01-03 Nov 2023

XIII Argentine Congress of Bioinformatics and Computational Biology (XIII CAB2C)

XIII International Conference of the Iberoamerican Society of Bioinformatics (XIII SoIBio)

III Annual Meeting of the Ibero-American Artificial Intelligence Network for Big BioData (III RiaBio)

Book of Abstracts

http://2023.a2b2c.org.ar

a2b2c

bioinformatica_ar

RiaBioNet

Electronic version of this book of abstracts:

http://2023.a2b2c.org.ar/BookOfAbstracts2023.pdf

ICB2C – Rosario 2023

Book of Abstracts

Este libro es una obra colectiva de los resúmenes enviados por sus autores y presentados en el 13vo Congreso Argentino de Bioinformática y Biología Computacional, la 13va Conferencia Iberoamericana de Bioinformática y la 3ra reunión anual de la red RiaBio realizada entre los días 01 y 03 de noviembre de 2023 en la ciudad de Rosario, Provincia de Santa Fe, Argentina.

Asociación Argentina de Bioinformática y Biología Computacional

Book of Abstracts ICB2C – Rosario 2023/ 1a ed ampliada. - Ciudad Autónoma de Buenos Aires: A2B2C, 2023.

Libro digital, PDF

Archivo Digital: descarga

ISBN: 978-987-48989-7-5

Asociación Argentina de Bioinformática
y Biología Computacional

Presidente A2B2C
Dr. Sebastián Fernández Alberti

Vice Presidenta A2B2C
Dra. Lucía B. Chemes

Secretario
Dr. Nicolás Palopoli

Tesorera
Dra. Georgina Stegmayer

Vocales
Dra. Elin Teppa
Dr. Diego Bustos

Vocales Suplentes
Dra. Cristina Marino
Dr. Flavio E. Spetale

Comité Organizador
Dr. Flavio E. Spetale
Dr. Gustavo Rodríguez
Dra. Dana Vazquez
Dra. Paolo Cacchiarelli
Lic. Ignacio Garcia Labari
Lic. Valentina Goytia Bertero

Comité Científico
Dra. Lucía B. Chemes
Dr. Esteban Mocskos
Dra. Elizabeth Tapia
Dr. Javier de las Rivas
Dra. Maribel Hernández-Rosales
Dra. Patricia Saragüeta
Dr. Gustavo Rodríguez
Dr. Flavio E. Spetale

01001
00100
10100

Our sponsors

Agencia I+D+i

CYTED

Santa Fe Provincia

SF500

ArgenTAG

HardCORE Computación

Toyoko

10100
10011
01000

01001
00100
10100

Our adherents

Municipalidad de Rosario

ROSARIO Convention & Visitors Bureau ARGENTINA

ENTE TURISTICO ROSARIO ARGENTINA

CONICET CIFASIS

FCEIA Facultad de Ciencias Exactas Ingeniería y Agrimensura

Facultad de Ciencias Agrarias Universidad Nacional de Rosario

UTN San Nicolás

IRAM

ISCB Regional Student Group Argentina

Universidad Nacional de Rosario — Confingere Hominem Cogitantem

ecu espacio cultural universitario Área Cultura UNR

Polo Tecnológico Rosario

10100
10011
01000

Table of contents

01001
00100
10100

Program

10100
10011
01000

Day 1: 1st November

09:00 to 09:20	Welcome Talk (Organizer Committee) Dr. Flavio E. Spetale - Responsible for the organizing committee Dra. Sandra Fernández - Director of the CCT CONICET Rosario Dr. Diego Moreno - Coordinator of the Area of Science, Technology and Innovation for Development of the UNR Marina Baima -Secretary of Science, Technology and Innovation of the Ministry of Production of Santa Fe
TRACK: Proteomics, Protein Structure, and Function	
09:20 to 10:00	***INVITED SPEAKER*. Gustavo Parisi (UNQ - CONICET, Argentina)** *Ghost interactions: revealing missing protein-ligand interactions using Alphafold predictions*
10:00 to 10:30	Coffee Break
10:30 to 10:45	***Lightning talk.* Matias Safranchik (IIB-UNSAM, Argentina)** *Comprehensive analysis of the pocket protein family interactions mediated by Short Linear Motifs*
10:45 to 11:00	***Lightning talk*. Franco Cuccovia Warletk (UNQ, Argentina)** *Functional characterization of baculoviral proteins with nuclear localization*
11:00 to 11:15	***Lightning talk.* Josefa Nuñez-Belmar (UMAYOR, Chile)** *New Bioinformatic Approaches in the Study of the Adhesive Fimbrial Systems (FimA) of the Human Oral Pathogen Porphyromonas gingivalis*
11:15 to 11:30	***Lightning talk.* Maribel Hernández-Rosales (CINVESTAV, México)** *Cellular Stress Response Mechanisms during Regeneration in Deroceras laeve: An Evolutionary Perspective*
11:30 to 11:45	***Lightning talk.* Luis Tataje Lavanda (UPSJB, Perú)** *Integrating Machine Learning for Enhanced Epitope Discovery and Vaccine Development in Avian Viral Pathogens*
11:45 to 12:00	***Lightning talk.* Sofia Rodriguez (USS, Chile)** *Analysis of regulatory networks based on single cell transcriptomics reveals Key Regulators in Rett Syndrome*
12:00 to 13:30	Break - Lunch
TRACK: Evolutionary, Comparative and Population Genomics	
13:30 to 13:45	***Lightning talk.* Mauro Ibañez (UNER, Argentina)** *Unraveling Key Features of Microbial Alpha Diversity Metrics and Their Practical Applications*
13:45 to 14:00	***Lightning talk.* José Gaete-Loyola (UMAYOR, Chile)** *Identification of microRNA families and their target genes involved dormancy stages into sweet cherry cultivars with contrasting chilling requirements*
14:00 to 14:15	***Lightning talk.* Mariano Elean (CERELA, Argentina)** *Genomic overview of Proteolytic Enzymes in Lactobacillus delbrueckii: Implications for Dairy Fermentation*
14:15 to 14:30	***Lightning talk.* Franco Fernández (CIAP-INTA, Argentina)** *Genomic approach to delimiting species boundaries in the phytopathogenic bacteria phytoplasmas: analysis of the X-disease clade*
14:30 to 14:45	***Lightning talk.* Marisol Navarro-Miranda (CINVESTAV, México)** *Modeling Microbial Competitive Interactions using Biosynthetic Gene Clusters (BGCs) and Network Features*

14:45 to 15:00	***Lightning talk.*** **Gabriel Krüger (UNAB, Chile)** *A genomic and artificial intelligence approach: adaptation and resistance of Salmonella in the poultry industry*
15:00 to 15:30	Coffee Break
15:30 to 17:00	***POSTER SESSION I***
17:00 to 17:40	***INVITED SPEAKER*****. Alejandra Cervera Taboada (INMEGEN, México)** *Analysis of human RNA-seq data from cardiac sarcoidosis, lymphocytic, and giant cell subtypes reveals novel insights into molecular mechanisms of myocarditis*

Day 2: 2nd November

09:00 to 09:20	RIABIO Session Welcome
TRACK: Machine Learning in Computational, Systems Biology and Bioinformatics	
09:20 to 10:00	***INVITED SPEAKER*****. Ignacio Ponzoni (ICIC - CONICET, Argentina)** *Explainable Artificial Intelligence: guidelines for its application to Drug Discovery*
10:00 to 10:30	Coffee Break
10:30 to 10:45	***Lightning talk.*** **Arlen Mosquera (PUJ, Colombia)** *Transposable elements classification using recurrent neural network*
10:45 to 11:00	***Lightning talk.*** **Matias Gerard (Sinc(i)-UNL-CONICET, Argentina)** *A novel approach for RNA folding inference based on message-passing graph neural networks*
11:00 to 11:15	***Lightning talk.*** **Ignacio Garcia Labari (CIFASIS-CONICET, Argentina)** *A music-inspired codification of ncRNAs with pseudoknots*
11:15 to 11:30	***Lightning talk.*** **Javier Sanchez-Galan (UTP, Panamá)** *Improvements in Recurrent Neural Networks for the Prediction of Molecular Properties and Characteristics of Natural Products*
11:30 to 11:45	***Lightning talk.*** **Leandro Bugnon (Sinc(i)-UNL-CONICET, Argentina)** *SincFold: a new tool for RNA folding based on deep learning*
11:45 to 12:00	***Lightning talk.*** **Paola Sosa Basso (FCIEN, Uruguay)** *Bioinformatics Analysis of lncRNAs in Kinetoplastid Parasites*
12:00 to 13:30	Break - Lunch
TRACK: Companies and startups bio-agro	
13:30 to 13:45	***Lightning talk.*** **María Laura Orcellet (SF-500, Argentina)** *How to create a science-based company?*
13:45 to 14:00	***Lightning talk.*** **Fernando Johann (Tell, Argentina)** *Language as a window to the early diagnosis of neurodegenerative diseases*
14:00 to 14:15	***Lightning talk.*** **Cristian Rohr (Heritas, Argentina)** *Data challenges*
14:15: to 14:30	***Lightning talk.*** **Alfredo García (Syngenta Agro, Argentina)** *Technology applied to agriculture: a solution to produce more and better*
14:30 to 14:45	***Lightning talk.*** **Iván Regali (DeepAgro, Argentina)** *Artificial Intelligence applied to agriculture*
14:45 to 15:00	***Lightning talk.*** **Eugenia Colomar (IRAM, Argentina)** *What's behind a product?*
15:00 to 15:30	Coffee Break

15:30 to 17:00	***POSTER SESSION II***
17:00 to 17:40	***INVITED SPEAKER*. Javier de Las Rivas (CiC-IBMCC, España)** *Bioinformatics & Machine Learning to analyse transcriptomic and clinical data from cancer patients*
21:00	***SOCIAL DINNER***

Day 3: 3rd November

09:00 to 09:20	Morning Welcome
TRACK: Genomics, Transcriptomics, and Metagenomic	
09:20 to 10:00	***INVITED SPEAKER*. Patricia Saragüeta (IIBYME-CONICET, Argentina)** *3D organization of the endometrial genome*
10:00 to 10:30	Coffee Break
10:30 to 10:45	***Lightning talk.* Kesia Barrows (UTP, Panamá)** *Comparison of bioinformatics tools for de novo genome assembly of Setophoma (Phaeosphaeriaceae), a genus isolated as an endophyte of Coffea arabica*
10:45 to 11:00	***Lightning talk.* Sabrina Costa-Tártara (UNLU, Argentina)** *Identification of regulatory motifs in putative coding genes of small Heat Shock Proteins in Chenopodium quinoa*
11:00 to 11:15	***Lightning talk.* Nicolas Lavatti (CONICET, Argentina)** *Transcriptomic meta-analysis to elucidate the presence and expression of neuropeptides in insect embryos*
11:15 to 11:30	***Lightning talk.* Danitza Silva García (UMAYOR, Chile)** *Transcriptomic exploration of dormancy mechanisms in sweet cherry varieties with contrasting cold requirements, under equitable agricultural conditions*
11:30 to 11:45	***Lightning talk.* Katia Avina Padilla (CINVESTAV, México)** *Understanding tomato Host Response to PSTVd-Infection: Insights from Network Analysis and Transcriptional Reprogramming*
11:45 to 12:00	***Lightning talk.* Santiago Prochetto (CONICET, Argentina)** *evolSOM: A Novel R Package for evolutionary conservation analysis using Self-Organizing Maps*
12:00 to 13:30	Break - Lunch
TRACK: Biomedical Omics, Biological databases and Bioimaging	
13:30 to 14:15	***INVITED SPEAKER*. Patricio Yankilevich (IBIOBA- MPSP-CONICET, Argentina)** *From data to diagnosis and treatment: An NGS data analysis pipeline*
14:15 to 14:30	Coffee Break
14:30 to 14:45	***Lightning talk.* Natalia Alonso-Moreda (CiC-IBMCC, España)** *Computational modeling and prediction of survival in breast cancer tumors based on myelo-lymphocyte ratios and hot-cold scores*
14:45 to 15:00	***Lightning talk.* Luciano Anselmino (CONICET, Argentina)** *Deciphering pharmacological targets via computational analysis for drug repositioning in treating 5-FU-resistant colorectal cancer*
15:00 to 15:15	***Lightning talk.* Luciana Ant (IByME, Argentina)** *Comparing cistromes from endometrial adenocarcinoma cell line and endometrial tumor samples*

15:15 to 15:30	***Lightning talk.*** **Alexander Mulet de los Reyes (UBA, Argentina)** *GBManalizerApp: A CAD tool for the three-dimensional automatic segmentation of the glioblastoma multiforme*
15:30 to 15:45	***Lightning talk.*** **Humberto Debat (INTA, Argentina)** *A Decade of Viral Exploration through High-Throughput Sequencing Data: Insights from the Global South*
15:45 to 16:00	***Lightning talk.*** **Juan Miguel Zuñiga-Umaña (CENAT, Costa Rica)** *Development of an automated system for determination of ultrastructural patterns in spores of pathogens and symbionts of agricultural relevance*
16:00 to 16:30	***Poster and Best Talk Prizes A2B2C - SoIBio Closing Ceremonies***

01001
00100
10100

Invited Speakers

10100
10011
01000

Gustavo Parisi

He has a Ph.D in Biochemistry from the Faculty of Exact Sciences. Currently, he is Principal Researcher (CONICET), Associate Professor (UNQ, UNLP) and director of the Structural Bioinformatics group. His main topics of interest are: Development of tools for the quality assessment of protein models; Study of the role of conformational diversity in protein evolution; Study of protein-protein interactions using structurally constrained models of sequence evolution; Study of the sequence-structure relationships in allergenic proteins.

Alejandra Cervera Taboada

She has a Ph.D in Biomedicine form University of Helsinki. Currently, she is Deputy Director of Population Genomics at the National Institute of Genomic Medicine of Mexico (INMEGEN). Her main topics of interest are: Computational Genomics and Expression Analysis; Genomics of Atopic Diseases: Genomics of Cardiovascular Diseases, Genomics of Liver Diseases, Genomics of Metabolic Diseases; Genomics of Autoimmune Diseases; Genomics of Psychiatric and Neurodegenerative Diseases; Cancer Genomics; Genomics of Bone Metabolism and Microbiome Genomics.

Ignacio Ponzoni

He has a Ph.D. in Computer Science from Universidad Nacional del Sur (UNS). Member of the Institute for Computer Science and Engineering (ICIC) as Independent Researcher at CONICET (National Council of Scientific and Technological Research). Full time Associated Professor at the Department of Computer Science and Engineering, UNS. Bahía Blanca, Argentina. Vice-head of the Laboratory for Research and Development in Scientific Computing (LIDeCC), Department of Computer Science and Engineering, UNS. His main topics of interest are: Development of Machine Learning and Evolutionary Computing techniques for Computacional Modeling in: Bioinformatics & Computational Biology; Chemioinformatics; Medical Informatics; Materials Science and Chemical Engineering.

Javier de las Rivas

He has MSc in Biology and PhD in Biochemistry and Molecular Biology by the University of the Basque Country (1990). In 1998 gained a position as "Cientifico Titular" in the CSIC and in 2006 he became "Investigador Cientifico". His research as experimental biochemist focused on biochemical studies on protein structure and function, mainly on redox proteins. Since 1998, he works in the new field of Bioinformatics and Systems Biology, with two stays for specialization in 1998 in the EMBL (Heidelberg, Germany) and in 2000-2001 in the Mount Sinai School of Medicine (New York, USA). Since 2002 he is staff scientist in the CIC (Salamanca, Spain) and leader of a research group in Bioinformatics, Functional Genomics and Systems Biology.

Patricia Saragüeta

She has a degree and doctorate in Chemistry. Currently, she is an Independent Researcher (CONICET) and a professor at the Department of Physiology and Molecular and Cellular Biology, FCEyN UBA. Her scientific work investigates cell biology, genomic and molecular mechanisms in reproductive systems. Since 2006 she has also dedicated herself to poetry, art and her connection with science. In 2007 she was selected to perform with Joan Jonas (USA).

Patricio Yankilevich

He has a Ph.D. in Molecular Biology, a MSc. in Bioinformatics & Neuroinformatics and a BSc. in Computer Science. Currently, he is an Independent Researcher and Bioinformatics Group Leader at Instituto de Investigación en Biomedicina de Buenos Aires – CONICET – Instituto Partner de la Sociedad Max Planck (IBioBA-CONICET-MPSP). Before this position I was a Product Manager at Integromics SL. and as Researcher at the Spanish National Center of Biotechnology (CNB). In Argentina I worked as Bioinformatics Laboratory Director at the Instituto de Agrobiotecnología de Rosario, and previously at BioSidus S.A. where I was the Bioinformatics unit P.I., and in Madrid at the Spanish National Cancer Research Centre (CNIO, Madrid) as Staff Scientist at the Bioinformatics unit. In 2017 I founded Genomap Bioinformatics.

01001
00100
10100

Lightning talks

TRACK: Proteomics, Protein Structure, and Function

10100
10011
01000

ID# 117

Comprehensive analysis of the pocket protein family interactions mediated by Short Linear Motifs

Matías Safranchik[1], Jacqueline Kramar[1], Johanna Kliche[2], Nicolás A. Garrone[1], Gustav Sundell[2], Lucía Alvarez[1], Carla Lorenze[1], Juliana Glavina[1], Norman Davey[3], Ylva Ivarsson[2] and Lucía B. Chemes[1]

[1]Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín, Buenos Aires, Argentina.
[2]Division of Cancer Biology, The Institute of Cancer Research, London, UK.
[3]Department of Chemistry - BMC, Uppsala University, Sweden.

Abstract

The pocket protein family, which includes retinoblastoma (pRb), p107 and p130, regulates cell cycle progression, differentiation and quiescence by binding to E2F transcription factors and chromatin regulators. Pocket protein interactions are mediated by the LxCxE and E2F short linear motifs (SLiMs). Although pocket proteins are highly relevant in cell cycle regulation, the number of known targets harboring these SLiMs is low. We performed a Proteomic Peptide Phage Display (ProP-PD) using the pocket domains of pRb, p107 and p130 as baits and libraries displaying intrinsically disordered regions of human and viral proteomes. ProP-PD returned over 500 hits including promising candidates with cell cycle, cell differentiation and transcriptional roles. Screens revealed a significant enrichment (>60%) in hits containing E2F-like and LxCxE-like SLiMs, with a high recall of true positive hits (35%) compared to the average recall for similar assays (19.3%). To test binding to the E2F cleft, we used pRb and p107 LxCxE cleft mutants. We tested 81 peptide hits by in vitro direct interaction assays. Over 50% of the hits were positively validated in pull down and Alpha Screen assays. To analyze binding mechanisms, we performed AlphaFold2 (AF2) in silico analysis by modeling pRb with 71 peptides including known true positive and novel ProP-PD candidate interactions. All true positives were accurately predicted: peptides docked to the expected cleft and side chains were correctly positioned, as shown by a low RMSD (70). True negatives and ProP-PD hits that were negatively validated in vitro showed a low confidence prediction in AF2 models. More than 80% of the peptides with a putative LxCxE motif docked to the LxCxE cleft with high confidence, as well as 30% of the peptides with the LxSxE variant. Last, 35% of the peptides matching the E2F motif docked to the E2F cleft with high confidence. All together these results suggest that novel functions can be identified for pocket proteins. We are currently working on AF2 modeling and trying to elucidate whether this strategy allows us to predict pocket protein SLiM-mediated interactions

ID# 112

Functional characterization of baculoviral proteins with nuclear localization

Cuccovia Warlet, FU[1]; Lassalle, FM[1]; Motta, LF[1]; Simonin, JA[1]; Belaich, MN[1]; Cerrudo, CS[1]

[1]Laboratorio de Ingeniería Genética y Biología Celular y Molecular –Área Virosis de Insectos (LIGBCM-AVI), Instituto de Microbiología Básica y Aplicada (IMBA), Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Roque Sáenz Peña 352 (B1876BXD). Bernal, Provincia de Buenos Aires. Argentina.

Background: Baculoviruses are insect pathogens with a double-stranded DNA genome and have multiple biotechnological applications, including heterologous protein expression systems, bioinsecticides, and the delivery of therapeutic genes. Several baculoviral proteins are crucial for viral DNA replication and transcription in the host cell nucleus or are necessary for nucleocapsid formation. Therefore, these proteins must be localized in the nucleus, imported through the nuclear import receptor (importin superfamily) using nuclear localization signals (NLSs), or positioned in the inner nuclear membrane (INM), guided there by the inner nuclear membrane sorting motif (INM-SM). Although several baculoviral proteins with NLSs and INM-SM have been identified, a better understanding of these baculoviral motifs is needed to determine all potential signals and pathways to the nucleus. In fact, only a few proteins with these motifs have been experimentally characterized and comparatively studied among different viruses.

Results: In our search for a generic sequential pattern associated with these two motifs, we conducted a study using an iterative approach. This involved enriching the NLSs of nine baculoviral proteins and the INM-SM of five baculoviral proteins (with experimentally characterized motifs) with the corresponding orthologous protein sequences from other baculoviruses, accomplished through our orthology pipeline. For each subset, we performed a multiple alignment (using MUSCLE) and, following manual verification and correction, generated syntactic patterns and sequence logos. Additionally, we conducted comparative bioinformatic characterizations using various tools (InterPro, Prosite, IUPred3, JPred4, AlphaFold2, CCTOP) to analyze the protein motifs and domains, secondary and tertiary structures, and the presence of post-translational modifications and specific topologies of the proteins under consideration. This comprehensive approach enabled us to structurally characterize the motifs under study and create syntactic patterns. Finally, we identified NLSs and INM-SM in the overall baculoviral proteome using ScanProsite and the corresponding patterns.

Conclusions: The study of the sequential and structural characteristics of baculoviral NLS and INM-SM allowed the generation of common patterns that describe the nuclear targeting signals in baculoviral proteins. This enhances our understanding of their proteomes' functionality and facilitates rational genetic interventions aimed at enhancing their biotechnological applications

ID# 070

New Bioinformatic Approaches in the Study of the Adhesive Fimbrial Systems (FimA) of the Human Oral Pathogen Porphyromonas gingivalis

Josefa Nuñez-Belmar[1,2], Mauricio Morales-Olavarría[1], Damariz Gonzalez[1], Cristian Cortez[3], Juan P. Cardenas[1,4], Jaime Andrés Rivas-Pardo[1,4]

[1]Centro de Genómica y Bioinformática, Facultad de Ciencias, Universidad Mayor, Santiago, Chile
[2]Doctorado en Genómica Integrativa, Universidad Mayor
[3]Escuela de Tecnología Médica, Facultad de Ciencias, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile [4]Escuela de Biotecnología, Facultad de Ciencias, Universidad Mayor, Santiago, Chile.

Abstract

Periodontitis is a chronic, immuno-inflammatory disease characterized by the loss of teeth-supporting tissues, which ultimately leads to bone loss. Porphyromonas gingivalis plays a crucial role in causing dysbiosis in periodontitis. The first step of host infection is the adhesion of fimbriae (FimA) to gingival cells. Previous studies have shown a correlation between the presence of different FimA genotypes and the capacity to adhere to the host epithelium. Type I FimA is related to general health, while Type IV is associated with periodontitis. Currently, there are few comparative studies linking FimA genotypes to the disease phenotype. The aim of this study is to determine the diversity and evolutionary properties of FimA.

To conduct phylogenetic and evolutionary analyses of FimA, we initially performed a phylogenomic analysis using 84 publicly available high-quality genomes of P. gingivalis. We extracted FimA sequences from the proteomes (annotated with Prokka). All phylogenetic analyses were performed using IQTREE, and the results were visualized with Figtree. Evolutionary pressure among sites was calculated using the FEL model from the Datamonkey Server. Allele diversity was calculated with Tajima's D value.

The phylogenomic analysis suggests that P. gingivalis genomes tend to form clusters between high and low virulent strains, but its relation with FimA genotyping is not monophyletic. This could also be related to virulence, as there are genotypes associated with high or low virulent strains. The phylogenetic tree of P. gingivalis FimA suggests that genotype IV of fimbriae is the most ancestral one, which is associated with several phenotypes of periodontitis and virulent strains. From datamonkey analyses, FimA has only two sites that are positively selected ($p \leq 0.05$), along with a set of 101 sites that are negatively selected. Tajima's D value was 0.89 ($p > 0.05$), suggesting that there is no excessive allele diversity detected in the population of fimA genes

This comparative genomics study allowed us to describe the differences between the sequences of FimA and relate them to the virulence of P. gingivalis strains. Our study strengthens the idea that FimA could be used as a new therapeutic target

ID# 102

Cellular Stress Response Mechanisms during Regeneration in Deroceras laeve: An Evolutionary Perspective

Luis Felipe Hernández Ramírez[1], Alfredo Varela Echavarría[2], Wilbert Gutiérrez[2], Emilio Ortiz[2], Katia Aviña[1], Maribel Hernández-Rosales[1]

[1]Laboratorio de Bioinformática y Redes Complejas, Departamento de Ingeniería Genética. Cinvestav-Unidad Irapuato, México
[2]Varela-Echavarría Lab, Instituto de Neurobiología, U.N.A.M. Campus Juriquilla, México

Background: Evidence from Sacoglossan sea slugs known for their unique ability to sequester photosynthetic-active plastids known as kleptoplasty and extreme autotomy of some species followed by high regeneration capabilities has recently emerged for the intricate interactions between kleptoplasts and the host's immune system, mediated by pattern-recognition receptors, scavenger receptors and C-type lectin receptors emerge as key players, not only in kleptoplast(plastids that have been sequestered) recognition but also in orchestrating molecular responses that safeguard genome stability and cellular integrity. Our hypothesis posits that stress response mechanisms not only defend against ROS but also activate DNA/protein repair, and genome stability, and that, this response could be conserved among the Mollusca phyla and other model organisms.

Results: Our findings transcend the specific context of Deroceras Laeve, suggesting the potential conservation of these mechanisms within the broader spectrum of Mollusca phyla and other well-established animal models for regeneration. In our pursuit of understanding the intricate interplay of cellular processes during regeneration, the assembly of a de novo reference transcriptome was pivotal. This resource provides a robust foundation for future investigations in regenerative biology, enabling a more comprehensive exploration of stress response mechanisms in diverse species and contexts.

Conclusions: This research, which bridges the knowledge gap within the field of regenerative biology, demonstrates the potential conservation of stress response mechanisms across different species, including mollusks and other established animal models for regeneration. Beyond regenerative biology, these findings have broader implications, particularly in the context of cancer research. The overlap between molecular mechanisms in regeneration and those associated with cancer cell growth hints at the possibility of identifying biomarkers or therapeutic candidates for cancer treatment.

ID# 035

Integrating Machine Learning for Enhanced Epitope Discovery and Vaccine Development in Avian Viral Pathogens

Luis Tataje-Lavanda[1,2,3], Edith Málaga[4], Manuela Verastegui[4], Egma Mayta[2], Eliana Icochea[5], Manolo Fernández-Díaz[1], Mirko Zimic[1,6]

[1]Research and Development Laboratories, FARVET SAC, Chincha Alta, Ica, Peru
[2]Laboratory of Clinical Molecular Virology, Faculty of Biological Sciences, National University of San Marcos, Lima, Peru
[3]School of Human Medicine, Private University San Juan Bautista, Lima, Peru
[4]Research Laboratory on Infectious Diseases, Cayetano Heredia Peruvian University, Lima, Peru
[5]Avian Pathology Laboratory, Faculty of Veterinary Medicine, National University of San Marcos, Lima, Peru
[6]Bioinformatics, Molecular Biology, and Technological Developments Laboratory, Faculty of Science and Philosophy, Cayetano Heredia Peruvian University, Lima, Peru

Background: Bioinformatics and computational biology have become indispensable in devising innovative strategies to combat infectious diseases. Immunoinformatics, in particular, has emerged as a cost-effective tool for vaccine development against viral pathogens. In this multidisciplinary study, we combine bioinformatics, machine learning, immunology, biology, and veterinary science to identify and assess immunogenic epitopes as potential vaccines for poultry-affecting viruses, specifically focusing on Newcastle disease virus (NDV) and avian infectious laryngotracheitis virus (ILTV).

