal role in viral entry, making it an ideal target for antivirals. The E protein comprises three structural ectodomains (DI, DII, DIII) and a transmembrane region. DIII is an immunoglobulin-like domain that contains receptor-binding sites. In this work we are developing tailor-made antiviral biologics that specifically target and bind to the E protein DIII, preventing viral entry into host cells. The methodology involves identifying epitope regions on the target surface regions, selecting binding motifs from a large “Atlas” containing a description of a large pool of protein binding motifs, and docking the binding motifs to the epitope. Binding affinity will be optimized using proteinMPNN, a deep learning-based protein sequence design method. Protein structure prediction tools, like AlphaFold2, will be used to filter the most promising designs. Additionally, molecular dynamics simulations will be employed for stability evaluation as an in silico control. Our collaborators will then experimentally evaluate the selected candidates for their binding affinity. This comprehensive approach seeks to redefine strategies in combating the Zika virus, holding the potential to enhance preparedness against emerging viral threats with pandemic potential.

Antiviral proteins targeting Influenza A hemagglutinin: design, production and characterization

Madalena C. Marques[1], Pedro Moreira[1,2], Rita I. Teixeira[1], Marta Alenquer[3], Maria João Amorim[3], Cláudio M. Soares[1], João B. Vicente[1] and Diana Lousa[1]

1. Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal
2. Centro de Engenharia Biológica, Escola de Engenharia da Universidade do Minho, Braga, Portugal
3. Instituto Gulbenkian de Ciência, Oeiras, Portugal

In recent years we have felt the devastating impact of viral pandemics, highlighting the need for preparation for future pandemics. Antiviral biologics, including small proteins that bind to and block viral targets, are promising therapeutic options that should be explored to increase pandemic preparedness.

One of the viruses with high pandemic potential is influenza, the causative agent of flu. Despite being characterized by annual seasonal epidemics, global pandemics caused by this virus have occurred sporadically and unpredictably1. In the Influenza virus, the fusion of the viral and host membrane (a crucial step in infection) is elicited by hemagglutinin A (HA), a homotrimeric glycoprotein, which engages the virus with sialic acid receptors at the host cell surface and is a privileged target for antivirals2. The focus of this work is the design of Virus-Targeting Antibody-like scaffolds (ViTAIs), which can bind to HA, thereby blocking Influenza A entry into host cells. The design is based on innovative strategies that combine knowledge-based and physics-based computational methods to generate tens of thousands of ViTAIs, which are ranked according to relevant parameters, such as binding free energy and shape complementarity. The folding stability and conformational dynamics of selected designs are studied through molecular dynamics simulations to obtain a deeper knowledge of their properties and discard candidates that are predicted to be unstable. The candidates that pass all the computational filters are then produced in bacteria and tested using a platform based on biolayer interferometry to assess their binding affinity for the target and on differential scanning fluorimetry to evaluate their thermal stability. Those that have a high binding affinity and high thermal stability are produced in higher amounts and characterized using biophysical techniques and will subsequently be validated using in vitro neutralization assays. This work contributes to the development and validation of an innovative strategy that can be applied to tackle a broad range of viruses, including influenza A.

POSTER COMMUNICATIONS

Session 1 - Nº 18

Artificial Intelligence-Based Design of Antibody-like Engineered Protein Scaffolds

André R. F. Salgueiro[1,2], Pedro Moreira[1,3], Cláudio M. Soares[1], Leonardo Vanescchi[2] and Diana Lousa[1]

1. Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal
2. NOVA Information Management School, Universidade Nova de Lisboa, Lisboa, Portugal
3. Centro de Engenharia Biológica, Universidade do Minho, Braga, Portugal

Throughout humanity history, viral outbreaks caused devastating epidemics, like the recent COVID-19, caused by the SARS-CoV-2 virus originating in China. As of December 13, 2023, over 772 million confirmed COVID-19 cases, with more than 6.9 million deaths, have been reported globally to the World Health Organization. Over time, SARS-CoV-2 has evolved into different strains with varied characteristics, including immune system evasion and increased infectivity. This highlights the unpredictable nature of future pandemics, needing the development of new methodologies to combat a wide range of viruses. Therapeutic monoclonal antibodies, with their adaptability, show potential in targeting a broad range of viruses. However, their intricate and costly development processes hinder widespread use. Engineered protein scaffolds, such as monobodies1–3, which are smaller and simpler than monoclonal antibodies, offer a promising alternative. Their expression in bacteria makes their production and development processes simple and cost-effective, presenting exciting prospects for innovative treatments in cancer, infectious diseases, and autoimmune disorders. Although these engineered protein scaffolds have the potential to revolutionize medicine, we need efficient strategies to enable the design of molecules with tailor-made properties. The combination of machine learning methods like ProtGPT24, RFdiffusion5 and MSA Transformer6 and molecular dynamics (MD) simulations can be a powerful strategy to address this problem. This work aims to create a computational framework integrating machine learning techniques and MD simulations to streamline the development of engineered protein scaffolds that combine a high affinity for the target with optimal developability properties (including efficient production in bacteria and high physical and chemical stability). As proof of concept, we are focusing on designing protein scaffolds, including monobodies and helical miniproteins, to neutralize SARS-CoV-2. This approach holds promise for the development of biopharmaceuticals that can be adapted to a broad range of pathogenic agents, contributing to the ongoing battle against infectious diseases.

