in person

11-12 Nov 2026

XV CAB2C

Congreso Argentino de Bioinformática y Biología Computacional

Book of Abstracts

2026.a2b2c.org.ar

a2b2c

bioinformatica_ar

#XVCAB2C

XV CAB2C

15vo Congreso Argentino de Bioinformática y Biología Computacional

Book of Abstracts

Este libro es una obra colectiva de los resúmenes enviados por sus autores y presentados en el 15vo Congreso Argentino de Bioinformática y Biología Computacional realizado los días 11 al 12 de noviembre de 2026 en la ciudad de Corrientes, Provincia de Corrientes, Argentina.

Asociación Argentina de Bioinformática y Biología Computacional - A2B2C
1ra Edición – 2026
Ciudad Autónoma de Buenos Aires
Libro Digital/PDF - Archivo digital: Descarga y Online

Asociación Argentina de Bioinformática
y Biología Computacional

DRAFT - UNDER REVIEW

Asociación Argentina de Bioinformática y Biología Computacional
Book of Abstracts XV CAB2C / 1a ed. - Ciudad Autónoma de Buenos Aires: A2B2C, 2026. Libro digital, PDF
ISBN XXX-XXX-XXXXX-X-X
CDD XXX.XX

President A2B2C

Dra. Lucía B. Chemes

Vice-President A2B2C

Dra. Georgina Stegmayer

Secretary A2B2C

Dra. Patricia Saragüeta

Treasurer A2B2C

Dra. Juliana Glavina

Board Members A2B2C

Dra. Elin Teppa

Dr. Diego Bustos

Alternate Board Members A2B2C

Dra. Cristina Marino-Buslje

Dr. Flavio Spetale

Organizing Committee

Chairs

Dra. Lucía Chemes

Dra. Patricia Saragüeta

Dra. Georgina Stegmayer

Other members

Dra. Juliana Glavina

Lic. Clara Campos

Scientific Committee

Dr. Silvio C. E. Tosatto

Dr. Gustavo Daniel Parisi

Dr. Sebastián Fernández Alberti

Dr. Diego Martín Bustos

Abstract Reviewers

Martin Abba, Estefania Mancini, Diego Bustos, Eliana K. Asciutto, Carolina Cerrudo, M. Florencia Martini, Patricio Yankilevich, Carolina Torres, Matias Gerard, Gustavo Parisi, Dolores Blanco Fernandez, Heli Garcia Alvarez, Monica Pickholz, Nicolas Bellora, Sebastián Fernández Alberti, Luciana Gavernet, Valeria Risso, Emilio Luis Angelina, Flavio Spetale, Leandro Bugon, Silvina Fornasari, Luciana Capece, Gabriela Merino

DRAFT - UNDER REVIEW

















DRAFT - UNDER REVIEW

Table of contents

Program **11**

Keynotes and Invited Speakers **16**
Ana Tereza de Vasconcelos . . . 17
Salvador Ventura . . . 17
Ariela Vergara . . . 17
Miguel A. Andrade-Navarro . . . 17
Pierre Luisi . . . 17
Juliana Vianna . . . 17
Rodrigo Nores . . . 18
Salvador Capella-Gutierrez . . . 18
Laura Kamenetzky . . . 18
Henning Hermjakob . . . 18
Cristina Marino-Buslje . . . 19
Silvio C. E. Tosatto . . . 19
Juan Cortes . . . 19
Gábor Erdös . . . 19
Nathalie Sibille . . . 20
Rodrigo F. Díaz . . . 20
Anna Strachotová . . . 20
Jimena Rinaldi . . . 20
Marcelo A. Martí . . . 20
Esteban Mocskos . . . 21
Jiri Vondrasek . . . 21
Gavin Farrell . . . 21
John Hancock . . . 22

Short talks **23**
SESSION 1: Proteomics and Protein Structure and Function . . . 24
SHORT TALK #127 - In silico prediction of (p)ppGpp alarmone binding to DksA and NodD1 in Bradyrhizobium diazoefficiens: a possible link between the stringent response and nodulation gene expression . . . 24
SHORT TALK #129 - A Novel Experimentally-derived Dataset for Structural and Sequence Analysis of Disordered Flexible Linkers . . . 25
SHORT TALK #122 - Simulation-guided structural modelling of amyloid-like fibrils. From isolated peptides to nanoscale assemblies. . . . 26
SHORT TALK #121 - Molecular dynamics simulations support a dynamic RNA-recognition hotspot in the intrinsically disordered N-terminal domain of the androgen receptor . . . 27
SESSION 2: Evolutionary, Comparative and Population Genomics . . . 28
SHORT TALK #137 - Assessing the Impact of Kinship on Population Genomic Analyses in the Jaguar (Panthera onca) . . . 28
SHORT TALK #135 - Protein intrinsic disorder is associated with replicon location in the multipartite prokaryotic genome of Sinorhizobium meliloti . . . 29
SHORT TALK #130 - Bioinformatics applied to the study of the population dynamics of carbapenem-resistant Klebsiella pneumoniae ST258 in Buenos Aires Hospitals. . . . 30
SESSION 3: Genomics, Transcriptomics, Epigenetics and Metagenomics . . . 31
SHORT TALK #128 - Epigenomic and transcription factor profiles distinguish Progesterone Receptor binding sites with and without the canonical motif . . . 31
SHORT TALK #123 - An attention-based deep model for RNA secondary structure prediction 32
SHORT TALK #120 - Temporal-Spatial genomic regulation of the human endometrium . . . 33

DRAFT - UNDER REVIEW

SHORT TALK #134 - FusionAR-KB: A Multi-Layered Knowledgebase for Transcript-Specific Architecture and Clinical Interpretation of Gene Fusions in Cancer. . . . 34
SESSION 4: Biological Databases and Bioimaging . . . 35
SHORT TALK #125 - COMPUTATIONAL ANALYSIS OF HUMAN IRIS COLOUR: FROM SEGMENTATION TO QUANTITATIVE PHENOTYPING . . . 35
SHORT TALK #132 - Automatic segmentation of glioblastoma and its hidden tumor infiltration 36
SHORT TALK #133 - Ordering the Disorder: Updates to the Intrinsically Disordered Proteins Ontology . . . 37
SESSION 5: Artificial Intelligence and Systems Biology . . . 38
SHORT TALK #131 - Making Models That Matter: How to Build Trustworthy and Useful Systems Biology Models . . . 38
SHORT TALK #126 - Network organization of protein binding-site landscapes across scales . 39
SHORT TALK #136 - From human DNA to plant regulatory regions: transfer learning with a Genomic Large Language Model . . . 40
SHORT TALK #124 - Monosemantic Decomposition of Protein Language Models via Sparse Autoencoders for Interpretable Missense Variant Classification . . . 41
SHORT TALK #138 - When perturbation queries fail: static gene and protein embeddings outperform simulated scFM perturbation responses . . . 42
SESSION 6: IDPfun2 Special Track . . . 43
SHORT TALK #141 - Transposable elements as evolutionary substrates of protein disorder in the human proteome . . . 43
SHORT TALK #140 - LINKER-Pred: A Deep Learning Method and Web-Server for the Prediction of Disordered Flexible Linkers in Proteins . . . 44
SHORT TALK #139 - PonTED: augmenting flexible-linker prediction with protein language models and inter-domain context . . . 45
SHORT TALK #142 - Structural Diversity and Conformational Plasticity of Plantaricin A Variants 46

Poster presentations **47**
POSTER #1 - Benchmarking molecular representations and sequence-guided evidence transfer in LigQ 2 . . . 48
POSTER #2 - Precision control of microbial cell therapies through AI-driven receptor engineering . . . 49
POSTER #3 - Fine-tuning of the ProGen2 protein language model for the generation of artificial beta-lactamases . . . 50
POSTER #4 - The Generalization Gap in Melanoma AI: Cross-Dataset Reliability and the Challenge of Underrepresented Latin American Data . . . 51
POSTER #5 - emb2bind: a deep learning binding predictor for IDRs based on protein language model and energy embeddings . . . 52
POSTER #9 - Repositioning approved drugs as α-Synuclein aggregation inhibitors through ligand-based virtual screening . . . 53
POSTER #10 - Characterizing Intrinsically Disordered Protein Regions Using ESM-2 Attention Patterns . . . 54
POSTER #11 - Deep Learning Recurrent Architecture for Intrinsically Disordered Protein Classification . . . 55
POSTER #12 - Incorporating Chemical Probing Data into Deep Learning for RNA Secondary Structure Prediction . . . 56
POSTER #13 - Identifying inflammatory Programmed Cell Death associated features to predict clinical outcome in patients with diffuse gliomas . . . 57
POSTER #14 - Héritas Second Brain: An Expert-Curated Knowledge Graph for Clinical Genomic Interpretation . . . 58
POSTER #15 - A three-dimensional video-tracking protocol for the assessment of neurobehavioral alterations in adult zebrafish (Danio rerio) . . . 59

DRAFT - UNDER REVIEW

POSTER #16 - Phthalate exposure and obesity in adults: a small but robust association, and three leakage mechanisms that inflate it. . . . 60
POSTER #17 - Probing a genomic language model for off-target, pseudogene and chimeric reads in COI eDNA metabarcoding . . . 61
POSTER #18 - AI-Based Extraction of Structured Lifespan Data from Scientific Literature . . . 62
POSTER #20 - Functional characterization of drug repurposing candidates through integrated analysis of molecular targets and biological pathways . . . 63
POSTER #21 - Segmentation metrics for the evaluation of Pfam domain annotation . . . 64
POSTER #22 - Modeling the occurrence of toxigenic Aspergillus carbonarius in grapes using regularized regression and neural networks . . . 65
POSTER #23 - AUTISM SPECTRUM DISORDER CLASSIFICATION USING GRAPH NEURAL NETWORKS FROM GUT MICROBIOTA PROFILES . . . 66
POSTER #24 - Machine learning in species distribution models: A case study between a cactophilic drosophilid and its hosts . . . 67
POSTER #25 - Simplifications and approximations in molecular systems biology: lessons from the toggle switch . . . 68
POSTER #27 - A Machine Learning-Derived Glycosignature Associated with Glioblastoma Stem-like Cell Subtypes and Patient Survival . . . 69
POSTER #28 - Inference-Time Toxicity Mitigation in Protein Language Models . . . 70
POSTER #29 - Gut Microbiome as a Source of Biomarkers in Neurodegenerative and Neurodevelopmental Disorders: A Machine Learning and Explainable AI Approach . . . 71
POSTER #31 - From protein embeddings to complex networks: multiscale organization of molecular recognition . . . 72
POSTER #32 - Integrating Multi-Omics and Cross-Species Knowledge Graphs to Identify Heat-Responsive lncRNA Regulatory Networks . . . 73
POSTER #34 - BioAIrepo: Making AI Resources Discoverable and Reusable Across the Life Sciences . . . 74
POSTER #35 - Glycosylated IDRs: Biological Implications and a Cross-Reference Between DisProt and GlyConnect . . . 75
POSTER #37 - Cross-Beta-DB: An Updated Resource for Naturally Occurring Amyloid Fibrils and Benchmarking Prediction Methods . . . 76
POSTER #38 - In silico evaluation of the role of BIRC6 in the DNA damage response in lung adenocarcinoma . . . 77
POSTER #41 - Bioinformatic Characterization of a Cohort of PD-L1-Negative Non-Small Cell Lung Cancer Patients Through Network Analysis . . . 78
POSTER #42 - Differences in the phyllosphere and endophyllosphere fungal communities among three Nothofagus species from the Andean-Patagonian forest. . . . 79
POSTER #43 - The Evolution of Disorder: From Prokaryotes to Eukaryotes in Universally Conserved Proteins . . . 80
POSTER #45 - Low-coverage sequencing as a strategy for biobanking of underrepresented populations. . . . 81
POSTER #46 - Evaluating translational rate conservation and cotranslational folding in legumes using minMax profiles . . . 82
POSTER #47 - A city-scale synthetic population with demographic, socioeconomic and genetic structure for the evaluation of forensic genetic genealogy methods . . . 83
POSTER #48 - How do reptiles survive hydric stress? A comparative review of the genomic signatures of adaptation in desert and marine environments . . . 84
POSTER #49 - Comparative pangenome analysis of Pasteurella multocida identifies conserved genes with potential as vaccine candidates . . . 85
POSTER #50 - Comparative genomics and genome mining of Antarctic marine Pseudomonas sp. m6p . . . 86