Results: Employing an in-silico approach, we selected peptides with high affinity for the chicken major histocompatibility complex type I (MHC-I). Our artificial neural network, CHANN, designed for the Cobb-500 chicken variety, successfully identified conserved peptides within NDV proteins (Polymerase L, HemagglutininNeuraminidase HN, Fusion F, and Matrix M) and ILTV proteins (Transcriptional Regulator ICP4, Glycoprotein G gG, and Glycoprotein B gB). Subsequent in vitro evaluations revealed the ability of five selected NDV peptides to induce a significant cellular immune response, as evidenced by interferon gamma (IFN-γ) production.

Conclusions: This study presents a novel perspective for developing multi-epitope vaccines against NDV, potentially surpassing conventional vaccines' limitations and offering a safer, more effective immunization approach. The successful synergy of bioinformatics, computational biology, and immunology, bolstered by machine learning, holds promise for avian pathogen vaccine design. The integration of computational and experimental methodologies is pivotal in enhancing the effectiveness of vaccines against pathogens like NDV, with potential benefits for food security and avian health, not only in Latin America but also globally.

ID# 028

Analysis of regulatory networks based on single cell transcriptomics reveals Key Regulators in Rett Syndrome

Sofia Rodriguez[1,4]; Camilo Villaman[2,4]; Mauricio Saez[3]; Alberto J. Martin[4]

[1]Programa de Doctorado en Biología computacional, Universidad San Sebastián, Centro Científico y Tecnológico de Excelencia Ciencia & Vida

[2]Programa de Doctorado en Genómica Integrativa, Universidad Mayor, Santiago, Chile., Centro Científico y Tecnológico de Excelencia Ciencia & Vida, Santiago, Chile

[3]Facultad de ciencias de la salud. Departamento de procesos diagnósticos y evaluación

[4]Facultad de Ingeniería, Arquitectura y Diseño, Universidad San Sebastián, Centro Científico y Tecnológico de Excelencia Ciencia & Vida, Santiago, Chile

Background: Rett syndrome (RS) is a rare developmental disorder primarily caused by mutations in the gene encoding the methyl-CpG binding protein 2 (MECP2) on the X chromosome. with Female patients accounting for 95% of the total diagnosis rate. The Compensatory mechanisms regulating the expression of the two X chromosomes lead to the stochastic silencing of genes in them (XCI), resulting in cellular mosaicism which has been linked to symptom severity for this condition depending on the distribution of MECP2+/- expressing cells; the functional role of MECP2 as a transcriptional regulator contributes to the complex nature of the condition, the combination of these factors results in a condition that requires both methods that can embrace the heterogeneous nature of the cells, and can allow us to evaluate this condition wholly, thus Single Cell RNA-seq in combination with Regulatory Network methods arise as a viable alternative to explore the data.

Results: Currently, we could generate Pseudo time trajectory and developmental timeframe-based clustering of the samples, obtaining a comprehensive trajectory for neuron development and differentiation processes within Rett syndrome, resulting in 31 unique developmental clusters, with 5 unique MECP2- developmental points, and differences in distribution and expression on MECP2- and MECP2- cells within the same developmental points.

Complementary to this process we performed a cellular type base clustering by annotating the groups formed and study the difference between MECP2+ groups and MECP2-negative groups, from which we were able to identify several regulatory and compensatory mechanisms for Rett syndrome, with the most significant results being a compensatory mechanism of compensatory master regulators in GABAergic network, ABCB1 associated regulatory networks absence in MECP2- cells, and a system of Non-coding RNA related regulatory mechanism within dopaminergic neurons in Rett syndrome.

Conclusions: Pseudotime calculations enabled the reconstruction of developmental sequences, providing a comprehensive understanding of the disrupted developmental pathways of Rett syndrome, emphasizing the importance of understanding the underlying cellular mechanisms.

Network analysis identified master regulators (MR) responsible for modulating expression in GABAergic neurons. Identifying a system with MR exclusive to MECP2+ neurons and MR exclusive to MECP2- neurons, co-regulating each other providing insight within Rett syndrome's compensatory mechanisms.

01001
00100
10100

Lightning talks

TRACK: Evolutionary, Comparative and Population Genomics

10100
10011
01000

ID# 008

Unraveling Key Features of Microbial Alpha Diversity Metrics and Their Practical Applications

Ignacio Cassol[1*], Mauro Ibañez[2*], Juan Pablo Bustamante[1,2]

* Equal contributions

[1] Facultad de Ingeniería, Universidad Austral, Argentina, LIDTUA, CIC

[2] Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Argentina

Abstract

Proteins that are strongly driven towards aggregation in the form of amyloid fibrils are called amyloidogenic. In this work, we study the evolutionary rates of 81 human proteins for which an in vivo amyloid state is supported by experiment-based evidence. We found that amyloidogenic proteins evolve faster than the reference dataset (~16500 proteins from *Homo sapiens* with known orthologs on *Mus Musculus*), although they are highly expressed and abundant. In spite of the lack of significant differences in the evolutionary rates of secreted and amyloidogenic proteins, we found substantial differences in other features such as in their tendency to aggregate.

Firstly, conformational diversity is higher in amyloidogenic proteins, also evidenced by the higher presence of disordered or highly flexible regions. The high conformational diversity could increase the chances of exposing amyloid-prone regions in slightly unfolded conformers driving the protein towards fibril formation. Secondly, amyloidogenic proteins are more expressed and abundant than secreted proteins. The impact of protein concentration and solubility on amyloid formation has been extensively studied. As it was described, to remain soluble, abundant proteins require the constant assistance of quality control mechanisms such as molecular chaperones.

Furthermore, we found a positive correlation between RMSD100 (a measure of the degree of conformational diversity) and evolutionary rates which indicates that the higher the conformational diversity, the higher the evolutionary rates, possibly indicating a higher propensity to aggregate. Thirdly, a linear model combining an intrinsic protein parameter, as the RMSD100, with a cellular condition, as the supersaturation score, better explains the variation in the evolutionary rates observed in amyloidogenic proteins but not in secreted ones.

Emerging evidence suggests that amyloidogenic proteins represent a “metastable subproteome” strongly driven towards the formation of amyloid fibrils. In the end, we showed that evolutionary rates reflect this particular behavior, showcasing the importance of metastability above other modulating factors. In the future, it could be interesting to evaluate protein metastability as a general modulating factor of evolutionary rates at the proteome level.

ID# 021

Identification of micro-RNA families and their target genes involvedin dormancy intwo sweet cherry cultivars with contrasting chilling requirements

José Gaete-Loyola[1], Humberto Prieto[2], Andrea Miyasaka Almeida[1,3]

[1]Centro de Genómica y Bioinformática, Facultad de Ciencias, Ingeniería y Tecnología, Universidad Mayor, Huechuraba, Santiago, Chile.

[2]Instituto de Investigaciones Agropecuarias INIA La Platina, La Pintana, Santiago, Chile.

[3]Escuela de Agronomía, Facultad de Ciencias, Ingeniería y Tecnología, Universidad Mayor, Huechuraba, Santiago, Chile.

Abstract

Sweet cherry is a deciduous tree that faces unfavorable seasonal climate conditions. In autumn these trees must enter a stage called dormancy, in which they cease growth to endure low temperatures. During this stage, trees must fulfill chilling requirement (CR) to release dormancy and resume active growth in spring. This process is crucial for optimal growth, flowering, and fruit production. Micro RNAs play an important role in the regulation of dormancy in this species through posttranscriptional gene silencing guiding mRNA cleavage or translational repression. Therefore, these molecules participate in plant development, reproduction, and genome reprogramming, thus contributing to the phenotypic plasticity of the plant. The objective of this study was to identify miRNA families and their target genes involved in stages of dormancy in two sweet cherry cultivars with contrasting CR.

To achieve this, small RNA libraries from floral bud samples at three stages of cumulative chilling hours were generated for Regina and Tulare, which corresponds to high and low CR cultivars, respectively. For Regina, these stages were 200, 1160 and 1700 chilling hours (CH), whereas for Tulare these were 200, 400 and 500 CH. After quality control of the libraries, reads were mapped to Prunus avium cv. Tieton reference genome v2.0. A total of 16 and 44 micro RNAs families were identified in Regina and Tulare, respectively. Fourteen families were shared between the two cultivars, 2 were identified exclusively in Regina and 5 in Tulare. miR396 family was found 4 times in each cultivar, followed by miR166 and miR167 with 3 each. A target prediction analysis allowed to identify 298 and 338 candidate target genes in Regina and Tulare, respectively.

The gene ontology analysis of these targets indicated that genes were enriched in categories of developmental process, regulation of transcription, regulation of gene expression, regulation of nucleic acid-templated transcription. Candidate genes were filtered for subsequent analysis and lab tests using mRNA-seq data publicly available in NCBI. These results will allow us to select micro RNAs to be tested in plants as a tool for specific gene silencing to modulate chill requirement completion and flowering time.

Funding: ANID/ACT21000

ID# 037

Genomic overview of Proteolytic Enzymes in Lactobacillus delbrueckii: Implications for Dairy Fermentation

Mariano Elean[1], Leonardo Albarracin[1], Julio Villena[1], Lucila Saavedra[1] and Elvira M. Hebert[1]

[1]Reference Centre for Lactobacilli (CERELA-CONICET), Tucumán 4000, Argentina

Background: Lactobacillus delbrueckii is widely recognized as the primary starter culture in the dairy industry for the elaboration of a variety of fermented dairy products, such as yogurts, fermented sour milks, mozzarella, and Swiss and Italian cheeses. This group of microorganisms has a limited capacity to synthesize amino acids and is therefore dependent on the use of exogenous nitrogen sources for optimal growth. As milk contains only small amounts of amino acids and short peptides, Lactobacillus depend on a complex proteolytic system to obtain essential amino acids from caseins during growth in this substrate. The cell envelope proteinase (CEP) stands as a pivotal enzyme, initiating the hydrolysis of casein. Strain-specific differences in proteolytic activity have been observed, prompting an investigation into potential disparities in CEPs.

Results: In this study, we analyzed the genomes of 27 L. delbrueckii strains. Despite all strains featuring a singular CEP gene, certain strains (KCTC 13731, NBRC 3202, and KCTC 3035) exhibited premature termination codons, leading to truncated proteinase variants. Two distinct proteinases were identified: PrtB in L. delbrueckii subsp. bulgaricus, and PrtL in L. delbrueckii subsp. lactis and L. delbrueckii subsp. delbrueckii. Remarkably, these proteinases shared over 95% amino acid identity, with L. delbrueckii subsp. bulgaricus variants even surpassing 98% identity. Most amino acid substitutions maintained the conservative nature of the three-dimensional structure, including an unchanged catalytic triad. The comparisons identified syntenic blocks common among the genomes highlighting significant conservation surrounding the prt gene, particularly within L. delbrueckii subsp. bulgaricus and L. delbrueckii subsp. lactis. Additionally, sequence conservation was observed in the promoter region of the prt gene.

The substantial conservation of the amino acid sequence of Prt and its promoters implies that disparities in activity among strains likely stem from regulatory variances. These variations might become evident through transcriptional, translational, or modulation of enzyme activity levels. Future research will aim to uncover these regulatory mechanisms.

Conclusions: By shedding light on the proteolytic intricacies of L. delbrueckii, this research advances our comprehension of the microbial mechanisms underpinning dairy fermentation processes. These insights hold the potential to improve both the quality of dairy products and production efficiency.

ID# 005

Genomic approach to delimiting species boundaries in the phytopathogenic bacteria phytoplasmas: analysis of the X-disease clade

Fernández, FD[1]; Alessio, FI[1]; Bongiorno, VA[1]; Conci, LR[1]

[1]Instituto de Patología Vegetal CIAP-INTA, Argentina

Abstract

Phytoplasmas are phytopathogenic bacteria with global relevance. Traditionally, phytoplasma taxonomy relied on 16S rRNA gene analysis, revealing limitations in distinguishing closely related strains. Progress in sequencing and genomics during the last decade has facilitated precise species differentiation through genome sequencing. Our study focuses on the X-disease group, a specific clade of phytoplasmas, due to its regional significance as a causal agent of diseases in different crops. We explored the connection between 16S rRNA-based classification and genomic homology parameters Average Nucleotide Identity (ANI) and digital DNA-DNA hybridization (dDDH), analyzing nine genomes from public databases.

Intriguingly, our genomic analysis unveiled the presence of four putative new candidate species, identified through ANI and dDDH scores. Additionally, we performed an identification of Single-Copy Orthologous Genes (SCOGs) for subsequent phylogenomic exploration. We found 263 SCOGs used to discern phylogenetic relationships. Notably, these phylogenomic results aligned seamlessly with the genomic homology findings. Furthermore, our efforts were dedicated to identifying novel marker genes that could enhance species differentiation accuracy. Analyzing polymorphism, phylogeny, and PCR amplification, we identified seven new marker genes (secA, secY, cimH, tuf, ctpE, mutM, and gpmI). Subsequent PCR and sequencing on X-disease local isolates confirmed these markers effectively delineate species, supported by whole-genome analysis.

In essence, our study bridges the gap between conventional 16S rRNA-based classifications and contemporary genomic analyses for phytoplasmas, specifically within the X-disease clade. The identification and validation of novel marker genes empower researchers with accurate tools for species differentiation, while the detection of candidate species underscores the need for ongoing genomic contributions to refine our understanding of these pathogens.

ID# 089

Modeling Microbial Competitive Interactions using Biosynthetic Gene Clusters (BGCs) and Network Features

Marisol Navarro-Miranda[1], Víctor M. Higareda Alvear[2], Obed Ramírez Sánchez[2], Gabriela Olmedo-Álvarez[1], Maribel Hernández-Rosales[1]

[1]Department of Genetic Engineering, Irapuato Unit, Center for Research and Advanced Studies of the National Polytechnic Institute (CINVESTAV), Irapuato, Mexico
[2]Department of Soil Genomics & Discovery, Solena Ag., Leon, Mexico

Abstract

In this study, we investigated a collection of 78 bacterial sediment isolates obtained from a shallow lagoon. We previously assessed their ecological traits through networks approach from pairwise interactions assays and classifying them taxonomically. Our research aimed to unveil intricate relationships between BGCs and observed microbial interactions, emphasizing the power of genomics in deciphering community dynamics. Understanding community assembly and stability remains a fundamental challenge in microbial ecology. We studied a well-characterized bacterial community consisting of 78 strains collected from sediment from a shallow lagoon in Cuatro Cienegas, Mexico.

Initially, we explored pairwise interactions with a network approach, classifying them according to their degree of antagonism or sensitivity. Building upon this foundation, our research had two main objectives. The first, to know if there was a link between the taxonomic classification of the strains and the role assigned on the basis of antagonism interactions. Second, to take advantage of the power of machine learning techniques in deciphering community dynamics, to uncover possible relations between BGCs and observed microbial interactions.

Our findings showed the ecological properties assigned to each strain were clearly a function of its taxonomic identity. Furthermore, we identified gene clusters encoding secondary metabolites as key drivers of community dynamics, illustrating their role in shaping the ecosystem. Additionally, by applying machine learning techniques to predict microbial interactions based on BGCs among the strains, we unveiled intricate relationships between BGCs and observed microbial interactions, emphasizing the power of genomics in deciphering community dynamics.

Understanding the importance of BGCs in competitive ecological interactions may offer new venues for investigating ecological dynamics in microbial communities where BGCs could play pivotal roles. Further research is essential to uncover the intricate mechanisms underlying BGCs as a valuable tool for predicting and manipulating microbial interactions within communities.

ID# 042

A genomic and artificial intelligence approach: adaptation and resistance of Salmonella in the poultry industry

Gabriel Krüger[1], Coral Pardo-Este[1,2], Phillipi Zepeda[1], Valentina Salinas[1], Valentina Pavez[1], Francisca Urbina[1], Juan Castro-Severyn[3], Jorge Olivares-Pacheco[4], Claudia Saavedra[1]

[1]Laboratorio de Microbiología Molecular, Facultad de Ciencias de la Vida, Universidad Andrés Bello, República 330, Santiago, Chile
[2]Laboratorio de Ecología Molecular y Microbiología Aplicada, Departamento de Ciencias Farmacéuticas, Facultad de Ciencias, Universidad Católica del Norte, Antofagasta, Chile
[3]Laboratorio de Microbiología Aplicada y Extremófilos, Departamento de Ingeniería Química, Universidad Católica del Norte, Antofagasta, Chile
[4]Grupo de Resistencia Antibacteriana en Bacterias Patógenas Ambientales GRABPA, Instituto de Biología, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

Background: The study of Salmonella resistance to different industrial environments is critical to understanding their adaptation to various stressors. Comparative genomics and advanced sequencing technology make it possible to track the spread of pathogenic bacteria. In the poultry industry, Salmonella enterica is notable for its multi-resistance and virulence. This research focuses on the adaptation of Salmonella isolates to multiple stress factors present in a chicken farm, where the constant use of disinfectants promotes bacterial antimicrobial tolerance and decreases their diversity. Using an artificial intelligence approach, we tried to identify which modifications are necessary in the Salmonella genome to survive the industrial stressful.

Results: We characterized the genomes of 186 Salmonella isolates, identifying some mobile genetic elements associated with antibiotic resistance. Subsequently, physiological characteristics of the strains were evaluated in response to sodium hypochlorite, acid and osmotic stress and antibiotic resistance. The results show that 75% of the isolates are resistant to sodium hypochlorite, exceeding permitted application limits, and 81% were classified as biofilm-forming, doubling or quadrupling their resistance. In particular, the identified isolates of serotype Infantis showed greater resistance to various types of stress, highlighting the slaughterhouse as a focus of resistance to antimicrobials and other stressors. In addition, we determined the transcriptomic profiles of two strains isolated from slaughterhouse and farm resistant to the stress factors evaluated, a classification of resistance markers was obtained, searching objectively for patterns of response to environmental stress at the level of expression of genes involved. Finally, a machine learning system based on a filter/wrapper strategy was developed by integrating genomic, transcriptomic and physiological data to identify Salmonella adaptation to antimicrobials in the production line of a broiler farm, with a precision of 97.67% and accuracy of 98.71% in isolate classification and Matthew's coefficient of 0.9678.

Conclusions: The importance of understanding Salmonella adaptation and resistance in the poultry industry is highlighted. The integration of advanced technologies and artificial intelligence tools holds promise for addressing these challenges. This study seeks to establish a predictive capability based on genomic surveillance of Salmonella, being identified through the application of artificial intelligence and genetic information.

01001
00100
10100

Lightning talks

TRACK: Machine Learning in Computational, Systems Biology and Bioinformatics

10100
10011
01000

ID# 009

Transposable elements classification using recurrent neural network

Arlen James Mosquera Ruiz[1,3], Luis E. Tobon[4], Fabian Tobar-Tosse[2,3]

[1]Doctoral Student in Engineering and Applied Sciences, Pontificia Universidad Javeriana Cali, Colombia
[2]Departamento de Ciencias Básicas de la Salud, Pontificia Universidad Javeriana Cali, Colombia
[3]Red Iberoamericana de Inteligencia Artificial para Big BioData (RIABIO)
[4]Departamento de Ingeniería y Ciencias, Pontificia Universidad Javeriana Cali, Colombia

Abstract

Transposable elements (TEs) are DNA sequences with the active or inactive ability to insert into different regions of the genome. Initially, TEs were considered as junk DNA with little functional and regulatory significance. However, with the availability of scientific evidence, they have gradually been recognized as transcendental components in the evolution, organization and stability of the genome.

As a result of the different evolutionary patterns, multiple origins and continuous diversification, it is possible to evidence a great variety of forms, structures and transposition mechanisms in TEs. This has led to difficulties in establishing a classification system to organize TEs in a hierarchical manner and with defined guidelines to establish relationships among them. Although some classification methodologies have been developed, most of them follow a homology-based approach, requiring more computational resources and databases with availability of correctly classified and annotated sequences. Recently, some machine learningbased approaches have been used to address the problem of TE classification, and their sorting as functional elements.

In this study, recurrent neural networks (RNN) with a long, short-term memory (LSTM) architecture were used to classify TEs at different levels of taxonomic classes. The results obtained show significant improvements in terms of evaluation metrics and time effectiveness compared to previous methodologies in order-level classification. However, there are areas for improvement, such as addressing overfitting and further exploring the use of others RNN architectures.

These findings provide a basis for future research in TE classification and the development of more efficient algorithms, especially nowadays, where due to the improvement of sequencing technologies, there is a greater amount of biological data that could influence the discovery of new TEs that challenge the traditional classification systems.

ID# 103

A novel approach for RNA folding inference based on message-passing graph neural networks

M. Gerard[1] and L. Di Persia[1]

[1]Research Institute for Signals, Systems and Computational Intelligence, sinc(i), FICH/UNL-CONICET, (3000) SF, ARG

Background: Nowadays, prediction of secondary structure of RNA is an open challenge. In simple terms, it consists of identifying which nucleotides in the sequence are paired, without considering the backbone. Classical methods based on thermodynamics are typically used for this prediction. More recently, a wide range of deep learning methods have appeared to compete with them, achieving increasingly better results. However, those methods need the training of millions of learnable parameters, which require vast amounts of data, and make those models prone to overfitting.

Results: Here, we present a novel approach for RNA secondary structure prediction from its sequence. In contrast to other deep learning methods, our proposal uses a small fraction of the learnable parameters used for those neural models. To achieve this goal, our approach first transforms the folding problem into a classification one, where the aim is to classify connections between pairs of nucleotides (nts) as feasible or not. First, the problem is modeled as a graph, where each node describes a particular nucleotide and has an associated set of features, and possible connections are modeled with edges linking them. Then this graph is inverted, turning nodes into arcs and vice versa. On this new graph, where the nodes now represent connections between nucleotides, a message-passing neural model [1] is applied that learns new features that are then used to classify the active connections between nucleotides (nodes on this graph). Preliminary results for this model (~40000 parameters), employing k-fold cross-validation (k=5) with archive II dataset with sequences up to 200 nts, yielded F1 = 0.848 (std: 0.015). In comparison, the state-of-the-art model UFold [2] (8641377 parameters, as stated in [2]) produces F1 = 0.914 (std: 0.036) for the same experiment.

Conclusions: Our novel approach for predicting RNA secondary structures demonstrates a good performance with minimal parameter usage (~0.4% of those used by UFold), showcasing efficiency in RNA structure prediction. This preliminary result shows our proposal is able to make good predictions using a very small number of parameters. However, it is clear that further work and experiments are needed to improve the performance obtained with this model.

[1] Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E. (2020). Message Passing Neural Networks. In: Schütt, K., Chmiela, S., von Lilienfeld, O., Tkatchenko, A., Tsuda, K., Müller, KR. (eds) Machine Learning Meets Quantum Physics. Lecture Notes in Physics, vol 968. Springer, Cham. https://doi.org/10.1007/978-3-030-40245-7_10

[2] Laiyi Fu, Yingxin Cao, Jie Wu, Qinke Peng, Qing Nie, Xiaohui Xie, UFold: fast and accurate RNA secondary structure prediction with deep learning, Nucleic Acids Research, Volume 50, Issue 3, 22 February 2022, Page e14, https://doi.org/10.1093/nar/gkab1074

ID# 068

A music-inspired codification of ncRNAs with pseudoknots

Garcia Labari Ignacio[1*], Spetale Flavio E.[1,2], Tapia Elizabeth[1,2]

[*] corresponding author

[1]CIFASIS-UNR/CONICET. Rosario, Argentina

[2]Facultad de Ciencias Exactas, Ingeniería y Agrimensura, UNR. Rosario, Argentina

Background: Non-coding RNA molecules (ncRNAs) play vital regulatory and structural functions within cells. These functionalities are encoded into complex 2D structures, posing a formidable challenge for their precise modeling and interpretation. In previous work, we introduced a coding-based approach to integrate primary sequence and secondary structure information of ncRNA molecules into sequences of a higher-order alphabet, suitable for ubiquitous machine learning processing. However, it is important to note that representing ncRNA molecules in this way compromises their interpretability.

Results: We have improved our coding approach to model ncRNA molecules, with a focus on including pseudoknots. These are intricate secondary structural motifs that exhibit base pair interactions between distant regions of RNA sequences. To facilitate the interpretation of these newly encoded sequences, which now utilize an abstract new alphabet of 28 symbols, we have adopted a music-inspired mapping approach, transforming them into sequences represented by seven musical notes and four distinct tones (piano notes). In this new scenario, ncRNA molecules can be analyzed, and interpreted, with well-established digital signal processing tools. To validate our approach, we undertake decision tree classification on a dataset containing nearly 4000 members belonging to 10 families of ncRNA molecules. These molecules vary in length, spanning from 28 to 2968 nucleotides, and their linear and 2D structures, including pseudoknots, are experimentally known. Each of these sequences is transformed into a Fourier spectrum which serves as an embedding layer that generates fixed feature vectors suitable for interpretable machine learning purposes. Beyond the 93% classification accuracy rate, our results confirm our intuitive expectations, i.e., short and simple molecules lacking pseudoknots are associated with low and medium notes, while long and complex molecules with pseudoknots are associated with high notes.

Conclusions: Collapsing the 1D + 2D structure of ncRNAs into musical notes preserves both predictive power and interpretability features. Our next step is to extend this methodology to predict the functionality of long non-coding RNAs with pseudoknots using Gene Ontology (GO) annotations.

ID# 066

Improvements in Recurrent Neural Networks for the Prediction of Molecular Properties and Characteristics of Natural Products

Naicolette A. Agudo-Muñoz[1], Juan F. Avellaneda-Tamayo[2], José L. López-Pérez[3], José L. Medina-Franco[2], Javier E. Sánchez-Galán[4, *]

[*] corresponding author

[1]Maestria en Ingeniería Matemática, Facultad de Ciencias y Tecnología, Universidad Tecnológica de Panamá

[2]Grupo de Investigación DIFACQUIM, Departamento de Farmacia, Escuela de Química, Universidad Nacional Autónoma de México

[3]Facultad de Farmacia, Universidad de Salamanca

[4]Grupo de Investigación en Biotecnología, Bioinformática y Biología de Sistemas, Universidad Tecnológica de Panamá.

Abstract

One of the ways to unambiguously identify the structures of chemical compounds and, particularly natural products, is through the linear notation of canonical SMILES. Due to their textual and linear nature, they store spatial information and some descriptors in an uninterrupted chain of characters, becoming a very common structural representation in cheminformatics. However, this information also belongs to a set representable through graphs, to which there may be interconnected properties (such as toxicity, activity, solubility or solvation energy) that are not appropriately identified or predicted using only Recurrent Networks. Even worse if we talk about relating continuous properties to multicategorical properties.

This work seeks to explore various modifications to the architecture proposed by the SMILES2vec network, capable of improving its performance for values labeled with more than two categories. The NAPROC-13 chemical library to be studied and the aim is to predict the type of deuterated compound by which a natural product can (in cases of chemoprospecting) or could have been dissolved (in unreported cases), when the respective nuclear magnetic resonance (NMR) is performed. Strings of up to 160 characters are evaluated, representing 96.1% of the natural products listed in the database, cured and in their chiral representation. Then, the constitutional descriptors, properties of interest and molecular fingerprints are determined for the characterization of the chemical space using the RDKit library in Python and the Molecular Operating Environment software, in order to select those descriptors that are most decisive for the differentiation of solvents to be predicted and which are will be implemented for improvements, contemplating their own architectures for Networks with Attention Mechanisms (GAT) and Messengers (MPNN).