POSTER COMMUNICATIONS

Session 1 - Nº 19

Addressing the challenge of oligomerization in computational protein design

Diogo Silva [1], Pedro Moreira [1,2], Rita I. Teixeira [1], Bárbara Fernandes [1], Mariana Parada [1], João B. Vicente [1], Diana Lousa [1], Cláudio M. Soares [1]

1. Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal
2. Centro de Engenharia Biológica, Escola de Engenharia da Universidade do Minho, Braga, Portugal

Computational protein design is a field of research with potential to greatly impact areas such as drug development or enzyme technology, by combining knowledge-based and physics-based methods to design tailor-made proteins. This concept was introduced by David Baker with the development of Rosetta Commons. This is an extensive framework containing tools that enable protein design using physics-based scoring functions. Recently, the field of computational protein design has had major breakthroughs with the introduction of artificial intelligence (AI)-driven tools, namely RFDiffusion1 and ProteinMPNN2, which, when combined with AI-based 3D structure prediction tools (AlphaFold23), provide more efficient protein design pipelines.

This project aims to combine AI-driven methods with physics-based methods to re-design a protein binder and increase its affinity for a given target, considering potential oligomeric states of the design candidates. For this purpose, the receptor binding domain (RBD) of the Sars-CoV-2 spike protein was utilized as proof of concept, by starting with a known binder, that was predicted to dimerize, and improving its binding affinity to the RBD. The protocol starts with a molecular dynamics (MD)-based analysis of the original protein binder in solution (in monomer and dimer forms), as well as of the Target:Bindermonomer and Target:Binderdimer complexes. This guides the subsequent design steps, where ProteinMPNN2 is used to improve the binder's affinity for the target, considering both the monomer and dimer states of the binder in parallel re-design runs. The best candidates obtained in each re-design run will then be produced in bacterial systems, and their binding affinity and stability evaluated through binding assays and biophysical characterization, respectively. Hopefully, this project will contribute to the validation of computational AI-driven protein design as an approach that holds promise for biopharmaceutical applications and show the importance of combining MD simulations with protein design methodologies to obtain robust protein structures.

Exploring the Genomic Diversity of Hepatitis E Virus in European Rabbits: A Search for Optimal Primers

Margarida Cardeano Pinheiro[1,2,3], Rafael Vieira [1,2],Joana Abrantes[3,4,5], Diogo Pratas[6,7,8], Ana Margarida Lopes[3,4,9], João Carneiro [2]

1. Faculty of Sciences, University of Porto, Porto, Portugal
2. Interdisciplinary Centre of Marine and Environmental Research (CIIMAR), University of Porto, Porto, Portugal
3. CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal
4. BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal
5. Department of Biology, Faculty of Sciences, University of Porto, Porto, Portugal
6. IEETA/LASI, Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro
7. DETI, Department of Electronics, Telecommunications and Informatics, University of Aveiro
8. DV, Department of Virology, University of Helsinki
9. UMIB - Unidade Multidisciplinar de Investigação Biomédica, ICBAS - Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal

The Hepatitis E virus (HEV) is a global health concern, with zoonotic transmission impacting epidemiology. The virus has eight recognized genotypes, with genotypes 1 and 2 being exclusive to humans, while genotypes 3 and 4 have a broader host range and potential animal-to-human transmission. While historically linked to human infections, wild and domestic animals should not be disregarded. Indeed, the European rabbits (Oryctolagus cuniculus) may serve as a reservoir, prompting investigations into genomic diversity and potential zoonotic transmission within this host.
Our work aims to study HEV genomic diversity. An initial alignment of all rabbit HEV sequences (HEV-3ra) from NCBI database revealed 46.2% identical sites with a pairwise identity of 83.8%. These results indicate substantial genetic similarities and conservation, offering insights for future molecular research. For example, conserved regions may facilitate primer design for PCR amplification and sequencing, contributing to our understanding of viral evolution. We employed a tool named AROLit for the retrieval of primers from publicly available literature associated with HEV-3ra. After converting papers to text file format, this tool was used to extract the oligonucleotide sequences in each paper. A thorough examination identified a comprehensive set of 477 primers, primarily distributed within the size range of 15 to 27 base pairs. Following a Blastn analysis against a HEV-3ra reference sequence (GenBank accession number FJ906895.1), 145 primers exhibited a significant match with an e-value of ≤ 0.01, and 9 primers surpassed a bit score of 50.
Further refinement of this search is required in the next steps, which also include estimating primer conservation, additional filtering (e.g., self-folding), and restricting primer minimal sizes. These procedures are necessary to focus on primers that are both unique to the target virus and have properties that facilitate successful molecular identification.