DRAFT – UNDER REVIEW

POSTER #52 - Genomic characterization of B-genome-specific pack-MULE transposons in the allotetraploid Arachis hypogaea . . . 87
POSTER #53 - Genome Sequencing and Functional Characterization of Extrachromosomal Elements in Rhodococcus erythropolis JCM 2893, a Petroleum-Derived Strain Capable of Growth under Oligotrophic Conditions . . . 88
POSTER #54 - Metagenomic profiling of the metal resistome of microbial communities from the Caviahue-Copahue Geothermal System . . . 89
POSTER #56 - Comparative Evaluation of Pathway Topology Analysis Methods for Reactome, with a Focus on Intrinsically Disordered Proteins . . . 90
POSTER #57 - Temporal and spatial persistence of a bacterial core across vineyard soil and wine in an emerging wine-producing region . . . 91
POSTER #59 - Revisiting historical 16S rRNA datasets: does improved bioinformatic resolution change our view of bacterial diversity in eutrophic lake sediments? . . . 92
POSTER #60 - DIFFERENTIAL EXPRESSION OF INFLAMMASOME RELATED GENES IN COLORECTAL CANCER AND ITS STRATIFICATION BY MICROSATELLITE INSTABILITY STATUS . . . 93
POSTER #61 - Pipeline for the taxonomic and functional analysis of plant growth-promoting bacteria from long-read sequencing data . . . 94
POSTER #62 - Cross-domain metabarcoding analytics: an integrated bioinformatics pipeline for bacterial and fungal communities in sugarcane soils . . . 95
POSTER #63 - Cake-cutting applied to genomic fragments with repeated sequences . . . 96
POSTER #64 - Genome-wide AP-1-directed enhancer demethylation, but a stableLRP1 locus, marks monocyte-to-macrophage differentiation: ananopore methylome analysis . . . 97
POSTER #65 - Genomic variation in Argentine isolates of Cercospora kikuchii: contrasting patterns among virulence, fungicide resistance, and phylogenetic genes . . . 98
POSTER #66 - Uncovering Peripheral Biomarkers of Celiac Disease via Integrative Transcriptomic Analysis . . . 99
POSTER #67 - Improving Microbiome Beta Diversity Analysis through a Parametric Model of Abundance Gradients . . . 100
POSTER #68 - Systematic Profiling of Expressed HRR Variants: An RNA-seq Analysis Pipeline Applied to Prostate Cancer Cohorts . . . 101
POSTER #69 - Inferring spatial gene expression in the non-model giant kidney worm Dioctophyme renale by extrapolation from Brugia malayi spatial transcriptomics . . . 102
POSTER #70 - Microbiome Composition and Predicted Functions Across Sequencing Platforms 103
POSTER #71 - Transcriptomic dynamics of the peripheral immune response to FMDV vaccination in cattle . . . 104
POSTER #72 - Annotating the secretome of an under-represented parasitic nematode, the giant kidney worm Dioctophyme renale . . . 105
POSTER #73 - Unveiling the structure of bacterial communities in groundwater using long-read metagenomic sequencing. . . . 106
POSTER #74 - Bioinformatic Identification and Experimental Validation of microRNAs Associated with Rejuvenation in Ilex paraguariensis . . . 107
POSTER #75 - High-resolution single-nucleus transcriptomics reveals striatal molecular remodeling during the off-L-DOPA state . . . 108
POSTER #76 - Plastic-degrading potential in the South Atlantic plastisphere: taxonomy and polymer targets . . . 109
POSTER #77 - Multi-Omic Characterization of Hepatocellular Carcinoma across Liver Fibrosis Stages: A Comparison of Ishak 0 and Ishak . . . 110
POSTER #78 - BAL-comp: an automated workflow for Lactic Acid Bacteria comparative genomic analysis . . . 111
POSTER #79 - Same genes, different crowd: host community turnover underlies seasonal shifts in the sugarcane soil resistome . . . 112

DRAFT – UNDER REVIEW

POSTER #80 - Shotgun Metagenomic Profiling of Pelagic and Benthic Microbial Communities Across the Argentine-Uruguayan Common Fishing Zone . . . 113
POSTER #81 - DESIGN AND IMPLEMENTATION OF A REPRODUCIBLE HPC PLATFORM FOR AUTOMATED RNA-SEQ DATA ANALYSIS . . . 114
POSTER #82 - Genome-wide quantitative profiling of human core promoters reveals sequence determinants of transcriptional activity and noise . . . 115
POSTER #83 - Sex-hormone crosstalk shapes gut microbiome-immune responses to microbial stimuli . . . 116
POSTER #85 - Transcript-anchored mapping and evidence-weighted curation of germline breast cancer predisposition variants in a Panamanian cohort . . . 117
POSTER #86 - METAGENOMIC CHARACTERIZATION OF MICROBIAL COMMUNITIES FROM THE AMBARGASTA AND HUYAMAMPA SALT FLATS IN NORTHERN ARGENTINA . . . 118
POSTER #89 - Can ML Diffusion Models Capture Intrinsically Disordered Protein Ensembles? 119
POSTER #91 - Annotating experimental conditions of amyloid structures to explore polymorph relevance . . . 120
POSTER #92 - A novel structure-driven strategy for Parkinson's disease immunotherapy . . . 121
POSTER #94 - Diverse sequences, equivalent ensembles: Understanding the molecular grammar underlying conformational buffering in IDRs . . . 122
POSTER #95 - Grafting Site Modulates Intramolecular Inhibition in Monomeric Domain-Linker-Motif Constructs . . . 123
POSTER #96 - Glycosylation across IDRs and non-IDR regions: patterns and functional associations . . . 124
POSTER #97 - Insights into the allosteric modulation of IDO1: molecular dynamics simulations reveal the binding mode of a novel inhibitor . . . 125
POSTER #98 - AmyloTools: a Python package and web application for exploration and quantitative characterization of amyloid fibril structures. . . . 127
POSTER #99 - Analysis of Antiplatelet Activity through Protein Disulfide Isomerase (PDI) Inhibition by Peptides . . . 128
POSTER #100 - Computational identification of structurally uncharacterized gamete fusogens in fungi . . . 129
POSTER #101 - Beyond Annotation: A Curated Proteomic and Bioinformatic Approach to Map the Small Protein Universe of Bordetella pertussis . . . 130
POSTER #102 - Impact of mutations at evolutionary conserved sites on the conformational dynamics of tandem-repeat proteins . . . 131
POSTER #104 - Detection of Bacteroides fragilis Division II (cfiA+) by Mass-Shift Analysis of MALDI-TOF MS Spectra and Machine Learning Classification . . . 132
POSTER #105 - Beyond the nucleolus: Integrative interactome and transcriptome analysis identifies MAGEB2 as a ribosome-associated regulator of p53 under nucleolar stress . . . 133
POSTER #106 - Proportional phylogenetic assignment of shared peptides for accurate microbial biomass estimation in quantitative metaproteomics . . . 134
POSTER #107 - Conservation of the catalytic core and active-site diversification in class A flavin-dependent monooxygenases . . . 135
POSTER #108 - Same motif, different partners: an evolutionary-structural view of human sHSP hetero-oligomers . . . 136
POSTER #109 - Disordered tails, conserved contacts: a proteome-wide search for non-catalytic autoregulation . . . 137
POSTER #110 - In silico Optimization of Artificial Cellulosomes: Simulating Multienzyme Cellulose Hydrolysis via Agent-Based Reaction–Diffusion Dynamics . . . 138
POSTER #111 - Uncovering Drug-Repurposing Opportunities across CCR5 and CCR3: A Structure-Based Strategy toward Chemokine Receptor Modulation in Cancer . . . 139

DRAFT - UNDER REVIEW

POSTER #112 - Methodological Optimization and Molecular Dynamics Validation of the Catalitic Core in Herpes Virus Type 1 DNA Polymeras: Preservation of the Mg(II) Coordination Sphere 140
POSTER #114 - A workflow for structural phylogenetic analysis of divergent protein domains 141
POSTER #116 - Backbone scaffold conformation influences the stability of amyloid cross-β structures: an all-atom molecular dynamics study 142
POSTER #118 - Experimental warming alters fungal transcriptional signals recovered from Colobanthus quitensis RNA-seq data 143
POSTER #120 - A Post-Hoc Analytical Framework for the Characterization of Disease Endotypes Applying Biclustering 144
POSTER #121 - 16S rRNA-based analysis of storage-associated shifts in the relative composition of the raw boar seminal microbiome at 17 °C 145
POSTER #122 - Emergent Time-Sharing in Competitive Systems with Cross-Inhibition and Slow Self-Adaptation 146
POSTER #123 - An Integrated Bioinformatics Framework for ORF Characterization and Visualization 147
POSTER #124 - Comparative genomic analysis of budded virion production in baculoviruses 148
POSTER #125 - Complete genome assembly of Argentinean Creole bovine using long reads . 149
POSTER #126 - Inferring microRNA expression patterns and regulatory circuits at single-cell resolution through pri-miRNA detection 150
POSTER #127 - Consistency and Stability of XAI Explanations to Understand QSAR Models trained for Predicting Developmental Toxicity 151
POSTER #130 - PHENOTYPING OF FUNGAL INFECTIONS IN TOMATO PLANTS USING MULTISPECTRAL IMAGING 152
POSTER #131 - Generation and evaluation of a bioinformatics pipeline for the detection of viral sequences in insects using public NGS data 153
POSTER #133 - Prophage profile of Staphylococcus aureus isolates from persistent infections.154
POSTER #134 - Identification of Novel Agricultural Antifungal Candidates Targeting Cytochrome b through Virtual Screening, Molecular Modeling and Resistance-Mutation Analysis. 155
POSTER #136 - A modular workflow for residue-level analysis of peptide–protein interactions from molecular dynamics trajectories 156
POSTER #138 - Identification of the heavy metal ATPases (HMAs) genes in Yerba Mate 157
POSTER #139 - Protein disorder in Rhizobium favelukesii LPU83 and its association with the acid stress response 158
POSTER #140 - Comparative Genomic Analysis of Lentilactobacillus kefiri CIDCA 8348 and CIDCA 83111: Probiotic Potential and Safety Assessment 159
POSTER #142 - Bioinformatic strategies to detect specific anti-enzyme antibodies of the primary immune response 160
POSTER #143 - Identification of Bacterial Species of the Genus Leptospira and Preliminary Pathogenicity Assessment Based on Sequencing Data from Environmental Samples Collected from the Salado, Coronda, and Paraná Rivers and Associated Lagoons in the Santa Fe Metropolitan Area 161
POSTER #144 - Contributions to the study of the genetic basis of the dehiscence trait in the forage grass Panicum Coloratum 162
POSTER #145 - in vitro caracterization of the activity of the CASK protein kinase domain . . . 163
POSTER #146 - Characterization of the Resistome and Mobilome in Microbiota Metagenomes of Wild Penguins from the Strait of Magellan. 164
POSTER #147 - GraphPert: Network propagation-aware drug repurposing from transcriptional165
POSTER #148 - Hydra effect emerges from optimal control in structured populations 166
POSTER #149 - Generating holo structures from apo conformations with AI structure-prediction models: pipeline design and a docking baseline on apo–holopairs 167

DRAFT – UNDER REVIEW

01001
00100
10100

Program

10100
10011
01000

Day 1: November 11th

8:00 to 9:00	**Registration**
9:00 to 9:15	**Opening Ceremony**
9:15 to 10:00	**KEYNOTE TALK. Ana Tereza Ribeiro de Vasconcelos (National Laboratory for Scientific Computing, Brazil)** *From Genomes to Knowledge: Bioinformatics at National Scale in Brazil*
10:00 to 10:15	**Coffee Break**
	SESSION 1: Proteomics and Protein Structure and Function
10:15 to 10:45	**INVITED SPEAKER. Ariela Vergara Jaque (Facultad de Ingeniería Talca, Chile)** *Computational strategies for designing modulators of ion channel protein–protein interactions.*
10:45 to 11:15	**INVITED SPEAKER. Miguel Andrade (Johannes-Gutenberg, University in Mainz, Germany)** *Sequence, context, and AI: Studying the functional landscapes of intrinsically disordered regions.*
11:15 to 11:55	**SHORT TALKS SELECTED FROM ABSTRACTS** **Ornella Falduti (IBBM, UNLa Plata, Argentina)** *In silico prediction of (p)ppGpp alarmone binding to DksA and NodD1 in Bradyrhizobium diazoefficiens: a possible link between the stringent response and nodulation gene expression* **Juliana Glavina (UNSAM, BA, Argentina)** *A Novel Experimentally-derived Dataset for Structural and Sequence Analysis of Disordered Flexible Linkers* **Exequiel Barrera (IHEM, UNCuyo, Mendoza, Argentina)** *Simulation-guided structural modelling of amyloid-like fibrils. From isolated peptides to nanoscale assemblies* **Galo Ballati (UNQ, BA, Argentina)** *Molecular dynamics simulations support a dynamic RNA-recognition hotspot in the intrinsically disordered N-terminal domain of the androgen receptor*
11:55 to 12:00	**Short Break**
12:00 to 12:45	**ROUND TABLE.** Chairs: Salvador Capella, Georgina Stegmayer *Perspectives of Artificial Intelligence in the life sciences* **Gabor Erdos (ELTE, Hungary)** **Jiri Vondrasek (ELIXIR Czech)** **Ana Tereza de Vasconcelos (NLSC, Brazil)** **Henning Hermjakob (EMBL-EBI)** **Rodrigo Diaz (UNSAM, Argentina)**
12:45 to 14:00	**Lunch Break**
	SESSION 2: Evolutionary, Comparative and Population Genomics
14:00 to 14:30	**INVITED SPEAKER. Pierre Luisi (UNC, Argentina)** *Ancient and Present-Day Genomes in the Southern Cone: Challenges for Reconstructing Population History*
14:30 to 15:00	**INVITED SPEAKER. Juliana Vianna (UNAB, Chile)** *From reference genomes to population genomes: evolution and adaptation of South American and Antarctic vertebrates*
15:00 to 15:45	**SHORT TALKS SELECTED FROM ABSTRACTS**

DRAFT – UNDER REVIEW

	Francisco Pisciottano (IBYME-CONICET) *Assessing the Impact of Kinship on Population Genomic Analyses in the Jaguar (*Panthera onca*)*
	Abril Pagnutti (UNLa Plata, Argentina) *Protein intrinsic disorder is associated with replicon location in the multipartite prokaryotic genome of Sinorhizobium meliloti*
	Francisco Gonzalez Espinosa (UBA, BA, Argentina) *Bioinformatics applied to the study of the population dynamics of carbapenem-resistant Klebsiella pneumoniae ST258 in Buenos Aires Hospitals.*
15:45 to 16:00	**Coffee Break**

SESSION 3: Genomics, Transcriptomics, Epigenetics and Metagenomics	
16:00 to 16:30	**INVITED SPEAKER. Rodrigo Nores (UNC, Argentina)** *Paleogenomic Analysis Reveals 8,000 Years of Indigenous Lineage Continuity in Human Populations of Central Argentina*
16:30 to 16:45	**INVITED SPEAKER. Salvador Capella Gutierrez (BSC, CNS, Spain)** *ELIXIR: Ten priority areas for AI in Life Sciences. How do we move forward?*
16:45 to 17:00	**INVITED SPEAKER. Laura Kamenetzky (UBA, Argentina)** *Consolidation of a long-read sequencing genomics center in southern South America*
17:00 to 17:45	**SHORT TALKS SELECTED FROM ABSTRACTS** **Carina Fariñas (IBYME, BA, Argentina)** *Epigenomic and transcription factor profiles distinguish Progesterone Receptor binding sites with and without the canonical motif*
	Bruno Breggia (sinc, UNL, Sta Fe, Argentina) *An attention-based deep model for RNA secondary structure prediction*
	Luciana Ant (IBYME, BA, Argentina) *Temporal-spatial genomic regulation of the human endometrium*
	Daniela Orchanski (UNC, Cordoba, Argentina) *FusionAR-KB: A Multi-Layered Knowledgebase for Transcript-Specific Architecture and Clinical Interpretation of Gene Fusions in Cancer*
17:45 to 17:55	**SPONSOR TALK.** *ZS*
17:55 to 19:30	**POSTER SESSION I**
19:30 to 20:30	**WELCOME COCKTAIL.** *Art, Science and Technology Event*