Among the techniques to evaluate its performance, the Mean Square is used for continuous values, and Categorical Cross Entropy and Categorical Precision for values in categories, so previously the data in SMILES format will be encoded as one hot vectors. The results, compared to Recurrent Networks, show improvements in prediction accuracy of up to ten percentage points with other steroid receptors, such as ERα and/or GR as putative mechanisms underlying LXR effects.

ID# 057

SincFold: a new tool for RNA folding based on deep learning

L.A. Bugnon[1], L. Di Persia[1], M. Gerard[1], J. Raad[1], S. Prochetto[1,2], E. Fenoy[1], U. Chorostecki[3], F. Ariel[2], G. Stegmayer[1], D.H. Milone[1]

[1]Research Institute for Signals, Systems and Computational Intelligence, sinc(i), FICH-UNL, CONICET, Ciudad Universitaria UNL, 3000 Santa Fe, Argentina

[2]Instituto de Agrobiotecnología del Litoral, CONICET-UNL, CCT-Santa Fe, Ruta Nacional N° 168 Km 0, s/n, Paraje el Pozo, 3000 Santa Fe, Argentina

[3]Faculty of Medicine and Health Sciences, Universitat Internacional de Catalunya, 08195 Sant Cugat del Vallès, Barcelona, Spain.

Background: Non-coding RNAs fold into well-defined secondary structures, which defines their functions. However, its computational prediction from a raw sequence is a long-standing unsolved problem, without significant changes in performance in the last decades. Traditional RNA secondary structure prediction algorithms are based on thermodynamic models and dynamic programming for free energy minimization. More recently, deep learning methods have appeared.

Results: We present sincFold, an end-to-end deep learning approach that predicts the contact matrix among nucleotides in a sequence using only the RNA sequence as input. The model is based on hierarchical 1D-2D residual neural networks that can learn short- and long-range interaction patterns. Extensive experiments on several benchmark datasets were conducted, comparing sincFold against classical methods and recent deep learning models. Results show that sincFold can outperform state-of-the-art methods on all datasets assessed. The source code is available at https://github.com/sinc-lab/sincFold and an online demo is provided at https://huggingface.co/spaces/lbugnon/sincFold

Conclusions: SincFold has proven to be better suited to identify structures that might defy traditional modeling. Even when there is a small number of examples of some RNA families to learn from (around hundreds of sequences), sincFold can learn the structures with high performance. As more examples are available, the generalization capability of the model is significantly better than other methods. Results also show that sincFold, thanks to its capability for capturing a

ID# 069

Bioinformatics Analysis of lncRNAs in Kinetoplastid Parasites

Paola Sosa Basso[1], José Sotelo-Silveira[2], Beatriz Garat[3], María Ana Duhagon[4] , Pablo Smircich[1]

[1]Laboratorio de Bioinformática, Departamento de Genómica - Instituto de Investigaciones Biológicas Clemente Estable (MEC). Sección Genómica Funcional - Facultad de Ciencias (UdelaR), Uruguay
[2]Departamento de Genómica - Instituto de Investigaciones Biológicas Clemente Estable (MEC) Sección Biología Celular - Facultad de Ciencias (UdelaR), Uruguay
[3]Sección Genómica Funcional - Facultad de Ciencias (UdelaR), Uruguay
[4]Sección Genómica Funcional - Facultad de Ciencias (UdelaR), Departamento de Genética- Facultad de Medicina (UdelaR), Uruguay

Abstract

Kinetoplastids are flagellated protists, including parasites causing Chagas disease, sleeping sickness, and leishmaniasis, by Trypanosoma cruzi, Trypanosoma brucei, and Leishmania spp., respectively. Long non-coding RNAs (lncRNAs), non-coding transcripts exceeding 200 nucleotides, play diverse roles in gene regulation, cell differentiation, and interactions with DNA, RNA, proteins, and chromatin across multiple organisms. However, lncRNAs in Trypanosoma cruzi and other kinetoplastids remain largely unexplored.

We present a bioinformatics pipeline designed to annotate lncRNAs in Trypanosoma cruzi, Trypanosoma brucei, and Leishmania major using transcriptomic data. The pipeline involves coding potential calculation, prediction employing Random Forest methods, and filtering by processing site identification. This approach led to the identification of several hundred putative lncRNAs in all examined trypanosomatids.

We conducted a conservation analysis of these lncRNAs through sequence alignment and functional classification based on k-mer content among the three organisms. Our analysis revealed the presence of conserved lncRNAs among the parasites. Additionally, we performed differential expression analysis of the identified lncRNAs across various parasite stages and biological conditions. This analysis shows that certain lncRNAs are differentially expressed under varying conditions, further underscoring their potential importance in parasite biology.

This work serves as a foundation for in-depth functional studies of lncRNAs in kinetoplastids and their potential roles in parasite biology, pathogenesis, and adaptability to different environmental and biological conditions.

01001
00100
10100

Lightning talks

TRACK: Genomics, Transcriptomics, and Metagenomic

10100
10011
01000

ID# 030

Comparison of bioinformatics tools for de novo genome assembly of Setophoma (Phaeosphaeriaceae), a genus isolated as an endophyte of Coffea arabica

Kesia Barrows[1,2], Librada Atencio[1], Javier E. Sanchez-Galan[3,4,5], Luis C. Mejía[1,6]

[1]Centro de Biodiversidad y Descubrimiento de Drogas, Instituto de Investigaciones Científicas y Servicios de Alta Tecnología de Panamá (INDICASAT-AIP)
[2]Doctorado en Biociencias y Biotecnología, Facultad de Ciencias y Tecnología, Universidad Tecnológica de Panamá
[3]Grupo de Investigación en Biotecnología, Bioinformática y Biología de Sistemas, Universidad Tecnológica de Panamá
[4]Facultad de Ingeniería de Sistemas Computacionales, Universidad Tecnológica de Panamá
[5]Sistema Nacional de Investigación (SNI), Senacyt, Panamá
[6]Smithsonian Tropical Research Institute, Panamá

Abstract

The genus Setophoma contains species frequently associated with plants of economic interest such as onion and tea tree. Interestingly, Setophoma species are usually associated to specific plants, except for Setophoma terrestris, which is a soil borne fungus that has been isolated from different crops. In a sampling of endophytic fungi of Coffea arabica in Panama, Setophoma sp. LCM1079 was isolated from coffee leaves.

To understand the symbiosis and possible functional roles of Setophoma and other endophytic fungi in C. arabica we initiated the development of genomic resources for species in clades without public information on genomes. As part of this work, different genome assembly methodologies for Setophoma sp. LCM1079 were compared with the purpose of defining a bioinformatics workflow that allows obtaining high-quality whole genome information. Specifically, a genomic library was sequenced on the Oxford Nanopore Technology MinION Mk1C system, and a library that was sequenced on the Illumina® NovaSeq 6000 platform (150bp with paired ends).

Genome assemblies were performed using long and short reads with de novo Canu assembler and three polishing steps using the tools Racon, Medaka, and Pilon. In addition, an assembly using only the short reads was performed with SPAdes assembler. The quality metrics of these assemblies were compared using the Quast and BUSCO tools, and it was determined that hybrid assembly using Canu is a good method to achieve the assembly of fungal genome

ID# 063

Identification of regulatory motifs in putative coding genes of small Heat Shock Proteins in Chenopodium quinoa

Costa Tártara, S.M.[1,2,*], Arce, D.P.[1,3], Cacchiarelli, P.[1,4], Tolosa, G, H.[2], Pratta, G.R.[1,4]
* corresponding author

[1]CONICET
[2]Departamento de Ciencias Básicas, Universidad Nacional de Luján, Av. Constitución y Ruta Nac. N° 5 (s/n), (6700), Luján, Buenos Aires
[3]Secretaría de Ciencia y Tecnología, Facultad Regional San Nicolás, Universidad Tecnológica Nacional, Colón 332, San Nicolás, Buenos Aires.
[4]Facultad de Ciencias Agrarias UNR, CC 14, 2125 Zavalla, Santa Fe

Background: Heat Shock Elements (HSE) are conserved motif sequences of DNA (GAA..TTC) present in gene regulatory regions of Heat Shock Proteins (HSPs). HSE interact with Heat Shock transcription Factors (Hsfs) to trigger HSP transcription depending on heat, abiotic or oxidative stress response. In plants, HSPs are chaperones that assist in protein folding, preventing their irreversible aggregation during stress. However, it is currently known that they are present in cells during other biological processes like germination and fruit maturation. Chenopodium quinoa (quinoa) is an allotetraploid (2n = 4x = 36) crop species belonging to the Chenopodiacea family, whose grains are consumed as cereal, although they are a good quality protein source. Quinoa represents a model to study HSPs because of an intrinsic tolerance to abiotic stress, like high salinity, water shortage and light intensity. At the same time, it is sensitive to high temperatures. We already reported 75 putative small (s) HSP coding gene sequences about genome (v2) and in this work we explored the upstream region, defined in 1000 pb, of a group of genes closely located in chromosome 4B (Cq4B). Each upstream sequence was retrieved and filtered with an ad-hoc Python script. HSE were scanned in each gene's putative promoter region using the RSAT platform.

Results: The high number of sHSP sequences reported is congruent with other plant species, which show 20 to 70 members, approximately. We also observed possible duplication events in some chromosomes, an evolutionary mechanism reported in the expansion of the HSP gene family. Among the tandem copy arranged selected genes CQ020271 (567 bp), CQ020272 (549 bp), CQ020274 (555 bp), CQ020275 (522 bp), CQ020276 (552 bp), CQ020278 (609 bp) located in Cq4B, HSEs were only detected in CQ020272 and CQ020276 promoters, indicating a possible Hsf-dependent regulation characterized in the HSP gene family.

Conclusions: These observations support previous findings on sHSP subfamily characterization in other plant species of agronomic interest including Arabidopsis sp., tomato, potato, cotton, wheat, barley or sorghum. Further approaches could be conducted to characterize gene expression profiles of nearly duplicated genes (or tandem copy-located genes).

ID# 098

Transcriptomic meta-analysis to elucidate the presence and expression of neuropeptides in insect embryos

Nicolas Lavatti[1], Agustín Baricalla[2], Daiana del Valle[2], Rolando Rivera-Pomar[1]

[1]UNNOBA
[2]CONICET

Abstract

Neuropeptides (NPs) are small peptides that act as neurotransmitters, neuromodulators and growth factors, and, in addition, modulate feeding, courtship, and circadian cycle. Despite its importance, little is known about the expression and function of NP precursor (NPP) genes during the embryonic development in animals. Among the very few examples, there is an early development regulator in Drosophila melanogaster larvae called Trunk, a peptide expressed in nurse cells and transferred to the mature oocyte, and its receptor tyrosine kinase Torso. They set the terminal system of the embryo. Later on, in the life of the insects, Torso also serves as a receptor for the prothoracotropic hormone (PTTH). PTTH and Torso were identified in Bombyx mori during its early embryonic development.

Here we study the expression of NPPs transcripts during embryonic stages in different insect orders (Coleoptera, Diptera, Hemiptera, Hymenoptera, Blattodea, Orthoptera) using public transcriptomic data across time series of embryogenesis. A quality check was carried out through FastQC, adapter removal with Trim Galore!, the alignments to the respective reference genomes were carried out with HISAT2, quality assessment with Qualimap and quantification with featureCounts. Differential gene expression analysis was performed using the edgeR library in RStudio. In parallel to this workflow, NPP present in each species were identified through Proteinortho, searching for sequence homology between a manually curated database containing sequences of interest and the proteome of each species. The results obtained were reviewed manually to avoid data loss and homogenize the NPP nomenclature across all studied species.

These results were then used to identify differentially expressed NPP genes. Among the NPP we were able to determine the expression of mostly those identified through this approach. In order to correlate our data to landmark events of embryogenesis and make comparison among different species possible we performed an exhaustive literature review. It was possible to detect the early expression and association of NPP (AST, AKH, CRZ, SIF) with specific stages of development. Interestingly, several NPPs (e.g., TRUNK, PTTH, IRP) are expressed in very early stages of development, before the emergence of the nervous system. To begin validating these results we performed qPCR on early stages of Oncopeltus fasciatus and show that NPA, TK, AT, NPLP, CCH and ETH are expressed in the early embryo as mRNA of maternal origin. Taken together, our results suggest that the expression of NPP widely occurs during embryonic development throughout the Insect a class, which is a novel observation that opens a new field of study in developmental biology.

ID# 034

Transcriptomic exploration of dormancy mechanisms in sweet cherry varieties with contrasting cold requirements, under equitable agricultural conditions

Giovana Acha[1], Claudio Urra[1], Gabriela Saavedra[1,2], José Gaete-Loyola[1], Danitza Silva-García[1], Andrea Miyasaka Almeida[1,3]

[1]Centro de Genómica y Bioinformática, Facultad de Ciencias, Ingeniería y Tecnología, Universidad Mayor, Huechuraba, Santiago, Chile

[2]Programa de Doctorado en Genómica Integrativa, Vicerrectoría de Investigación, Universidad Mayor, Huechuraba, Santiago, Chile

[3]Escuela de Agronomía, Facultad de Ciencias, Ingeniería y Tecnología, Universidad Mayor, Huechuraba, Santiago, Chile

Abstract

Sweet cherry fruit has being highlighted as one of the main exported fruit in Chile. As other perennial trees, it enters a state of dormancy in autumn, as a defense mechanism to avoid adverse cold weather conditions. Dormancy is divided into fundamental stages: endodormancy, marked by reduced metabolism and growth arrest, where the accumulation of cold is crucial to pass to ecodormancy, a stage in which growth is reactivated in response to warm weather conditions in spring.

The purpose of this study, was to understand how each dormancy mentioned stage works and what are the molecular components that generate the differencs between them. For this aim, samples of flower buds from sweet cherry varieties with low (Santina) and high (Regina) cold requirements were collected during endodormancy and ecodormancy stages. Using RNA sequencing technology (RNA-seq), the changes in the transcriptome were analyzed.

In late endodormancy, 4458 differentially expressed genes were identified in Regina, with respect to early endodormancy, (1809 overexpressed and 2649 underexpressed). while in Santina, 3834 differentially expressed genes were identified (1940 overexpressed and 1894 underexpressed). For both sweet cherry varieties differentially expressed genes were enriched for cell wall biogenesis, cell cycle, and carbohydrate metabolism. In ecodormancy, 6330 differentially expressed genes were identified in Regina, with respect to early endormancy (1126 overexpressed and 533 underexpressed). While in Santina, 6872 differentially expressed genes were detected (1235 overexpressed and 561 underexpressed). For both sweet cherry varieties differentially expressed genes were enriched for photosynthesis, cell wall biogenesis, and various enzymatic activities.

This study is allowing to elucidate the intricate molecular mechanisms behind sweet cherry dormancy, that will allow the development of methodologies and approaches for the control of flowering in this species.

Funding: ANID/ACT210007

ID# 074

Understanding tomato Host Response to PSTVd-Infection: Insights from Network Analysis and Transcriptional Reprogramming

Katia Aviña-Padilla[1], Octavio Zambada-Moreno[1], Marco A. Jimenez-Limas[2], Rosemarie W. Hammond[3], Maribel Hernández-Rosales[1]

[1]Centro de InvesAgación y de Estudios Avanzados del I.P.N Unidad Irapuato
[2]Centro de InvesAgación en Computación, Instituto Politécnico Nacional, Ciudad de México
[3]USDA, Agricultural Research Service, Beltsville Agricultural Research Center, Beltsville Maryland

Background: Viroids, as the smallest pathogens of angiosperms, manifest significant threats to crops by inducing severe diseases. Their impact on angiosperms, particularly tomato, has garnered increased aUenAon in recent research. These non-coding RNA pathogens lead to notable developmental alterations in infected plants, predominantly due to their effect on genetic regulation. The current study was undertaken to unveil the overarching mechanisms governing the host's response to viroid infection, with a keen focus on the differentially expressed transcriptional responses in tomato infected with both mild and severe PSTVd variants.

Results: Our comprehensive approach combined root and leaf transcriptomic data, network analysis, and functional characterization of gene modules, leading to the identification of master transcription regulators (MTRs). Notable findings include the roles of bHLH, MYB, and ERF transcription factors in the regulation of pivotal signaling pathways. Particularly, bHLH-MTRs were associated with metabolism and defense mechanisms, whereas MYB-MTRs played roles in signaling and hormone processes. A bHLH-TF family member emerged as a potential microprotein, offering insights into post-translational hormone signaling regulation. Furthermore, using co-expression analysis, we uncovered and characterized differentially expressed gene modules, revealing the vital roles of bHLH TFs in managing metabolic functions in both roots and leaves**.**

Conclusions: Our findings present a comprehensive understanding of the nuanced interactions between PSTVd and its tomato host, emphasizing the significant roles played by various transcriptional factors in determining the host response. The identification of these key genes and transcriptional pathways not only deepens our grasp on the molecular mechanisms underlying viroid-induced symptom development but also opens avenues for potential interventions and strategies for disease management in agronomic settings.

ID# 090

evolSOM: A Novel R Package for evolutionary conservation analysis using Self-Organizing Maps

Prochetto S.[1,2], Stegmayer G.[2], Reinheimer R.[1]

[1]Instituto de Agrobiotecnología del Litoral (UNL-CONICET), CCT-Santa Fe, Argentina
[2]Sinc(i), Research Institute for Signals, Systems and Computational Intelligence (UNL-CONICET), CCT-Santa Fe, Argentina

Background: Gene regulatory networks (GRNs) are crucial for understanding cellular processes and diseases. Discovering the conservation of GRNs across species and integrating phenotypic and genotypic information are essential steps toward comprehending network dynamics and identifying key gene drivers. Self-organizing maps (SOMs) are powerful tools for analyzing complex biological data due to their ability to capture high-dimensional relationships in lower dimensions. The use of SOMs may facilitate the discovery of candidate genes as potential enablers of specific phenotypic traits, or to detect displacements in GRNs between species.

Results: We present here evolSOM, a novel R package that leverages SOMs to investigate and visualize features conservation while facilitating the integration of phenotypic and genotypic attributes. Using SOMs, the package enables the projection of multi-dimensional expression profiles onto simple-to-read two-dimensional grids, aiding in the identification of conserved gene modules across multiple species or conditions. Building several species-specific or condition-specific SOMs, it is possible to analyze features displacement (changes in expression patterns such as early/delayed/flip) between species or conditions. These types of displacements are calculated automatically by the model and shown graphically (see Figure). This allows the user to efficiently compare features, revealing conserved and displaced features and species-specific patterns. The package facilitates the integration of diverse phenotypic data types, such as morphological data or metabolomics, enabling the exploration of potential gene drivers underlying observed phenotypic changes. The package's user-friendly interface and visualization capabilities enhance the accessibility of complex network analyses. We have used evolSOM to study the displacement of genes and phenotypic traits, identifying drivers of phenotypic differentiation in grass leaves.

Conclusions: Our package represents a valuable contribution to biologists by offering a powerful toolset for the exploration of GRNs evolution and the integration of diverse phenotypic data types. By employing SOMs, the package provides an intuitive means to study and visualize complex regulatory networks, aiding in the identification of conserved modules and potential gene drivers. The package's efficiency, ease of use, and production of informative visualizations make it a promising asset for researchers seeking to unravel the intricacies of GRNS and their conservation across species

01001
00100
10100

Lightning talks

TRACK: Biomedical Omics, Biological databases and Bioimaging

10100
10011
01000

ID# 016

Computational modeling and prediction of survival in breast cancer tumors based on myelo-lymphocyte ratios and hot-cold scores

Alonso-Moreda, N.[1,2], Sánchez-Santos, J.M.[1,2], De Las Rivas, J.[1]

[1]Cancer Research Center (CiC-IBMCC, CSIC/USAL/IBSAL), Consejo Superior de Investigaciones Científicas (CSIC), University of Salamanca (USAL) & Instituto de Investigación Biomédica de Salamanca (IBSAL), 37007 Salamanca, Spain
[2]Department of Statistics, University of Salamanca (USAL), 37008 Salamanca, Spain

Background: Breast cancer is the most common cancer found in women and the second most prevalent cancer worldwide, 70% of which show high infiltration of immune cells. The tumors are not only made up of malignant cells but they are involved in a complex dynamic of interactions with other cell types (such as immune and stromal cells) within the Tumor Microenvironment (TME). It seems that the infiltration level of these non-malignant cells influences tumor progression and therapy response. Recent studies defined two types of tumors based on immune cell infiltration: Hot, inflammatory tumors with many immune cells; Cold, non-inflammatory tumors that contain fewer immune cells and other cell types with immunosuppressive phenotype. Typically, Hot tumors have better prognosis than Cold. Understanding the infiltration of different cells in the TME is crucial to predict the response to antitumor treatments based on the cellular and immunological characteristics of each tumor.

Results: We present a series of computational analyses of the prognosis of breast cancer patients combining full transcriptomic data from the tumors, plus data about the degree of immune cells infiltration in the tumors and survival data of the patients. Analyses using Kaplan-Meier estimator shown better survival for individuals with tumors of high immune cell infiltration. We also observed better survival of patients with larger numbers of neutrophils than lymphocytes (high NLR, Neutrophil-Lymphocyte Ratio), and for individuals with low levels of monocytes (low MLR, Monocyte-Lymphocyte Ratio). We also studied tumor prognosis using Cox regression as a function of Hot/Cold tumors and gene expression. Significant results (p-value< 0.05) were obtained showing better survival for Hot tumors. Furthermore, we found 10 genes upregulated in Hot tumors that had a positive influence on survival: BCL8, NEUROG2, SLITRK3, TDRD12, UGT2B4, HRASLS2, MYT1L, DGS1, TUBB1, KRT13.

Conclusions: The results revealed that the presence of monocytes in breast tumors may be a poor prognostic factor, whereas elevated neutrophils may indicate good prognosis. In addition, a signature composed of 10 genes could be used to build accurate predictors associated with the prognosis of breast cancer, marking tumors with a high infiltration of lymphocytes and thus Hot tumors with good prognosis.

ID# 032

Deciphering pharmacological targets via computational analysis for drug repositioning in treating 5-FU-resistant colorectal cancer

Luciano Anselmino[1], Florencia Malizia[1], Aylén Avila[2], Nahuel Cessatti Laluce[1], Macarena Mamberto[1], Lucia Zanotti[1], Mauricio Menacho Márquez[1]

[1]CONICET
[2] CIPREB

Abstract

Colorectal cancer (CRC) is a globally prevalent and deadly disease, with its treatment primarily relying on cytotoxic drugs, notably 5-fluorouracil (5-FU). However, the development of resistance to these treatments and tumor recurrence is a common phenomenon. Drug repositioning seeks to uncover new drug uses, both repurposing existing ones and considering new compounds. Employing genetic signatures, computational drug repositioning predicts drug interactions within disease pathways. Here, we employ public gene expression data to identify potential drugs for treating recurrent CRC post 5-FU-based chemotherapy. By analyzing two independent studies, common differentially expressed genes (FDR<0.05 and |logFC|>1) between recurrence-present and -absent patients were selected using non-parametric algorithms. These genes underwent functional enrichment analysis and were matched with a drug repositioning database to pinpoint drugs capable of countering the recurrence-associated expression pattern.

Additionally, through batch effect reduction methods, we constructed a merged gene expression matrix for patients subjected to 5-FU-based chemotherapies. Applying feature selection algorithms, we identified key genes associated with recurrent phenotype and carried out functional enrichment studies.

Both approaches indicated potential involvement of Rho GTPase protein pathway in upregulated instances during tumor recurrence. Immune response-related pathways showed significant downregulation (adjust p-value<0.05). Analyzing promoter regions revealed enrichment of binding sites for Serum Response Factor family and zinc finger proteins within overexpressed genes.

Guided by these insights and drug repositioning database results, we shortlisted compounds for in vitro testing on 5-FU-resistant CRC cells models. Two compounds showed promise and were validated in murine models, revealing reduced tumor growth, resensitization to 5-FU, and fewer lung micrometastases. Among the tested compounds, the RAC1 inhibitor 1A-116 yielded the most significant reversal of resistance.

ID# 080

Comparing cistromes from endometrial adenocarcinoma cell line and endometrial tumor samples

Luciana Ant[1], Alejandro La Greca[2], Nicolás Bellora[3], Patricia Saragüeta[1]

[1]CONICET-Instituto de Biología y Medicina Experimental, Capital Federal, Argentina
[2]LIAN, Fleni Institute-CONICET, Buenos Aires, Argentina
[3]Institute of Nuclear Technologies for Health, INTECNUS-CONICET, Bariloche, Argentina

Background: Tamoxifen is an effective and widely applied therapy in breast cancer that acts through competitive inhibition of estrogen (E2), affecting estrogen receptor (ERα) interactions with other nuclear receptors, chromatin modulators, genomic structural proteins and coregulators. Even though tamoxifen inhibits ERα in breast cells, inhibiting the progression of breast cancer, it stimulates ERα in certain other tissues, including the endometrium, specifically in a low estrogen environment. This increases the risk of endometrial cancer particularly in postmenopausal women. Our laboratory had previously analyzed the progesterone receptor (PR) and ERα cistrome of Ishikawa cells when treated with ovarian hormones E2 and progestins.

Results: Here, we compared publicly available data of ERα and H3K27ac cistromes of endometrial tumors of both tamoxifen users and non-users with ERα and PR cistromes of Ishikawa adenocarcinoma cell line generated at our laboratory. We re-processed publicly available data with the same parameters for quality control, alignment and peak calling previously used in cistromas of Ishikawa cells. Then data were compared. This analysis included the intersection of peaks, correlation of peaks with genes and ontology analysis of those genes using deepTools, bedTools, GREAT and Metascape. ER binding sites were enriched in cells of tamoxifen users while PR binding sites were enriched in endometrial cancers of tamoxifen non-users. Our results indicate that tamoxifen treatment could change the chromatin landscape of endometrial tumor cells that can be occupied by PR in response to progestins.

Conclusions: We were able to process publicly available data through a simple pipeline in order to gain new insights by comparing them with in-house generated data using only a bioinformatic approach, without the need of any wet lab approach. Our analyses indicate that tamoxifen treatment could change the chromatin landscape of endometrial tumor cells that can be occupied by PR in response to progestins. These results propose a possible mechanism to explain a context dependent response and genomic regions critical to be taken into account to prevent a deregulation of endometrial cells under tamoxifen treatment.

ID# 040

GBManalizerApp: A CAD tool for the three-dimensional automatic segmentation of the glioblastoma multiforme

Alexander Mulet de los Reyes[1,2], María Elena Buemi[3], Julio Jacobo Berlles[3], Daniel Gandía[4], Karina Breitburd[5], Cecilia Suárez[1,2]

[1]CONICET - Universidad de Buenos Aires, Instituto de Física Interdisciplinaria y Aplicada (INFINA). Buenos Aires, Argentina
[2]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Física. Buenos Aires, Argentina
[3]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Computación. Buenos Aires, Argentina
[4]Centro de Especialidades Médicas. Provincia de Buenos Aires, Argentina
[5]Servicio de Neurocirugia, Hospital Alemán, Buenos Aires, Argentina

Abstract

The glioblastoma multiforme (GBM) is the most prevalent and aggressive primary brain tumor and of worst prognosis. Although it very rarely metastasizes, its high invasive capacity is considered one of the reasons of the almost inevitable recurrence, that renders the mean survival time to nearly 15 months. In this work we present the GBManalizer: a Windows App for the three-dimensional automatic segmentation of the GBM, able to discriminate the active tumor mass from its inner necrosis and the extended diffuse peripheral edema which is partially infiltrated by tumor cells.