Medical Procedure Recognition and Entity Linking in Spanish

Rafaela Lopes [1], Pedro Ruas [1], André Neves [2], and Francisco M. Couto [1]

1. Faculdade de Ciências da Universidade de Lisboa
2. Unicage Europe

MedProcNER (Medical Procedures Name Entity Recognition)1 is a shared task that aims to give a boost to the processing of clinical case records written in Spanish. Although there have been advances in terms of procedure mention detection extraction and normalization in English, there's still a big gap when it comes to other languages, such as Spanish. For example, this excludes valuable medical information from providing early diagnosis or improving epidemiological research.

The goal of the study is to adapt and improve the entity extraction tool BENT 2 for the MedProcNER task. BENT is a Biomedical Entity Annotator with the ability to perform Named Entity Recognition (NER) and Named Entity Linking (NEL) in a biomedical context. The application of BENT to this task will require several improvements, such as the addition of the terminology SNOMED CT, to deal with a new branch of information such as medical procedures and to work with Spanish text. Additionally, we plan to integrate the Unicage methodology and commands in the developed solution. The methodology allows the user to create and organized layered data system, whilst the commands are used to build highly efficient programs that can be combined in a modular way to build robust and flexible big data processing pipelines.

The evaluation benchmark includes annotations manually done by clinical experts. This evaluation uses precision, recall and F1-score to compare different approaches. BENT is an open source, and will continue to be, as the main goal is to help and reach as many people as possible. All information about MedProcNER is also freely available.

POSTER COMMUNICATIONS

Session 1 - N° 22

Discovery of new Ligands for Biopharmaceuticals Purification

Jéssica Rodrigues[1], Ana Cecília Roque[1], Arménio J. M. Barbosa[1]

1. Associate Laboratory i4HB - Institute for Health and Bioeconomy, Chemistry Department, NOVA School of Science and Technology, Campus Caparica, 2829-516 Caparica, Portugal UCIBIO – Applied Molecular Biosciences Unit, Chemistry Department, NOVA School of Science and Technology, 2829-516 Caparica, Portugal

Affinity Chromatography is widely used in biopharmaceuticals purification for therapeutics and diagnostics, a market expected to reach 300 billion USD in 2025. High-scale production of antibodies is very expensive due to its affinity chromatography step. Synthetic affinity ligands have garnered increased attention and been developed for biopharmaceutical purification: A2P, 22/8, DAAG, among others. Though, expensive biological ligands, like Protein A, are still the standard for antibody purification.

The current project sets to conduct the discovery of new synthetic affinity ligands binding to antibodies and understand their properties in affinity chromatography at a molecular level. Using the Petasis-Ugi scaffold, with 4 R-groups, it is possible to fine-tune affinity to antibodies. As example, the the Petasis-Ugi based B1Al2A2 ligand demonstrated affinity for IgG, Fab and Nanobody structures.

By screening virtual compound libraries, it’s possible to reduce the number of compounds requiring experimental testing while maximizing the chemical diversity. By using the Petasis-Ugi scaffold with commercially available building blocks, from chemical vendors, an in silico combinatorial library with approximately 800 million structures was enumerated. The combinatorial library will be screened towards a target antibody through molecular docking calculations. Afterwards, to extract critical information on the protein-ligand interactions Molecular Dynamics simulations of the best hits will be performed.

By integrating experimental design, virtual screening, and molecular dynamics simulations, this project is poised to expedite the discovery of new synthetic affinity ligands with robust binding affinity to antibodies, while striving to low the cost of biopharmaceutical production.

POSTER COMMUNICATIONS

Session 1 - N° 23

Characterization of metabolic phenotypes in Genome-scale models of complex diseases through Machine Learning

Rigoberto Rincón Ballesteros [1], Germán Andrés Preciat [1], Francisco Javier Álvarez Padilla [1]

1. Department of Translational Bioengineering, University of Guadalajara, Guadalajara,México

Complex diseases present a significant challenge due to their multifactorial origin, involving a combination of genetic, environmental, and lifestyle factors, many of which remain unidentified. The intricate interplay of these factors contributes to the inherent complexity of these conditions, making traditional paradigms of diagnosis and treatment insufficient to address these issues efficiently. In response to this challenge, personalized medicine emerges as a promising solution, by adapting diagnoses and treatments on the unique genetic and physiological characteristics of each patient. Bioinformatics tools play an essential role in generating computational models in this context and are commonly employed in personalized medicine. They integrate vast amounts of information to study the mechanisms and factors leading to certain pathologies. In this field, genome-scale models are utilized to explore the metabolic capabilities of a biological system. These network-based models are frequently employed in metabolic engineering, integrating information from genes, enzymes, reactions, and metabolites. They estimate metabolic flux of reactions, which collectively represent the metabolic phenotype, the outcome of interactions among various biological components. As part of the research project, a methodology is proposed to implement machine learning algorithms, specially focusing on clustering algorithms such as k-means, hierarchical clustering and DBSCAN. These algorithms will aim to characterize phenotypes in patient-specific genome-scale models by identifying and grouping them based on similar patterns of metabolic flux. To validate this methodology, a comprehensive strategy that combines cross-validation technique and performance metrics will be used to quantitatively assess the quality of clustering outcomes. By implementing this novel approach, the goal is not only to enhance the understanding of underlying metabolic mechanisms but also to establish the groundwork for diagnostics and the design of personalized therapies.