Day 2: November 12th

SESSION 4: Biological Databases and Bioimaging	
9:00 to 9:30	**INVITED SPEAKER. Henning Hermjakob (EMBL-EBI, United Kingdom)** *The Complex Portal: From Mass Spectra to Structures*
9:30 to 10:00	**INVITED SPEAKER. Cristina Marino (F. Leloir, Argentina)** *Exploring the conformational landscape of the amylome*
10:00 to 10:30	**SHORT TALKS SELECTED FROM ABSTRACTS** **Martina Ceballos (UC Cordoba, Argentina)** *Computational analysis of human iris colour: from segmentation to quantitative phenotyping*

DRAFT – UNDER REVIEW

	Alexander Mulet (UBA, BA, Argentina) *Automatic segmentation of glioblastoma and its hidden tumor infiltration*
	M. Victoria Nugnes (UPadova, Italy) *Ordering the Disorder: Updates to the Intrinsically Disordered Proteins Ontology*
10:30 to 10:45	**Coffee Break**

SESSION 5: Artificial Intelligence and Systems Biology

10:45 to 11:15	**INVITED SPEAKER. Silvio Tosatto (U. Padova, Italy)** *The DOME ecosystem: from recommendations to applications in Artificial Intelligence*
11:15 to 11:45	**INVITED SPEAKER. Juan Cortés (CNRS, France)** *Generative models for the design of disordered proteins*
11:45 to 12:40	**SHORT TALKS SELECTED FROM ABSTRACTS** **John Hancock (University of Ljubljana, Slovenia)** *Making Models That Matter: How to Build Trustworthy and Useful Systems Biology Models*
	Ariel Chenomoretz (UBA, F. Leloir, BA, Argentina) *Network organization of protein binding-site landscapes across scales*
	Leonardo Pedraza (INTECNUS, Rio Negro, Argentina) *From human DNA to plant regulatory regions: transfer learning with a Genomic Large Language Model*
	Santiago Britez (F. Leloir, BA, Argentina) *Monosemantic Decomposition of Protein Language Models via Sparse Autoencoders for Interpretable Missense Variant Classification*
	Nicolas Yanovsky (F. Leloir, BA. Argentina) *When perturbation queries fail: static gene and protein embeddings outperform simulated scFM perturbation responses*
12:40 to 12:45	**Short Break**
12:45 to 13:30	**ROUND TABLE.** Chairs: Lucia Chemes, John Hancock *Leveraging research infrastructures for bioinformatics: the ELIXIR experience* **Jimena Rinaldi (CEBEM & UNSAM, Argentina)** **Salvador Capella Gutierrez (BSC, Spain)** **Jiri Vondrasek (ELIXIR Czech)** **Anna Strachotova (ELIXIR Czech)** **Marcelo Martí (FCEN-UBA, Argentina)** **Esteban Mocskos (FCEN-UBA, Argentina)**
13:30 to 14:30	**Lunch Break**

SESSION 6: IDPfun2 Special Track

14:30 to 15:00	**INVITED SPEAKER. Gabor Erdos (ELTE, Hungary)** *A zero-shot thermodynamic potential for protein mutational stability*
15:00 to 15:30	**INVITED SPEAKER. Nathalie Sibille (CNRS, France)** *Functional Disorder in Cell Signaling: Molecular Mechanisms and Case Studies*
15:30 to 16:15	**SHORT TALKS SELECTED FROM ABSTRACTS** **Juan Mac Donagh (UNQ, BA, Argentina)** *Transposable elements as evolutionary substrates of protein disorder in the human proteome*
	Heli García Álvarez (UNSAM, BA, Argentina) *LINKER-Pred: A Deep Learning Method and Web-Server for the Prediction of Disordered Flexible Linkers in Proteins*

DRAFT - UNDER REVIEW

	Leandro A. Bugnon (sinc, UNL StaFe, Argentina) *PonTED: augmenting flexible-linker prediction with protein language models and inter-domain context*
	Marina Navarro (UNQ, BA, Argentina) *Structural Diversity and Conformational Plasticity of Plantaricin A Variants*
16:15 to 16:25	**SPONSOR TALK.** *Fundación Sadosky*
16:25 to 16:40	**Coffee Break**
16:40 to 18:00	**POSTER SESSION II**
18:00 to 18:45	**KEYNOTE TALK. Salvador Ventura (UAB, Spain)** *Targeting Protein Aggregation at the Interface of Experiment and Computation*
18:45 to 19:15	**PRIZES AND CLOSING CEREMONY**
19:15 to 19:30	**A2B2C MEETING.** *Election & change of Authorities*

Day 3: November 13th

WORKSHOPS	
9:00 to 17:00	**WORKSHOP.** *ELIXIR-GOBLET - TtT (Unit I & II)*
9:00 to 10:30	**WORKSHOP.** *IDPfun2 - Workshop 1. Computational tools for the prediction of intrinsic disorder in proteins*
11:00 to 12:30	**WORKSHOP.** *IDPfun2 - Workshop 2. Molecular Network Resources of the European Bioinformatics Institute*
9:00 to 17:00	**WORKSHOP.** *Language Models Applied to Biological Sequence Analysis*

Day 4: November 14th

WORKSHOPS	
9:00 to 17:00	**WORKSHOP.** *ELIXIR-GOBLET - TtT (Unit III & IV)*

DRAFT - UNDER REVIEW

01001
00100
10100

Keynotes and Invited Speakers

10100
10011
01000



Ana Tereza de Vasconcelos

National Laboratory of Scientific Computing, Brazil

Researcher at the National Laboratory of Scientific Computing and former President of AB3C. Her work focuses on genomics, metagenomics, transcriptomics, and computational methods for genomics.



Salvador Ventura

Universitat Autònoma de Barcelona, Spain

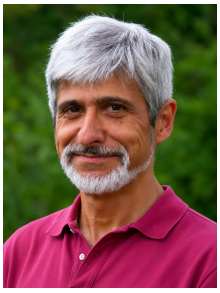
Professor of Biochemistry and Molecular Biology at the Universitat Autònoma de Barcelona. His research focuses on protein structure, protein misfolding, conformational diseases, and the development of therapeutic molecules.



Ariela Vergara

Universidad de Talca, Chile

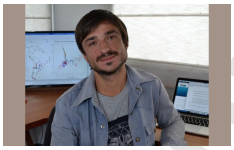
Researcher at the Center for Bioinformatics, Simulation and Modeling at Universidad de Talca. Her research focuses on computational structural biology of membrane proteins, molecular modeling, docking simulations, and molecular dynamics.



Miguel A. Andrade-Navarro

Johannes Gutenberg University Mainz, Germany

Professor of Bioinformatics at Johannes Gutenberg University Mainz. His research interests include computational biology, protein and gene function analysis, and data mining approaches in molecular biology.



Pierre Luisi

CONICET / Universidad Nacional de Córdoba, Argentina

Human Population Geneticist, CONICET Researcher and Full Professor at Universidad Nacional de Córdoba. His research focuses on human population genetics, ancient DNA, population genomics, and the application of bioinformatics and machine learning to evolutionary and biomedical studies.

DRAFT - UNDER REVIEW



Juliana Vianna

Universidad Andrés Bello (UNAB) / Centro de Regulación del Genoma (CRG), Chile

Juliana Vianna is a researcher at the Molecular Biodiversity Laboratory at Universidad Andrés Bello (UNAB) and the Center for Genome Regulation (CRG), Chile. She holds a PhD in Biological Sciences, specializing in Ecology, from Pontificia Universidad Católica de Valparaíso (2010). Her research examines the effects of glacial cycles on Antarctic and sub-Antarctic biodiversity. She leads projects in genomic regulation and long-read DNA sequencing, and has participated in graduate teaching and supervision in ecology, evolutionary biology, and conservation medicine.



Rodrigo Nores

IDACOR / CONICET / Universidad Nacional de Córdoba, Argentina

Rodrigo Nores is a bioanthropologist specializing in paleogenomics. He received his PhD in Chemical Sciences from the Universidad Nacional de Córdoba (UNC) in 2006. He currently works as a CONICET Researcher at the Instituto de Antropología de Córdoba (IDACOR) and is a Full Professor in the Anthropology undergraduate program at the Faculty of Philosophy and Humanities (UNC). His work focuses on reconstructing the evolutionary history of human populations and pathogens in South America through ancient DNA analysis, conducting research in collaboration with international reference centers such as the Department of Genetics at Harvard University and the Max Planck Institute for Evolutionary Anthropology (MPI-EVA).



Salvador Capella-Gutierrez

BSC / INB–ELIXIR, Spain

Salvador Capella-Gutierrez, PhD, leads the Technologies for Biomedical Research Laboratory at the Barcelona Supercomputing Center (BSC), as well as the Coordination team of the Spanish National Bioinformatics Institute (INB). INB is the Spanish node of ELIXIR, the European research infrastructure for life science data. Salvador joined the BSC in July 2017 from the Spanish National Cancer Research Center (CNIO) in Madrid. A bioinformatician with a solid technological background in both Computer Engineering and Health Sciences, Salvador received his PhD in Bioinformatics in 2012, awarded by the Center for Genomic Regulation and Pompeu Fabra University in Barcelona, Spain.

His current focus is on developing long-term infrastructure to facilitate the scientific benchmarking and technical evaluation of research software in Life Sciences, including tools, web services, and workflows. More recently, he is interested in how to further elevate those efforts to continue serving the research community in the fast-evolving AI context.



Laura Kamenetzky

iB3 / Universidad de Buenos Aires / CONICET, Argentina

Laura Kamenetzky graduated in Biology in 2001 and earned her PhD in Molecular Biology in 2007 from the Faculty of Exact and Natural Sciences, University of Buenos Aires (UBA). She conducted postdoctoral research at the National Institute of Agricultural Technology (INTA). She is a Principal Researcher at CONICET, based at the Institute of Biosciences, Biotechnology and Translational Biology (iB3-UBA). Her research focuses on genomics and bioinformatics of eukaryotes. She has coordinated research projects funded by Argentine and European institutions and organized scientific events and workshops in helminth research, bioinformatics, and genomics.

DRAFT - UNDER REVIEW



Henning Hermjakob

EMBL-EBI, United Kingdom

Leader of the Molecular Systems Cluster at EMBL-EBI. His work focuses on systems biology resources, molecular interaction databases, pathways, and biological data standardization.



Cristina Marino-Buslje

Fundación Instituto Leloir / CONICET / ITBA, Argentina

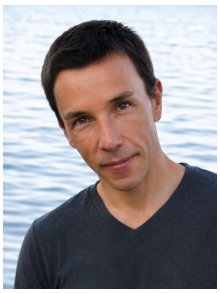
Principal Investigator at CONICET and Head of the Structural Bioinformatics Unit at Fundación Instituto Leloir. Her research focuses on protein structure and evolution, intrinsically disordered proteins, aggregation, and machine learning applications in bioinformatics.



Silvio C. E. Tosatto

Università degli Studi di Padova, Italy

Full Professor in Bioinformatics and Principal Investigator at the BioComputing UP lab of the Department of Biomedical Sciences, Università degli Studi di Padova. His research focuses on bioinformatics and computational biology, particularly on the structural and mechanistic aspects of complex biological systems.



Juan Cortes

CNRS, France

Computer scientist specialized in algorithmic robotics and Research Director at CNRS. His interdisciplinary research combines robotics, artificial intelligence, and structural bioinformatics approaches.



Gábor Erdös

ELTE, Hungary

Assistant Professor and Principal Investigator at the Department of Biochemistry at Eötvös Loránd University (ELTE), Hungary, where he leads the Protein Dynamics Research Group. Trained as a bioengineer, he holds a PhD in structural biochemistry and has a decade of experience in bioinformatics. His career spans ten years of academic research alongside specialized experience in the pharmaceutical industry, with professional activity in Argentina and extensive collaborations across Europe, particularly in Italy and Israel. His research expertise centers on the characterization of intrinsically disordered proteins through the integration of biophysical concepts and advanced computational approaches. His work focuses on developing novel predictive methods for both general and functional protein disorder, with the goal of bridging artificial intelligence and biophysics. To this end, he employs a broad range of methodologies that includes machine learning, molecular dynamics simulations, and statistical mechanics to investigate the complex behavior and functional roles of intrinsically disordered proteins. He serves as Co-lead of the ELIXIR IDP Community and as the Hungarian Representative for COST. He is also a core member of the ML4NGP COST action, IDP2Biomed twinning project and a contributor of the IDPFun2 project.

DRAFT - UNDER REVIEW



Nathalie Sibille

CNRS / University of Montpellier, France

Research Director at CNRS and Team Leader at the University of Montpellier. Trained in Physics and an expert in NMR spectroscopy for Structural Biology, her research focuses on intrinsically disordered proteins (IDPs), post-translational modifications, and the molecular mechanisms regulating GPCR and arrestin interactions.



Rodrigo F. Díaz

ITBA / CONICET / UNSAM, Argentina

Rodrigo F. Díaz is Associate Dean of the School of Management and Technology at ITBA, Independent Researcher at CONICET, and professor at UNSAM. His research applies statistical and machine learning methods to detect exoplanets in radial velocity data, separating planetary signals from stellar activity. He founded the Bachelor's in Data Science at UNSAM and led the creation of ITBA's Master's in Artificial Intelligence.