This CAD (Computed Assisted Diagnosis) tool is able to: 1) load MRI (magnetic resonance image) images in their four basic modalities: T1, T1c (T1 with gadolinium contrast), T2 and FLAIR (FluidAttenuated Inversion Recovery); and in their two most-used medical formats (DICOM and NifTi). 2) anonymize the images. 3) modify the image visualization changing their size and/or contrast. 4) pre-process the images (co-register to the same anatomical template, interpolate to the same resolution and extract the skull) by the Matlab-associated SPM12 software. And 5) segment the images with an accuracy of 83.9%, by an algorithm that combines classic image-processing methods (multi-threshold Otsu's, Chan-Vese's active contours and morphological erosion) with a Perceptron neural network trained by preprocessed images from the BraTS (Brain Tumor Segmentation) challenge database.

The network consists of one hidden layer with 80 neurons, is fed by 30 selected radiomics features and classifies each image voxel into one of four classes: active tumor, necrosis, edema or normal tissue (network 30:80:4). The complete segmentation algorithm achieves Dice similarity coefficients of 89.3%, 80.7%, 79.7% and 66.4% for the whole ROI (region of interest), active tumor, edema and necrosis segmentation volumes, respectively. When our own pre-processing method is added prior to the segmentation, a maximum correlation of 81% is obtained for the whole ROI volume, when a pattern-matching process is applied between our segmentations and the ground truths offered by the database.

The GBManalizerAPP presented here offers the potentiality to become a new CAD tool of clinical utility in the therapeutic option selection for the glioblastoma multiforme local management.

ID# 004

A Decade of Viral Exploration through High-Throughput Sequencing Data: Insights from the Global South

Humberto Debat[1,2]

[1]Instituto Nacional de Tecnología Agropecuaria (IPAVE - CIAP - INTA), Córdoba Argentina
[2]Consejo Nacional de Investigaciones Científicas y Técnicas, Unidad de Fitopatología y Modelizacion Agrícola, Argentina

Abstract

Viruses, as the most ubiquitous biological entities on Earth, constitute a realm of paramount significance. Current estimates conservatively indicate that an overwhelming majority, surpassing 99.9% of the virosphere's constituents remains elusive. Leveraging the invaluable resource of publicly accessible databases, data mining emerges as an indispensable instrument for unlocking the viral dark matter. This approach stands as an efficient and sustainable strategy, facilitating secondary analyses of publicly available data to expedite virus discovery. Within this talk, I will delineate a series of strategic paradigms poised to illuminate the global viral landscape, underpinned by open sequencing data. I will summarize our experience in virus hunting from our very first Roche 454 and sRNA libraries to our most recent Nanopore runs, and the use of NCBI resources such as TSA and SRA to identify novel viruses. I will introduce compelling instances of robust detection and characterization of new members of virus species from our group, nestled within diverse niches such as gastropod neuronal networks, spider venom glands, firefly lantern organs, the parasitized hosts of a zombie fly fungus, cannabis plants, amphibian brains, malaria-vector mosquitoes indigenous to the Amazonas, and Argentinian mice. The manifold challenges inherent to virus discovery within the Latin American context and potential impact in virus emergence and pandemic prediction will be discussed.

ID# 071

Development of an automated system for determination of ultrastructural patterns in spores of pathogens and symbionts of agricultural relevance

Juan Miguel Zúñiga-Umaña[1], Jordán Ledezma-Hernández[1,2], Fabricio Quirós-Corella[2], Andrés Gatica-Arias[3], José Roberto Vega-Baudrit[1]

[1]National Laboratory of Nanotechnology (LANOTEC), National Center of High Technology (CeNAT)
[2]Colaboratorio Nacional de Computación Avanzada (CNCA), National Center of High Technology (CeNAT)
[3]Laboratory of Plant Biotechnology, School of Biology, University of Costa Rica

Abstract

Microscopy techniques are broad and their approaches diverse, despite they have the common characteristic of providing massive information, the extraction of which is often tedious and exhaustive. In this sense, biological images have a complex nature: morphological patterns can be a consequence of ecological-evolutionary events or experimental manipulations, the solution of hypotheses in this direction requires the collection of sufficient evidence to test the assumptions. When it comes to microorganisms, their complexity goes beyond their life cycle, which is subject to selective pressures at different stages. The interest in these life forms lies in their pathogenic role in both plants and animals, with direct repercussions on the quality of life of human beings. Given the rapid mutation rates in these organisms, the time required in determining variations can be crucial. Considering that one of the biggest limitations is the existence of computer programs that speed up data capture at higher orders, the need arises to establish a mechanism that speeds up this process, in addition to giving objectivity to data capture, given the establishment of measurements. punctual samples adjusted to mathematical criteria which reduces the introduction of error, common in manual sampling. Machine learning techniques prove their validity for the analysis and segmentation of biological electron microscopy images; from a deep learning approach, the search for patterns and classification of geometries is projected towards greater efficiency. The mass of potentially available data can then be involved in mathematical, ecological/epidemiological models, which has a direct impact on decision-making in the face of problematic events related to the biology of these pathogens.

Considering that Costa Rica is a country whose culture and economy partially revolves around agriculture, it is worth noting that coffee cultivation is affected by orange rust (Hemileia vastatrix Berkeley & Broome), a pathogenic organism of great interest due to its virulence, which serves as a precursor of this proposal. Thus, the objective of the proposed project is to develop a system for automatic extraction of microstructural features in biological electron microscopy images, including a procedure that is responsible for determining ultrastructural variations in pathogens of agricultural importance through automatic analysis that will provide an effective tool for the identification and segmentation of relevant patterns, in order to robustly and quantitatively characterize their morphology, investigate the search for response patterns to the environment and establish management strategies based on the conjugation of information.

01001
00100
10100

Poster presentations

10100
10011
01000

ID# 003

Embracing Open Science in Bioinformatics: A perspective from the Global South

Humberto Debat[1,2]

[1]Instituto Nacional de Tecnología Agropecuaria (IPAVE - CIAP - INTA), Córdoba Argentina
[2]Consejo Nacional de Investigaciones Científicas y Técnicas, Unidad de Fitopatología y Modelizacion Agrícola, Argentina

Abstract

In the era of rapidly advancing bioinformatics, the principles of Open Science have emerged as crucial drivers for accelerating research, fostering collaboration, and promoting knowledge dissemination. While these concepts have gained significant momentum in the global scientific community, their impact and implementation have been unevenly distributed, often leaving researchers from the Global South at a disadvantage.

This talk aims to shed light on the unique perspective of Open Science and Open Access from the Global South. By exploring the challenges faced by researchers in these regions, we will uncover the barriers hindering the full potential of open practices and access to scientific resources. We will delve into the factors contributing to the digital divide, such as limited infrastructure, funding constraints, and discrepancies in technological resources.

The presentation will highlight case studies of researchers and institutions from the Global South that have championed Open Science and Open Access principles, overcoming obstacles to drive meaningful change within their communities. Moreover, it will emphasize the positive impact of open practices in fostering international collaboration, knowledge exchange, and capacity building.

The presentation will revolve around the potential solutions that can be implemented to bridge the gap between regions. Collaborative efforts, capacity-building programs, and initiatives promoting equitable access to data and research publications will be explored as potential pathways for empowering researchers and institutions in the Global South.

ID# 006

Genome Sequencing and Characterization of Bacillus velezensis strain NKG50

Fernández, FD[1]; Sardo MF[1]; Monteoliva, M[2]; Valetti, L[1]

[1]IPAVE-CIAP-INTA
[2]IFRGV-CIAP-INTA. Córdoba Capital (5020)

Abstract

Bacillus velezensis is a Gram-positive bacterium that colonizes the plant rhizosphere and has been widely studied for its growth-promoting and biological control effects on fungi, bacteria, and nematodes. This study aims to sequence and characterize the genome of the Bacillus velezensis strainNKG50, which was isolated from chickpea plants and selected for its antifungal capacity. Genomic DNA was isolated from pure cultures using a commercial kit. Sequencing libraries were constructed using the Rapid Barcoding Kit (SQK-RBK004) and sequenced on the Oxford Nanopore (ONT) MinION Mk1b platform.

The raw signals obtained from ONT were processed using Guppy v 6.4.2 (HAC model), and filtered reads were used for de novo assembly using Flye and Medaka. The final assembly resulted in a single circular contig with a size of 4,123,916 nt (GC 46%). A total of 4,048 CDSs were annotated, including 462 hypothetical proteins, 9 complete rRNA operons, and 85 tRNA. The BUSCO evaluation revealed 94.4% completeness of the assembly. The phylogenomic analysis involved comparing the NKG50 genome with 18 selected genomes from other Bacillus species.

The study of orthologs identified 58 single-copy genes shared among all species studied. The phylogenetic tree constructed from the DNA sequence of these 58 SCGs (40,806 bp) showed that the NKG50 genome clustered within the B. velezensis species with high support. Additionally, to confirm the taxonomic identification and assess the genomic similarity with other strains, a digital DNA-DNA hybridization (dDDH) comparison and average nucleotide identity (ANI) evaluation were performed.

The results demonstrated that the B. velezensis NKG50 genome showed values higher than 96% in ANI and above 70% in dDDH, confirming its identification as B. velezensis. This study provides a valuable genomic resource that enhances our understanding of the biology of the B. velezensis NKG50 bacterium and its relevance as a biocontrol agent.

ID# 011

Transposable element annotation and polymorphism identification in wild strawberry (Fragaria Vesca)

Rocio Tolley[1], David Posé-Padilla[3], Pablo Manavella[2], Ileana Tossolini[1]

[1]Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Oro Verde, Argentina
[2]Instituto de Agrobiotecnología del Litoral (CONICET-UNL), Universidad Nacional del Litoral, Santa Fe, Argentina
[3]Instituto de Hortofruticultura Subtropical y Mediterránea: Málaga, España

Background: Strawberries constitute a crop of economic and nutritional value worldwide and display a wide geographic distribution, suggesting a high degree of adaptability. Transposable elements (TEs) are repetitive DNA sequences capable of translocating and multiplying within a host genome. They are a source of genetic variability, impacting epigenetic processes and adaptive mechanisms in plants. Studying TEs in strawberry could help identify regions with biotechnological potential which, upon editing, may result in favorable phenotypic traits, such as quicker growth rate or improved organoleptic properties. However, TEs have not been annotated in the reference genome of the wild strawberry Fragaria vesca, nor has their variability across populations been analyzed.

Results: TEs annotation was carried out on F. vesca reference genome through a combination of both structural and homology-based methods. This resulted in a total of 146,005 transposable elements, which cover approximately 30% of the genome. Among these, around 37% belong to Class I, 52% to Class II (DNA transposons), and 11% to Class II (miniature inverted-repeat transposable elements, MITEs). By analyzing 210 wild F. vesca accessions for presence/absence polymorphisms, we identified 187 variable annotated TEs (minor allele frequency ≥ 5), which accounted for a total of 20 presence and 278 absence variants. Notably, 15 of the presence variants were located within 3,000 base pairs (bp) of 24 distinct annotated genes. Conversely, among the absence variants, 241 were located within 3,000 bp of 481 genes. This indicates a greater frequency of TE deletions in gene-rich areas compared to insertions

Conclusions: TEs annotation and polymorphism detection was carried out for the diploid strawberry Fragaria vesca, which serves as model plant for cultivated strawberry as well as the Rosaceae family. These results offer opportunities for studying the role and impact of TEs in the strawberry genome. Future analyses will enable the determination of the effects of these variable TEs on F. vesca phenotypic traits, potentially leading to the identification of non-coding sequences that may be used as biotechnological targets by CRISPR-Cas9 seeking to generate plants with altered fruit characteristics.

ID# 012

Functional Redundancy and Diversification and Concerted Evolution Shaped the Complex BAHD Acyltransferase Superfamily in Land Plants

Fernando Villarreal[1], Agustin Amalfitano[2], Hugo Marcelo Atencio[3], Nicolas Stocchi[1], Arjen ten Have[1]

[1]Instituto de Investigaciones Biológicas, IIB-UNMdP-CONICET, Mar del Plata, Argentina
[2]Laboratorio de Procesamiento de Imagenes, ICyTE-CONICET-UNMdP, Mar del Plata, Argentina
[3]Banco Activo de Germoplasma de Papa, EEA Balcarce, INTA (BAG BAL), Argentina

Background: BAHD acyltransferases are key enzymes in the specialized metabolism from land plants. These enzymes can transfer an acyl group from an coenzyme A (CoA) activated molecule (donor) to an acceptor molecule. Due the large diversity of both donors and acceptors these enzymes can process, BAHD superfamily constitute an interesting model to study evolutionary mechanisms leading to functional diversification in specialized metabolism. As such, we study the BAHDome in over 200 land plants complete proteomes using a computational approach.

Results: A high-quality initial training group of 1805 BAHD homologues from 27 land plant proteomes and SwissProt entries was divided into 16 clusters with 100% precision and recall (100% P&R). However, these clusters don't always represent functional families (FunFams) due to variations in biochemical properties, taxonomy, and sequence identity (which can be as low as 20% intragroup). Expanding the analysis to include 191 additional land plant proteomes allowed for re-clustering of some 100%-P&R groups, confirming their status as FunFams. One example is G1, the largest and most diverse group, dominated by dicotyledons sequences, suggesting functional diversification through concerted evolution. G10 is another example, with a core role in suberin production across land plants. However other groups, like G4, must be sub-clustered to detect potential FunFams (such as subgroup G4-1, associated with chlorogenic acid and lignin biosynthesis). The search for equivalog (functionally equivalent homologues) supports the identification of potential FunFams. Finally, functional redundancy and diversification also contributes to BAHD evolution. Identifying specificity determining positions (SDPs) in G4-1 compared to major super-clusters (SC) reveals more about functional diversification. While SDPs related to the binding pocket are rare when compared to sequences using similar acceptors and donors (SCI), they are more common when compared to SCII, which handles more dissimilar substrates compared to G4-1.

Conclusions: Our study allows to shed light into the complex evolutionary pathways that lead to functional diversification in land plants BAHD acyltransferases. This is driven not only by reshaping the binding pocket, but also by other strategies that can, for example, alter the protein dynamics affecting substrate specificity. Additional studies are required to further study this hypothesis.

ID# 013

Genome sequencing and analysis of black flounder (Paralichthys orbignyanus) reveals new insights into Pleuronectiformes genomic size and structure

Fernando Villarreal[1], Germán F. Burguener[2], Ezequiel J. Sosa[2], Nicolas Stocchi[1], Gustavo M. Somoza[3,4], Adrián Turjanski[2], Andrés Blanco[5], Jordi Viñas[6], Alejandro S. Mechaly[7,8,9]

[1]Instituto de Investigaciones Biológicas (IIB-CONICET-UNMdP), UN de Mar del Plata, Mar del Plata, Argentina
[2]Plataforma de Bioinformática Argentina, Instituto de Cálculo, UBA, Buenos Aires, Argentina
[3]Instituto Tecnológico de Chascomús (CONICET-UNSAM), Chascomús, Buenos Aires, Argentina
[4]Escuela de Bio y Nanotecnologías (UNSAM). Argentina
[5]Universidade de Santiago de Compostela, Facultade de Veterinaria, Santiago de Compostela, Lugo, Spain
[6]Laboratori d'Ictiologia Genètica, Departament de Biologia, Universitat de Girona, Girona, Spain
[7]Instituto Nacional de Investigaciones y Desarrollo Pesquero (INIDEP), Mar del Plata, Argentina
[8]Instituto de Investigaciones en Biodiversidad y Biotecnología (INBIOTEC-CONICET), Mar del Plata, Argentina
[9] Fundación para Investigaciones Biológicas Aplicadas (FIBA), Mar del Plata, Argentina

Background: Black flounder (Paralichthys orbignyanus, Pleuronectiformes) is an economically important marine fish with aquaculture potential in Argentina. In this study, we sequenced the whole genome of this species using an Illumina sequencing technology, revealing a small genome size of ~538 Mbp compared to most teleosts. Black flounder and Pleuronectiformes appear to have smaller genomes than most other teleost groups. Thus, the main objective of the present study was to determine whether the small genome size of Black flounder and other Pleuronectiformes can be explained by the amount of transposable elements (TEs) and other repetitive elements, as well as the size of introns and exons.

Results: The Black flounder's genome (libraries from one male and one female were used) was sequenced with > 35 X-fold coverage. Analysis of the assemblies at the core gene level revealed that more than 98% of the core genes were present, with more than 78% of them having more than 50% coverage. This indicates a reasonably complete and accurate genome at the coding sequence level. This genome contains 25,231 protein-coding genes, 445 tRNAs, 3 rRNAs, and more than 1,500 non-coding RNAs of other types. Next, we performed a comparative genome analysis between Pleuronectiformes and other teleost orders to determine whether the small genome size can be explained by gene features, including whole genome genes and introns sizes, and as well as differential contents of repetitive elements. We show that the genome size shrinkage in flounder is indeed due to several factors, including changes in the number of repetitive elements (particularly TEs) and smaller gene size, mainly due to a reduced number of very large and small introns. Thus, these components appear to be involved in the reduction in Black flounder genome size.

Conclusions: Black flounder genome's resolution is an important tool for future aquaculture strategies for this species. The reduction of its genome size compared to other teleost is achieved by a mechanisms that, combined, are novel compared to those from another teleost, in general, and Pleuronectiformes, in particular.

ID# 015

Characterization of the HSP20 subfamily in Cannabis sativa

Arredondo S.[1*], Lozano A.[1*], Calua B.[1], Gullino F.[1], Chávez U.[1], Mantovani L.[1], Furque I.[1] and Arce D.P.[2]
* equal contribution

[1]Instituto Privado "Fray Luis Beltrán" Av. Moreno 401, San Nicolás, Buenos Aires, Argentina
[2]Secretaría de Ciencia y Tecnología, Facultad Regional San Nicolás, Universidad Tecnológica Nacional (FRSN-UTN) Colón 332, San Nicolás, Buenos Aires, Argentina

Background: HSP20s are 20 kDa heat shock proteins involved in preventing protein aggregation during heat and drought stress responses in numerous plant species. They are conserved across all living organisms, yet their study in Cannabis sativa var. hemp (Cs) has been limited. Cs is one of the earliest domesticated crops, finding applications in various industries such as fiber, food, and oil production. It holds particular significance in medical contexts, especially for pain treatment and addressing refractory epilepsy stemming from neurological disorders. While genotyping and phenotyping of Argentine varieties have only recently commenced, the characterization of CsHSP20 in particular remains ambiguous.

Results: This study involves the analysis and characterization of the CsHSP20 subfamily through diverse bioinformatics tools. These tools encompass sequence analysis, chromosomal mapping, gene structure examination, and the utilization of publicly accessible RNAseq data from the Cs genome. Forty members of the CsHSP20 subfamily were identified and annotated. Noteworthy is the discovery of eight CsHSP20s situated on chromosome 06, forming a cluster of tandemly arranged genes. Moreover, the scrutiny of public RNA-seq data unveiled six CsHSP20s displaying upregulation under conditions of drought-induced stress.

Conclusions: This investigation represents the initial instance of characterizing the CsHSP20 subfamily in Cannabis sativa, shedding light on its involvement in stress responses. The findings underscore the transcriptional activity of these members and the presence of tandemly arranged genes within the Cs genome. These results align with previously reported data in numerous other plant species such as Arabidopsis, tobacco, tomato, potato, and grape.

ID# 017

Advancing Computational Modeling of Protein-Carbohydrate Interactions using Water Sites Biased Docking

Jorge 0. Lannot[1,2,#], Esteban Rey[1,2,#], Marcelo Gamarra[1,2], Marcelo A. Martf[1,2], Carlos P. Modenutti[1,2,*]

[#] equal contribution

[*] corresponding author

[1]Departamento de Quimica Biol6gica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pab. II (CE1428EHA), Buenos Aires, Argentina

[2]Institute de Quimica Biol6gica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN) - CONICET. Ciudad Universitaria, Pab. II (CE1428EHA), Buenos Aires, Argentina

Background: Molecular docking is a widely used computational method for modeling ligand-receptor interactions, primarily focusing on small rigid drug-like molecules. However, its application to flexible carbohydrate molecules often leads to suboptimal performance due to their hydrophilic nature, low affinity, and ligand conformational flexibility. To address this, our study aims to develop a precise method capable of predicting both oligosaccharide conformations and relative binding energies to their receptors. This is achieved by leveraging solvent structure information at the ligand binding site to modify the scoring function of AutoDock suite programs. We assembled a diverse dataset comprising various oligosaccharides and sugar-binding proteins, encompassing differing binding site sizes, shapes, and polarities. This dataset incorporates an array of lectins, antibodies, enzymes, and carbohydrate-binding motifs. The systems were sourced from the Protein Data Bank (PDB), both in their "Halo" (ligand-bound) and "Apo" (ligand-free, for solvent structure information) forms. The performance of the programs was evaluated using Root Mean Square Deviation (RMSD).

Results: The Water Sites Biased Docking Method (WSBDM) exhibited significant enhancement in precision and accuracy across the analyzed programs for the studied systems. By incorporating solvent structure data from the receptor's ligand binding site, our method yielded predictions closer to experimental observations. Most notably, substantial improvements were evident in systems involving large ligands, where other programs tend to underperform.

Conclusions: WSBDM not only enhances the comprehension of protein-carbohydrate interactions but also improves the reliability, quality, and reproducibility of computational predictions. Its utilization promises a more nuanced understanding of molecular-level protein-ligand interactions, significantly impacting the field of glycobiology. Furthermore, it holds potential for improving carbohydrate docking predictions in homology-based models when crystal structures are unavailable.

ID# 018

Clade-wide proteome analysis shows widespread non-canonical DCR proteins in Fungi

Lorena I. Melet[1,2,3], Nathan R. Johnson[1,2], Elena A. Vidal[1,2]

[1]Centro de Genómica y Bioinformática, Facultad de Ciencias, Ingeniería y Tecnología, Universidad Mayor, Chile
[2]ANID-Millennium Science Initiative Program-Millennium Institute for Integrative Biology (iBio), Chile
[3]Programa de Doctorado en Genómica Integrativa, Vicerrectoría de Investigación, Universidad Mayor, Chile

Abstract

Dicer (DCR) is a class-III ribonuclease with a fundamental role in the synthesis of small RNAs and the regulation of gene expression. DCRs can be found in many eukaryotic organisms, including fungi, where DCRs are shown to be essential for defense against virus, transposon regulation and development. To date, fungal DCRs have mainly been identified in the most studied phyla: ascomycetes, basidiomycetes, and chytridiomycetes. However, there are still many fungal species where these proteins have not yet been described. Moreover, there is not a clear understanding of the structural domains present on fungal DCRs. In recent years, there has been an increase in the availability of fungal genomes and proteomes, with more than 1,500 annotated and reference genomes that have annotated proteomes.

In this study, we characterize and identify fungal proteins in silico on a large scale in more than 1,423 species, that span 9 phyla of the fungal kingdom. We found that a small fraction of proteins contains all the canonical domains described for plants and animals. Remarkably, the majority of fungi do not possess a PAZ domain, whose distance from the RNAseIII catalytic core governs the length of the sRNAs produced by the DCR protein.

Our analysis shows that PAZ domains are only present in DCR proteins of the Mucoromycota and Microsporidia phyla. Interestingly, we found PAZ in fungal species that are known symbionts of plants. Considering the emerging relevance of sRNAs for establishing interspecies interactions between fungi and plant hosts, our results suggests that the presence of PAZ in these species might be related to the generation of sRNAs of similar length to those produced by plants, that probably serve a role as communicating molecules during the interaction.

Funding information: ANID-Millennium Science Initiative Program ICN17_022, ANID-FONDECYT 11220727 And Beca Doctoral Universidad Mayor

ID# 022

Hemoglobinopathies vector cluster based predictor algorithm

Franco Salvatore[1], Franco G. Brunello[1,2], Marcelo A. Martí[1,2]

[1]Departamento de Química Biológica - Facultad de Ciencias Exactas y Naturales - UBA, Ciudad Universitaria Pabellón 2
[2]Instituto de Química Biológica IQUIBICEN - CONICET - Ciudad Universitaria Pabellón 2

Background: Hemoglobinopathies are hemoglobin type diseases which occur frequently in the human being and most of the pathogenic variants have lethal potential on life. Understanding better the nature of these variants may help and contribute to the study of mammals globins and potential treatments for the diseases if possible. It's already known that globins play a fundamental part in the proper function of most living life forms yet their study is truly complex. Thus, this research also aims to help the evolutionary study of mammals globins. The main focus of this work is to develop a high accuracy clinical significance predictor for any hypothetical hemoglobin variant based on the biological study of the protein and training the algorithm with that information and then test it against popular and well-known random forest algorithms to see if we are capable to build a more accurate model when training it with specific information.

Results: Once able to develop the model and trained it with a cross-validation data set we got that 75% of the cases in general were accurate when using a k-means type algorithm. Nevertheless, when focusing on the two different possible outputs of the predictor (wether pathogenic or benign prediction), the pathogenic accuracy was next to 80% whilst the benign accuracy was slightly higher than 50%. This tells us that the benign variants are predicted somewhat randomly and when seeing different plots made during the research it's visible that benign registered variants are extremely difficult to cluster away from pathogenic ones, but a significant group of pathogenic variants are feasible to cluster with almost no noise from benign variants. This explains why benign variants are so hard to predict and biologically we understand it to be because of how crucial and delicate the stability of the protein truly is, indicating that it's not likely to accept most kinds of mutations. Nowadays most of the well-respected pathogenicity predictors round between 70% to 80% of accuracy ratio so we think of our model to be more than acceptable when predicting pathogenic variants but we are still challenged when predicting benign ones.

Conclusions: Benign variants are a challenging group to study in the globin world and we aim to further understand it in the near future and refine the model as much as possible. Pathogenic variants are mostly covered but there is still room for improvement as we also aim to maximize the accuracy of the predictor. Aside from the clinical significance prediction it's also a must do in our future goals to achieve a phenotype predictor as well. This means that once predicted as pathogenic, we also aim to predict its phenotype which can be an unstable protein, a higher or lower oxygen affinity, methemoglobinemia between other phenotype classifications.

ID# 029

Tissue-specific Gene Regulatory Network models for Solanum lycopersicum

José D. Fernández[1,2,3], Jonathan Canan[1,2], Nathan R. Johnson[1,2], Javier Canales[2,4], José M. Álvarez[2,5], Elena A. Vidal[1,2]

[1]Centro de Genómica y Bioinformática, Facultad de Ciencias, Ingeniería y Tecnología, Universidad Mayor, Santiago, Chile
[2]ANID-Millennium Science Initiative Program-Millennium Institute for Integrative Biology (iBio)
[3]Doctorado en Genómica Integrativa, Vicerrectoría de Investigación, Universidad Mayor, Santiago, Chile
[4]Instituto de Bioquímica y Microbiología, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile
[5]Centro de Biotecnología Vegetal, Facultad de Ciencias, Universidad Andrés Bello, Santiago, Chile

Background: Tomato (Solanum lycopersicum) is one of the most relevant vegetable crops cultivated worldwide; nonetheless, while being considered a model plant for physiological research, molecular-level studies in tomato are still scarce. Coexpression networks have been utilized to capture the tomato gene interactions, but they lack the essential regulatory relationships required to decode the regulatory cascades affected by gene responses to environmental cues. Thousands of transcriptome datasets are available in online repositories, providing an opportunity to use machine learning algorithms in a systems biology uptake to generate Gene Regulatory Networks (GRNs) and discover the transcription factors (TFs) responsible for transcriptional cascades affected by plentiful stimuli.