POSTER COMMUNICATIONS

Session 1 - N° 24

Exploring methane mitigation strategies in photosynthetic microorganisms through genome-scale metabolic models

Gonçalo Apolinário [1], Joana Gonçalves [1], Emanuel Cunha [1], Leandro Madureira [1], Filipe Maciel [1,2], Pedro Geada [1,2] and Oscar Dias [1,2]

1. CEB - Centre of Biological Engineering, University of Minho, Braga, Portugal;
2. LABBELS — Associate Laboratory, Braga/Guimarães, Portugal

The problematic of greenhouse gas (GHG) emissions is a global environmental challenge that has raised concerns in the past few decades. Particularly, the increase in atmospheric concentrations of carbon dioxide, methane, and other damaging gases can lead to catastrophic repercussions to life as we know it. Therefore, reducing GHG emissions and fomenting strategies for their mitigation are crucial steps that need to be taken in order to meet the Paris Agreement and ultimately ensure a sustainable future for our planet and those that inhabit it.

Over the last 20 years, metabolic models have been widely used as a source of information for metabolic engineering, drug targeting, metabolic pathway analysis, and process optimization. Genome-scale metabolic (GSM) models allow the in silico simulation and prediction of metabolic fluxes on a large scale, providing a powerful tool for optimizing and designing metabolic engineering methods. By integrating high-throughput data with genome-scale models, a comprehensive understanding of cellular metabolism and identification of strategies to improve a certain objective function can be obtained. The importance of this emerging technology in industry stems from its ability to offer a faster and more cost-effective approach, surpassing the efficiency of traditional methods.

Taking this into account, as well as the urgent search for sustainable solutions addressing GHGs mitigation, our work aims at identifying the metabolic capabilities of photosynthetic microorganisms to reduce methane emissions. In this regard, we investigate the potential of these microorganisms, using genome-scale metabolic (GSM) models to understand their metabolic networks in detail. Therefore, herein we describe the reconstruction of GSM models for the microalga Chlorella vulgaris sp. – iGA1305, includes 2635 reactions and 1305 genes – and for the cyanobacterium Synechocystis sp. – iJG708, includes 2165 reactions and 708 genes –, which will be applied to understand their CH4-related metabolic networks in detail. Both GSM models provide a powerful tool for metabolic improvement, allowing predictions and simulations of CH4 metabolism in response to different culture conditions and genetic modifications.

An Efficient Toolkit for Large-Scale Genomic Analysis and Its Application to Multi-Resistant Bacteria

Jorge Miguel Silva[1] ; José Luís Oliveira[1]; Armando Pinho[1] ; Diogo Pratas[1,2]

1. IEETA/DETI/LASI, University of Aveiro, Portugal.
2. DoV, University of Helsinki, Finland.

We present a comprehensive computational tool specifically developed to tackle the challenges of analyzing large-scale genomic datasets and include its application in studying multi-resistant bacteria. With the increasing prevalence of antibiotic resistance, understanding the genomic intricacies of these bacteria is crucial. Our toolkit addresses this need by enabling efficient analysis of vast genomic datasets, commonly comprising millions of sequences, which traditional alignment-based methods struggle to process due to their time-intensive nature and scalability issues.

The toolkit offers a robust, user-friendly platform that surpasses traditional alignment-based methods in efficiency, especially for datasets comprising millions of sequences.

Its standout feature is the efficient processing of large datasets without sacrificing computational speed, surpassing traditional methods that often require segmenting analyses to manage data effectively. An automatic filtering mechanism crucially maintains the accuracy and relevance of analyses by eliminating low-quality, biased, or aberrant sequences. This ensures the integrity and reliability of results, which is particularly vital in bacterial genomic studies where data quality is paramount.

The toolkit's core functionality extends to detailed sequence data analysis. It adeptly identifies unique and common patterns within high-quality bacterial sequences, facilitating an in-depth examination of their characteristics. This includes exploring static properties and temporal dynamics, offering insights into the evolution and behaviour of multi-resistant bacteria over time.

In summary, this toolkit emerges as a powerful, efficient resource for genomic research. Its alignment-free approach, coupled with stringent quality control and speed optimization, makes it invaluable in studying large datasets, including those of multi-resistant bacteria. Its open-source availability further contributes to its significance in the scientific community.