Anna Strachotová

ELIXIR CZ / IOCB Prague, Czech Republic

Anna Strachotová is a Czech scientific project manager and the National Node Coordinator for ELIXIR CZ (the Czech National Infrastructure for Biological Data), based at the Institute of Organic Chemistry and Biochemistry (IOCB) of the Czech Academy of Sciences in Prague. Since 2018, she has served as Node Coordinator and Project Manager for ELIXIR CZ, working within Jiří Vondrášek's bioinformatics research group at IOCB Prague.



Jimena Rinaldi

CEBEM / UNSAM, Argentina

Jimena Rinaldi is a structural biochemist and group leader of the Photobiology and Optogenetics Laboratory at the Institute of Nanosystems, Universidad Nacional de San Martín (UNSAM), Argentina. She earned her degree in Biological Sciences and her PhD in Biological Chemistry at the University of Buenos Aires, studying the structure of the PKA regulatory subunit. As a postdoctoral fellow and later CONICET researcher at the Instituto Leloir, she focused on how bacteria sense light, uncovering the molecular mechanisms of the Brucella blue-light histidine kinase and of bacteriophytochromes, red/far-red photoreceptors. Her work combines X-ray crystallography, cryo-EM, biochemistry, and spectroscopy. Her lab now also develops optogenetic tools based on bacteriophytochromes to control proteins of biotechnological and clinical interest. She has led national research grants, supervises doctoral and undergraduate students.

DRAFT - UNDER REVIEW



Marcelo A. Martí

UBA / CONICET, Argentina

Marcelo A. Martí holds a PhD from the University of Buenos Aires (UBA), is a Senior Researcher at CONICET and an Associate Professor in the Department of Biological Chemistry at the Faculty of Exact and Natural Sciences, UBA. He also serves as Secretary of Graduate Studies at the faculty. His work focuses on bioinformatics, with applications in protein biochemistry and clinical genomics. He combines molecular simulations and genomic data analysis to study protein function, contribute to drug design, and improve the interpretation of variants associated with rare diseases.

He has authored more than 200 scientific publications and led numerous national and international research projects, with a strong commitment to training researchers. He contributed to the creation of bioinformatics courses and the UBA School of Clinical Genomics. His distinctions include the 2018 Houssay Award and recognition in MIT Technology Review's TR35 Argentina-Uruguay in 2012. He is a scientific co-founder of Bitgenia, an Argentine genomics and bioinformatics company focused on diagnostic and precision medicine applications.



Esteban Mocskos

UBA / CONICET, Argentina

Esteban Mocskos is an Assistant Professor in the Department of Computer Science at the Faculty of Exact and Natural Sciences, University of Buenos Aires, and an Independent Researcher at CONICET. He directs the Interdisciplinary High-Performance Computing Laboratory (LICAR), whose research spans computer architecture, distributed systems, and large-scale applications.

He participates in the management and operation of HPC infrastructure at the Center for Computational Simulation for Technological Applications (CSC-CONICET), which hosted the TUPAC supercomputer for several years. He currently coordinates the governance committee of Clementina XXI, the high-performance computing system that marked Argentina's first entry in the TOP500 ranking.

His experience combines the acquisition, administration, and management of high-performance computing systems, the coordination of technical teams, and research in computer architecture and distributed systems.

He has also contributed to the regional and international HPC community as a co-founder and Steering Committee member of CARLA (the Latin American and Caribbean High-Performance Computing Conference) and as General Co-Chair of ISCA 2024, the leading international computer architecture conference, held in Latin America for the first time in its 50 editions.



Jiri Vondrasek

Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences (IOCB Prague); Head of ELIXIR CZ, Czech Republic

Head of the Bioinformatics Group at the Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences (IOCB Prague) and Head of ELIXIR CZ. His research focuses on structural bioinformatics, computational structural biology, molecular modeling, protein evolution, and biomolecular interactions.

DRAFT – UNDER REVIEW



Gavin Farrell

ELIXIR, United Kingdom

Technical Officer at ELIXIR, supporting the activities of the Data, Interoperability and Compute Platforms. He holds a BSc in Biotechnology and an MSc in Biomedical Genomics, with research experience focused on the gut microbiome and colorectal cancer.



John Hancock

ELIXIR, United Kingdom

ELIXIR Communities and Services Coordinator. He promotes collaboration between ELIXIR Nodes and the ELIXIR Hub and leads the development of research communities associated with ELIXIR use cases. He has extensive experience in computational biology, bioinformatics leadership, and research infrastructure management.

DRAFT – UNDER REVIEW

01001
00100
10100

Short talks

10100
10011
01000

SESSION 1: Proteomics and Protein Structure and Function

SHORT TALK #127

In silico prediction of (p)ppGpp alarmone binding to DksA and NodD1 in Bradyrhizobium diazoefficiens: a possible link between the stringent response and nodulation gene expression

Source: 04_short_talks/ST127.tex · Excel row 128

Falduti O[1], Torres Tejerizo G[1], Pérez Jiménez J[1], Rodriguez Sawicki L[2,3]

[1] Instituto de Biotecnología y Biología Molecular (IBBM), Facultad de Ciencias Exactas, Universidad Nacional de La Plata [2] Structural Bioinformatics Group (SBG), Department of Science and Technology, National University of Quilmes, Argentina [3] Instituto de Investigaciones Bioquímicas de La Plata (INIBIOLP), CONICET, Facultad de Ciencias Médicas, Universidad Nacional de La Plata

Bradyrhizobium diazoefficiens is a soil bacterium that forms nitrogen-fixing nodules on soybean roots. Nodulation genes are controlled by NodD1, a transcriptional regulator that, upon sensing plant released flavonoids, activates the nodYABC operon, whose products synthesize the Nod factors that trigger nodule organogenesis. During this process, rhizobia may face nutritional stress, which triggers the stringent response (SR), mediated by the alarmones (p)ppGpp. In B. diazoefficiens, under nutritional stress, pGpp levels increase while ppGpp levels decrease. In Escherichia coli, the SR relies on DksA, a small transcription factor that binds the secondary channel of RNA polymerase (RNAP): (p)ppGpp binds directly to DksA, and the complex alters the stability of RNAP-promoter complexes, amplifying the transcriptional response. Alarmones can also bind other regulatory proteins, modulating their activity according to the metabolic state of the cell. Understanding how nutritional signals are integrated with symbiotic gene expression is key to improving biological nitrogen fixation in soybean crops. This work aims to provide computational evidence for a molecular link between nutritional stress signaling and the regulation of symbiotic nodulation genes in B. diazoefficiens. To predict whether these alarmones interact with DksA and/or NodD1, we performed in silico analyses. The three-dimensional structures of both proteins were modeled from their amino acid sequences using AlphaFold Colab, and molecular docking was carried out with UCSF Chimera 1.19. In addition, the domain architecture of NodD1 was analyzed using Pfam and the Conserved Domains Database. We predicted a putative binding site for both pGpp and ppGpp on DksA (ΔG = −5.973 and −6.265 kcal/mol for pGpp and ppGpp, respectively), formed by similar amino acid residues in both cases. NodD1 contains a LysR-type domain with significant homology to type 2 periplasmic binding proteins, which has been reported to bind ppGpp. Molecular docking predicted a putative ppGpp binding site in NodD1 (ΔG = −7.228 kcal/mol), whereas no pGpp binding site could be predicted with the same approach. Taken together, these preliminary results suggest that one or both alarmones may promote nodulation gene transcription in B. diazoefficiens, either through DksA or through direct interaction with NodD1. Further in vitro and in vivo experiments will be required to validate these predictions.

DRAFT – UNDER REVIEW

SHORT TALK #129

A Novel Experimentally-derived Dataset for Structural and Sequence Analysis of Disordered Flexible Linkers

Source: 04_short_talks/ST129.tex · Excel row 130

Glavina J[1,2], Meng D[3], García-Alvarez HM[1,2], Leonetti CO[4,5], Pollastri G[3], Chemes LB[1,2]

[1] Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín (UNSAM) – Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) [2] Escuela de Bio y Nanotecnologías (EByN), Universidad Nacional de San Martín [3] School of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland [4] Universidad Abierta Interamericana, Facultad de Tecnología Informática, Buenos Aires, Argentina [5] Universidad Abierta Interamericana, Centro de Altos Estudios en Tecnología Informática, Buenos Aires, Argentina

Disordered flexible linkers (DFLs) are intrinsically disordered protein regions that function by connecting domains and/or short linear motifs, regulating the spatial positioning of the elements they connect. Despite their functional importance our understanding of DFL features is limited by a lack of comprehensive datasets. The DisProt database, considered the gold standard in the field, contains fewer than 400 linkers. Here, we develop a workflow for the systematic identification of DFLs connecting folded domains from protein structures and create a comprehensive dataset known as the DLD dataset.

First, we downloaded all structurally defined domain annotations from SCOP2. We improved the quality of the dataset by ensuring continuity between domains, discarding chimeric proteins, and restricting to PDB chains with two or more domains of 30 or more residues from high-resolution (< 5 Å) structures. Next, we annotated secondary structure at the residue level using DSSP and performed a two-step smoothing process that improved the annotation of longer DFLs. A second strategy to recover longer linkers was the annotation of missing residues as coils. This strategy enabled the redefinition of domain boundaries and the extraction of over 2000 DFLs in addition to loops, disordered N- and C-termini and folded domains.

T-SNE visualization of different regions using ProtTrans sequence embedding revealed that DFL features overlapped with those of DisProt linkers. DFLs were also clearly distinguishable from other disordered and folded protein regions. A principal component analysis using 13 non-redundant sequence and structural features revealed that disordered regions were distinctly separated from folded regions by hydrophobicity. Within disordered regions, N-termini showed a higher aliphatic fraction, a lower proline fraction, and a moderately higher net charge per residue (NCPR) compared to DFLs. C-termini were primarily distinguished from linkers by a lower proline fraction and higher fractions of charged residues (FCR) and NCPR. The DLD dataset is available at https://dld.chemeslab.org/, a web server where linkers are annotated with sequence and structural features and can be visualized using a structure viewer.

In conclusion, we developed a novel DLD dataset that expands DFL annotation and reveals features that distinguish DFLs from other protein regions. This dataset offers a valuable resource for developing models to predict DFLs at the proteome-scale.

DRAFT - UNDER REVIEW

SHORT TALK #122

Simulation-guided structural modelling of amyloid-like fibrils. From isolated peptides to nanoscale assemblies.

Source: 04_short_talks/ST122.tex · Excel row 123

Barrera EE[1], Faccio R[2], Pantano S[3]

[1] Instituto de Histología y Embriología (IHEM) - Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), CC56, Universidad Nacional de Cuyo (UNCuyo), Mendoza, Argentina. [2] Área Física, Departamento de Experimentación y Teoría de la Estructura de la Materia y sus Aplicaciones (DETEMA), Facultad de Química, Universidad de la República, Montevideo C.P. 11800, Uruguay. [3] Institut Pasteur de Montevideo, Mataojo 2020, 11400, Montevideo, Uruguay.

Aromatic-rich peptides are well known for their ability to assemble into amyloid-like aggregates of ordered cross-β-sheet fibrils, with β-strands running perpendicular to the fibril axis. These structures exhibit special mechanical properties, presenting great potential within the nanomaterials field. Nevertheless, defects in their long-range arrangement may pose difficulties for structural determination, thereby impairing their practical use in nanotechnology. Here, we employed a multiscale molecular dynamics (MD) approach to describe the aggregation mechanism of tetraphenylalanine (Phe4), a short peptide whose fast aggregation kinetics and pseudo-crystalline nature obstruct the acquisition of high-resolution structures. The SIRAH coarse-grained (CG) force-field allowed us to simulate successive rounds of peptide aggregation, starting from monomers in solution, generating aggregation seeds, and observing unbiased amyloid-like fibrillization. We decided to stop the fibril-growing process after 60 μs of accumulated trajectories and the addition of 209 Phe4 monomers, and observed a coaxial architecture with 3 central plus 6 peripheral (3C6P) protofibrils packed against each other with a characteristic helical twist, also presenting branching points. Based on this aggregate, we employed a Cartesian helix parametrization to build an ~13 nm long idealized unbranched model, and proceeded to evaluate the stability of the fibril stoichiometry (3C6P) using the same CG force field and the all-atom (AA) Gromos 54a7 force field. Both CG and AA simulations (1 and 0.3 μs, respectively) confirmed the stability of the fibril model, presenting low RMSD values, not observing dissociation events, and maintaining the original architecture.

Complementarily, we validated our Phe4 aggregate model against experimental X-ray powder diffraction data, showing strong agreement with observed lattice parameters. We further demonstrate that unbranched Phe4 fibrils, the smallest reported to date, can be selectively obtained by harvesting supernatant fractions from low-concentration assemblies, as confirmed by transmission electron microscopy. We consider that this ab initio CG simulation protocol constitutes an attractive and unbiased strategy to generate early aggregation species of biotechnological interest.

DRAFT – UNDER REVIEW

SHORT TALK #121

Molecular dynamics simulations support a dynamic RNA-recognition hotspot in the intrinsically disordered N-terminal domain of the androgen receptor

Source: 04_short_talks/ST121.tex · Excel row 122

Balatti GE[1,2], Farahi N[3,4], Tompa P[3,4,5,6], Loris R[3,4]

[1] Universidad Nacional de Quilmes/CONICET, Roque Sáenz Peña 352, 1826 Bernal, Argentina [2] Universidad de Belgrano, Villanueva 1324, 1426 Ciudad Autónoma de Buenos Aires, Argentina [3] VIB-VUB Center for Structural Biology, Vlaams Instituut voor Biotechnologie (VIB), Pleinlaan 2, 1050 Brussels, Belgium [4] Structural Biology Brussels, Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium [5] Institute of Molecular Life Sciences, HUN-REN Research Centre for Natural Sciences, 1117 Budapest, Hungary [6] HUN-REN Office for Supported Research Groups (TKI), Cell Cycle Laboratory, National Institute of Oncology, 1122 Budapest, Hungary

The androgen receptor (AR) is a ligand-regulated transcription factor whose intrinsically disordered N-terminal domain (NTD) mediates transcriptional activation, interdomain communication and condensate formation. Long non-coding RNAs (lncRNAs), including HOTAIR and SLNCR1, have been implicated in AR signaling, but the molecular basis of their interaction with disordered AR regions remains poorly understood. Here, we used molecular dynamics (MD) simulations to characterize how short lncRNA-derived fragments interact with AR residues 1–100 and to interpret experimental binding data.