Results: We collected publicly available RNA-Seq data from tomato roots, leaves, flowers, and fruits in any experimental scenario for more than 5600 libraries to develop gene regulatory models for Tomato at the organ level. For each organ dataset, transcriptomic data was used to infer transcription factortarget regulatory interactions using the GENIE3 random-forest-based algorithm. By comparing predicted gene regulatory networks to available ChIP-Seq data for tomato TFs, we reveal substantial enrichment and accuracy in our networks, supporting their ability to capture confirmed regulatory interactions. We found important differences between the targets of each TF in different organs and generated a list of the most relevant TFs for each tissue regulation by network attributes. As a proof of concept, we studied the response to sulfur (S) deficiency, a common crop stressor, using the GRNs obtained from leaves and roots. We generated S-context-specific GRNs by using the differentially expressed genes by S deprivation from RNA-Seq data. The networks' analysis allowed us to propose new key TFs that could control the sulfate deficiency regulatory cascades.

Conclusions: Our work provides the first GRN models for tomato, that can be used to determine potential key regulators of plant responses to environmental signals or under any experimental conditions. We provide substantial background information on tomato regulatory cascades that might be utilized to determine the impact of several TFs under the influence of multiple stress sources.

Funding information: This work was supported by ANID-FONDECYT 1211130, ANID-Millennium Science Initiative Program ICN17_022 and Beca doctoral ANID, Anillo ACT210007

ID# 033

Comparative structural analysis of the oligomeric state of bacterial glutamate dehydrogenases

Lic. Nicolás Chamorro[1], Dra. María Natalia Lisa[1,2]

[1]Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET)
[2]Plataforma Argentina de Biología Estructural y Metabolómica (PLABEM)

Background: Glutamate dehydrogenases (GDHs) are ubiquitous oligomeric enzymes that are classified into the subfamilies of small GDHs (S-GDHs), which are hexamers composed of 50 kDa monomers, and large tetrameric GDHs (L-GDHs) with 115 or 180 kDa subunits. S-GDHs have been the subject of diverse biochemical and structural studies whereas L-GDHs have been less studied.

The first experimental structure of a bacterial L-GDH_{180} (from M. smegmatis) was recently obtained by our group. The mycobacterial L-GDH_{180} (mL-GDH_{180}) consists of monomers that contrast with those of S-GDHs by containing long N- and C-terminal extensions flanking the catalytic domain. Such regions are modular and provide the surfaces for oligomerization. Here we report three experimental models of mL-GDH_{180} in different conformations that reveal transitions in the quaternary structure of the enzyme.

Results: We have made a detailed comparison of the catalytic domain of mL-GDH_{180} by aligning its sequence and 3D structure with those of diverse S-GDHs. This analysis revealed that secondary structure motifs involved in the oligomerization of S-GDHs are absent in mL-GDH_{180}, while recently identified contact motifs in mL-GDH_{180} are lacking in S-GDHs. Similarly, a comparison of the different conformations of mL-GDH_{180} allowed to recognize contact areas previously not evidenced.

Finally, we used the bioinformatic tool PeSTo to find putative sites for protein-protein interactions. The results obtained suggest that the GarA regulatory protein would bind to the catalytic domain of mL-GDH_{180}.

Conclusions: Based on the results obtained, we conclude that, contrary to S-GDHs, the catalytic domain of mL-GDH_{180} would not be essential for the stabilization of the quaternary structure, but it would fulfill functions in signal transduction by allosteric mechanisms, beyond the N- and C-terminal extensions. These findings contribute to understanding the molecular mechanisms that modulate glutamate metabolism in bacterial species relevant to health and biotechnology.

ID# 036

Ligand-Protein Interactions of Cymbopogon citratus Compounds and Their Implications for Chagas Disease Treatment

J. Leonardo Gómez Chávez[1], Germán A. Conti[1], Matías O. Miranda[1], Emilio L. Angelina[1], Nelida M. Peruchena[1]

[1]Lab. Estructura Molecular y Propiedades, IQUIBA-NEA, Universidad Nacional del Nordeste, CONICET, FACENA, Av. Libertad 5470, Corrientes 3400, Argentina

Background: The chronic stage of Chagas disease is characterized by severe cardiomyopathy caused by infection with the parasite Trypanosoma cruzi. Through molecular dynamics simulations, compounds derived from *Cymbopogon citratus* have demonstrated promising potential as ligands for proteins involved in this stage of the disease, displaying favorable binding energies. These interactions between the ligand-protein complexes may explain the observed effects of relieving this pathology by reducing amastigote nets and inflammatory infiltrates in the cardiac tissue of mice.

In this study, we analyzed the key interactions between compounds derived from *Cymbopogon citratus* and the most significant proteins associated with Chagas disease in mice.

Results: Ptgs2, Hck, and Csf1r complexes have demonstrated excellent binding free energies (ΔGbind) compared to specific inhibitors targeting these proteins. An analysis based on Quantum Theory of Atoms in Molecules (QTAIM) revealed that, in the case of Ptgs2, it exhibits a high affinity for binding to molecules with both a polar and non-polar (unsaturated) moiety, such as certain terpenes. This is attributed to the characteristic triad in its active site, consisting of arginine, tyrosine, and aspartic acid, which can attract the polar part of ligands. Furthermore, due to the presence of numerous non-polar residues in the active site, a significant number of non-polar interactions are formed, stabilizing the interaction with the formed complexes.

Similarly, Hck and Csf1r also show a strong tendency to bind to terpenes with structural unsaturations, leading to the formation of numerous non-polar interactions within the complexes. Although these non-polar interactions are weaker compared to polar interactions, they still contribute to stabilizing and forming a high affinity with these complexes.

Conclusions: This study highlights the importance of interactions between compounds derived from *Cymbopogon citratus* and key proteins involved in Chagas disease in mice. Furthermore, the fact that multiple compounds bind to different target proteins suggests that the observed alleviation of symptoms in the chronic phase of Chagas disease may be due to a collective action of multiple molecules on different targets. These findings encourage further investigation of *Cymbopogon citratus* as a potential alternative for Chagas disease treatment.

ID# 038

Exploration of heavy metal resistance in the yeast Wickerhamomyces anomalus: implications for bioremediation

Garolera B.[1], Pajot H. P.[1, 2], Fernández P. M.[1, 2]

[1]PROIMI-CONICET (Planta Piloto de Procesos Industriales Microbiológicos, Consejo Nacional de Investigaciones Científicas y Técnicas)

[2]FACEN-UNCA (Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Catamarca)

Abstract

The inadequate disposal of wastewater containing toxic heavy metals and industrial contaminants has become a critical issue today, posing serious risks to the health of humans, animals, and the environment. Numerous studies have investigated how heavy metals are absorbed, accumulated, and transformed using microorganisms. The yeast strain Wickerhamomyces anomalus M10 has demonstrated notable resistance to several heavy metals, showing promising potential for bioremediation.

Therefore, through computational methods (In Silico studies) to investigate metal resistance mechanisms, relevant genes related to this resistance were identified. Specialized databases like BacMet and InterPro were used, compiling genes from bacteria and fungi that enhance survival under environmental stress. These databases categorize resistance genes based on their function and induced phenotypes. Using BacMet, we identified 59 genes, including 11 ABC-type transporters, 10 related to resistance and the efflux of multiple drugs, as well as proteins for copper (5), nickel (10), arsenic (2), mercury (5), and silver (1). These proteins perform functions in binding, reduction, and transport, along with superoxide dismutase proteins, DNA regulation, and repair genes. Additionally, through InterPro, we found 394 genes, including 257 transporters, 25 copper-related proteins, and 23 zinc-related proteins. Fifteen multidrug resistance proteins and other DNA regulation and repair genes (10) were detected.

In summary, our bioinformatics analysis of Wickerhamomyces anomalus highlights its potential for bioremediation in metal-contaminated environments. It excels particularly in mitigating nickel and arsenic pollution due to resistance-related genes. However, comprehensive in vivo studies are required to confirm its survival and transformation capabilities against these metals. These findings underline the significant bioremediation potential of yeast strains. The study not only reveals genes that expand the scope of microbial bioremediation but also advances the concept of the substantial role of yeast strains in mitigating environmental pollution challenges

ID# 039

Genomic diversity and security profile of Corynebacterium *pseudodiphtheriticum* species

Mariano Elean[1], Ramiro Ortiz Moyano[1], Fernanda Raya Tonetti[1], Julio Villena[1]

[1]Laboratory of Immunobiotechnology, Reference Centre for Lactobacilli (CERELA-CONICET), Tucumán 4000, Argentina

Background: Bacteria belonging to the genus *Corynebacterium* have usually been associated with infectious diseases. C. diphtheriae and C. ulcerans produce toxins that can induce upper respiratory tract illness characterized by sore throat, fever, and nasal discharge. However, not all the species of this genus are pathogenic. C. *pseudodiphtheriticum* is usually found as a member of the normal microbiota of the upper respiratory tract and some studies have indicated that these bacteria exert beneficial effects for the host. In this regard, we have shown that the strain C. *pseudodiphtheriticum* 090104 is able to beneficially modulate immune responses in the respiratory tract, improving the resistance to pathogens such as *Streptococcus pneumoniae, Klebsiella* pneumoniae and respiratory syncytial virus. Considering the positive effects of this species, we aimed to carry out a comparative genomic study of C. *pseudodiphtheriticum* strains to characterize their security profile in terms of virulence genes

Results: The genomes of 29 *C. pseudodiphtheriticum* strains, including the 090104 strain, were selected for the study. The virulent strains *C. ulcerans* 809, *C. diphtheriae* NCTC13129 and *C. pseudotuberculosis* PAT10 were also included for comparisons. The average nucleotide identity analysis revealed two large clusters. One constituted by the strains MSK305, 090104, CPD, CP10, MSK092, MSK080, DSM44287 and MSK037; and the second constituted by the remaining strains. All the *C. pseudodiphtheriticum* strains show more than 90% identity between them. The pangenomic analysis revealed a pangenome of 5050 genes, a core-genome of 1471 genes, a shell-genome of 1118 genes and a cloud-genome of 2461 genes. The presence of virulence genes *dip0733, dtxr*, and genes linked to secretion systems (*sec* genes and *t2sf*) was observed in all strains. The virulence gene *hmuv* was found to be present in 27 of 29 strains (absent in strains MSK037 and MSK092).

Conclusions: *C. pseudodiphtheriticum* strains have low potential to generate infections due to the low number of virulence genes. However, although the risk for health is small, it is advisable to perform detailed in vitro and in vivo studies of the strains intended to be used in probiotic formulations, to ensure its complete safety.

ID# 041

Characterization of RNA and DNA viromes in sewage samples collected from Buenos Aires, Argentina

Zambrana Montaño R.[1,2], Blanco Fernández M.D.[1,2], Torres C.[1,2]

[1]Universidad de Buenos Aires, Facultad de Farmacia y Bioquímica, Instituto de Investigaciones en Bacteriología y Virología Molecular (IBaViM), Buenos Aires, Argentina
[2]Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina

Background: The utilization of next-generation sequencing (NGS) techniques on environmental samples plays a crucial role in public health to identify the presence of viral species that are responsible for a wide variety of human diseases. The aim of this work was to conduct an initial approach into the RNA and DNA virome of two sewage samples collected in Buenos Aires province (Argentina) during two seasons, Summer (February) and Spring (October) 2022.

Results: A virus concentration method based on prefiltering a 2L sample using 0.45 and 0.22 µm filters, followed by ultrafiltration (50 kDa) and PEG precipitation was used. RNA and DNA extraction was carried out separately using commercial kits. To facilitate viral detection in NGS, sequence-independent single-primer-amplification (SISPA) was implemented on RNA and DNA samples. Paired-end sequences were obtained from the Illumina platform (2x150). Adaptors and primers were removed, short reads were trimmed (< 75 bp) and reads with low-quality score (Phred score below 30) were eliminated using Trim Galore. Then, the taxonomic classification of reads was performed using Kraken2 and Bracken against the NCBI RefSeq viral database.

RNA sample obtained in Summer was composed of viral (30.3%), bacterial (20.6%) and unclassified sequences (48.8%), whereas that obtained in Spring showed viral (24.8%), bacterial (48.3%) and unclassified (26.1%) sequences. Among the viruses, despite the overrepresentation of *Virgaviridae* (94.8-99.3%), our analysis revealed species from more than 30 other families, including *Astroviridae*, and detected reads corresponding to the *Picornaviridae* family in Summer and *Caliciviridae* in Spring viromes.

DNA samples resulted in a low recovery of viral sequences (< 10% of total reads). Most of the viruses found belonged to the *Duplodnaviria* domain (91.4%-97.7%), dominated by dsDNA phages of the order *Caudovirales*. In both Summer and Spring samples, a low abundance of Herpesviridae, Circoviridae, *Parvoviridae, Poxviridae, and Adenoviridae* families was found, among others. Besides, the Summer samples also had reads classified as belonging to *Papillomaviridae* and *Polyomaviridae* families.

Conclusions: The strategy implemented in this work allowed the detection of several viral families from different host species in sewage samples in Argentina. Subsequent analyzes will allow a deep characterization of viral species and their diversity.

ID# 043

WS-YOLO: an agronomical and computer vision-based framework to detect water stress in lettuce seedlings using IR imaging and YOLOv8

Wolter-Salas, S.[1]; Canessa, P.[2,3]; Campos-Vargas, R.[4]; Opazo, M.C.[5]; Sepúlveda, R.V.[1]; Aguayo, D.[6]

[1]Center for Bioinformatics and Integrative Biology, Facultad de Ciencias de la Vida, Universidad Andrés Bello, Av. República 330, 8370186, Santiago, Chile

[2]Centro de Biotecnología Vegetal, Facultad de Ciencias de la Vida, Universidad Andrés Bello, Av. República 330, 8370186, Santiago, Chile

[3]ANID–Millennium Science Initiative–Millennium Institute for Integrative Biology (iBIO), Av. Libertador Bernardo O'Higgins 340, 7500565, Santiago, Chile

[4]Centro de Estudios Postcosecha, Facultad de Ciencias Agronómicas, Universidad de Chile, Av. Santa Rosa 11315, 8831314, Santiago, Chile

[5]Instituto de Ciencias Naturales, Facultad de Medicina Veterinaria y Agronomía, Universidad de Las Américas, Av. Manuel Montt 948, 7500000, Santiago, Chile

[6]Agricultura Digital, Servicio Agrícola, Salinas y Fabres S.A., Ruta 5 Sur, Parcela #165, 9540000, Maipo, Chile

Background: The agricultural sector has been severely affected by water shortages impacting horticultural production. The effects of water stress on horticultural crops can induce physiological stress in plants, leading to stunted growth, diminished produce quality, and increased susceptibility to pests and diseases. Lettuce (*Lactuca sativa L*.) is highly susceptible to drought and water deficits, resulting in lower crop yields, unharvested areas, reduced crop health and quality. To address this, we developed a High-Throughput Phenotyping platform using Deep Learning and infrared images to detect stress stages in lettuce seedlings, which could help to apply real-time agronomical decisions from data using variable rate irrigation systems.

Results: Accordingly, a comprehensive database comprising infrared images of lettuce grown under drought-induced stress conditions was built. In order to capture the required data, we deployed a Raspberry Pi robot to autonomously collect lateral infrared images of lettuce seedlings during an 8-day drought stress experiment. This resulted in the generation of a database containing 2119 images through augmentation. Leveraging this data, a YOLOv8 model was trained (Water Stress – YOLO, or WS-YOLO), employing instance segmentation for accurate stress level detection. The results demonstrated the efficacy of our approach, with WS-YOLO achieving a mean Average Precision (mAP) of 93.62% and an F1 score of 89.31%. Particularly, high efficiency in early stress detection was achieved, being a critical factor for improving food security through timely interventions.

Conclusions: This research contributes to the field of agricultural technology and stress detection in lettuce. By introducing a novel High-Throughput Phenotyping platform that leverages Deep learning, Robotics, and Computer Vision, the study addresses the critical challenge of early stress detection through infrared imaging in lettuce, crucial for ensuring food security and mitigating yield losses. These findings showcase the efficacy and potential of AI-driven solutions in tackling pressing challenges in food production and sustainability. Moreover, the creation of a comprehensive database of IR images through autonomous data collection further enriches the scientific knowledge base and opens opportunities for further research in cutting-edge Deep Learning techniques for stress detection in crops.

ID# 045

Bioinformatic study of large glutamate dehydrogenases and their signaling pathway in Actinobacteria

Baffo, Natacha[1]; Marthey, Sylvain[2]; André-Leroux, Gwenaelle[2]; Lisa, María-Natalia[1,3]

[1]Laboratorio de Microbiología Estructural y Bio-diseño, Instituto de Biologia Molecular y Celular de Rosario (IBR, CONICET-UNR), Rosario, Argentina
[2]MaiAGE, INRAe, Jouy-en-Josas, France
[3]Plataforma de Biología Estructural y Metaboĺ ómica (PLABEM), Rosario, Argentina

Background: Glutamate dehydrogenases (GDHs) are ubiquitous oligomeric enzymes with key roles in amino acid homeostasis. The subfamily of large GDHs (L-GDHs) includes L-GDHs$_{115}$ and L-GDHs$_{180}$ (115 and 180 kDa monomers, respectively). We have recently solved the 3D structure of the mycobacterial L-GDH$_{180}$ (mL-GDH$_{180}$), the first experimental structure available for a L-GDH. Notably, mL-GDH$_{180}$ adopts a unique quaternary architecture and includes domains possibly involved in sensing metabolic signals. Furthermore, L-GDHs$_{180}$ from Mycobacterium species are modulated by the regulator GarA, responding to extracellular nutrient availability. Thus, the tubercle bacillus could integrate intracellular and host metabolic information through mL-GDH$_{180}$. It has been proposed that GlnX monitors the concentration of L-glutamate in the mycobacterial periplasm and that a protein-protein interaction would transduce such information through the membrane via GlnX, to then activate the kinase PknG that phosphorylates GarA. This cascade, extending from GlnH to GarA, potentially mediates the regulation of nutrient utilization based on amino acids availability in the extracellular medium. Our investigation is aimed at examining the ubiquity of this signaling pathway across actinobacteria. Notably, our analysis includes an exploration of the subfamilies of L-GDHs, as previous studies in this context are lacking.

Results: We identified 21 L-GDHs$_{115}$ and 140 L-GDH$_{180}$ in a database of 247 complete and referenced genomes available for Actinobacteria. Sequences were grouped based on primary structure conservation and a set of 33 representative L-GDHs was selected to perform structural predictions. These calculations were carried out using ab initio algorithms. Altogether, our findings indicate that the tertiary structure of the modules constituting monomers is maintained. Concerning the pathway exploration, our approach employs a novel tool enabling gene co-occurrence analysis, interspecies comparisons and pathway prevalence evaluation. This search allowed the categorization of 247 genomes into 23 groups based on enzyme combinations observed in the pathway. This analysis requires further in-depth exploration, and our efforts in this regard remain ongoing.

Conclusions: This work contributes to support the hypothesis that the structural characteristics found for mL-GDH$_{180}$ constitute a common theme in the L-GDHs subfamily. We have observed at least partial conservation of the pathway in the most prominent genera of the phylum.

ID# 046

New set of classes for fruit shape classification in tomato based on machine learning

Vazquez, Dana V.[1,2]; Spetale, Flavio E.[3,4]; Tapia, Elizabeth[3,4], Rodríguez, Gustavo R.[1, 2]

[1]Cátedra de Genética. Facultad de Ciencias Agrarias. UNR, Zavalla, Argentina
[2]IICAR-CONICET-UNR
[3]Facultad de Ciencias Exactas, Ingeniería y Agrimensura, UNR, Rosario, Argentina
[4]CIFASIS-CONICET-UNR

Background: Tomato (*Solanum lycopersicum L.*) is the second most consumed global vegetable. Fruit shape significantly impacts on yield, quality, consumer preference, and commercial usage. Despite of the digital advancements in precision agriculture, the determination of fruit shape still relies predominantly on visual assessment, and there are no standardized approaches. Classification criteria often vary among experts, and exist for tomato four: "Rodriguez2011", "Visa2014," "UPOV," and "IPGRI". They define eight, nine, ten, and eight classes, respectively) and do not present consensus. This study aims to develop a machine-learning model for automated tomato shape classification and establish a “gold standard”.

Results: Using the Solanaceae Genomic Network Repository, a total of 1424 longitudinal-sectioned tomato fruit images were examined, and 41 numerical variables were obtained from Tomato Analyzer software. The fruits were visually classified using the four known criteria. Additionally, a novel set of classes was introduced, merging the rectangular class from Rodriguez2011 method into the ellipsoid class. The data set was split for train (80%) and test (20%). Variables were standardized using Z-Score. Four highly correlated (>0.95) variables were removed. The key variables were identified for each method by Recursive Feature Elimination, ultimately keeping 12 ones that were common across all methods. The supervised classification methods employed were multinomial logistic regression, random forest, and support vector machine. The models did not show significant differences in mean accuracy ($p>0.05$). However, substantial differences were noted among the methods ($p<0.01$) for all models. Wilcoxon-Mann-Whitney test showed that mean accuracy for UPOV and IPGRI was not significantly different and exhibited the lowest values, Rodriguez2011 and Visa2014 showed no significant differences for accuracy and intermediate values, and the novel set of classes yielded the highest mean accuracy values across all four models, i.e., 85%.

Conclusions: The results demonstrate the new set of classes enhances classification accuracy and is a "gold standard" for future shape studies. Finally, the novel method proposes seven shape classes: flat, round, ellipsoid, heart, oxheart, obovoid, and long, achieving 85% accuracy. This "gold standard" for fruit shape facilitates precise tomato cultivar description and consensus among researchers, aiding genetic understanding.

ID# 047

A comparative assessment on Fungal genome annotation

Andreiu, Erika María Martha[1]; Chelaliche, Aníbal Sebastian[1,2]; Benítez, Silvana Florencia[1,2]; Zapata, Pedro Dario[1,2]; Fonseca, María Isabel[1,2]

[1]Universidad Nacional de Misiones. Facultad de Ciencias Exactas, Químicas y Naturales. Instituto de Biotecnología Misiones. Laboratorio de Biotecnología Molecular. CP3300 Posadas, Misiones, Argentina
[2]CONICET. Buenos Aires, Argentina

Background: Genome annotation techniques and methodologies used in the assessment of fungal genomes play a crucial role in understanding the genetic background and functionalities of these organisms. There are various approaches to this complex process, and it is important to evaluate the quality and reliability of the annotation performed on fungal genomes. In this context, *Pleurotus pulmonarius* LBM 105, a fungus isolated from Misiones (Argentina), has demonstrated qualities suitable for environmental biotechnology applications. Therefore, progress in annotating its genome represents a significant advancement in both scientific and technological understanding.

Results: In this study, we examined various annotation methods applied to the de novo assembly of *P. pulmonarius* LBM 105 in order to evaluate their accuracy and completeness. We compared a reference annotation generated with CLC Genomic Workbench to classic ab initio annotation using the Augustus software and fungal optimized annotation software called Funnanotate. Additionally, we incorporated two different masking approaches for the ab initio annotations - one using RepeatModeler to generate repeat sequences from the genome, and another utilizing the Dfam repeat database. The most complete annotation was achieved by the reference annotation yielding a total of 13560 genes, 79283 CDS and 12694 mRNAs. However, using Augustus we achieved similar results with 13331 genes, 57933 CDS and 13331 mRNA

Conclusions: This level of completeness suggests that the annotation pipeline used in this study successfully captured a significant portion of the fungal genome, highlighting the importance of using a high-quality reference for the genomic studies. However, the ab initio annotation results showed promising results, allowing to better capture the strain-specific characteristics of the genome.

ID# 051

RAC1 as a therapeutic target in colorectal cancer

Katia Otterstedt[1], Florencia Malizia[2], Luciano E. Anselmino[2], Mauricio Menacho Márquez[3]

[1]Centro de Investigación y Producción de Reactivos Biológicos, Argentina
[2]Instituto de Inmunología Clínica y Experimental de Rosario, Argentina
[3]CONICET, Argentina

Background: Colorectal cancer (CRC) is the third most commonly diagnosed type of cancer worldwide. Rac1 is a key member of the Rho GTPases family. It modulates cell adhesion and movement, and is highly expressed in tumors. Researchers are increasingly exploring Rac1 as a potential target for tumor therapy.

The aim of this work was to evaluate the impact of Rac1 in CRC. Given that approximately 30–40% of CRC patients carry a KRAS mutation, we began evaluating the relevance of RAC1 expression in KRAS wild-type/mutated CRCs. We first downloaded the TCGA-COAD dataset and separated patients according to RAC1 expression levels using the "Survminer" package. In turn, we added an additional filter further separating patients according to KRAS gene status. We extended this evaluation for Rac1 guanine nucleotide exchange factors (GEFs).

To continue characterizing the role of Rac1 in CRC, the study aimed to identify genes that exhibited differential expression between patients with high and low RAC1 expression. The analysis was refined by filtering patients based on their KRAS gene status. By identifying these differentially expressed genes, we created a genetic signature for each patient group. These signature genes were then subjected to overrepresentation analysis and gene set enrichment analysis, enabling the identification of key pathways associated with each phenotype.

Results: Differential genes obtained as a result of this study are strongly linked to cellular metabolism, proliferation, amyloid fiber formation and programmed cell death. Moreover, we found that high expression of RAC1 was associated with poor prognosis in patients where KRAS is mutated and that several RAC1 GEFs display different survival patterns depending on the presence of KRAS mutations in their genomes.

Conclusions: All our data allowed us to postulate that targeting Rac1 represents a promising approach for developing novel therapies for CRC patients, particularly those with mutated KRAS gene.

ID# 052

A workflow integrating R/Bioconductor, GSEA, and TIMER 2.0 to explore the role of the Vav protein family in cutaneous melanoma

Aylén Avila[1], Luciano E. Anselmino2, Nahuel Cesatti2, Macarena Mamberto[2], Lucía C Zanotti[2], Florencia Malizia[2], Mauricio Menacho Márquez[3]

[1]Centro de Investigación y Producción de Reactivos Biológicos, Argentina
[2]Instituto de Inmunología Clínica y Experimental de Rosario, Argentina
[3]CONICET, Argentina

Abstract

Vav proteins are RHO guanine nucleotide exchange factors (GEFs). This family consists of three members, which typically exhibit functional redundancy and are associated with proactive functions in cancer. However, their role in melanoma remains largely unexplored.

Our aim was to establish a systematic approach, utilizing bioinformatic techniques, to investigate the role of each member of the Vav family in melanoma.

Gene expression data from cutaneous melanoma patients were obtained from the 'Cancer Genome Atlas' database. Raw counts were subsequently normalized to counts per million (CPM) using the 'edgeR' package. The patient cohort (n=460) was stratified based on high or low expression levels of VAV1, VAV2, and VAV3. Survival plots were generated using the Kaplan-Meier estimator and 'survminer' package. The log-rank test revealed an association between high VAV2 expression and poorer prognosis, whereas elevated VAV1 and VAV3 expressions correlated with increased patient survival probability ($p<0.05$ in each case).

Gene set enrichment analysis was conducted for each comparison group using the GSEA software. To assess immune and stromal cell infiltration in tumor tissues, Immune Score and Microenvironment Score were calculated based on gene expression profiles of the tumor microenvironment, employing the ESTIMATE and xCell algorithms. Both Scores showed a strong and positive association with VAV1 and VAV3 expressions ($p<0.001$). Then, using eight different algorithms, with the 'estimate' package and the TIMER2.0 application, correlation with some cell types was evaluated. A robust positive correlation was identified between VAV1 expression and some types of immune cell signatures ($p<0.001$). Conversely, no significant correlation was observed between VAV2 or VAV3 expression and cell types.

Our findings suggest that a favorable prognosis in melanoma is linked to elevated expressions of VAV1 and VAV3, coupled with reduced VAV2 expression. This prognosis may arise from Vav1's impact on intercellular communication within the tumor microenvironment, while heightened VAV3 expression could regulate the activation of tumor cell signaling pathways, thereby promoting greater immunogenicity.