Development and deployment of an infrastructure for genomic health data

Miguel Santos[1,2,3], Jorge S. Oliveira[1,2,3], Daniel Faria[1,2,3], José Borbinha[1,2,3], Fernando Mira da Silva[1,2,3], Ana Teresa Freitas[1,2,3], Jorge Silva[1,4], José Luís Oliveira[1,4], Astrid Vicente[5], Hugo Martiniano[5], Mário Silva[1,2,3]

1. BioData.pt | ELIXIR-Portugal
2. Instituto Superior Técnico, Universidade de Lisboa
3. Instituto de Engenharia de Sistemas e Computadores - Investigação e Desenvolvimento,
4. Universidade de Aveiro
5. Instituto Nacional de Saúde Doutor Ricardo Jorge

Genomics is a scientific field with the potential to revolutionize healthcare by leading the development of precision medicine practices, that adjust the treatment of a patient based on his characteristics. It could also enhance the quality of diagnostics and prevention practices, and increase the efficiency of the usage of limited resources. This can be a game changer in the treatment of diseases with genetic predisposition, such as cancer, rare diseases, or neurodiseases, greatly improving the health conditions of citizens. To take advantage of this, researchers need access to a vast database of genomic data, annotated with visible characteristics of the individuals sampled, such as age, height, gender, and diseases or anomalies that are diagnosed. It is thus essential to create an infrastructure for European health research and healthcare that respects security and privacy laws and regulations across different Member States. The present document focuses on the deployment of a genomic data infrastructure, in the context of the 1 + Million Genomes (1+MG) initiative and the Genomic Data Infrastructure (GDI) project, whose purpose is to provide secure and authorized access to genomic data as well as the associated clinical data throughout Europe to improve research, healthcare, and health policymaking. The main result of this work was the deployment of the core services for the Portuguese node of the federated genomic data infrastructure developed by European partners. This includes services to securely archive and distribute genomic and clinical data, manage the access to it, search for relevant characteristics in existing datasets, and perform analysis pipelines remotely.

Acknowledgments:
This work was supported by national funds through FCT, Fundação para a Ciência e a Tecnologia, under project UIDB/50021/2020 (DOI:10.54499/UIDB/50021/2020).

Creating a Comprehensive Database of Plastic Degrading Enzymes for Machine Learning Applications

Clara Cerqueira[1,2], Mariana Fernandes[1,2], Sérgio Sousa[3,4], Diogo Pratas[5,6,7], João Carneiro[2]

1. Department of Computer Science, FCUP – Faculty of Sciences of the University of Porto, University of Porto,
2. BBEH, Blue Biotechnology, Environment and Health Group, CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto,
3. BioSIM, Department of Biomedicine, Faculty of Medicine, University of Porto,
4. UCIBIO, Applied Molecular Biosciences Unit,
5. IEETA/LASI, Institute of Electronicsand Informatics Engineering of Aveiro, University of Aveiro,
6. DETI, Department of Electronics, Telecommunications and Informatics, University of Aveiro,7DV,Department of Virology, University of Helsinki

The accumulation of plastics in the biosphere is one of the greatest concerns among scientists worldwide. Eventually, these plastics disintegrate into microplastics and enter both terrestrial and marine ecosystems, negatively affecting flora and fauna. These synthetic long-chained polymers are highly durable and resistant, and their basic materials are normally extracted from oil, coal and natural gas.As traditional plastic recycling processes have many disadvantages, such as being harmful to the environment and endangering public health, a more sustainable solution has emerged – biodegradation. Biodegradation occurs naturally without the need for human intervention through specific enzymes with bioremediative potential against plastic particles. These are microorganism produced enzymes that participate in the cleavage of plastic polymers into monomers. Excreting these enzymes enables microorganisms to use degradation products as a carbon or energy source. As the study of these types of enzymes has increased, a large amount of data has been produced regarding their physical and chemical properties. However, this information is widely dispersed across different articles and databases, making it our goal to unify it into a single database.

In this work, a custom workflow was developed to automatically retrieve enzyme-specific information from the PAZy database. PAZy is a database that exclusively lists biochemically characterized plastic-active enzymes that act on 13 different synthetic polymers, 4 of which still have no enzymes listed. Overall, activity, gene and protein data were collected for 205 unique enzymes that are listed in PAZy. The three most common host organisms in the database were T. fusca, Paenarthrobacter ureafaciens strain K172 and Amycolatopsis orientalis. As a result, a final table including the newly retrieved protein sequences and PDB structures was generated. Additionally, a unique identifier was created for the entire database. For prospects, the new database will be merged with preexisting online databases and others, while the new information will be used to obtain features essential to the implementation of machine learning (ML) models. The ML models will be used to predict the most promising variants of each plastic degrading enzyme.