All-atom MD simulations were performed for AR1–100 in complex with HOTAIR-20nt, SLNCR1-20nt and a PolyA-20nt control. Unbiased simulations were complemented with docking-derived MD refinements to explore alternative RNA-binding registers. Trajectories were analyzed using contact maps, residue–nucleotide occupancies, qualitative Coulomb and Lennard-Jones profiles, and non-covalent interaction classification. The results were interpreted together with electrophoretic mobility shift assays, microscale thermophoresis, nuclear magnetic resonance perturbations and mutational analysis.

The simulations showed that HOTAIR-20nt and SLNCR1-20nt engage overlapping but non-identical regions of the AR-NTD, consistent with a dynamic and fuzzy RNA-binding mode rather than a single rigid interface. HOTAIR-20nt formed a relatively compact interaction pattern, with contacts around the experimentally mapped RNA-sensitive region and a stronger relative electrostatic contribution during stabilization. SLNCR1-20nt sampled a broader and more heterogeneous contact network involving hydrogen bonds, van der Waals contacts, cation–π interactions and π-stacking. Docking-derived refinements supported recurrent AR hotspots while revealing shifts in the contacted RNA nucleotides, suggesting register-variable recognition. In contrast, the PolyA-20nt control also contacted AR1–100, consistent with nonspecific RNA–IDR interactions, but did not reproduce the same hotspot architecture observed for the lncRNA-derived sequences.

Overall, these results show that MD simulations can help distinguish nonspecific RNA binding from sequence-dependent interaction patterns in dynamic protein–RNA complexes. Our analysis supports a model in which HOTAIR and SLNCR1 bind a flexible RNA-recognition region within the AR-NTD, providing a molecular framework to interpret how lncRNAs may modulate AR interdomain communication.

DRAFT - UNDER REVIEW

SESSION 2: Evolutionary, Comparative and Population Genomics

SHORT TALK #137

Assessing the Impact of Kinship on Population Genomic Analyses in the Jaguar (Panthera onca)

Source: 04_short_talks/ST137.tex · Excel row 138

Pisciottano F[1], Campos MC[1], Gabaldón T[2,3], Saragüeta P[1]

[1] Instituto de Biología y Medicina Experimental (IBYME) - CONICET, Argentina [2] Barcelona Supercomputing Centre (BSC-CNS), España [3] Institute for Research in Biomedicine (IRB), España

The jaguar (Panthera onca) is an emblematic species of the Neotropics and an important target of conservation and reintroduction programs. Genomic characterization of individuals included in these programs is increasingly relevant for assessing genetic diversity, population structure, and inbreeding. However, the inclusion of closely related individuals may influence population genomic analyses and lead to different interpretations of the underlying genetic structure and diversity of a dataset. This is particularly relevant for reintroduction programs, where founder individuals and their descendants may be represented together in genomic datasets.

In this work, we evaluated empirically how the inclusion of related individuals affects the genomic characterization of jaguars. For this purpose, we generated whole-genome sequencing (WGS) data at high depth (>20X) for 9 Argentinean jaguar samples, and integrated them with publicly available WGS data from 18 additional jaguars from Argentina and Brazil. Short reads were processed and aligned against an improved high-quality chromosome-level reference genome obtained from local Argentinean samples.

Using this combined dataset, we compared two analytical schemes derived from the same collection of individuals. The first dataset comprised 27 jaguars, including individuals with first-, second-, and third-degree relationships identified using KING/PLINK. The second dataset comprised 20 unrelated jaguars, generated by excluding individuals involved in first- and second-degree relationships. After variant curation and linkage disequilibrium filtering, we retained 744,569 independent biallelic SNPs for the 27-individual dataset and 627,596 for the 20-individual dataset. We used principal component analysis (PCA) and STRUCTURE to investigate population structure, and estimated genomic heterozygosity (H) and inbreeding coefficients (F) to characterize genetic diversity and inbreeding in both datasets.

By comparing these two schemes, we assessed how the presence of related individuals influences genomic inferences of population structure and genetic diversity, highlighting the importance of accounting for relatedness when analyzing jaguar genomic datasets, particularly those associated with conservation and reintroduction programs.

DRAFT – UNDER REVIEW

SHORT TALK #135

Protein intrinsic disorder is associated with replicon location in the multipartite prokaryotic genome of Sinorhizobium meliloti

Source: 04_short_talks/ST135.tex · Excel row 136

Pagnutti AL[1], Zapata G[1], Erdozain Bagolín SA[2], Fornasari MS[3], Parisi GD[3], Andrade MA[4], Lozano MJ[1], Lagares A[1]

[1] Universidad Nacional de La Plata [2] Philipps-Universität Marburg [3] Universidad Nacional de Quilmes [4] Johannes Gutenberg-Universität Mainz

Bacterial genomes with multipartite architectures provide a great model to study protein evolution across replicons with different ancestries and evolutionary contexts. While intrinsically disordered regions (IDRs) are well characterized in eukaryotes, their distribution, functional implications and evolutionary constraints in bacteria remain poorly understood. In this work, we carried out a genome-wide bioinformatic analysis to evaluate how intrinsic disorder correlates with genomic compartmentalization, functional specialization and translational features in the plant symbiont Sinorhizobium meliloti.

We analyzed a dataset of ~60,000 proteins from the reference strain 1021 and 14 wild-type isolates carrying diverse accessory plasmids. Intrinsic disorder was predicted using AIUPred and integrated with Pfam functional annotations to assess disorder distribution, IDR lengths, and functional domains across replicons. Across all genetic compartments, proteins with higher disorder content tended to be shorter, and long disordered segments were mostly concentrated near the N- and C-termini. Comparative analysis revealed a clear evolutionary gradient: accessory plasmids carry a higher density of disorder and long IDRs compared to the more conserved chromosome and symbiotic megaplasmids. Functional annotation revealed a contrasting pattern in accessory plasmids: despite having lower overall Pfam coverage, their predicted domains were significantly enriched in disordered residues relative to those in core replicons, supporting a distinct contribution of accessory replicons to protein structural and functional diversification. Together, our results reveal that protein disorder is differentially distributed across the S. meliloti genome and associated with replicon-specific features.

Previous studies have reported lower codon usage optimization in IDPs than in ordered proteins. To further investigate the evolutionary implications of protein disorder, we are currently characterizing codon usage in IDRs and ordered regions across the S. meliloti genome, as synonymous codon usage can influence translational efficiency. We are also collecting proteome data to explore whether there is a differential expression pattern between proteins containing IDRs and those that do not.

The integration of disorder, codon usage and proteomic information will provide a broader view of how genomic context may shape the evolution and expression of disordered proteins in bacterial genomes.

DRAFT – UNDER REVIEW

SHORT TALK #130

Bioinformatics applied to the study of the population dynamics of carbapenem-resistant Klebsiella pneumoniae ST258 in Buenos Aires Hospitals.

Source: 04_short_talks/ST130.tex · Excel row 131

Gonzalez-Espinosa F[1,2], Ciminello S[1], Poklepovich T[3], Haim M[3], Gutkind G[1,2], Gross A[4], Reuter S[4], Radice M[1,2], Cejas D[1,2]

[1] - Universidad de Buenos Aires, Facultad de Farmacia y Bioquímica, Instituto IBaViM. Buenos Aires, Argentina. [2] -CONICET. Buenos Aires, Argentina. [3] -Unidad Operativa Centro Nacional de Genómica y Bioinformática (UOCNGyB), ANLIS Dr. Carlos G. Malbrán. Buenos Aires , Argentina. [4] -Institute for Infection Prevention and Control, Medical Center – University of Freiburg. Freiburg, Germany.

Carbapenem-resistant K. pneumoniae ST258 (CR-Kp-ST258) is a globally disseminated high-risk lineage comprising two major genetic clades, CI and CII, associated with capsular loci KL106 and KL107, respectively. CR-Kp-ST258 has circulated in Argentina since 2010, but its population structure and evolutionary dynamics remain poorly understood. Here, we used bioinformatic tools applied to bacterial genomics to investigate the diversity, temporal dynamics, and evolutionary relationships of Argentine CR-Kp-ST258 within a global context.

We analyzed 86 CR-Kp-ST258 isolates collected from two hospitals in the Buenos Aires Metropolitan Area between 2019 and 2022, together with two KPC-2-producing ST258 isolates from 2010. Whole-genome sequencing was performed using Illumina technology. Three additional Argentine genomes and 410 international genomes were included for comparative analysis. Reads were mapped to CP006923.1 using Snippy, recombinant regions were removed with Gubbins, and phylogenetic relationships were reconstructed with IQ-TREE3. Transmission clusters were defined using a $\leq$21-SNP threshold. Genomic characterization was performed using SPAdes, Kleborate, AMRFinderPlus, and MOB-suite.

Two major Argentine populations corresponded to CI and CII, each comprising three transmission clusters. CI was predominantly associated with KPC-2 (n=41) and included the 2010 isolates. Within CI, genomic diversification was associated with recombination and the emergence of KL20- and KL13-defined subgroups, including isolates from 2021–2022. CII (n=47) was mainly associated with the worldwide carbapenemase NDM-1/-5 and, to a lesser extent, KPC. Some Argentine genomes also carried determinants associated with resistance markers to arsenic, silver, copper, and tellurium. In the global phylogeny, CI genomes clustered with isolates from Ecuador and Colombia, whereas CII genomes were closely related to genomes from the United States.

Our genomic data reveal a dynamic CR-Kp-ST258 population in Argentina shaped by lineage replacement, diversification, and recombination. The contemporary population reflects the expansion of an NDM-1/-5-associated CII background alongside the persistence and diversification of KPC-2-associated CI, including a recombinant KL20 lineage. These findings provide genomic evidence of the evolutionary dynamics underlying the changing population structure of CR-Kp-ST258 in Argentina.

DRAFT - UNDER REVIEW

SESSION 3: Genomics, Transcriptomics, Epigenetics and Metagenomics

SHORT TALK #128

Epigenomic and transcription factor profiles distinguish Progesterone Receptor binding sites with and without the canonical motif

Source: 04_short_talks/ST128.tex · Excel row 129

Fariñas CE[1], Ant L[1], Saragüeta PE[1]

[1] CONICET-Instituto de Biología y Medicina Experimental

BACKGROUND. To carry out its transcriptional function in the endometrium, the Progesterone Receptor (PR) binds to regulatory regions (enhancers and promoters) to shape its cistrome. Like other nuclear receptors, the PR recognizes these sites by binding directly to DNA through its specific sequence motif (PRE, Progesterone Response Element).However, it has been reported that it also employs a mechanism called tethering, where it is indirectly recruited to chromatin through interactions with other DNA-binding proteins. To characterize these two binding modes, we used PR ChIP-seq data generated in our laboratory in Ishikawa cells, an endometrial cancer model.

RESULTS. Motif analysis (using the MEME suite) of the experimental PR peaks revealed that many sites lack the canonical PRE motif, suggesting that PR associates with chromatin via tethering compensating for sites with lower affinity. To identify potential factors mediating this non-canonical recruitment, we integrated Ishikawa ChIP-seq data from public databases for various transcription factors and epigenetic marks evaluated in the absence of hormone treatment. Using feature selection algorithms (Boruta), this machine learning approach ranked the top predictors differentiating putative tethering from direct binding sites. Notably, binding sites lacking PREs displayed a higher signal intensity of co-regulators, particularly SOX17, compared to canonical sites.

CONCLUSION. In endometrial cancer, tethering likely constitutes an important mechanism for PR binding. Our computational integration suggests that in regions lacking a PRE sequence, PR recruitment is sustained by a strong local enrichment of specific transcription factors, such as SOX17, which appear as co-localizing partners. These findings propose a model where PR compensates for the lack of direct DNA affinity by relying on protein complexes that are present prior to hormonal stimulation to achieve chromatin binding.

DRAFT – UNDER REVIEW

SHORT TALK #123

An attention-based deep model for RNA secondary structure prediction

Source: 04_short_talks/ST123.tex · Excel row 124

Breggia B[1], Bugnon LA[1], Stegmayer G[1], Milone DH[1]

[1] sinc(i) - Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional (UNL-CONICET)

Computational prediction of RNA secondary structure remains an open challenge in bioinformatics, with important applications in drug design and functional genomics. Current approaches include thermodynamic models, based on free energy minimization, and deep learning (DL) methods. The latter have shown remarkable progress in recent years (SPOT-RNA, E2Efold, MXfold2, Ufold, REDfold and sincFold), although a key limitation persists: learning-based models generalize well within a single RNA family but perform poorly when are tested across families. To address this challenge, we propose a k-mer tokenization schema that guarantees a coarse- to fine-grained approach, based on attention maps. Attention maps are the core elements of transformer models, with the capacity to relate distant elements in a sequence. In our case, the attention matrix will help the model learn how distinct parts of RNA attend to each other whilst generating a two-dimensional array from a one-dimensional sequence. The goal is to provide models with better abstraction capacity to enhance RNA cross-family generalization. With that in mind, we design a hybrid model combining two stages. Firstly, the sequential input is tokenized with different k-mers and shifts, compressing its original length, and for each shift we calculate the corresponding self-attention scores in matrix format. Such matrices represent coarse-grained possible connection density. After expansion and realignment, the second stage processes the data at fine-grained level, with convolutional ResNet2D blocks. The attention blocks are pre-trained with synthetic data to recognize all possible nucleotide connections. Then, the convolutional stage is trained to filter out the ones that do not actually correspond. The final output is a nucleotide-resolution contact matrix. This approach seeks to better capture long-range dependencies and opens the possibility of incorporating self-attention mechanisms analogously to large language models. Preliminary results show that attention layers can memorize complementary segment pairs across different tokenization sizes (k-mers), over a single reduced synthetic dataset in a relatively small number of epochs during supervised pretraining. Moreover, the estimation error barely increases with sequence length beyond the ones seen in training, being able to correctly predict connection densities for sequences up to 10 times the length seen in the train data (see Figure 1).