Our study presents a comprehensive pipeline that could serve to explore the implications of other proteins in diverse disease.

ID# 054

Unsupervised Clustering and Convolutional Neural Networks for Learning Structures in Bioinformatics

Ignacio Fucksmann[1], Leandro A. Bugnon[1], Diego H. Milone[1]

[1]Research Institute for Signals, Systems and Computational Intelligence, sinc(i), FICH-UNL, CONICET, Ciudad Universitaria UNL, 3000 Santa Fe, Argentina

Background: The study of ribonucleic acids (RNAs) greatly relies on computational prediction of sequence folding, as it plays a crucial role in determining their functions. Artificial intelligence, particularly deep learning, has begun to be employed for the prediction of RNA structure based on its sequence. However, state-of-the-art methods provide results below 70% of F1 on average, so there is still a lot of room for improvement. This is mainly due to the structural diversity found on RNAs.

Results: We propose a novel method to improve learning of RNA structures by combining supervised and unsupervised learning. We started from a reference model based on convolutional neural networks (CNNs), and integrated information regarding the clustering of sequences. This was achieved by training a number of submodels in correspondence to each cluster and assembling the predictions afterwards. We have studied the relationship between clusters based on the secondary structure and the ones based on the sequence distance by measuring their mutual information. This was achieved through the generation of a collection of clusters based on the two distances, and measuring similarity of partitions using the normalized mutual information (NMI). Our results achieved a value of NMI=0.79, showing that, for a certain number of clusters, based on structure are very correlated to the ones based on sequence. Then, we trained prediction models using CNNs independently for each cluster based on sequence distances (as structure distances are not available on test). These models were then assembled to obtain the final prediction, resulting in a significant improvement in F1 score (5.20%) compared to the reference model.

Conclusions: We achieve better results using the hybrid unsupervised/supervised approach. The information about structure, inferred only from sequences, can be used to divide the learning problem into several less-complex subproblems and obtain better structure predictions by training specific models for each structure type.

ID# 055

SNP identification in two local C. cardunculus genotypes by GBS

Cacchiarelli, Paolo[1]; Zayas, Aldana[2]; Cravero, Vanina[1]; Leskovar, Daniel[3]; Martin, Eugenia[1]

[1]Instituto de Ciencias Agrarias de Rosario (IICAR-CONICET-UNR), Zavalla, SF, Argentina
[2]Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET), Rosario, SF, Argentina
[3]Texas A&M AgriLife Research (Texas A&M University), Uvalde, Texas, USA

Abstract

The genetic diversity analysis coupled with new generation sequencing technologies such as Genotyping by Sequences (GBS) provides a valuable tool for breeding programs of crops. The species Cynara cardunculus (2n=2X=34) includes three taxa: the globe artichoke, the domesticated cardoon, and the wild cardoon. The globe artichoke is traditionally cultivated for its immature flower heads worldwide, especially in Mediterranean countries, USA, Peru, and Argentina, whereas the wild cardoon, considered the ancestor of both domesticated forms, is cultivated for its biomass as energy crop.

The objective of this study was to identify polymorphisms in two *C. cardunculus* genotypes by applying GBS technology. The plant material was one genotype of Estrella del Sur FCA (ES) globe artichoke cultivar and a local genotype of wild cardoon (WC), both previously used as parental in local genetic maps of the species. Genomic DNA was isolated from young leaves of each genotype using DNeasy Plant mini Kit (QIAGEN) and quantified with Qubit 2.0 fluorometer (Thermo Fisher). DNA libraries were prepared according to a double restriction system and were sequenced using a Hiseq2500 (Illumina) at the Texas A&M Genomics and Bioinformatics Service. A total of 5.292.112 raw data was obtained for the ES genotype and 6.581.974 raw data for *WC*. The raw reads obtained from both genotypes were cleaned and mapped on the globe artichoke reference genome V2 using fastp and *BWA*, respectively. After filtering with fastp we obtained 3.033.318 reads from ES genotype and 3.749.488 reads from *WC*. Variant calling was performed using lofreq tool and 25.128 SNP variants were detected in the ES genotype and 25.352 SNPs in *WC*. The distribution of each type of SNP was analyzed and classified as transitions or transversions.

In conclusion, the GBS technology successfully revealed thousands of polymorphisms in *C. cardunculus* and can be used in the development of high-density genetic maps to assist breeding programs of the species

ID# 058

Wickerhamomyces anomalus, a Biotechnological Yeast... an Opportunistic Pathogen?

Garolera B[1], Pajot H. P[1,2], Fernández P. M[1,2]

[1]PROIMI-CONICET (Planta Piloto de Procesos Industriales Microbiológicos, Consejo Nacional de Investigaciones Científicas y Técnicas)

[2]FACEN-UNCA (Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Catamarca)

Abstract

Wickerhamomyces anomalus is a yeast with significant biotechnological potential, utilized in the production of fermented beverages and ethanol, as well as in the biocontrol of postharvest diseases, bioremediation of metals and organic contaminants, and surfactant production, among other applications. Its utility arises from its capacity to thrive on various carbon sources, tolerate a wide range of pH values, and survive temperatures ranging from 20 to 40°C. Nevertheless, documented cases of fungal infections in immunocompromised patients, including infants and adults, suggest its potential as an opportunistic pathogen. In our laboratory, we isolated a strain of Wickerhamomyces anomalus with potential for hexavalent chromium bioremediation. The aim of this study was to in silico assess the presence of virulence factor genes in the Wickerhamomyces anomalus M10 strain, which could confer opportunistic pathogenicity in humans.

Using PHI-base, we identified a total of 65 virulence factors, categorized based on their phenotypic impact in mutations. Loss-of-pathogenicity factors (13) are primarily associated with genes involved in hyphae or pseudohyphae formation, such as Cas5, CDC42, and VMA4, whereas virulence reduction factors (37) are linked to genes related to biofilm formation, like BMH1 and RPS41. Furthermore, we identified genes that may enable the yeast to evade host defenses, such as an alpha-mannosyltransferase (MNN10), all belonging to proteins involved in signaling pathways, transport, and transcription regulation.

In summary, based on this initial bioinformatic analysis of the Wickerhamomyces anomalus genome, we suggest that the yeast possesses genes that could facilitate its invasion and dissemination within the host, along with the ability to evade host defenses, potentially conferring opportunistic pathogenicity. It is crucial to conduct in vivo assays to validate these findings, given the potential risk to human health when used in biotechnological processes.

ID# 059

An in-silico analysis of possible inhibitors for the nsP2 protein contained in Chikungunya virus

Mauricio Gonzalez[1], Edgar Cardozo[1], Ivana Fernandez[1], Ana Gómez[1], Silverio Andrés Quintana[1]

[1]Universidad Nacional de Asunción, Facultad de Ciencias Exactas y Naturales, Departamento de Biotecnología, San Lorenzo, Paraguay

Abstract

Chikungunya fever, caused by the CHIKV alphavirus, is a disease that is rapidly expanding in tropical areas like South America. Chikungunya virus is a small positive sense RNA virus containing 11.8 kb nucleotides, of which its domains consist of 4 non-structural genomes (nsP1-4) and 5 structural proteins. This virus infects fibroblasts and dermal macrophages and later myocytes, the complete response may cause persistent arthralgia for up to a few years. For this reason, we started to analyze some compounds that originate from plants and bacteria, compounds that may serve as inhibitors for the replication machinery of the virus, in this case the nsP2 domain which contains the helicase of the viral particle.

The CHIKV viral helicase (PDB ID: 6jim, chain B) was tested against cured ligands in Dockthor's website. The nature of ligands is diverse, Reserpine, Ganoderic Acid A, Artemisinin, Prostratin, Eugenol and Lucidumol are extracted from plants and fungi, while Papain Inhibitor and Bestatin are from bacteria. These ligands were retrieved from NCBI Pubchem and were treated for energy optimization with Avogrado2's OpenBabel plugin.

Results from docking shows scores of ΔG=-9.323 kcal.mol^{-1} for Reserpine; ΔG=-8.244 kcal.mol^{-1} for Artemisinin and Prostratin; ΔG=-8.495 kcal.mol^{-1} for Ganoderic Acid A; ΔG=-7.837 kcal.mol^{-1} for Bestatin; ΔG=-7.859 kcal.mol^{-1} for Lucidumol; ΔG=-7.61 kcal.mol^{-1} for Papain Inhibitor and ΔG=-7.434 kcal.mol^{-1} for Eugenol, while 4alpha-TPA, Ganodermanontriol, Oxytetracycline, Atrolactic acid and Phorbol failed the docking procedure. Later the results were taken to AutoDockTools 1.5.7 for verification and imaging. The scores show that using these compounds as inhibitors may work, making Reserpine a good candidate for further testing.

In conclusion, there are a lot of possible candidates that may serve as inhibitors in silico, but experimental testing is necessary, there's also the probe designing, which is the next step necessary to reach the replication complex and prevent viral particle replication, because of this, it's necessary to focus on development of new countermeasures against the virus

ID# 060

Sequential Assembly of Large Bromodomain Containing Complexes in Trypanosomatids

Rodríguez Araya, Elvio[1,2]; Serra, Esteban[1,2]

[1]Laboratorio de Tripanosomatidos, Instituto de Biología Molecular y Celular de Rosario, Rosario, Argentina
[2]Area Parasitología, Universidad Nacional de Rosario, Rosario, Argentina

Background: Acetylation signaling pathways in trypanosomatids, a group of early branching organisms, are poorly understood due to highly divergent protein sequences. Also, there is virtually no structural evidence of the complexes involved. Recently, Staneva et al systematized the analysis of the proteins present in many putative chromatin regulators, using co-immunoprecipitation followed by mass spectrometry assays to identify protein networks. In this work, we took these interactomic datasets and made a subset focusing on those that contains bromodomain factors (BDFs), relevant to recognize and transduce the signals of acetylated-lysines, and predicted all the pairwise proteinprotein interactions present in the subset using AlphaFold2-multimer and high-performance computing. We then performed another round of complex structure prediction to assemble trimers, tetramers and pentamers with high probability of being interactors, followed by a sequential assembly using CombFold to assemble the higher order complexes.

Results: The results showed 6 distinct complexes that assemble with different degrees of shared subunits. Two of them are partially conserved complexes present in human and yeast, the NuA4 complex and Swr1 complex, both involved in chromatin remodeling mediated by acetylation. Another complex shares many structural similarities with NuA4, containing several paralog proteins that assemble in a completely novel complex, but that do not seem to share subunits. One contains 4 BDFs interacting together, potentially taking similar functional roles like those of large proteins with tandem bromodomains observed in other organisms (poly-bromo), not present in trypanosomatids. Another forms a complex containing a histone deacetylase that shares BDF2 as a subunit with Swr1. The last one contains BDF7, which is homolog of human ATAD2. It arranges as a hexamer, forming a ring with a central pore like what has been observed for these kinds of proteins.

Conclusions: Overall, these findings shed light on the acetylation signaling pathways in trypanosomatids, providing valuable insights into the organization and structure of bromodomain factor containing complexes. The identification of novel complexes and their potential functional roles expands our understanding of chromatin remodeling mechanisms via acetylation in these early branching organisms and offers exciting prospects for further research in epigenetic regulation.

ID# 062

Analysis of CIS-regulation in kinase proteins

Navarro, Álvaro Martin[1*]; Alonso, Macarena[1*]; Gibson, Toby[2]; Iserte, Javier[1]; Tompa, Peter [3*]; Marino-Buslje, Cristina[1*]

[*] equal contribution

[1]Bioinformatics Unit, Fundación Instituto Leloir (FIL), Buenos Aires, Argentina
[2]Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
[3]Structural Biology Brussels (SBB), Bioengineering Sciences Department, Vrije Universiteit Brussel (VUB), Brussels, Belgium

Background: Protein activity can be regulated through intramolecular interactions given by cis-regulatory elements (CRE). This type of regulation can lead to auto-inhibition or auto-activation. Kinases, important enzymes in many biological processes, are frequently regulated via inhibitory interactions in cis. Dysregulation of the kinase activity has been implicated in a variety of diseases. In this work, we investigate the structural mechanisms of cis-regulation in CRE-containing kinases proteins.

Results: We compiled a dataset of 280 kinases from literature and UniProtKB. In a second stage, we retrieved their orthologous from OmaDB and OrthoInspector databases. We constructed multiple sequence alignments (MSAs) and downloaded the structures from the Protein Data Bank when available. We measured the interatomic distances between the CRE and the kinase domains (KD) within a protein, as well as between the CREs of two proteins in pairs and calculated the contact maps.

Based on the CRE and KD distances of the different structures in an MSA, we find two typical scenarios. In the first one, all the structures are in the same conformation, for example in the cell-surface receptor tyrosine-protein kinase FLT3 family. In this case, most of the CRE residues are in contact distance with the KD and therefore inhibit its activity. In the second scenario, the CRE is in contact with the KD in some conformers (denoting a specific condition) while it is not in others. Such is the case of the calcium/calmodulin-dependent protein kinase family (CAMK).

We saw that CREs disorder content show a great variability, suggesting the presence of ordered as well as disordered CREs modulating the protein function.

Conclusions: With the available information, we conclude that the mechanisms of cis regulation are different among kinase families. They can carry different classes of CREs, from very short motifs to large fully folded domains, and engage minimal movement as well as large conformational changes. In all cases to achieve the fine-tuned process of self-regulation.

We believe that further experimental data is needed to have a better understanding of the kinases auto-inhibitory mechanism through the lens of the structure.

This work may help advance our knowledge of biological processes and disease, and could have implications for the development of therapeutic interventions.

ID# 065

Obtention and characterization of the genomes of mitochondria and chloroplast from two tomato genotypes

Nicolas Defarge[1], Vladimir Cambiaso[1,2], Gustavo Rubén Rodriguez[1,2]

[1]Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR-CONICET-UNR), Argentina
[2]Grupo Genética y Mejoramiento del Tomate, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Zavalla, S2125ZAA Santa Fe, Argentina

Abstract

In tomato (*Solanum lycopersicum L.*), two types of crossings are used to improve the quality and quantity of production. Intervarietal crossings are made between different cultivars to obtain hybrid plants with higher performance than their parents, while crossings with wild species are used to transfer beneficial genes into the crop. Reciprocal effects (ER) refer to changes in the hybrid's phenotype when the sexual role of the parents is reversed. Our research showed that reciprocal effects exist for fruit attributes in tomatoes, and the maternally inherited mitochondrial and chloroplastic genes could be the cause. In order to develop molecular markers that allow us to track the inheritance of mitochondria and chloroplasts, we reconstructed the cytoplasmic genomes aligning reads obtained from whole genome sequences of the *S. lycopersicum* cultivar *Caimanta (C)* and the *S. pimpinifolium* accession LA0722.

Illumina reads were aligned to the available mitochondrial (MF034193, accession LA1479) and chloroplast (NC_007898.2, accession LA3023) reference genomes using Bowtie2. This resulted in an alignment rate of 5.23% and 12.11% for C and 5.39% and 17.61% for LA0722, respectively. Cytoplasmic genomes were entirely covered with a mean mapping quality of 34.53 and 29.73 for C (mitochondria and chloroplast, respectively), and 34.49 and 29.74 for LA0722. Four insertion/deletions (INDELs) and nine single nucleotide polymorphisms (SNPs) were found to discriminate the mitochondrial genomes of the two tomato genotypes, while we found 18 INDELs and 57 SNPs at the chloroplast genomes comparations.

We aligned Illumina reads from two parental genotypes used in our tomato breeding program to the available mitochondrial and chloroplast reference genomes. This allowed us to assemble usable cytoplasmatic reference genomes for C and LA0722, and identify and map INDELs and SNPs that discriminate between both accessions. These results will serve as a basis for developing molecular markers that allow us to track the inheritance of mitochondria and chloroplasts hybrids and segregating populations derived of their cross. This will help us verify if cases of paternal/maternal inheritance in tomatoes. Furthermore, this work may reveal that QTLs for agronomic traits of interest are linked to chloroplasts and mitochondrial genomes.

ID# 072

Integrating omic data to detect association between sHSP transcripts and phenotypic variability in tomato fruits

Cacchiarelli Paolo[1], Spetale Flavio[2,3], Tapia Elizabeth[2,3], Pratta Guillermo[1]

[1]Instituto de Ciencias Agrarias de Rosario (IICAR-CONICET-UNR), Zavalla, SF, Argentina
[2]CIFASIS-CONICET-UNR, Rosario, Santa Fe, Argentina
[3]Facultad de Ciencias Exactas, Ingeniería y Agrimensura, UNR. Rosario, Argentina

Background: Fruit ripening is a complex developmental process highly coordinated by different gene families such as small Heat Shock Proteins (sHSP), which maintain cell homeostasis and are a superfamily of chaperones that have been characterized in other organisms. In plants, sHSP promotes protein folding and disaggregation during stress or developmental changes. Previously, we obtained the transcriptome of cv. *Caimanta* (C, *S. lycopersicum*), the exotic LA0722 (P, *S. pimpinellifolium*), and their interspecific hybrid (CxP) in 3 main stages of ripening. Also, agronomic characterization of fruit attributes was achieved in the three genotypes.

Results: The objective of this communication was to evaluate the association between both datasets. A Generalized Procrustean Analysis (GPA) was applied to gene expression levels of two clusters of sHSP located on chromosomes 6 and 9, respectively, as transcriptomic data, and 11 quantitative fruit traits as phenomic data. Also, an estimation of the degrees of dominance (d/a) was carried out for the levels of gene expression and the fruit traits. Principal Components 1 (PC1) and 2 (PC2) from the GPA explained 77.4% and 22.6% of the total variability, respectively. In the biplot, the consensus positions of P and CxP were close at positive values of PC1, while C was located at negative values. Contrarily, P and CxP were discriminated against by PC2 because the exotic parent was on its negative values and the hybrid on the positive ones. C was located at 0 value. Nevertheless, for each genotype positions according to transcriptomic and phenomic characterizations were similar, indicating a high association among both datasets. Also, d/a for all traits evidenced a high dominance of the wild genome, which agreed to the proximity of P and CxP in the biplot.

Conclusions: The integration of transcriptomic and phenomic data by PGA and d/a estimation allows the identification of a high degree of association among expression levels of both sHSP genes and quantitative traits. This information can be applied in breeding programs to get optimal use of exotic genes obtaining new varieties with adequate fruit quality traits

ID# 073

Comprehensive Analysis of Intronless Genes: Unique Characteristics and Their Implications in Human Biology

Katia Aviña-Padilla[1], José A. Ramírez-Rafael[1], Octavio Moreno Zambada[1], Gabriel E. Herrera-Oropeza[2], Alfredo Varela-Echavarría[3], Javier de las Rivas[4], Maribel Hernández-Rosales[1]

[1]Centro de Investigación y de Estudios Avanzados del IPN, Unidad Irapuato, Guanajuato, Mexico
[2]Centre for Developmental Neurobiology, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, United Kingdom
[3]Instituto de Neurobiología, Universidad Nacional Autónoma de México, Querétaro, Mexico
[4]Centro de Investigación del Cáncer, Salamanca, España

Background: While the majority of eukaryotic genes exhibit complex structures involving exons and introns, IGs, constituting approximately 5% of human genes, stand out for their simplicity, as they lack introns, prompting scientific curiosity.

Results: Human IG-encoded proteins were notably involved in chromosome and genetic material organization, as well as displaying heightened responses to sensory stimuli in various biological processes. Remarkably, their promoter sequences exhibited significant enrichment in bHLH and Forkhead motifs compared to Multiple Exon Genes (MEGs), suggesting unique transcriptional regulation mechanisms for IGs. IGs displayed a higher GC content, indicating potential effects on gene expression stability. Moreover, they possessed longer 3' untranslated regions (UTRs) and coding sequences, hinting at intricate post-transcriptional regulation, influencing mRNA stability, microRNA interactions, and complex protein structures. Additionally, IGs exhibited genomic clustering patterns on chromosomes 4, 6, 9, 11, and 21, suggesting potential functional associations within these chromosomal regions. Furthermore, our exploration of IG expression profiles in normal tissues, leveraging the GTex database, revealed highly tissue-specific expression patterns. Brain tissue displayed distinct transcriptional profiles, while renal cortex, lung, prostate, breast, esophageal mucosa, and colon transverse tissues exhibited unique clustering patterns, indicative of specialized roles. Conversely, gastrointestinal tissues shared similar expression profiles, and bladder samples lacked specific tissue expression patterns.

Conclusions: IGs, although constituting a minority within the human genome, appear to play a significant role in genetic organization and sensory response. Their diverse tissue-specific expression patterns in normal tissues hint at their potential relevance in various physiological processes. Understanding the functions of IGs can provide valuable insights into human evolution and may offer new avenues for investigating diseases where IGs could play pivotal roles. Further research is essential to uncover the intricate mechanisms underlying IG contributions to human health and disease.

ID# 077

Bioinformatic identification of sequence variants by integrating omic data to design functional DNA markers in tomato

Valentina Goytia Bertero[1], Paolo Cacchiarelli[2], Guillermo Pratta[2], Débora Arce[1]

[1]Universidad Tecnológica Nacional. Facultad Regional San Nicolás, Buenos Aires, Argentina
[2]Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR-CONICET/UNR), Sante Fe, Argentina

Abstract

The cultivated tomato (*Solanum lycopersicum*) is one of the most agriculturally important vegetables. Due to its limited genetic variability, related wild species are often used in breeding programs to expand the available diversity. From previous experiments, we were able to obtain the genome and transcriptome of the cv. *Caimanta* (*S. lycopersicum*, C) and LA0722 (*S. pimpinellifolium*, P) progenitors. As an original contribution of this experiment, we analyzed proteins obtained through GC-MS and subsequently assigned their respective expression levels, obtained from a previous RNAseq experiment, allowing us to identify induced genes (IG) during the fruit maturation process in both genotypes (C and P).

The objective of this study was to develop molecular markers (MMs) from the detected IG, with a particular focus on Heat Shock Proteins (HSPs), a biologically significant superfamily that functions as molecular chaperones, folding other proteins and preventing their denaturation. To achieve the proposed objective, protein sequences (https://solgenomics.net/) of 9 selected chaperones (SOLYC03G113930, HSP20 - SOLYC04G082720, HSP20 - SOLYC05G014280, HSP20 - SOLYC06G076570, HSP20 - SOLYC07G006180, RIPENING REGULATED PROTEIN - SOLYC07G042250, CHAPERONINE21 - SOLYC09G011030, HSP70 - SOLYC09G075950, HSP70 - SOLYC12G056780, HSP20) were obtained. Subsequently, a BLASTn was performed on both progenitors to detect the start and end positions of the genes. 1000 bases were added to each start coordinate to extract the DNA sequences corresponding to the chaperones from the genomes of C and P.

Comparing these regions through sequence alignment using the Needle program (https://www.ebi.ac.uk/Tools/psa/emboss_needle/) revealed numerous Single Nucleotide Polymorphisms (SNPs) in the 9 analyzed genes. Developing MMs from omics data is expected to yield functional markers that can be used

ID# 079

Molecular evolution analysis workflow tutorial/for dummies

María Clara Campos[1], Francisco Pisciottano[1], David Bruque[2], Patricia Saragüeta[1]

[1]IBYME-CONICET, Argentina
[2]UTCH-SAMIC

Abstract

Creating a workflow for molecular evolution analysis can be time-consuming and requires significant resources, including bioinformatics expertise and human involvement. To develop an effective workflow, it is crucial to have experience with relevant tools.

We present a meticulously designed workflow for testing positive selection hypotheses for the target protein in the taxonomic group of interest. This workflow exclusively employs freely available tools and facilitates the assessment of conservation, as well as the identification of specific sites responsible for the selection signatures within the protein's structure.

The workflow consists of several stages: Retrieve the coding sequences of the protein of interest's orthologs from public databases for the relevant species. Using MEGAX, construct a multiple sequence alignment. Evaluate the degree of conservation and sequence identity percentages using SIAS. Reconstruct the gene phylogeny using maximum likelihood methods in MEGAX. Calculate synonymous and non-synonymous substitution rates in order to identify positive slection signatures through site-lineage-specific positive selection analysis using the codeml program from the PAML4 package implemented in multiPAML2.5, a tool created in our laboratory that automates PAML analysis.

Predict the three-dimensional structure of the protein using AlphaFold2 via a command-line interface in ColabFold, which can be executed from a local computer. Visualize and edit the top-ranked three-dimensional model using Chimera. Locate protein domains cross-referenced with specific literature, and adjust domain positions using secondary structures predicted by ColabFold and visualized with PDBsum. In the case of glycosylated proteins, also identify glycosylation sites using the UniProt database.

By following this simple workflow, researchers can easily analyze molecular evolution for their protein of interest, test evolutionary hypotheses in a given clade and eventually identify the positive selected residues involved in their protein molecular adaptation or rapid evolution.

ID# 081

Molecular dynamics to identify key residues in disordered protein complexes

Bastien, María Clara[1*]; Tavolaro, Franco Giuliano[1*]; Fernández, María Laura[2,3], Marino-Buslje, Cristina[4]
[*] equal contribution

[1]Life Sciences Department, Instituto Tecnológico de Buenos Aires (ITBA), Buenos Aires, Argentina
[2]CONICET - Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Instituto de Física Interdisciplinaria y Aplicada (INFINA), Buenos Aires, Argentina
[3]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Física, Buenos Aires, Argentina
[4]Bioinformatics Unit, Fundación Instituto Leloir (FIL)/IIBBA, Buenos Aires, Argentina

Background: Nuclear Magnetic Resonance (NMR) spectroscopy has long been a valuable tool for characterizing intrinsically disordered proteins (IDPs) molecular conformations, yet it often provides limited information about structural conformation of their complexes. These proteins are associated with several key biological processes; therefore, their malfunction could trigger diseases such as cancer, Parkinson, Alhzeimer, etc. In this study, we address the challenge of increasing the conformational information from NMR data to characterize IDPs complexes, by employing molecular dynamics simulations (MD). The NMR technique results in an ensemble of structures, whereas MD provides information on the behavior of the complexes over time, improving our understanding of their structural dynamics.

Results: Our investigation involved 9 protein-protein complexes, for each one, 2 MD simulations were performed using two different NMR conformers (complexes) as starting points. In a second stage, the results were analyzed revealing that each complex exhibits residue interactions with varying probabilities together with a wide range of conformational states. Notably, we identified core interactions within these complexes, shedding light on the key factors that influence their stability and eventually, their functionality. Furthermore, we identified 3 complexes where the proteins separated enough to lose the core interactions, more data will be needed to determine if the function associated requires these complexes to be transient or labile.

Conclusions: Our findings demonstrate that molecular dynamics simulations complement NMR data by providing a more comprehensive view of complex structures. The molecular dynamics approach refines the information obtained from NMR, yielding a more accurate count of core contacts within a protein-protein interaction over time. Moreover, our study highlights the pivotal role of specific residues in maintaining the stability and integrity of complex interactions. Overall, the combination of NMR and molecular dynamics proves to be a powerful strategy for unraveling the intricacies of biomolecular complexes.

Taken together, this study allows for a deeper understanding of the behavior and function of IDP complexes and eventually, this information could be used to explore new disease treatment strategies.