Bridging Aptamer Information: Standardization and Curation of Diverse Databases

Rafael Vieira[1,2], Margarida Cardeano Pinheiro[1,2], Diogo Pratas[3,4,5], João Carneiro[2], Sérgio F. Sousa[6,7]

1. Department of Computer Science, Faculty of Sciences, University of Porto
2. Interdisciplinary Centre of Marine and Environmental Research (CIIMAR), University of Porto
3. IEETA/LASI, Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro
4. DETI, Department of Electronics, Telecommunications and Informatics, University of Aveiro
5. DV, Department of Virology, University of Helsinki
6. LAQV/REQUIMTE - BioSIM, Department of Biomedicine, Faculty of Medicine, University of Porto

Aptamers are small single stranded nucleotides (usually between 20 and 100 units) that have the ability to bind with relative high affinity and specificity to different molecular targets (including proteins and small organic molecules). This ability makes aptamers an useful tool as biomarkers, in detection of heavy metals and pollutants, and as intermediates for targeted drug delivery. Unlike traditional antibodies, which also display high specificity and affinity towards specific targets, aptamers do not trigger an immune response, and show higher malleability from a chemical perspective (which reflects in easier modifications without compromising function). Aptamers are typically obtained through an experimental technique, SELEX (Systematic Evolution of Ligands by Exponential Enrichment), which consists of a repetitive cycle of binding, washing and amplification of aptamers for a specific target.

Despite the effectiveness of the experimental methodologies, more recent developments of in-silico techniques have improved the understanding of aptamer-target interactions. With that in mind, the goal of this work is to integrate current known information of different aptamer databases into a standardized one. A well-organized database that follows the FAIR principles, will then allow for a deeper statistical study of factors such as: “how does nucleotide distribution influence the type of target a given aptamer binds to”.

So far, a database with more than 2300 aptamer-target pairs has been compiled from three different sources: IITG, University of Texas, and Aptagen databases. These databases contain aptamers descriptors (such as sequence and chemistry), the associated target (and its chemistry) and the corresponding binding affinity.

Moving forward, the goal is to further characterize the aptamer-target pair, adding features such as the target sequence, in case of proteins the pseudo amino acid composition, and hydrophobicity, and possibly, the three-dimensional descriptors of the aptamers. This work will culminate in the development of a machine learning model that will allow the prediction of the occurrence of binding to different types of targets.

Unifying Information on Plastic Degrading Enzymes Across Different Databases

Mariana Fernandes[1,2], Clara Cerqueira[1,2],Diogo Pratas[3,4,5], Sérgio F. Sousa[6], João Carneiro[2]

1. Department of Computer Science, Faculty of Sciences, University of Porto
2. Interdisciplinary Centre of Marine and Environmental Research (CIIMAR), University of Porto
3. IEETA/LASI, Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro
4. DETI, Department of Electronics, Telecommunications and Informatics, University of Aveiro
5. DV, Department of Virology, University of Helsinki
6. LAQV/REQUIMTE – BioSIM. Department of Biomedicine, Faculty of Medicine, University of Porto

Plastic pollution is one of today's major environmental challenges due to its ever-growing production and mismanagement combined with its long-term durability. Available plastic management methods include physical and chemical recycling, landfilling, incineration, and the adoption of biodegradable polymers; however, most are considered harmful to the environment and can compromise the quality of the resulting products. The enzymatic biodegradation of plastics is a sustainable solution for plastic elimination and has received increased amounts of attention from the scientific community in recent years. In this process, microorganisms breakdown polymers into smaller molecules, which can ultimately be assimilated and mineralized into unharmful compounds or reutilized in the plastic industry.

Our current goal is to build a database of enzymes reported to biodegrade plastics, integrating and updating information contained across several available data sources. In the initial stage of this task, we focused on consolidating the online database PlasticDB. We utilized Octoparse for web scraping relevant data from the database into a CSV file, including organism, enzyme, degraded polymer and protein sequences. Additional features were later added through data mining techniques. Furthermore, a script was created to export the protein sequences into FASTA files, according to the plastic types they degrade, to conduct some preliminary phylogenetic analysis.

We extracted a total of 212 entries from the PlasticDB, with cutinases, polyhydroxybutyrate depolymerases and esterases being the predominant enzymes. Among the 100 microorganisms capable of using their enzymatic apparatus for plastic degradation reported in the data, Thermobifida fusca, Pseudomonas chlororaphis and Cupriavidus necator were the most common.

The next steps in this work include merging all collected data with our pre-existing database and other databases available online. These data will be used to select the best phylogenetic variants based on key characteristics and structural data of plastic-degrading enzymes. These variants will be analysed mainly through 3D modelling techniques and optimized to improve enzymatic degradation activity.

POSTER COMMUNICATIONS

Session 2 - Software - N° 30

A metagenomics pipeline for the characterization of human gut microbiome in colorectal cancer

Rafaela Andrade[1,2], Francisco Fernandes[2], Gracinda M. M. Sanches-Fernandes[4], Ana Margarida Almeida[3], Ana Teresa Freitas[1,2]

1. Instituto Superior Técnico (IST), Universidade de Lisboa, Lisbon, Portugal.
2. Instituto de Engenharia de Sistemas e Computadores: Investigação e Desenvolvimento em Lisboa (INESC-ID), Lisbon, Portugal.
3. Instituto de Medicina Molecular João Lobo Antunes (iMM), Lisbon, Portugal.
4. BioChromoGene, Lisbon, Portugal.