DRAFT - UNDER REVIEW

SHORT TALK #120

Temporal-Spatial genomic regulation of the human endometrium

Source: 04_short_talks/ST120.tex · Excel row 121

Ant L[1], Le Dily F[2], Saragüeta P[1]

[1] Instituto de Biología y Medicina Experimental (IBYME-CONICET), Buenos Aires, Argentina. [2] Centro de Regulación Genómica (CRG), Barcelona, Spain.

Decidualization is the trans-differentiation of endometrial stromal fibroblasts into decidual stromal cells, a process essential for implantation and pregnancy. This stable change in cell identity involves coordinated transcriptional and nuclear reorganization. To characterize the temporal dynamics of transcriptional regulation and chromatin organization during decidualization, we performed an integrative analysis of transcriptomic and epigenomic data from transformed human endometrial stromal cells (tHESC) at different time points following induction of decidualization.

We treated tHESC cells with a combination of estradiol (E2), cyclic AMP (cAMP) and either progesterone (P4) or the synthetic progestin R5020, or with a control treatment of ethanol (EtOH). We performed RNA-seq (3 replicates) and Hi-C (2 replicates) before treatment and at 1 hour, 3 days and 6 days after treatments, when the cells are expected to be fully decidualized. We applied bioinformatic and large-scale computational approaches, including differential gene expression, clustering and enrichment analyses, chromatin state and topological associated domain (TAD) conformation analysis to characterize temporal changes in gene regulation and genome organization.

Our results show both decidualizing treatments produced highly overlapping responses. At 1 hour we found an enrichment for TNFα/NFκB (tumor necrosis factor alpha/nuclear factor kappa B) signaling for both decidualizing treatments (FDR<0.01). By day 3, most regions that change compartment status (85% of the compartments that changed states) shift from closed (B) to open (A), coinciding with downregulation of cell-cycle genes (adj pval<0.01). This change is followed by a partial closure of the chromatin at 6 days, where 87% of the compartments that changed states did so from an A to a B state. Furthermore, by day 6 cell-cycle programs were no longer differentially regulated, and epithelial-mesenchymal transition (EMT) and extracellular matrix genes were upregulated.

These results reveal dynamic changes in chromatin organization during decidualization and provide a framework to further investigate their relationship with transcriptional regulation. Moving forward, we plan to integrate this data with publicly available HiChIP data, to examine promoter-enhancer interactions and better characterize how changes in chromatin organization relate to transcriptional changes throughout the decidualization process.

DRAFT – UNDER REVIEW

SHORT TALK #134

FusionAR-KB: A Multi-Layered Knowledgebase for Transcript-Specific Architecture and Clinical Interpretation of Gene Fusions in Cancer.

Source: 04_short_talks/ST134.tex · Excel row 135

Orschanski D[1,2,5], Nibeyro G[1,2], Martinetto H[3], Fernandez EA[1,2,4,5]

[1] ScireLab, Fundación para el Progreso de la Medicina, Córdoba, Argentina [2] CONICET, CCT Córdoba, Córdoba, Argentina [3] Laboratorio de Enfermedades Neurodegenerativas, Departamento de Neuropatología y Biología Molecular, FLENI, Buenos Aires, Argentina [4] Universidad Argentina de la Empresa (UADE). Instituto de Tecnología (INTEC). Buenos Aires, Argentina [5] Facultad de Ciencias Exactas Físicas y Naturales, Universidad Nacional de Córdoba, Departamento de Computación, Córdoba, Argentina

Background: Gene fusions are oncogenic drivers and clinically relevant biomarkers for cancer diagnosis, prognosis, and targeted therapy. Although fusion detection is increasingly incorporated into clinical practice, current resources primarily catalog gene-pair events and lack integration of transcriptomic, structural, and clinical information, limiting the interpretation of fusion biological and therapeutic consequences. Methods: FusionAR-KB, integrates genomic, structural, and clinical data from 1,420 RNA-sequencing samples across 25 cohorts representing 14 cancer categories. Transcript-specific fusion architectures were reconstructed to characterize breakpoint diversity, coding potential, reading frame, and protein domain preservation, while clinical relevance was assessed through integration with the CIViC database. Results: FusionAR-KB reconstructed transcript-specific architectures for 72.7K unique gene fusions, revealing that 7.5% displayed multiple breakpoint-transcript architectures and 36% generated distinct functional configurations predicted to alter reading frame, protein-coding potential, or protein domain preservation. Overall, 45.7% of fusions retained protein-coding potential, 3.6% involved kinase-associated genes, and only 0.4% corresponded to clinically relevant fusions. We identified 205 patients harboring clinically relevant gene fusions, including 41 patients with therapeutically actionable fusions that were not reported by CIViC for the corresponding tumor types, highlighting actionable opportunities beyond currently available knowledgebases. Furthermore, FusionAR-KB enables the investigation of associations between tumor fusion burden and clinical outcomes. Conclusions: By integrating transcript architecture, predicted functional consequences, structural features, and clinical evidence into a unified framework, FusionAR-KB transforms fusion catalogs into an interpretable resource for precision oncology, supporting biomarker discovery, therapeutic prioritization, and translational cancer research across diverse populations.

DRAFT – UNDER REVIEW

SESSION 4: Biological Databases and Bioimaging

SHORT TALK #125

COMPUTATIONAL ANALYSIS OF HUMAN IRIS COLOUR: FROM SEGMENTATION TO QUANTITATIVE PHENOTYPING

Source: 04_short_talks/ST125.tex · Excel row 126

Ceballos ML[1,2,3], Modesti NM[1], Volpini X[4]

[1] Forensic Genetics Center, Judicial Services Department, Judiciary of Córdoba, Córdoba, Argentina.
[2] Bioinformatics II Course, Faculty of Engineering, Universidad Católica de Córdoba, Córdoba, Argentina. [3] BSc in Bioinformatics Program, Universidad Siglo 21, Córdoba, Argentina. [4] BSc in Bioinformatics Program; Cell and Molecular Biology and Biochemical Techniques Chairs, Faculty of Engineering, Universidad Católica de Córdoba, Córdoba, Argentina.

In forensic genetics, eye colour prediction is commonly based on broad phenotypic categories such as blue, intermediate and brown. However, intermediate iris colours show greater variability and remain difficult to characterize reliably. To improve phenotypic resolution, quantitative and reproducible characterization of iris colour from real images is required. This study aimed to develop and evaluate a quantitative framework across different imaging conditions.

Iris regions were isolated using an original and previously validated U-Net segmentation model trained on heterogeneous ocular image datasets (independent test: Dice=0.959; IoU=0.922). Chromatic analysis included UPOL (384 standardized images; 128 eyes) and our GenFor dataset of images acquired with informed consent (300 images; 34 eyes, 17 individuals). Iris colour was described using CIELAB, HSV and chromaticity features. Six normalization strategies were compared, and repeated images of the same eye were used to assess chromatic repeatability. Multivariate structure was explored by principal component analysis (PCA), followed by cross-dataset comparison and exploratory clustering.

CLAHE applied to the L* channel showed the best balance between repeatability and preservation of chromatic differences. In GenFor, mean intra-eye CIEDE2000 distance decreased from 3.465 to 2.895, with improved repeatability in 88.2% of eyes. Independent PCA showed that PC1 explained 59.0% of variance in UPOL and 48.9% in GenFor. Despite acquisition-domain differences, PC1 loadings were similar (cosine similarity=0.841; Spearman ρ=0.827), and the ordering of GenFor eyes using the UPOL-derived PC1 was highly correlated with the independently derived GenFor PC1 (ρ=0.984). Exploratory clustering supported K=2 in UPOL and K=5 in GenFor, with high leave-one-subject-out stability for GenFor (mean ARI=0.931).

Overall, these findings support the representation of iris colour as a quantitative, continuous chromatic phenotype rather than predefined categories. Relative chromatic ordering was highly preserved across datasets despite differences in acquisition conditions, whereas absolute colour measurements remained acquisition-dependent. This suggests a common chromatic structure may be retained across acquisition conditions, although dataset-specific calibration is still required. This framework provides a consistent basis for future integration of quantitative iris phenotypes with genetic data for forensic DNA phenotyping.

DRAFT - UNDER REVIEW

SHORT TALK #132

Automatic segmentation of glioblastoma and its hidden tumor infiltration

Source: 04_short_talks/ST132.tex · Excel row 133

Mulet de los Reyes A[1], Jacobo Berlles J[2], Breitburd K[3], Suárez C[1]

[1] Instituto de Física Interdisciplinaria y Aplicada (INFINA), Departamento de Física, FCEyN, UBA-CONICET
[2] Departamento de Computación, FCEyN, UBA [3] Hospital Alemán

Glioblastoma multiforme (GBM) is the most prevalent and aggressive primary malignant brain tumor in adults. Currently, despite numerous clinical efforts to provide an effective combination therapy, this type of glioma continues to carry a poor prognosis and a median overall survival of 15 months. Tumor recurrence is almost inevitable and mainly driven by the microscopic infiltration of tumor cells into the surrounding edematous brain tissue, which is not visible by current imaging technology. The lack of tools able of delineating this infiltration zone prior to initial treatment (typically surgery combined with radiotherapy and chemotherapy) significantly limits therapeutic efficacy. This study develops a deep learning algorithm designed to precisely segment the tumor core (active tumor plus internal necrosis) and, more challenging, the volume of hidden infiltration within the peritumoral edema, relying exclusively on preoperative multimodal magnetic resonance imaging (MRI) data (11 modalities) from the public University of Pennsylvania (UPenn-GBM) dataset. Following an initial pre-processing phase involving registration, standardization, and skull stripping, the segmentation strategy employs a combination of two pipelines based on Swin-UNETR transformers and contrastive networks. The first pipeline was trained on 361 cases to achieve robust segmentation of the tumor core and peripheral edema, while the second pipeline applied transfer learning to 30 cases (using ground-truth data derived from progression or recurrence images) to fine-tune the network and specifically distinguish between infiltrative edema from purely vasogenic edema. The combination of both pipelines generates probability maps and corresponding segmentations (at both voxel and region levels) of the proposed treatment zone (tumor core plus infiltration) and the purely vasogenic edema. The combined model demonstrated high performance in segmenting the proposed treatment zone, with an area under the ROC curve of 99.7 ± 0.3%, a Dice similarity coefficient of 69.7 ± 9%, a sensitivity of 73.2 ± 17%, a precision of 69.2 ± 9.2%, and an overall accuracy of 98.6 ± 0.4%. This combined deep learning-based approach enables the non-invasive identification of hidden GBM infiltration at the time of diagnosis, which can contribute to the design of personalized, computer-guided treatments to delay recurrence and prolong survival.

DRAFT – UNDER REVIEW

SHORT TALK #133

Ordering the Disorder: Updates to the Intrinsically Disordered Proteins Ontology

Source: 04_short_talks/ST133.tex · Excel row 134

Castro Naser XA[1], **Nugnes MV**[1], Bouhraoua KEA[1], Zoubiri M[1], Hoyt CT[2], Tompa P[3], Tosatto SCE[1,4], Aspromonte MC[1]

[1] Department of Biomedical Sciences, University of Padova, Padova 35131, Italy [2] Institute of Inorganic Chemistry, RWTH Aachen University, Aachen, NRW, 52074, Germany [3] Institute of Molecular Life Sciences, HUN-REN Research Centre for Natural Sciences, Budapest 1117, Hungary [4] Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council (CNR-IBIOM), Bari 70126, Italy

Intrinsically disordered proteins and regions (IDPs/IDRs) break the classical link between a stable fold and biological function. Without adopting a fixed three-dimensional structure, they act in many cellular processes like signalling, transcriptional regulation and the formation of biomolecular condensates. Existing ontologies, such as Gene Ontology, allow databases to capture general aspects of protein function, but they were not designed to represent the conformational plasticity, state transitions and context dependence that define disordered proteins. Capturing this complexity in a structured, standardized form is essential for databases to store, compare and query disorder-related data consistently.

To address this gap, the IDP community developed the Intrinsically Disordered Proteins Ontology (IDPO) as a controlled vocabulary for describing the states, transitions, and functional roles of IDPs and IDRs. Here we present a major update of IDPO, substantially expanded and restructured to align with OBO Foundry principles and mapped to the Basic Formal Ontology, ensuring interoperability with other biomedical ontologies and supporting more precise, machine-readable annotations.

The most significant additions target the biology of biomolecular condensates: a new branch describes condensate material states, the transitions associated with liquid–liquid phase separation and maturation, and the roles of IDPs/IDRs in driving and regulating condensate formation. IDPO now also includes SLiM terms derived from the Eukaryotic Linear Motif (ELM) classification, connecting sequence-level interaction modules to processes. The ontology is developed openly at https://github.com/BioComputingUP/idpo and offers a biologically grounded framework to annotate, query and interpret the functional landscape of disordered proteins.