ID# 082

Evaluation of RNA-Seq assemblies of Matricaria chamomilla for the definition of a workflow in the construction of de Novo transcriptomes

Maggio J. F.[1], García L.[2,3], Costa Tártara, S.M.[1,4]

[1]Departamento de Ciencias Básicas, Universidad Nacional de Luján. Av. Constitución y Ruta Nac. N° 5 (s/n), Luján, Buenos Aires, Argentina

[2]IBAM-UNCuyo, CONICET. Alte Brown 500 Chacras de Coria, Luján de Cuyo, Mendoza, Argentina

[3]Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo. Padre Jorge Contreras 1300, Mendoza, Argentina

[4]CONICET, Argentina

Background: An organism's transcriptome represents the portion of its genome expressed at a specific moment under particular conditions and can be constructed using RNA sequences (RNA-Seq). This approach helps to generate primary gene expression information of a species, like some medicinal plants. Various tools exist to assemble short RNA reads in consensus sequences and generate de Novo transcriptomes (without a reference genome). Although the workflow for this analysis is defined, the parameters under different tools used are only sometimes reported. The work presents the analysis of different Matricaria chamomilla transcriptome constructions beginning from reads of Roche's 454 using four combinations of quality control parameters for pre-assembly processing and three assemblers (NEWBLER, SOAPDeNovo-Trans, and Trinity) that are based on two types of assembling algorithms (de Bruijn Graphs and Overlap Layout Consensus). We analyze the quality of outputs through different metrics using rnaQUAST, DETONATE and BUSCO tools.

Results: The results showed the same BUSCO pattern for every assembler transcriptome. Stricter combinations of quality values in trimming raw reads impacted generating possible losses of key sequences from the library. Completed and Single-copy genes and Completed and Duplicated genes from the Viridiplantae database were the minor classes represented compared to the Fragmented genes and Lost genes. Transcriptomes assembled with Trinity showed higher frequencies for all classes, followed by NEWBLER and the last SOAPDeNovo-Trans. The rnaQUAST indicates longer consensus sequences and greater N50 values for Trinity transcriptomes than other assemblers. RSEM-EVAL values calculated with DETONATE indicate stricter trimmed levels positively in the level of reads transformation during the transcriptome construction. These results also reflect better transcriptomes constructed using Trinity

Conclusions: Despite the small size of the library, it was a helpful input to test different pre-assembly treatments and assemblers. The output analysis with different tools allows for analyzing sequence parameters, the transformation of the library at the time of assembly and the integrity of the transcriptome. Trinity proved to be the most effective tool for assembling the library, surpassing even NEWBLER, the assembler recommended by Roche.

ID# 083

Novel bioinformatic approaches for the profiling of abundant sRNA identified by sRNA-seq

Mauro Ibañez[1], Clara Claus[2,3], Julieta Belén Grosso[2,4], Carla Borini Etichetti[2,3], Silvana Spinelli[2,4], Uciel Chorostecki[5,6]

[1]Faculty of Engineering (UNER), Entre Rios, Argentina
[2]Institute of Clinical and Experimental Immunology of Rosario, IDICER, CONICET-UNR, Rosario, Argentina
[3]Faculty of Biochemical and Pharmaceutical Sciences (UNR), Rosario, Argentina
[4]Faculty of Medicine (UNR), Rosario, Argentina
[5]Barcelona Supercomputing Center, Barcelona, Spain
[6]Institute for Research in Biomedicine, Barcelona, Spain

Abstract

Small RNAs are key players in the control of gene expression in humans and they have been extensively studied as potential new biomarkers for a wide range of diseases. Due to the advent of NGS approaches, screening of all known human biofluids and data accumulation became possible. However, while traditional bioinformatical analyses of sRNA-SEQ samples are usually focused towards the identification of miRNAs, the profiling of sRNA samples show that most of the small RNAs present correspond to other small RNAs, such as Piwi-interacting RNAs, small Nuclear RNAs, Small Nucleolar RNAs, ribosomal RNAs, transfer RNA's fragments and Y-RNAs.

We developed a bioinformatics platform to optimize the analysis of sRNA's. Our platform is highly efficient in terms of storage and computational costs, as it was optimized to easily process new samples of well-studied and novel species. It facilitates the analysis of sample composition, and contains tools for sample comparison and structural analysis of new sequences.

ID# 084

Assessment of Large Language Models for protein domain annotation

Rosario Vitale[1], Leandro Bugnon[1], Emilio Fenoy[1], Diego H. Milone[1], Georgina Stegmayer[1]

[1]Research Institute for Signals, Systems and Computational Intelligence, sinc(i), FICH-UNL, CONICET, Santa Fe, Argentina

Background: There are more than 248 million protein entries in UniProt but less than 1% of them have been annotated and classified into one of the 19,632 Pfam domains. This is a huge breach between sequencing and annotation capabilities, which exists due to the high speed of experimental data obtention and the very low and time-consuming manual curation of results. The annotation process today is done using sequence alignments and hidden Markov models (HMM). This approach has been successful for growing the Pfam annotations, however at a low rate in comparison to novel protein discovery.

Results: To tackle this issue related to the lack of annotations, we propose to use Transfer Learning (TL) from Large Language Model (LLMs). The TL approach involves self-supervised learning on large and unlabeled protein datasets to generate a numerical embedding for each sequence. This representation learned by a LLM can then be used with supervised learning on a small labeled dataset for a specific classification task, such as protein domain classification. We have compared 5 different LLMs, using 1,339,083 sequences for training several machine learning and deep learning models, and testing them in the annotation of 21,293 proteins. The baseline HMM models have a 18.10 % error rate for this task while a DL method without TL, one-hot input encodings and convolutional layers has a 12.20 % error rate. With TL, an ensemble of multilayer perceptrons can achieve a 15.08 % error rate, while a simple k-nearest neighbors model can have, in the best case, 8.63 % error. When TL is combined with a DL model, the error rate can be as low as 7.23 %.

Conclusions: In this work we have tested cutting-edge transfer learning techniques together with deep learning to improve the actual prediction of protein domain annotations. The results obtained indicate that this approach has effective predictive advantages over existing methods and it could become part of future PFam annotation tools.

ID# 085

Integration of Genomic Data Unveils Potential PGPT Activities in Antarctic Endophytic Bacteria

Orlowski, J.[1,2]; Massot, F.[1,2,3]; Basile, C.[1,2,3]; Ruberto, L.[1,2,3]

[1]NANOBIOTEC-CONICET, Argentina
[2]FFyB-UBA, Argentina
[3]IAA, Argentina

Abstract

In the context of an Antarctic soil bioremediation project targeting hydrocarboncontaminated soils, we isolated endophytic bacteria from one of the two indigenous plant species in the continent, *Deschampsia antarctica*. This plant species will play a crucial role in the setting-up of ecopiles for onsite soil remediation.

The isolated bacteria were subjected to Illumina sequencing, followed by genome assembly and gene/protein annotation. This study focuses on six isolates belonging to the Agreia genus, considering its high abundance in culture-dependent methods. By combining these isolated genomes with existing database entries, we constructed the Agreia pangenome, encompassing the gene pool of this genus. Subsequently, a bioprospecting and genome-mining effort was conducted to identify genes associated with plant growth-promoting traits (PGPT-proteins) activities, potentially enhancing the growth of these two plant species.

In the pursuit of bioremediation, understanding the genetic underpinnings of beneficial plant-microbe interactions is pivotal. The endophytic bacteria within indigenous Antarctic plants offer unique insight into their adaptive capabilities in extreme conditions. Through Illumina sequencing, we acquired comprehensive genomic data from the isolated bacteria. Following assembly and annotation, we directed our attention to six Agreia genus isolates. By amalgamating these genomes with public databases, we constructed the Agreia pangenome, a collection of genes characterizing this genus. Subsequent genome-mining efforts identified a plethora of genes potentially associated with PGPT activities, pivotal for enhancing the settle and growth of the target plant species in the remediation process.

This study unearths a wealth of potential PGPT-associated genes within the genomes of endophytic bacteria. By constructing the Agreia pangenome and pinpointing genes linked to beneficial activities, we shed light on the mechanisms that could enhance the growth of the indigenous Antarctic plant species involved in soil bioremediation efforts. The combination of culture-dependent methods with genomic studies has become crucial for the strategy of defining plant-microorganism combinations in applications within the field of environmental biotechnology, among others. These findings hold significance for the success of the bioremediation project and provide a stepping stone for further research into the genetic factors driving plant-microbe interactions in extreme environments.

ID# 086

Bioinformatics Analysis Reveals Potential Microorganisms and Enzymes for Microplastic Remediation in Antarctic Waters

Orlowski, J.[1,2], Bernard, N.[1,2,3], Cuba, F.[4], Poklepovich, T.[4], Campos, J.[4], Ruberto, L.[1,2,3]

[1]NANOBIOTEC-CONICET, Argentina
[2]FFyB-UBA, Argentina
[3]IAA, Argentina
[4]UOCNGyB ANLIS – Malbrán, Argentina

Abstract

The global issue of microplastic pollution necessitates urgent remediation measures. This study focused on bioremediation of microplastic-contaminated waters, specifically targeting the Antarctic region. Samples of plastic exposed to environmental conditions in seawater were collected, and metagenomic analysis was performed on cells recovered from these samples, DNA was extracted and sequenced using Illumina technology. The resulting reads were assembled into contigs and bins, yielding eight high-quality bins.

Gene annotation and proteomic analysis were conducted, followed by bioprospecting and genome-mining using as query The Plastics-Active Enzymes Database. The presence of PET-degrading enzymes (PETases) was investigated, along with other polymers such as Polyethylene terephthalate (PET), Polyurethane (PUR), Polyethylene (PE), Polyamide (PA), Polystyrene (PS), Polyvinylchloride (PVC), Polypropylene (PP), Other types of polymers, Polymers from mainly renewable resources, Polylactic acid (PLA), Polyhydroxyalkanoates (PHA), Polybutylene adipate terephthalate (PBAT), Polybutylene succinate (PBS), and Natural rubber (NR).

Using protein BLAST analysis against a database of petroleum-derived polymer-degrading proteins, 138 significant matches were obtained. Further motif and domain analysis narrowed down the results to 35 enzymes with hydrolytic activity and potential for degrading petroleum-derived polymers. Signal peptide searches were also conducted to identify 5 potentially exoenzymes capable of extracellular degradation since bacteria cannot incorporate microplastics into their cytoplasm.

Taxonomic analysis revealed the presence of Helicobacter, Flavobacteriaceae, Rhodobacteraceae, and other unidentified microorganisms. This study highlights the potential of Antarctic waters as a source of microorganisms and enzymes with the ability to degrade microplastics. The identification of PETases and other polymer-degrading enzymes opens up possibilities for the development of biotechnological solutions for microplastic pollution remediation. Further research is needed to explore the full potential of these microorganisms and enzymes in practical applications.

ID# 091

Genomics regions related to genic male sterility in globe artichoke

Zayas, Aldana[1]; Martina, Matteo[2]; Acquadro, Alberto[2]; Polli, María Francesca[2]; Comino, Cinzia[2]; Portis, Ezio[2]; Martin, Eugenia[3]

[1]Instituto de Biología Molecular y Celular de Rosario (IBR-CONICET), Rosario, SF, Argentina
[2]DISAFA, Plant Genetics and Breeding, University of Torino, Italy
[3]Instituto de Ciencias Agrarias de Rosario (IICAR-CONICET-UNR), Zavalla, SF, Argentina

Abstract

The identification and characterization of male-sterility genes in several plant species provides a valuable tool for crop hybrid breeding. In globe artichoke (*C. cardunculus var. scolymus* L, allogamous, 2n=2X=34), F1 hybrids have been successfully introduced in the market based on the exploitation of male-sterility but the genetic and molecular mechanism is still unclear. In this context, the aim of this study was identify genomic regions carrying genes affecting male sterility using bulked-segregant analysis coupled to DNA sequencing (BSA-seq). An F2 population, derived from an initial cross between a male sterile (MS) globe artichoke and a male fertile (MF) cultivated cardoon, was phenotypic characterized for pollen sterility. A 3:1 ratio was identified on the population, fitting a monogenic Mendelian segregation model, where the homocigous recesive “msms” are male sterile plants. Genomic DNA from 15 MS and 15 MF F2 plants was extracted, bulked in two groups (MS and MF), and sequenced by Illumina technology, as well as the two parental genotypes.

The QTLseq analysis reveal four genomic regions related to male sterility, mapped on the chromosomes 4, 12, 14 (2 peaks), containing several genes potentially involved in male sterility. On chromosome 4, sixty genes were found within the peak region. On chromosome 12, surrounding the peak were identified eighty-two genes. Finally, on chromosome 14, at the first peak were found fifty-five genes, whereas around the second peak a total of forty-three genes were identified. In summary, our results provide a first list of candidate genes that might play a pivotal role in the mechanisms leading to male sterility in globe artichoke.

ID# 095

Exploring MicroRNAs in S. frugiperda Larvae Infected with SfMNPV through miRNA-seq

Gómez Bergna, S.M.[1]; Tongiani SE.[1], Salvador R.[2], Romanowsk,i V.[1], Pidre, M.L.[1], Ferrelli, M.L.[1]

[1]Instituto de Biotecnología y Biología Molecular (IBBM, UNLP-CONICET), Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Buenos Aires, Argentina

[2]Instituto de Microbiología y Zoología Agrícola (IMyZA), Centro de Investigaciones en Ciencias Agronómicas y Veterinarias (CICVyA), Instituto Nacional de Tecnología Agropecuaria (INTA), Nicolás Repetto y de los Reseros s/n, Hurlingham, Buenos Aires, Argentina

Background: *Spodoptera frugiperda* Multiple Nucleopolyhedrovirus (SfMNPV) is a baculovirus that infects the larval stage of the moth *Spodoptera frugiperda*, commonly known as the fall armyworm. *S. frugiperda* is a very important agronomic pest of maize and other crops and SfMNPV a good candidate as bioinsecticide.

MicroRNAs are small non-coding RNAs that regulate gene expression in a sequence specific manner and are certainly involved in the host-pathogen interaction. In order to better understand the cross-talk regulation between the SfMNPV and its host, we set out to identify novel miRNAS expressed by SfMNPV and *S. frugiperda* larvae in the context of infection by means of small RNA-seq analysis.

Results: Larvae of *S. frugiperda* were infected with the Argentinean SfMNPV M isolate and total RNA was extracted 48 hours post-infection. High-quality RNA samples, along with RNA from uninfected control larvae, served as inputs for small RNA sequencing using Illumina technology, performed by Novogen Inc®. The resulting data was subjected to quality control and further trimming to eliminate adapter sequences, low quality reads, and short reads. Subsequently, the reads were collapsed and mapped to the reference genomes, resulting in 0.39% of the reads mapped to SfMNPV genome and 54.71% mapped to the *S. frugiperda* genome. This data was then processed using miRDeep2 to identify existing and putative novel microRNAs encoded by the host and the virus. As a result, we could identify *S. frugiperda* miRNAs already deposited in miRBase or reported by other authors, but also detected several putative novel miRNAs expressed by this insect. These miRNAs shared seed sequences with miRNAs from other species. Finally, we could also identify putative miRNAs coded by SfMNPV.

Conclusions: We successfully identified the majority of previously reported microRNAs and uncovered several putative novel microRNAs in *S. frugiperda*. Additionally, we detected novel putative miRNAs in SfMNPV. These findings significantly contribute to our understanding of the response of *S. frugiperda* to SfMNPV infection.

ID# 096

RIPK1 as a potential therapeutic target against Diffuse Gliomas

Leslie C. Amorós Morales[1]; Santiago M. Gómez Bergna[1]; Abril Marchesini[1]; Ma. Luján Scalise[1]; Nazareno González[2]; Marianela Candolfi[2]; Víctor Romanowski[1]; Matías L. Pidre[1]

[1]Instituto de Biotecnología y Biología Molecular (IBBM-CONICET-UNLP), Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina

[2]Instituto de Investigaciones Biomédicas (INBIOMED-CONICET-UBA), Facultad de Ciencias Médicas, Universidad de Buenos Aires, Buenos Aires, Argentina

Background: Diffuse gliomas (DG) are aggressive brain tumors that can be divided based on the presence of a mutation in the enzyme isocitrate dehydrogenase gene (IDH1). In general, mutated IDH1 (mIDH) correlates with a better prognosis compared with wild type IDH1 (wtIDH1). RIPK1 (receptor-interacting protein kinase 1) is an enzyme involved in several signaling pathways. Although the dysregulation of RIPK1 activity has been linked to various diseases, its involvement in glioma needs further investigation. The aim of this work was to characterize the role of RIPK1 in tumor progression associated pathways of DG through in silico analyzes from patient databases.

Results: We analyzed public datasets containing clinical, genomic and transcriptomic information from 661 patient samples. For this, two bioinformatic platforms were employed (cBioPortal and Xena). Samples corresponding to DG were filtered, classified by IDH status and divided into two subgroups according to the median value of RIPK1 expression, as required. Clinical and molecular attributes were then evaluated in conditions of high and low mRNA RIPK1 levels. The results showed a lower survival probability in patients belonging to the high RIPK1 expression subgroup compared to those samples with low RIPK1 expression. We also observed a higher expression of RIPK1 in wtIDH samples compared to those with mIDH. In order to further characterize the role of RIPK1 in DG, we performed a Gene Ontology and Pathway Enrichment Analysis using the Xena platform's differential expression tool. The results showed that RIPK1 is involved in inflammatory and immune responses. Hence, the expression levels of some of the genes involved in the following molecular processes crucial for cancer progression were studied: proliferation, epithelial-mesenchymal transition, immune cell infiltration and cell death pathways. Briefly, the results showed significant differences in genes related to increased cellular dedifferentiation, proinflammatory cell death pathways and tumor infiltrating immune cells gene signatures (Welch's t-test).

Conclusions: RIPK1 over-expression is associated with a poor prognosis in DG. This fact, together with our in-silico results suggest that RIPK1 may play a crucial role for glioma progression, positioning it as a promising candidate for therapeutic targeting.

ID# 097

The inhibitor of apoptosis BIRC6 and its role in non-small lung cancer

Abril Marchesini[1], Santiago Gómez Bergna[1], Leslie C. Amorós Morales[1], María Florencia Gottardo[2], Víctor Romanowski[1], Matías L. Pidre[1]

[1]Laboratorio de Virología Molecular, Instituto de Biotecnología y Biología Molecular (IBBM, UNLP-CONICET), Argentina
[2]Laboratorio de Oncología Molecular, Universidad Nacional de Quilmes (UNQ), Argentina

Background: Lung cancer is the main cause of cancer-related deaths worldwide, comprising two main morphological groups: small cell lung cancer (SCLC, 15% of cases) and non-small cell lung cancer (NSCLC, 85% of cases). Inhibitors of apoptosis (IAP) have been identified as central players in the development and aggressiveness of various tumors. Tumors resistant to conventional treatments, like chemo and radiotherapy, could take advantage of pathways such as the DNA damage response (DDR). The aim of this work was to characterize the role of the IAP family in NSCLC, especially in lung adenocarcinoma (LAC) and lung squamous cell carcinoma (LSCC). With the goal of proposing a new possible therapeutic target.

Results: Two TCGA transcriptomic databases were analyzed and seven IAP were queried (cBioPortal and Xena platform). Our results demonstrated that at least two (BIRC5 and BIRC6) of the seven IAP have a higher expression in tumor compared to normal tissue in LAC and LSCC (ANOVA). Also, our results showed that in LAC patients with a higher BIRC6 copy number is associated with resistance to radiotherapy and tumor recurrence ($\chi 2$). In order to further characterize the role of BIRC6, we run a Gene Ontology (GO) and Pathway Enrichment Analysis using the Xena platform's differential expression tool. We compared the results with those obtained for BIRC5, XIAP and NAIP. These results show that BIRC6 could be involved in DDR, specifically via the ATM and ATR pathways. To go deeper into these findings, we analyzed the expression levels of some of the genes involved in the ATM and ATR pathways. As result we observed a higher expression in most of the genes involved in both pathways, in tumor samples that present a high expression of BIRC6. On the other hand, only most of the ATR genes were up-regulated in samples with high BIRC5 expression (Welch's t-test).

Conclusions: BIRC6 over-expression is known to be associated with poor prognosis in different tumors. This fact, together with our results, are encouraging and open the way to future preclinical studies, postulating BIRC6 as a promising therapeutic target.

ID# 101

Stitched together: in silico optimization of a potent neutralizing anti-malarial antibody

Martina Wecera[1,2], Gordon Dale[1], Facundo Batista[1]

[1]Ragon Institute of MGH, MIT and Harvard
[2]Universidad Nacional de La Plata (UNLP), Argentina

Abstract

Malaria is a tropical infectious disease caused by the parasite Plasmodium falciparum. Globally, malaria is estimated to have a disease burden of 241 million cases and ~627,000 deaths, primarily in children, in 2020. For more than 40 years, work has been conducted on generating an effective and durable vaccine. Despite the recent approval of the first malaria vaccine (RTS, S), lasting protection remains low, with protective antibody titers found for less than 1 year. Thus, complementary approaches are necessary to reduce disease burden.

One such approach is the use of passive immunization. Indeed, clinical trials using the potent neutralizing antibody CIS43 have demonstrated significant protection for up to one year in malaria endemic regions. Although this is a promising strategy, it is limited by cost-effectiveness due to the relatively large quantity of antibody that needs to be transferred. In order to address this, the Batista Lab and colleagues, recently developed a CIS43 variant, D3, that has a significant increase in affinity and protection in preclinical models of malaria infection. This is critical because increases in affinity allow for a decreased dose and cost for equivalent protection, enabling passive immunization to be a viable strategy to prevent malaria infections.

In this work, we sought to apply rational design to further improve this antibody and in doing so achieve two goals: (1) increase the affinity of CIS43 to its cognate antigen, and (2) provide proof of principle for rational in silico antibody design. Using a novel workflow, we designed several CIS43 variants. These variants were cloned, expressed, and purified. Using endpoint ELISAs we, screened promising candidates for further interrogation and affinity measurement using biolayer interferometry. As a result of this work, we produced several candidate monoclonal antibodies that have greater affinity than D3, including a new best-in-class antibody, Frankie.

ID# 104

Bioinformatics exploration of LTR retrotransposons and in silico development of inter primer binding site primers for evaluation of genetic variability of white carob somaclones (Neltuma alba (Griseb.) C.E. Hughes & G.P. Lewis)

Pablo Vélez[1], Diego Lisa[2], Lorena Palacio[2]

[1]Molecular Biology Unit. Ceprocor. Ministry of Science and Technology of the Province of Córdoba
[2]Plant Genetic Resources Unit. Ceprocor. Ministry of Science and Technology of the Province of Córdoba

Background: *Neltuma alba* (white carob) is a native tree species of significant forestry interest in north-central Argentina. The study focuses on the micropropagation of white carob, which offers a valuable method for multiplying selected genetic material. Analyzing genetic variability between obtained clones is crucial in this process. Retrotransposons, elements capable of moving within the genome, are useful markers for studying genomic variations in somaclonal cultures. Among various techniques, the iPBS (PCR amplification between primer binding sites) method is chosen for its universality, reliability, and efficiency. This technique involves PCR reactions using oligonucleotides flanking the primer binding site (PBS) of the LTRs (Long Terminal Repeats) of retrotransposons.

Results: We employed two bioinformatics tools, LTR Detector and LTR Harvest, to identify retrotransposons within the white carob genome. LTR Detector located 13,552 retrotransposons, while LTR Harvest found 2,879. By comparing the outputs of both tools, 1,088 matches were found with a deviation of +/- 50 nt at the 3' end of the 5' LTR portion. From these matches, 1,995 sequences were extracted, yielding 860 unique sequences of 12 nt and 1,094 of 18 nt. In silico PCR was carried out to evaluate the frequency of sequences found by the tools. Interestingly, a greater number of in silico fragments did not necessarily come from the most frequent sequences detected by LTR Detector/LTR Harvest, on the contrary many low frequency sequences showed a significant number of fragments. In vitro tests were conducted with the synthesis of seven oligonucleotides. Sequences with higher frequencies in detections with LTR Detector and LTR Harvest exhibited a greater number of bands, validating the effectiveness of these oligonucleotides for iPBS PCR.

Conclusions: The bioinformatic exploration of LTR retrotransposons using LTR Detector and LTR Harvest successfully provided valuable oligonucleotides for applying the iPBS technique to the white carob genome. These informative markers will be instrumental in evaluating genetic variability in somaclones of this species. The study showcases the potential of micropropagation combined with molecular techniques to advance the forestry interests in north-central Argentina, specifically in the context of *Neltuma alba*.

ID# 106

Enhancing Compound Similarity Prediction: A Novel Approach

E. Borzone[1], L. Di Persia[1], M. Gerard[1]

[1]Research Institute for Signals, Systems and Computational Intelligence, sinc(i), FICH/UNL-CONICET, (3000) SF, Argentina

Background: Predicting similarity between compounds is a challenge when the structure of the compounds involved is unknown. To address this issue, we have proposed a model based on graph neural networks (GNNs) to generate embeddings that capture compound relationships. Despite the good results obtained up to now, our analyses have shown that embeddings are not clustered according to the similarity of the compounds they represent.

Results: In order to improve the embeddings, we introduced a modification in the cost function We expect that by encouraging the formation of more well-defined and cohesive clusters within the latent space, we could establish a robust negative correlation between spatial distance and compound similarity. Preliminary results provide compelling evidence of the effectiveness of our proposed modification. The embeddings produced by our modified cost function now exhibit significantly improved clustering, resulting in spatial similarity that closely aligns with compound similarity, expecting an improvement for compounds with unknown structure due to the better correlation between the measured distance in embedding space and the similarity index used (Tanimoto index). This crucial enhancement has translated into a boost in the accuracy of similarity predictions, with an increase of 0.54%. The tighter clustering and enhanced spatial organization suggest that our approach effectively addresses the limitations encountered in previous models. Furthermore, a notable challenge tackled in this study pertains to predicting similarity among compounds lacking known structural information. To assess the practical implications of our model, we performed an analysis by plotting the distances in the embedding space against the Tanimoto index. In this analysis, we aimed to observe the presence of a negative correlation, as enhanced embeddings should ideally exhibit this relationship. The results clearly demonstrate substantial improvements compared to previous methodologies. This breakthrough represents a major step forward in our ability to predict the similarity, even when dealing with compounds lacking known structures. These findings underscore the potential utility of our model in diverse applications across various domains, including drug discovery and metabolic pathway analysis.

Conclusions: Our study marks a significant leap forward in the prediction of compound similarity within metabolic pathways. Leveraging Graph Neural Networks (GNNs) and introducing an innovative embedding modification, we have achieved substantial progress. Our modified cost function has substantially boosted the accuracy of compound similarity predictions, resulting in more coherent and interpretable compound representations. This breakthrough not only promises a deeper understanding of metabolic pathways but also holds great potential for practical applications. Looking ahead, our future work will focus on expanding the scope of our research. Incorporating additional metabolic pathways into our dataset will enrich our model's ability to handle diverse biochemical contexts. Furthermore, we plan to explore the incorporation of edge features to capture finer nuances in compound interactions. These steps will not only enhance the comprehensiveness of our model but also open doors to a broader range of applications, reinforcing our commitment to advancing the field of compound similarity prediction within metabolic pathways.

ID# 107

Accurate Detection of RNA Modifications on Nanopore Sequencing Data at Signal-level with Machine Learning

Servi, L.[1,2]; Petrillo, E.[1,2]; Stegmayer G.[3]

[1]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Fisiología, Biología, Molecular, y Celular, Buenos Aires, Argentina

[2]CONICET-Universidad de Buenos Aires, Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE), Buenos Aires, Argentina

[3]Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional, sinc(i), FICH-UNL/CONICET, Argentina

Background: The accurate identification of RNA modifications is essential for advancing our understanding of gene regulation and it's the functional implications of RNA modifications in biological and medical research. In this work study, we address the challenge of precise RNA modification detection by leveraging Oxford Nanopore sequencing data. RNA modifications play a pivotal role in regulating gene expression and post-transcriptional processes, making their accurate identification crucial in various fields of biology and medicine. The emergence of Oxford Nanopore sequencing technology has opened new avenues for RNA modification detection. However, the noisy data generated by nanopore sequencing poses a significant challenge for accurate RNA modification detection. Most available methods are very simple and look only at the reading errors as indicators of, for example, RNA methylation, with low accuracy and not being able to distinguish the exact modification.