Metagenomic studies have emerged as a primary method for understanding the human microbiome. They are essential for uncovering how changes in the microbiome's composition are related to various diseases. One possible example is colorectal cancer (CRC).

Differences in gut microbiota diversity throughout all CRC stages highlight the essential role of the microbiome in comprehending the development and evolution of this disease.

An extensive exploration of metagenomic samples involves a sequence of computational steps, collectively known as a metagenomics pipeline. This pipeline encompasses diverse stages, including data preprocessing, quality checks, assembly, binning, taxonomic or functional classification, and data interpretation. The main goal is to convert the initial raw reads obtained from sequencing platforms into what are known as Metagenome-Assembled Genomes (MAGs). The critical aspect lies in choosing and combining informatics tools tailored to our specific sample type.

In this work we have explored different bioinformatic tools to define a metagenomics pipeline that enhances the characterization of the gut microbiome, thereby enriching our insights.

Initially, FastQC, Cutadapt, bowtie2 and fastp were utilized for preliminary quality assessment and data preprocessing. Subsequently, high-quality reads from these steps were assembled into contigs using Megahit— a specialized assembler designed for complex samples like metagenomic datasets, utilizing succinct de Bruijn graphs for its assembly process.

Binning, a pivotal stage, involved the use of two distinct tools—Metabat2 and Semibin2—whose performances were compared. Following binning, quality evaluation using Checkm categorized bins based on contamination and completeness, retaining high and medium quality MAGs. Taxonomic classification of these MAGs was executed using GTDB-tk database.

Visual representation via bar and pie charts delineated the relative abundance of taxonomic groups, complemented by statistical metrics. The pipeline's application encompassed raw reads from 10 samples of healthy patients and 10 samples of stage III and IV CRC patients. The main goal was to compare the MAG sets from each group to discern differences in species abundance and diversity.

During the binning stage, Semibin2 generally yielded a higher number of bins for both healthy and diseased samples. Moreover, it generated a greater quantity of high-quality bins from these samples.

As preliminary results, from the 10 samples of stage III/IV CRC patients, Semibin2 identified a total of 10 distinct phyla and 183 unique species, whereas Metabat2 detected 9 phyla and 167 distinct species.

In the forthcoming investigation, one of the goals is to explore potential associations between the occurrence of antibiotic resistance genes and CRC. This examination seeks insights into potential connections between antibiotic resistance and this disease.

POSTER COMMUNICATIONS

Session 2 - Software - Nº 31

Precision Genome Analysis: Unraveling SNVs and CNVs with a Multi-Variant Caller WGS Workflow

MARTA FERREIRA[1,2,3,4], CELINA SÃO JOSÉ[2], FRANCISCO ALMEIDA[1,2,4], JOAQUÍN MAQUEDA[1,2], RITA MONTEIRO[1,2,5], PEDRO FERREIRA[1,2,4], CARLA OLIVEIRA[1,2,6]

1. Instituto de Investigação e Inovação em Saúde, Porto, Portugal
2. Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal
3. Doctoral Program in Computer Sciences Faculty of Science, University of Porto, Porto, Portugal
4. Department of Computer Science, Faculty of Science, University of Porto, Porto, Portugal
5. Currently at: Inovretail, SA, Maia, Porto, Portugal
6. Department of Pathology, Faculty of Medicine, University of Porto, Porto, Portugal

Exome analyses fail to detect single nucleotide variants (SNVs) or structural variants (SV) occurring outside the coding sequence. Genome-wide analysis emerges as a solution to understand genome variation, and Whole Genome Sequencing (WGS) the preferred method for that purpose. We aimed at developing a WGS analysis workflow using a combination of multiple variant callers for SNV, CNV and SV calling.

We used a gold standard sample from the Genome in a Bottle project (GIAB) to access the performance of a novel pipeline encompassing alignment (BWA mem); post-processing (GATK tools); CNV (integration of LUMPY, Delly and GRIDSS caller outputs) or SNV (integration of HaplotypeCaller-GATK (HC) and DeepVariant (DV) outputs) calling; and Merging of multiple called variants (overlap analysis of SNV, CNV and SV calls). We calculated recall, precision and F1 scores by comparing our outputs with those from GIAB for SNVs. As no gold standard is available in GIAB for CNVs, we compared our CNV outputs with a GIAB pool of high confidence and with variant calls from Manta, to improve our pipeline.