DRAFT – UNDER REVIEW

SESSION 5: Artificial Intelligence and Systems Biology

SHORT TALK #131

Making Models That Matter: How to Build Trustworthy and Useful Systems Biology Models

Source: 04_short_talks/ST131.tex · Excel row 132

John M. Hancock[1], Mihail Anton[2,3], Frank T. Bergmann[4], Irina Balaur[5], Carissa Bleker[6], Thomas C. Collin[7], Oscar Dias[8,9], Elena Domínguez-Romero[10,11], Chris Evelo[10], Gavin Farrell[12], Martina Kutmon[13], Vitor Martins dos Santos[14,15], Anna Matuszyńska[16], Sébastien Moretti[17], Sara Morsy[18], Anna Niarakis[19,20], Marek Ostaszewski[21], Miguel Rocha[8,9], David Safranek[22], William Scott[23,24,25], Rahuman Sheriff[26,27,28,29], Silvio Tosatto[12], Dagmar Waltemath[30], Ulrike Wittig[31], Jan Zrimec[6], Anže Županič[6]

[1] Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia [2] Department of Life Sciences, Chalmers University of Technology, Gothenburg SE-412 96, Sweden [3] ELIXIR, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, United Kingdom [4] BioQUANT, Heidelberg University, Heidelberg, Germany [5] Luxembourg National Data Service, Luxembourg [6] Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia [7] Frontiers Media SA, Publishing Development, Lausanne, Switzerland [8] Centre of Biological Engineering (CEB), University of Minho, Portugal [9] LABBELS Associate Laboratory, Braga, Guimarães, Portugal [10] Department of Translational Genomics (TGX), Maastricht University, Universiteitssingel 40, 6229 ER Maastricht, the Netherlands. [11] Université de Lorraine, Laboratoire d'Ingénierie des Biomolécules (LIBio), Nancy F-54000, France. [12] Department of Biomedical Sciences, University of Padova, Via Ugo Bassi 58/B, 35131 Padova, Italy [13] Maastricht Centre for Systems Biology and Bioinformatics (MaCSBio), Maastricht University, The Netherlands [14] Dept Bioprocess Engineering, Wageningen University and Research, The Netherlands, [15] LifeGlimmer GmbH, Berlin, Germany [16] Computational Life Science, Department of Biology, RWTH Aachen University, Germany [17] SIB Swiss Institute of Bioinformatics, Vital-IT group, Lausanne, Switzerland [18] Faculty of Health, Medicine, Life Sciences, University of Hertfordshire, U.K. [19] Molecular, Cellular and Developmental Biology Unit (MCD), Centre for Integrative Biology (CBI), University of Toulouse, UPS, CNRS, Toulouse, France [20] Lifeware Group, Inria, Saclay-île de France, Palaiseau, France [21] Luxembourg Centre for Systems Biomedicine, University of Luxembourg [22] Masaryk University, Faculty of Informatics, Brno, Czech Republic [23] Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN, United States [24] UNLOCK, Wageningen University & Research, Wageningen, The Netherlands [25] Laboratory of Systems and Synthetic Biology, Wageningen University & Research, Wageningen, The Netherlands [26] European Bioinformatics Institute, European Molecular Biology Laboratory (EMBL-EBI), Cambridge UK [27] Department of Surgery and Cancer, Faculty of Medicine, Imperial College London, London, UK [28] Department of Medicine, Division of Pulmonary Systems Medicine, University of Florida, Gainesville, USA [29] Earlham Institute, Norwich, UK [30] Department of Medical Informatics, University Medicine Greifswald, Greifswald, Germany [31] Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

Computational models supporting mechanistic understanding of (complex) biological systems, systems behavior prediction, and experimental design are becoming more and more embedded in research on complex biological systems. Reuse and refinement of models, rather than continuous reinvention, is becoming increasingly important as models' demands on computational infrastructure increase. However published models - despite the variety of efforts taken so far - are frequently difficult to reproduce or reuse, substantially limiting their scientific value. Here we address the requirements for model reusability in the light of the field-specific CURE framework (Credible, Understandable, Reproducible, Extensible) and the more general FAIR principles (Findable, Accessible, Interoperable, Reusable). Considering published guidance we identify broad agreement on requirements for findability, accessibility, and interoperability, but continued lack of clarity and consensus around reusability. Focusing on the scientific quality and usability of computational models we discuss six key practices underpinning model sharing and re-use. Mapping the FAIR and CURE principles onto the model lifecycle we propose ten recommendations for building and sharing systems biology models that are both FAIR- and CURE-compliant.

DRAFT - UNDER REVIEW

SHORT TALK #126

Network organization of protein binding-site landscapes across scales

Source: 04_short_talks/ST126.tex · Excel row 127

Giordano G[1,3,4], **Chernomoretz A**[1,2,3]

[1] Depto Fisica, FCEN UBA, [2] INFINA (UBA/CONICET), [3] Fundacion Instituto Leloir, [4] IIBBA (CONICET)

Background: Protein binding sites provide a local description of molecular recognition that may remain conserved even when the proteins hosting them are globally dissimilar. We asked whether protein language-model embeddings can capture these local environments and whether their organization reveals biologically meaningful structure beyond conventional sequence- and domain-based classifications.

Methodology: Experimentally resolved protein–ligand binding sites from BioLiP2/PDB were represented by averaging ProtT5 residue embeddings over the residues defining each site. Cosine similarity in embedding space was compared with structure-based local alignments using ProBiS. A mutual k-nearest-neighbor network was then built, with binding sites as nodes and similarity relationships as edges. Network organization was analyzed across scales using ligand annotations, InterPro homologous superfamilies, Infomap communities, kinase-focused subnetworks, and a filtered network connecting highly similar binding sites hosted by globally distant proteins.

Principal results: Binding-site embedding similarity increased with local structural similarity. At the global scale, the network showed strong homophily by ligand class. At the mesoscale, Infomap communities subdivided conventional structural and evolutionary classes into more compact binding-site niches. Homologous superfamilies displayed distinct landscape organizations, ranging from several discrete compact niches to more continuous diversification around conserved structural frameworks. Also at this level, we considered the example of kinases to illustrate how intra-protein binding-site communities detected in embedding space can separate canonical ATP-associated, back-pocket, distal and allosteric environments, while also resolving multiple recognition microenvironments within the broader ATP-binding cleft. Finally, after excluding relationships attributable to global protein similarity or shared InterPro annotations, highly similar binding sites still connected proteins from different structural and functional families.

Conclusions: Protein language-model embeddings provide an informative representation of experimentally defined binding sites, and their network organization captures a multiscale structure of molecular recognition complementary to whole-protein classifications. This framework can reveal recurrent local recognition environments and unexpected relationships across distant regions of protein space.

DRAFT - UNDER REVIEW

SHORT TALK #136

From human DNA to plant regulatory regions: transfer learning with a Genomic Large Language Model

Source: 04_short_talks/ST136.tex · Excel row 137

Pedraza Pérez LA[1,2,3], Bellora N[2,4]

[1] Laboratorio de Genómica Computacional, Instituto de Tecnologías Nucleares para la Salud (INTECNUS), San Carlos de Bariloche, Argentina [2] Centro Atómico Bariloche, Comisión Nacional de Energía Atómica (CNEA) [3] Instituto Balseiro, Universidad Nacional de Cuyo [4] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Background: Mapping regulatory regions experimentally is costly and specific to each species and condition. Here we address the challenge of predicting them from the sequence. Genomic large language models (GLLMs) capture long-range sequence patterns, but most were pretrained on animal genomes and their transfer to plants is poorly explored. We asked whether HyenaDNA, a long-context GLLM with single-nucleotide resolution pretrained on the human genome, can be adapted to predict regulatory regions in Arabidopsis thaliana, and whether it transfers to plant species not seen during fine-tuning.

Methods: Regulatory and non-regulatory regions (50 to 1,000 bp) were defined by the presence or absence of transcription factor (TF) binding in chromatin immunoprecipitation sequencing (ChIP-seq) data across tissues and conditions. We replaced the language-modelling head of the pretrained HyenaDNA-tiny checkpoint (16 kb context) with a per-base linear classifier. Each base of the TAIR10 genome was labeled as regulatory, non-regulatory or unannotated (masked from the loss). A contiguous region per chromosome was held out for testing. The model was fine-tuned on 16 kb windows with a class-weighted cross-entropy loss; the learning rate was tuned and robustness assessed by leave-one-chromosome-out (LOCO) cross-validation. Genome-wide region calls came from a sliding-window scan retaining the central 8 kb of each window. Finally, the Arabidopsis model was applied without retraining to accessible chromatin regions (ACRs) of seven plant species that diverged 15 to 150 million years ago.

Results: On the held-out test set the model reached an area under the receiver operating characteristic curve (AUC-ROC) of 0.94, an area under the precision-recall curve (AUC-PR) of 0.88 and, at a 0.5 threshold, an F1 score of 0.78 (precision 0.68, recall 0.91). LOCO cross-validation yielded a mean AUC-ROC of 0.91 and AUC-PR of 0.73. Genome-wide, the post-processed calls recovered 90% of annotated regulatory bases. In other species (recall only, as no negative set exists), the model detected 46% (cotton) to 81% (rice) of ACRs, compared with 95% of Arabidopsis test regions.

Conclusions: A GLLM pretrained on human DNA was successfully fine-tuned to predict regulatory regions in plant genomes. The genome-agnostic framework yields single-base regulatory scores that agree with experimental TF occupancy, and it transfers across a wide phylogenetic range with no monotonic decline with divergence time.

DRAFT - UNDER REVIEW

SHORT TALK #124

Monosemantic Decomposition of Protein Language Models via Sparse Autoencoders for Interpretable Missense Variant Classification

Source: 04_short_talks/ST124.tex · Excel row 125

Britez S[1,2], Marino-Buslje C[1,2], Fernández M[2,3,4]

[1] Fundación Instituto Leloir. Buenos Aires, Argentina [2] Instituto Tecnológico de Buenos Aires. Buenos Aires, Argentina [3] Departamento de Física, FCEyN, UBA, Buenos Aires, Argentina [4] CONICET - UBA, Instituto de Física Interdisciplinaria y Aplicada (INFINA), Buenos Aires, Argentina

It is estimated that only 2% of the 4 million single nucleotide variants (SNVs) in the human genome have been clinically classified, with the vast majority falling into the category of Variants of Uncertain Significance (VUS). To overcome this barrier, it is crucial to develop computational methods capable of inferring the impact of these variants at scale.

Currently, protein Language Models (PLMs) have enabled significant advances by learning physicochemical and structural rules through self-supervised training. However, these models suffer from poor interpretability: their vector representations contain overlapping information, with individual dimensions responding to multiple features that often lack clear biological meaning. Sparse Autoencoders (SAEs) address this limitation by decomposing representations into sparse, largely monosemantic features, isolating individual dimensions that respond to specific biological properties. This approach enables researchers to understand and validate the features driving model predictions, thereby improving the biological interpretability of PLM-based inferences.

In this work, we analyzed 37,902 ClinVar variants (20,940 pathogenic, 16,962 benign) in 2,852 human proteins. For each SNV, we extracted the 16,384 SAE features of the ESMC-6B AI language model specialized in biology and proteins. Subsequently, a linear probe evaluated with stratified k-fold cross-validation reached an AUC of 0.827 ± 0.023 (F1 0.759 ± 0.028), and 0.835 (F1 0.752) on a held-out set.

In another analysis, the five features with the largest attribution in each variant were retained for descriptive evaluation. Of the 13,718 distinct features in these positions, 1,221 ranked among the top five in at least 30 variants and accounted for 49.2% of the observations. Of these 1,221 features, 475 (38.9%) ranked mostly in pathogenic variants, 454 (37.2%) mostly in benign variants and 292 (23.9%) in both. Notably, predictions were rarely driven by a single feature: on average, the top-ranked feature accounted for 13% of the total attributions for each variant, and the ten top-ranked for 55%. Only in 100 out of 37,902 variants did a single feature exceed 50% of the attribution. While language models excel at predictive tasks, interpretability is fundamental to bridge the gap between performance and biological understanding. Further analysis of the biological meaning of these features may reveal which protein properties underlie variant pathogenicity

DRAFT - UNDER REVIEW

SHORT TALK #138

When perturbation queries fail: static gene and protein embeddings outperform simulated scFM perturbation responses

Source: 04_short_talks/ST138.tex · Excel row 139

Yanovsky N[1], Chernomoretz Ariel[1,2]

[1] Fundacion Instituto Leloir [2] Departamento de Fisica FCEN UBA

Predicting how cells respond to perturbations has drawn growing attention in AI for biology over the past few years, particularly for its potential to advance drug discovery through therapeutic target identification and treatment-response prediction, while also clarifying gene function. However, it remains unclear how well models trained on observational expression data capture the consequences of intervention. A recent framework achieves strong regulatory-link prediction by simulating perturbations in single-cell foundation models (scFMs), changing one gene's input expression and tracking how the model's outputs respond. We adapt this framework to ask whether the same query-derived features transfer from identifying regulatory links to predicting experimentally measured knockdown responses. Using two genome-scale CRISPRi Perturb-seq screens in HepG2 and Jurkat, we extract zero-out, virtual-value and gradient-trajectory features from scGPT, scPRINT and TranscriptFormer-sapiens, and compare them with static gene embeddings, ESM2 protein embeddings, control-cell co-expression, expression statistics and random features. Every representation is passed through the same small neural network to predict the perturbation-specific response after removing the mean response across training knockdowns, with evaluation on held-out groups of perturbations that have similar measured effects. In both cell lines, static gene and protein embeddings substantially outperform output-query features, which perform around or below the co-expression baseline, while the tested scGPT fine-tuning procedure provides no detectable improvement. Hidden-state analysis in scGPT and scPRINT shows that perturbation identity is recoverable from the internal representations of the models, and, when used to predict measured responses, these internal features outperform output-query, although still fall short of static gene embeddings. Our results show that encoding an intervention is insufficient to predict its biological consequences, and that strong regulatory-link prediction does not by itself establish the transferability of simulated perturbations to response prediction

DRAFT · UNDER REVIEW

SESSION 6: IDPfun2 Special Track

SHORT TALK #141

Transposable elements as evolutionary substrates of protein disorder in the human proteome

Source: 04_short_talks/ST141.tex · Excel row 142

Mac Donagh J[1], Vergesio N[1], Aguilar A[2], Nores R[3], Lagares A[4], Fornasari MS[1], Parisi G[1]

[1] Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Buenos Aires, Argentina. [2] Instituto de Investigaciones en Microbiología y Parasitología Médica, Facultad de Medicina, Universidad de Buenos Aires – Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPAM, UBA-CONICET), Buenos Aires, Argentina. [3] Instituto de Antropología de Córdoba (CONICET/UNC), Departamento de Antropología, Facultad de Filosofía y Humanidades, Universidad Nacional de Córdoba, Córdoba, Argentina. [4] Instituto de Biotecnología y Biología Molecular (CONICET CCT La Plata) Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina.