Results: Our approach focuses on the development of a Machine Learning (ML) model based on Natural language Processing (NLP) models, harnessing the electrical current generated during sequencing and taking advantage of basecalling errors as fundamental features. For training, we will employ synthetic aptamers with modifications mimicking those found in naturally occurring RNA in addition to me-RIPseq combined with direct RNA sequencing (DRS) of plants in the same conditions. Through a rigorous training process, this model will be fine-tuned to detect with high accuracy modified RNA bases. We propose to use synthetic aptamers with known modifications that closely resemble those present in natural RNA in addition to a vast repertoire of public data.

Conclusions: Our study presents a novel and deeper approach to the precise detection of RNA modifications using a ML model based on NLP and Oxford Nanopore sequencing data. The potential implications of this research are far-reaching, as it provides a deeper understanding of genetic regulation and the functional consequences of RNA modifications. The accurate identification of RNA modifications will undoubtedly enhance our knowledge of gene regulation and open up new avenues for biotechnological applications and crop improvement.

ID# 108

Role of the intra-tumor microbiome in the non-small cell lung cancer immune microenvironment through a multi meta-omics analysis in Chilean patients

Ivania Valdés[1,2], Alberto J. Martin[3,4], J. Eduardo Martínez[5], Erick Riquelme[2]

[1]Universidad Mayor, Facultad de Ciencias, Vicerrectoría de investigación, Center for Integrative Biology, Santiago, Chile
[2]Pontificia Universidad Católica de Chile, Departamento de Enfermedades Respiratorias, Facultad de Medicina, Santiago, Chile
[3]Fundación Ciencia & Vida, Centro Científico y Tecnológico de Excelencia Ciencia & Vida, Laboratorio de Redes Biológicas, Santiago, Chile
[4]Universidad San Sebastián, Facultad de Ingeniería, Arquitectura y Diseño, Escuela de Ingeniería, Santiago, Chile
[5]Corporación CGNA, Temuco, Chile

Abstract

Lung cancer is the leading cause of death by cancer in Chile with a 5-years survival rate of 19.4% after the diagnosis. Research on the role of the tumor microbiome in cancer has shown that the microbiome can affect cancer cells and modulate cancer immunosurveillance. However, the underlying functional mechanisms have not been clearly defined. Here, we describe how the tumor microbiota could modulate the host's immune response to lung cancer. Using 16S rRNA sequencing, metaproteomics, and metabolomics, we characterized tumor and non-tumoral tissue microbiome of lung adenocarcinoma of Chilean patients using fresh-frozen and formalin-fixed paraffin-embedded samples. We observed an intra-tumoral microbiome enriched in members from class *Gammaproteobacteria, Flavobacteriia, and Actinobacteria*, related to drug resistance, and pathogenicity. Moreover, a set of intra-tumoral bacterial peptides were found enriched in tumor samples related to Major facilitator superfamily (MFS) transporters, and Glutathione S-transferase N-terminal domaincontaining protein, both related to transport activity. Furthermore, tumor-enriched bacterial metabolites were found such as guanosine, inosine, and adenosine, all of which were related to purine metabolism. In the host, we found tumor-enriched proteinsto be down-regulated, such as T−cell surface glycoprotein CD8 alphachain related to anticancer immune response, and up-expressed proteins such as macrophage migration inhibition factor (MIF), and eIF-2-alpha kinase GCN2 known to stimulate cell proliferation and drug resistance. Furthermore, metabolites from the riboflavin and glutathione metabolism were enriched in tumor samples. Using different omics and bioinformatic analyses, our results suggest a characteristic intra-tumoral microbiome that could be involved in tumor cell resistance through glutathione transport that allows cancer cells to cope with hypoxia and drugs, and tumor cell proliferation through modulation of the immune response by purine metabolism, and riboflavin transport through MFS to cancer cells. Increased riboflavin metabolism could also play a role as a substrate for auxotrophic and pathogenic tumoral bacteria for their outgrowth. Finally, the up-regulation of MIF could inhibit T cell activation, giving an immunosuppressive state. Consistently, by combining these analyzes, we provide an overview of the potential interactions between tumor microbiome and host immunity and how their crosstalk could synergistically favor a local immunosuppressive microenvironment with a poorer outcome for patients.

Funding information: This work was supported by FONDECYT1231629 and ANID-Subdirección de capital humano/Doctorado Nacional/2023/21231386.

ID# 110

Multiple resistance in a poly-extremophilic Nesterenkonia strain isolated from High-altitude Andean Lakes

Natalia N. Alvarado[1], Virginia H. Albarracín[1], Nicolás Palopoli[2]

[1]Centro Integral de Microscopía Electrónica (CIME), CCT-CONICET, UNT, Tucumán, Argentina
[2]Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes - CONICET, Bernal, Buenos Aires, Argentina

Background: Nesterenkonia is a genus of halophilic Actinobacteria, with its most representative microorganisms isolated from extreme environments.We are particularly interested in the strain Nesterenkonia Act 20 (Act20), which thrives in the extreme conditions of the high-altitude Andean Lakes (HAAL) in the central Andes of South America, at altitudes over 3,500 meters above sea level, where the highest UV-B radiation in the world is received. Microbiodiversity develops there with unique mechanisms of adaptation to hypersalinity, desiccation, high UV radiation and the presence of heavy metals. In previous works we demonstrated that the Act20 strain exhibits resistance profiles to heavy metals and UV-B radiation superior to those of Nesterenkonia halotolerans (NH), a reference strain from high salinity soils in China.

Here, we compare the computational genome annotations of Act20 and NH between them and with those of other reference Nesterenkonia strains. This comparison aims to identify candidate genes that can explain multi-resistance to extreme conditions in Nesterenkonia Act20.

Results: In our search of public genome databases at NCBI, we retrieved data from a total of 207 Nesterenkonia strains. Comparative gene analysis revealed that the ectoine biosynthesis system, which includes the proteins ect A, B, C, and D that are crucial for osmoprotection and UV-B absorption, is conserved in Act20. We also found that resistance to heavy metals could be linked to the presence of 8 conserved genes that are in proximity in the genome, and encode putative proteins copB; copZ; an EamA family transporter; csoR; copC, and copD. We analyzed the conservation across Nesterenkonia strains of thedomains on every one of these proteins using the Interpro database which demonstrated a conserved domain architecture their proximity into the genome.

Phylogenetic analysis of these proteins in 19 Nesterenkonia strains that have all 4 conserved proteins related to ectoine biosynthesis and the 8 conserved proteins associated with heavy metal resistance, reveals a close relationship between Act 20 and NH. These strains also exhibit proximity to Nesterenkonia lutea, with all of them sharing a similar relationship with N. jeotgali and N. sandarakina. These relationships are conserved in all protein-based phylogenetic trees and the resulting consensus tree.

Conclusions: Our results support the notion that Nesterenkonia Act20 possesses dedicated genes to survive in extreme conditions, a fact that is further supported by prior in vitro resistance screening experiments conducted in our laboratory. Despite the geographical distances, Act 20 and its closer strains exhibit environmental similarities in their respective isolation environments that suggest similar mechanisms evolved to sustain life in extreme conditions.

ID# 111

The ABC Transporter Classification Quest

Erika Ripani[1*], Cecilia Barbieri[1*], Fernando Villarreal[2*], Mariano Vero[1], Arjen ten Have[2]
* equal contribution

1INBIOTEC-QUIAMM, Argentina
2IIB-UNMdP-CONICET, Argentina

Background: In recent years we have developed HMMERCTTER, a tool for the clustering and classification of protein superfamilies. HMMERCTTER outperfoms Panther in terms of precision and recall and shows less over-partitioning. A remaining important problem is that it can't cluster protein superfamilies with different domain architectures. We study the superfamily of ABC transporters in eukaryotes in order to understand its evolution and find methods for reliable classificaction of multidomain proteins.

ABC transporters are membrane pumps, infamous for causing antibiotic and pesticide resistance as well as problems in chemotherapy. These ATP activated channels consist of two modules that each combine a transmembrane region (TMR) paired with an ATPase or ATP Binding Cassette (ABC). The family forms a hard case for clustering and classification since: 1 ABC transporters can occur as a single-chain polypeptide but also as either a hetero- or a homodimer; 2 Modules can have one of two architectures, ABC-TMR or TMR-ABC; 3 Homology does not necessarily follow architecture; 4 The TMRs are low complexity regions with poor discriminative power in similarity-based methods. There is indeed no consensus on how this superfamily evolved nor is there a consensus on classification We study the superfamily with interest in biology and bioinformatics.

Results: We performed datamining on five different taxon-sets: 1 Mammals; 2 Animals; 3: Plants; 4 Fungi; and 5 Other eukaryotes. Mining was supported by manual clustering, basically how protein families were studied 20 years ago. Sequence scrutiny by Seqrutinator was hampered by the different domain architectures as well as the broad taxonomic sampling. We developed a scrutiny based on dynamic pattern development using conservation as a guide resulting in taxon specific filters. Specific HMMER profiles were then used to classify sequences according to HMMERCTTERs 100% P&R principle.

Conclusions: ABC transporters are hard to classify and form a good case for HMMERCTTER multidomain algorithm development. TMR regions may require the use of additional dedicated substitution matrices.

ID# 113

Modelling a closed connexin hemichannel

Felipe Villanelo[1], Yasna Barrera[2], Joaquín Jensen-Flores[3], Fabian Parra[4], Tomas Perez-Acle[3]

1 Universidad San Sebastian, Chile
[2]FCV, Chile
[3]USS-FCV, Chile
[4]FCV-UNAB, Chile

Abstract

Connexin are proteins that form two types of channels: gap-junction channels (GJC) and hemichannels (HC). While the first are important to cell-to-cell communication because they connect the cytoplasm of two adjacent cells; the latter can provide a dangerous communication between cytoplasm and the external millieu. HCs have a wide pore with low selectivity, that allows the passive diffusion of solutes, from ions through small peptides, including ATP molecules. Open HCs can lead to apoptosis in several cases, and some researchers say that the opening of connexin HC are involved in every inflammatory process. That is the reason why HC are closed in physiological conditions, which are mediated by factors like external calcium concentration, membrane electrical potential, pH, among others. Sadly, the structure of a closed HC has remained elusive for the research community.

Most HC structure obtained by x-ray diffraction or cryo-electron microscopy are in an open conformation. In our research, we have been using molecular dynamics simulation to obtain a closed structure for Cx26-HC, mainly based on the published evidence regarding N-terminal helix (NTH) and parahelix (PH) and their role on different types of gating. We have used non-equilibrium simulation techniques like metadynamics and umbrella sampling to force the Cx26-HC to close, and then assess the thermodynamic of the process. Our results indicate that the forced-closed Cx26-HC are very unstable, except for one meta-stable conformation with a semi-closed pore.

Finally, there is a network of interactions between residues around NTH and PH which are key to maintain the stability of the open and close conformation. We expect that our ongoing research can help to understand the structure-function relationship of Cx26-HC.

ID# 114

Proteomic study of the Oclussion Bodies of Spodoptera frugiperda Granulovirus

Tomás Masson[1], Matías L. Pidre[1], Ricardo Salvador[1], Víctor Romanowski[1], M. Leticia Ferrelli[1]

[1]Instituto de Biotecología y Biología Molecular (IBBM), Facultad de Ciencias Exactas, Universidad Nacional de La Plata – CONICET, Argentina

Background: Spodoptera frugiperda Granulovirus (SpfrGV) infects the fall armyworm Spodoptera frugiperda, an important pest for maize and other crops. SpfrGV belongs to the Baculoviridae family, an atractive group of insect viruses with potential applications in biological pest control. Baculoviruses produce two virion morphotypes, being the Occlussion Derived Virus (ODV) the responsible for primary infection within the larval midgut. These virions are found in the environment embedded in a proteinaceous matrix that forms occlusion bodies (OBs), a phenotype that confers protection from adverse environmental conditions. As OBs play a key role in horizontal transmission, we used mass spectrometry (MS) to investigate the protein composition of the OBs of an Argentinean isolate of SpfrGV and analyzed the detected proteins.

Results: The peptides produced by MS were succesfully mapped to the theoretical proteome of SpfrGV ARG. This permitted the identification of 72 proteins within the OB, constituting 47% of the theoretical proteome encoded in the virus genome. Twenty-eight proteins can be categorized within the essential 38 core gene group of Baculoviridae. Among the 44 remaining proteins, several correspond to conserved structural and accessory proteins commonly found in baculovirus. However, there are at least 18 of them with no known associated function, including ORF 28 and ORF 40, that are unique to SpfrGV. In order to find more information on the proteins associated with SpfrGV OB we used ColabFold to predict all 72 protein structures. Upon visual examination, we observed some proteins with apparent similarity. In order to asses structural similarity among SpfrGV OB proteins we used DALI allagainst-all comparison and found that 6 proteins shared domain similarities. Interestingly, 3 of this group are well characterized baculovirus proteins.

Conclusions: The identification of 72 proteins within SpfrGV OBs, along with novel insights from structural comparisons, enhances our understanding of these OBs and associated proteins. This knowledge has implications for virus transmission dynamics and for the development of biological pest control strategies.

ID# 115

Exploratory analysis of omics data to understand heterosis in tomato

Mariano Vasulka[1], Fernando Carrari[1], Nicolas Bellora[2], Luisa Fernanda Bermudez[3]

[1]IFIBYNE, Argentina
[2]INTECNUS
[3]INTA, Argentina

Background: Heterosis is a complex trait well known, although the underlying mechanism is not well understood yet. Gene models as Dominance, Overdominance, Pseudo-overdominance, and/or Epistasis could explain hybrid vigor. Gene product dosage can determine a heterotic trait. Protein level shows transgressive modification in hybrids also sugar content. Goff hypothesized that hybrids could select more favorable alleles, redirecting energy flow to biomass/productivity. siRNA accumulation correlates positively with methylation levels. Different developmental stages in plant can have heterotic behavior. Making metabolic state a very important contribution resulting very important a study contemplating fruit development stages, expression patterns, metabolic, proteomic, and methylation levels.

Results: PCA of genotypes and their hybrids of data at two different fruit developmental stages. Barplot of Cytocine methylation level for differential methylated cytocines between at least two parental lines surrounding Solyc08g076360 gene. The datasets shows that all genotypes are different. Context dependent cytocine methilation shows differences between almost all parental lines in CpG and CHH contexts.

Conclusions: The datasets are capable to study heterosis. Differential methylation levels between parental genotypes explains intrinsic, genotype dependent methylation levels, that may explain vitamin E different content between genotypes, an interesting metabolic trait for human diet. Summarizing it is important to explore the selected dataset. We further investigate near Solyc08g076360 gene to show natural variability in methylation levels, that corresponds with tochopherol content, and transcription levels diversity. This work can shed light to the heterosis complex heritability, improving 'know-how' to breeders, saving time and resources from basic science.

ID# 116

Molecular dynamics analysis of the electroporation process in normal and tumor membrane models exposed to high electric fields

Paulo Ojeda[1], Cecilia Suárez[2,3], María Laura Fernández[2,3]

[1]Instituto Tecnológico de Buenos Aires (ITBA), Buenos Aires, Argentina
[2]CONICET - Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Instituto de Física Interdisciplinaria y Aplicada (INFINA), Buenos Aires, Argentina
[3]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Física, Buenos Aires, Argentina

Abstract

Electrotherapies are a group of techniques that apply electric pulses to permeabilize the plasma membrane in order to force the cell uptake of a nonpermeant permeable cytotoxic agent (electrochemotherapy) or a gene of therapeutic interest (electrogene transfer); to induce apoptosis (irreversible electroporation); to facilitate the release from inside the cell of a product of commercial interest (food technology). All these processes are possible by inducing pores in the cell membrane. Based on bibliography we developed in silico two lipid bilayer models to mimic the normal and tumor cell membranes by varying the lipid composition.

Both models are composed by: cholesterol (CHOL), dioleoilphosphatidyletanolamine (DOPE), dioleoilphosphatidylserine (DOPS), dioleoilphosphatidylcholine (DOPC) and sphingomyelin (SM). While in tumor bilayer systems both leaflets have identical lipid composition, normal bilayer systems present different lipid proportions in each leaflet. By molecular dynamics simulations, the membrane models were exposed to a high electric field normal to the bilayer, applied in one of the two possible directions, to open a single pore. The porated configuration obtained is then exposed to a lower electric field, in the same direction, to keep the pore open for 100 ns. Membrane response to the sustained electric field was evaluated considering: lipid redistribution between bilayer leaflets; pore stability, size and lipid composition; and ion passage through the pore.

Obtained results indicate that: 1) The electric field range appropriate to keep the pore open is very narrow (0.050 +/- 0.001 V/nm). 2) The pore structure remains constant after 100 ns considering both lipid composition and final pore size, for both types of membrane models and electric field directions. 3) The tumor bilayer response to the electric field is different from the normal bilayer, being the last one also dependent on the field direction. 4) Pore stability is related to DOPS translocation. 5) The velocity of ion passage through the pore depends on the pore size up to a maximum. These results suggest that the response of the tumor cell would be different from the normal one during an electrotherapy.

ID# 118

Amyotrophic lateral sclerosis: in the hunt for new disease-related genes, a network-based approach

Miguel Ángel Rubilar[1], Ute Woehlbier[1], Patricio Manque[1]

[1]Center for Integrative Biology, Universidad Mayor, Santiago, Chile

Abstract

Amyotrophic lateral sclerosis is a rare motor neuron disease with no cure or treatment, that significantly impacts the patient lives by severing the connection of these neurons with their innervated muscle, resulting in a complete paralysis of the patient. The onset of this disease has a wide range of ages, with some few cases of early onset up to reaching a peak at 60-70 years old, where most of the cases are diagnosed. ALS incidence was estimated to be 1.9 per 100,000 people worldwide, with a projected increase of incidence of 69% by 2040, due in part to an increasingly aging population. There have been several environmental and lifestyle-related risk factors that have been linked to the disease, such as exposure to heavy metals and cyanotoxins, smoking habits or even head traumas resulting from a lifetime of sports or sporadic accidents. It has been established that around 90% of the cases are sporadic with the rest being considered familiar cases, with a punctual mutation in specific genes being passed down in families. Hundreds of genes haven been associated to ALS, with specific single nucleotide polymorphisms, insertions, deletions, and duplications of segments of DNA being related in different population to the onset and development of ALS. Alongside of these, bulk and single cell RNA-seq analysis have also found differentially expressed genes in a context of ALS by interrogating mouse models and/or cell lines that attempt to represent the disease. Although these genes may appear to play an individual role in the disease, a systems biology approach and the use of biological networks would allow us to potentially recognize the connection between all these genes, finding genes that might otherwise go undetected because they fall below the statistical cutoff threshold of other approaches.

In this work, by cross-referencing information from various databases, it was possible to compile about 3,000 diverse ALS-related genes. These genes were considered "seed genes", which were used to generate a biological network allowing us to discover 180 new candidate genes previously unrelated to this disease which comprise gene ontology terms and KEGG pathways that the original seed list was unable to capture. By testing the transcript levels 10 of these genes in the spinal cord of an ALS mouse model, it was possible to determine transcriptional dysregulation at different stages of disease (pre-symptomatic versus symptomatic). With further validation in progress by Western blot and cellular studies.

This work provides new candidate genes that could play a role in ALS pathogenesis, which could help in the fight against the disease by expanding our knowledge and allowing us to develop genetic markers, treatments, or preventive measures.

Fondo Nacional de Desarrollo Científico y Tecnológico, Fondecyt, 1200459 Doctorado, Genomica Integrativa, Universidad Mayor, Santiago, Chile.

ID# 119

Bioinformatic pipeline design for the annotation of fungal genomes

Giordano L.J.[1], Ceschin D.G.[2], Merino M.C.[3]

[1]UFYMA - IPAVE. CIAP. INTA-CONICET, Argentina
[2]CIMETSA-IUCBC
[3]CIAP-INTA, Argentina

Background: Exploring and identifying genes in fungal genomes provides valuable insights into their features and metabolic pathways, contributing to a better understanding of the biology of these microorganisms. Advances in sequencing technologies have significantly increased the number of sequenced genomes across various species. However, in most of these genomes, the functional regions and genetic elements present have not yet been identified and characterized, meaning they are not annotated. This is due to fact that the annotation process is complex and requires several additional steps and techniques, beyond the assembled genomic sequence. These stages and methods include gene prediction, identification of regulatory regions, and function assignment, among others, which are essential for research in biology and genetics of organisms.

The goal of this work was to design a bioinformatic pipeline for the annotation of the available genomes of two agronomically important fungal pathogens.

Results: In this pipeline, RNAseq data obtained at our laboratory, and RNAseq data from public databases were first assembled by using Trinity. The assembled transcriptomes, along with public genomes databases, and fungal proteins from UniProt and OrthoDB, were used as input for Funannotate, BRAKER1, BRAKER2, and BRAKER3, resulting in different structural annotations. Finally, the AGAT toolkit was employed to check, fix, pad missing information of any kind of GTF and GFF file types creating a complete, organized, and standardized gff3 file. Thus, obtaining a final structural annotation for each fungus.

Conclusions: Comprehensive annotation of fungal genomes is a key procedure for studying evolutionary processes and regulatory regions in these microorganisms. This information is also useful for designing specific primers, enabling the analysis of gene expression under different conditions, and contributes to the development of tools for sustainable disease control in agriculture.

ID# 120

Machine Learning to Medical imaging study for cancer prognosis

Marcio Aparecido Bulla Junior[1], Fabiano Yokaichiya[1], Israel Tojal[2]

[1]Universidade Federal do Paraná, Brazil
[2]AC Camargo - Brazil

Abstract

The incidence of breast cancer has been a major global public health challenge, resulting in many deaths. However, one way to reduce fatalities is through the use of technologies that offer a more efficient prognosis, allowing a more appropriate treatment for the patient to be found. In this context, quantitative characteristics obtained from magnetic resonance images were used to predict the type of breast cancer. With the expansion of treatments aimed at deficiencies in repair mechanisms, including homologous recombination, it is relevant to evaluate whether radiomics can be used to evaluate homologous recombination (HR).

During the research work, data from The Cancer Imaging Archive was used, which contained images of patients with H2 made available by Polak. Some data had quantitative information about the images, while others only had patient images. It was necessary to classify the samples into two categories: HRD and HRP. Different families of features were also explored, such as Size, Enhancement-variance, Shape and others, which are features extracted from magnetic resonance images. Furthermore, differences between samples from the two categories were analyzed using the Mann-Whitney U hypothesis test. Machine learning models were created to differentiate the previously defined classes.

Our methodology involved collecting and pre-processing data, extracting features, classifying samples, training and evaluating machine learning models, carrying out hypothesis testing and creating visualizations to help interpret results.

ID# 122

Microbiome composition in Antarctic and non-Antarctic sponges of several ecoregions

Manrique-de-la-Cuba Maria[1], Parada-Pozo Génesis[1], Rodríguez-Marconi Susana[1], Abades Sebastián[1], Trefault Nicole[1,2,3]

[1]Centro GEMA - Genómica, Ecología & Medio Ambiente, Facultad de Ciencias, Universidad Mayor, Santiago, Chile
[2]Millenium Nucleus in Marine Agronomy of Seaweed Holobionts (MASH), Chile
[3]FONDAP Center IDEAL- Dynamics of High Latitude Marine Ecosystem, Chile

Abstract

Marine sponges and their microbiomes are ecosystem engineers of benthic habitats with a global distribution. In Antarctica, sponges dominate the benthos, an environment with harsh conditions and geographical isolation that could promote symbiont specialization. However, most knowledge of sponge microbiomes corresponds to surveys of sponges from tropical and temperate zones, overlooking sponges from polar zones. Here, we compare the microbial composition patterns of Antarctic and non-Antarctic sponges of several ecoregions using the inference of amplicon sequence variants (ASVs).

For this purpose, we selected and processed a subset of the 16S rRNA gene dataset from the Sponge Microbiome Project, which matched in taxonomy to available Antarctic sponges to at least at the family level. The community composition of sponge microbiomes differed between ecoregions under dissimilarity-based distances, where the Antarctic sponge habitat shows a remarkable difference from other sponge habitats. We found that Antarctic sponge microbiomes have higher similarity to each other and share more microorganisms with Hawaiian sponges than with sponges from other ecoregions, suggesting that similar environmental conditions and isolation may play a role in microbiome selection. Also, we identified that habitat-specific and habitat-generalist microorganisms have different taxonomy and abundance patterns in Antarctic and non-Antarctic sponges.

These results highlight the differences between Antarctic and non-Antarctic sponges and the strong influence of the ecoregion over microbiome composition, similar to that of the sponge phylogeny.

ID# 124

Assembly and annotation of diploid and tetraploid genomes of Bahiagrass (Paspalum notatum Flügge)

Vega, J.M.[1]; Podio, M.[1]; Orjuela, J.[2]; Siena, L.A.[1]; Mariac. C.[2]; Pessino, S.C.[1]; Leblanc, O.[2]; Ortiz, J.P.A.[1]

[1]Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR-CONICET-UNR), Zavalla, Argentina
[2]DIADE, Université de Montpellier, IRD, CIRAD, Montpellier, France

Abstract

Paspalum notatum is a subtropical forage grass native of South America. The species forms a multiploid complex, composed of sexual self-sterile diploids (var. saurae) and pseudogamous apomictic self-fertile tetraploids (var. notatum). It is currently one of the most important grasses for pastures and lawns in the southeastern USA. *P. notatum* is considered one of the models for the study of apomixis in grasses because of the large amount of information on the reproductive biology, genetics and breeding that has been generated over the last 50 years.

The objective of this work was to generate a high quality, fully characterized genome assembly of the species at the diploid level, to be used as a reference for disclosing the genomic structure of the apomictic tetraploid cytotypes. A natural biotype (#R1) from the center of origin and diversity of the species was used as plant material. Using more than 200x Oxford Nanopore long reads and 100x Illumina short reads, we generated a new chromosome level 564 Mb genome assembly (N50 = 56.1 Mb) of the species with high completeness (BUSCO score = 98.73%). The genome heterozygosity estimated with GenomeScope resulted of 1.73% while the repetitive elements detected by Repeatmaker, accounted for 57.86 % of the total genome. Gene annotation carried out using the MAKER pipeline predicted 45,074 gene models of which 32,101 were classified as high confidence due to their homology hits in InterProScan databases and having start and stop codons in appropriate positions.

Overall, a higher gene density was found in distal chromosome regions, while a higher frequency of transposable elements was found in pericentromeric regions. Alignment of the #R1 assembly with the reference and other grass genomes revealed a high degree of sequence conservation. Despite some large structural variations compared to rice and *Setaria italica* were found. Genome characterization included the identification of 59 miRNA precursors together with their putative targets and the generation of a set of 4,774 SSRs useful for breeding. Moreover, a preliminary assembly of the natural apomictic tetraploid genotype Q4117 was constructed using 80x Oxford Nanopore Technology long reads and 70x Illumina short reads. This assembly focused on preserving haplotype information, resulted in 5,500 contigs, spanning 2,345. Reference based annotation of this assembly identified 170,000 gene models exhibiting homology with #R1 gene models. Mapping the Q4117 assembly against the #R1 reference demonstrated an average coverage of 4.1 ± 2.1.

The present work provides a comprehensive genomic resource for analyzing the structure and evolution of the *P. notatum* agamic complex and the apomixis locus as well as a framework for stablishing molecular breeding strategies.

Asociación Argentina de Bioinformática
y Biología Computacional

ISBN 978-987-48989-7-5

a2b2c

bioinformatica_ar

RiaBioNet