We called 4.309.554 SNVs with DV and 4.578.886 with HC, and 4.109.099 were common. The performance of DV alone (F1 score=0.9819, 3.802.474 true positives) was better than SNV calling with HC alone (F1 score=0.9522, 3.759.774 true positives), but worse than the combination of DV and HC outputs (F1 score=0.9841, 3.836.476 true positives variants). Our CNV calling pipeline called a higher number of variants than the, as well as a set of inversions confirmed by visualization in samplot and supported paired-end and/or split-reads, that were not called by GIAB or Manta.

Our WGS-pipeline shows high performance and improves the likelihood of finding true positive germline SNVs, captures high confidence CNVs and uniquely calls inversions.

Bioinformatics Open Days 2024

ANNEX I: SCIENTIFIC SUBMISSIONS

N°	Authors	Title	Accept as
1	M. Pilar Cabezas et al	Get to know MIMt! A new, smaller and curated 16S rRNA reference database with less redundancy and higher accuracy at species-level identification	Oral
2	Joana Silva et al	Prediction of novel small RNAs from Pseudomonas aeruginosa	Oral
3	Cátia Santos-Pereira et al	Finding novel genes coding for eco-friendly surfactants from hypersaline Iberian locations using metagenomic approaches	Oral
4	João C. Sequeira et al	Individual-based modelling elucidates about the role of nanomaterials on methane production	Oral
5	João Correia et al	An Automated Structure-Based Platform for Gene Discovery and Enzyme Engineering	Oral
6	Pedro Moreira et al	Computationally-designed miniproteins showing neutralization activity against SARS-CoV-2	Oral
7	Javier Montoya et al	Deep Learning on Chaos Game Representation for Resistome	Oral
8	Daniel Martins et al	Modeling deep neural network architectures to improve schizophrenia predictability using genotype data	Oral
9	Darmit Kumar et al	A network science approach to interpret multidimensional associations between human transcriptome, tissues and traits	Oral
10	Mariana Ribeiro et al	The Portuguese Variome from over 12,000 Exomes: Streamlining Clinical Diagnostics and Translational Research	Oral
11	Susana Valente et al	Runs of homozygosity: bioinformatic approaches for diagnostic purposes and population analysis in 12,000 exomes	Oral
12	Hugo Magalhães et al	Sequence to graph alignment based copy number calling using a flow network formulation	Oral
13	Patrícia Ataíde et al	Plugging the holes: an intuitive software tool for reproducibility in Molecular Biology	Oral
14	Sónia Silva et al	Exploring bioinformatics tools to characterize a new regulator of Candida glabrata biofilm matrix	Poster
15	Ríos-Castro, R et al	A preliminary metabarcoding study of the diversity of Eukaryotic and Prokaryotic communities in the socio-ecological system from Rias Baixas (NW Spain)	Poster
16	Maria Benedita Pereira et al	Computational design of antiviral biologics targeting Zika virus envelope protein	Poster
17	Madalena C. Marques et al	Antiviral proteins targeting Influenza A hemagglutinin: design, production and characterization	Poster
18	André R. F. Salgueiro et al	Artificial Intelligence-Based Design of Antibody-like Engineered Protein Scaffolds	Poster
19	Diogo Silva et al	Addressing the challenge of oligomerization in computational protein design	Poster
20	Margarida Cardeano Pinheiro et al	Exploring the Genomic Diversity of Hepatitis E Virus in European Rabbits: A Search for Optimal Primers	Poster
21	Rafaela Lopes et al	Medical Procedure Recognition and Entity Linking in Spanish	Poster
22	Jéssica Rodrigues et al	Discovery of new Ligands for Biopharmaceuticals Purification	Poster
23	Rigoberto Rincón Ballesteros et al	Characterization of metabolic phenotypes in Genome-scale models of complex diseases through Machine Learning	Poster

ANNEX I: SCIENTIFIC SUBMISSIONS

24	Gonçalo Apolinário et al	Exploring methane mitigation strategies in photosynthetic microorganisms through genome-scale metabolic models	Poster
25	Jorge Miguel Silva et al	An Efficient Toolkit for Large-Scale Genomic Analysis and Its Application to Multi-Resistant Bacteria	Software
26	Miguel Santos et al	Development and deployment of an infrastructure for genomic health data	Software
27	Clara Cerqueira et al	Creating a Comprehensive Database of Plastic Degrading Enzymes for Machine Learning Applications	Software
28	Rafael Vieira et al	Bridging Aptamer Information: Standardization and Curation of Diverse Databases	Software
29	Mariana Fernandes et al	Unifying Information on Plastic Degrading Enzymes Across Different Databases	Software
30	Rafaela Andrade et al	A metagenomics pipeline for the characterization of human gutmicrobiome in colorectal cancer	Software
31	Marta Ferreira et al	Precision Genome Analysis: Unraveling SNVs and CNVs with a Multi-Variant Caller WGS Workflow	Software

XIII EDITION BIOINFORMATICS OPEN DAYS

Thank you for joining us at XIII BOD!
Your participation and support are greatly appreciated. We look forward to seeing you at future events and continuing to grow our Bioinformatics community.

Bioinformatics Open Days
2024

/bioinformaticsopendays