Intrinsic disorder is usually considered as an inherited property of the proteins, but not as the outcome of an insertion event with an identifiable donor. In this work, we show that disorder is repeatedly injected into already existing human proteins, and that donors are our own genome DNA “parasites”. Transposable elements (TEs) are the most abundant insertional sequences in the human genome and carry strong compositional biases; when they are exonized they do not deliver a neutral sequence, they deliver disorder.

To achieve this, we mapped TE-derived segments onto every Homo sapiens protein-coding transcript and isoform, intersecting RepeatMasker/Dfam annotations with the GRCh38.p14 assembly and projecting insertion coordinates onto translated UniProt sequences. Disorder was assigned per residue from AlphaFold pLDDT, PUNCH2-light relative solvent accessibility (RSA)-pLDDT , giving complementary read-outs of disorder and structural context. Insertions were split by TE class and family and by evolutionary age (pre-primate versus primate-specific), and expression was taken from the Genotype-Tissue Expression (GTEx) project.

We found that TE-derived segments are strongly enriched in disorder relative to the non-TE regions of the same proteins: transposition is a recurrent and quantifiable supply line of new intrinsically disordered regions (IDRs). What an insertion delivers is dictated by its donor, since GC-rich SINE/Alu segments and AT-rich LINE and LTR segments produce distinct structural outcomes, so the compositional bias of a selfish element is written directly into the proteome. Disorder is not the endpoint either: recent, primate-specific insertions preferentially yield disordered segments, whereas older insertions occupy ordered contexts, meaning TE-derived disorder is a transient state that surviving insertions leave behind. The transition carries a cost, as isoforms bearing TE-derived segments are expressed below their TE-free counterparts, most sharply when the insertion is young and disorder-rich.

With this approach, we systematically mapped TE insertions, dated them relative to the primate divergence, and unveiled their disorder contribution with a clear progression towards order. TEs act as a disorder source for the human proteome, contributing to the genesis of new IDRs.

DRAFT - UNDER REVIEW

SHORT TALK #140

LINKER-Pred: A Deep Learning Method and Web-Server for the Prediction of Disordered Flexible Linkers in Proteins

Source: 04_short_talks/ST140.tex · Excel row 141

Garcia Alvarez HM[1,2], Meng D[3], Glavina J[1,2], Leonetti CO[4,5], Pollastri G[3], Chemes LB[1,2]

[1] Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín (UNSAM) – Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). [2] Escuela de Bio y Nanotecnologías (EByN), Universidad Nacional de San Martín. [3] School of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland. [4] Universidad Abierta Interamericana, Facultad de Tecnología Informática, Buenos Aires, Argentina. [5] Universidad Abierta Interamericana, Centro de Altos Estudios en Tecnología Informática, Buenos Aires, Argentina.

Disordered Flexible Linkers (DFLs) are unstructured protein regions that connect two folded domains and/or short linear motifs. Tethering by DFLs mediates key functional features of proteins such as multivalency and autoinhibition. Despite their biological significance, accurate identification of DFLs remains challenging due to limited experimental annotations and the scarcity of dedicated prediction tools. Here, we introduce LINKER-Pred, a publicly available web server for residue-level DFL prediction.

LINKER-Pred predictors apply convolutional neural networks to protein language model representations and were trained on a novel curated dataset comprising structurally defined linkers connecting folded domains (DLD dataset) and experimentally annotated DisProt linkers. LINKER-Pred2 combines ProtTrans and MSA-Transformer embeddings with an ensemble CNN framework and achieves performance comparable to or exceeding leading methods on the independent CAID2 and CAID3 benchmarks. LINKER-Pred-Lite uses only ProtTrans embeddings, improving speed and thus enabling rapid multi-sequence predictions while maintaining competitive predictive accuracy. Both methods are implemented through an accessible web interface for prediction, visualization and download of residue-level linker scores, providing a scalable solution for DFL annotation across proteomes. The LINKER-Pred web server and associated resources are freely available at https://linkerpred.chemeslab.org/.

DRAFT – UNDER REVIEW

SHORT TALK #139

PonTED: augmenting flexible-linker prediction with protein language models and inter-domain context

Source: 04_short_talks/ST139.tex · Excel row 140

Bugnon LA[1], Monzon A[2]

[1] sinc(i) - Research institute for signals, systems and computational intelligence, UNL - CONICET, Santa Fe, Argentina [2] Department of Biomedical Sciences, University of Padova, Padova, Italy

Flexible linkers are disordered segments that separate adjacent folded domains and let them move relative to each other. They are a functionally distinct subclass of intrinsic disorder. In the Critical Assessment of Intrinsic Disorder (CAID) challenge it was shown that predictors specialized in disorder still do not perform well on identifying linker residues. Because a linker is delimited by the domains around it, a predictor needs a longer context than disorder prediction. Moreover, the data scarcity on this task is notorious, with less than 350 proteins annotated in DisProt.

We present PonTED, a per-residue flexible-linker predictor. A frozen protein language model (ProtT5 or ESM-2 650M) encodes the sequence, and a one-layer transformer head with global attention integrates that context. The transformer head replaces the usual convolutional heads used by earlier linker predictors. With the backbone fixed at ESM-2 650M, it raises average precision (APS) on the CAID3 linker reference from 0.323 to 0.469. We built an in-silico inter-domain training corpus from The Encyclopedia of Domains (TED): 6,814 candidate linkers in 6,236 curated Swiss-Prot proteins. Each segment was trimmed to its disordered core with Boltz-2 pLDDT, and features from UNIPROT to discard intra-membrane segments. Models were selected by 5-fold homology-aware cross-validation (MMseqs2, 25\% identity), with CAID3 targets. A per-protein median filter halves fragmentation, from a median of five predicted segments per reference linker to two.

On the CAID3 linker reference (31 proteins), PonTED with ProtT5 (PonTED-XL) reaches 0.619 APS and 0.917 area under the ROC curve (AUC-ROC), versus the best score in CAID3 that reaches 0.474 APS and 0.897 AUC-ROC. Whole-sequence context and an in-silico inter-domain corpus separate linker residues from the rest of the disordered fraction better than any CAID3 portal entry. We release three variants (Ponte-S, PonTED, PonTED-XL) as a container for offline CPU inference, and will be evaluated in the CAID4 challenge.

DRAFT – UNDER REVIEW

SHORT TALK #142

Structural Diversity and Conformational Plasticity of Plantaricin A Variants

Source: 04_short_talks/ST142.tex · Excel row 143

Navarro ME[1], Tymczyszyn EE[1], Bravo-Ferrada BM[1], Vranken W[2], Parisi G[3], Fornasari MS[3]

[1] Laboratorio de Microbiología Molecular, Universidad Nacional de Quilmes. [2] Bio2Byte, Interuniversity Institute of Bioinformatics in Brussels (IB)2, ULB-VUB. [3] Unidad de Bioinformática Estructural, Universidad Nacional de Quilmes.

Plantaricin A (PlnA) is a key antimicrobial and signaling peptide produced by Lactiplantibacillus plantarum, an intrinsically disordered protein that undergoes a disorder-to-order transition upon membrane association, adopting a more defined structure that facilitates its interaction with the bacterial membrane and contributes to its biological activity. To better understand its structural properties, this study investigates the PlnA variant from the oenological strain Lpb. plantarum UNQLp11 (PlnA4), while also performing a comparative analysis across the four existing PlnA types.

Amino acid sequences of PlnA variants were retrieved from public databases to perform multiple sequence alignments and analyze conservation patterns. Three-dimensional structures were predicted using AlphaFold3, while physicochemical properties were evaluated with ProtParam. Intrinsic disorder and conformational dynamics were assessed using FuzPred, PONDR, and the Bio2Byte tools DisoMine and DynaMine; and liquid-liquid phase separation propensities were analyzed through FuzDrop and CatGranule. Structural comparisons across the four variants were conducted using FoldMason.

Results revealed high primary sequence conservation, particularly within the N-terminal leader peptide and specific regions of the mature effector peptide. Physicochemical characterization of UNQLp11 PlnA4 demonstrated a basic mature peptide enriched in leucine and lysine residues. Structural modeling and disorder analyses revealed substantial conformational plasticity among PlnA variants, while Bio2Byte analyses showed consistently low early-folding propensities. Moreover, all variants exhibited low propensity for liquid-liquid phase separation. Together with previous experimental evidence showing disorder-to-order transitions toward amphipathic α-helical conformations in membrane-mimetic environments, these findings support membrane-context-dependent structural behavior of PlnA4.

Overall, our results suggest that the conserved sequence features and conformational plasticity of PlnA may facilitate its adaptation to different molecular environments, potentially supporting its dual role in signaling and membrane interaction. This comparative structural framework provides new insights into the sequence–structure–function relationships of Plantaricin A and contributes to understanding the molecular basis of its functional versatility.

DRAFT – UNDER REVIEW

Poster presentations

POSTER #1

TRACK: Artificial Intelligence and Systems Biology

Benchmarking molecular representations and sequence-guided evidence transfer in LigQ 2

Source: 05_posters/P001.tex · Excel row 2

Schottlender G[1], Costa M[3], de Kemmeter M[1], Varela Gamarnik S[1], García LG[1], Carvalho de Abreu A[3], Álvarez M[2], Marti MA[2], Fernández Do Porto D[1,2,3]

[1] Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires (FCEyN-UBA)–CONICET, Pabellón 2, Ciudad Universitaria, Ciudad Autónoma de Buenos Aires C1428EHA, Argentina. [2] Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires (FCEyN-UBA), e Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), CONICET, Pabellón 2, Ciudad Universitaria, Ciudad Autónoma de Buenos Aires C1428EHA, Argentina. [3] Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, Belém, PA, Brasil.

Sequence-guided ligand prioritization requires transferring bioactivity evidence from related proteins while preserving target specificity and selecting an appropriate definition of chemical similarity. To guide methodological choices in LigQ 2, we systematically evaluated these design decisions using custom target-specific benchmarks for 61 proteins spanning seven categories from the Directory of Useful Decoys, Enhanced (DUD-E). Active ligands were obtained from a curated Protein Data Bank (PDB)/ChEMBL resource, clustered to reduce near-duplicate redundancy, and divided into 10% known seed ligands and 90% evaluation actives. To avoid the known chemical biases of the original DUD-E decoys, putative inactives were randomly sampled from compounds with no reported activity against either the target or proteins sharing its domains or functional family. Evaluation sets were assembled at a 1:200 active-to-putative-inactive ratio, and enrichment was summarized across five random partitions using protein-category-balanced medians. Among the evaluated molecular representations, extended-connectivity fingerprints with diameter four (ECFP4) and Tanimoto similarity achieved the strongest early enrichment. Functional-class fingerprints with diameter four (FCFP4) and Topological Torsion showed similar profiles, whereas ChemBERTa embeddings did not outperform classical fingerprints. We next generated comparable ligand seed sets from the 2, 3, 5, 10, or 15 closest proteins in sequence space. Using five neighbors produced the best overall enrichment and was optimal in all evaluated protein categories except proteases, for which three neighbors performed best. Neighbor rank was strongly negatively associated with sequence identity (median Spearman's rho = −0.876), supporting the progressive loss of biological specificity as additional proteins were included. Finally, union-based combinations of representations increased recovery at the 99.5th percentile: ECFP4, FCFP4, and Topological Torsion together recovered 63.9% of evaluation actives, 5.8 percentage points more than ECFP4 alone, although precision among retrieved compounds decreased to 41.9%. These results show that intermediate-scale neighbor transfer optimizes evidence diversity and specificity, while representation combinations provide a controllable trade-off between enrichment and total active recovery.

DRAFT – UNDER REVIEW

POSTER #2

TRACK: Artificial Intelligence and Systems Biology

Precision control of microbial cell therapies through AI-driven receptor engineering

Source: 05_posters/P002.tex · Excel row 3

Nguyen TTH[1,2], Grosjean E[3], Pajikos M[1], Cambray G[3,4], Bonnet J[3,5], Cortes J[1]

[1] Université de Toulouse, CNRS, LAAS, France [2] L'Institut national des sciences appliquées de Toulouse, France
[3] Centre de Biologie Structurale (CBS), INSERM U1054, CNRS UMR5048, University of Montpellier, France
[4] Diversité des Génomes et Interactions Microorganismes Insectes (DGIMI), University of Montpellier, INRAE UMR1333, France [5] INSERM ART Synbio, Technology Research Accelerator for Synthetic Biology, France

Microbial cell therapies are emerging as a powerful strategy for targeted disease treatment, yet their efficacy remains limited by the lack of robust control systems enabling precise sensing and context-dependent responses. The Centre de Biologie Structurale (CBS, Montpellier) has recently developed a modular bacterial receptor capable of detecting clinically relevant biomarkers. However, the functional role of the juxtamembrane (JM) linker—a flexible and intrinsically disordered region (IDR) essential for signal transduction—remains poorly understood. High-throughput screening of ~210,000 synthetic JM variants revealed that classical sequence descriptors do not capture the complexity underlying linker functional behavior. In the framework of a collaboration between the LAAS-CNRS (Toulouse) and the CBS, we develop machine-learning-based approaches to understand and predict how JM linker sequences modulate receptor activation. More precisely, we leverage protein language model (pLM) embeddings, which have been shown to efficiently predict structural properties of IDRs. In this study, we demonstrate the capacity of pLMs to forecast the functional properties of JM linkers. Using a pre-trained (generic) ProtT5 model, we got an accurate prediction on linkers' performance up to 90%. We are currently developing a specific model for bacterial IDRs aiming to yield additional improvements. Overall, our studies highlight that pLMs with their rich, high-dimensional representation of protein sequences have a strong potential to reveal the complex sequence-function relationship in disordered protein.

DRAFT – UNDER REVIEW

POSTER #3

TRACK: Artificial Intelligence and Systems Biology

Fine-tuning of the ProGen2 protein language model for the generation of artificial beta-lactamases

Source: 05_posters/P003.tex · Excel row 4

Boggione MJ[2], Pontes CFT[1], Cicekyurt E[1], Tubio G[2], Valencia A[1,3]

[1] Barcelona Supercomputing Center, Barcelona, Spain [2] CONICET, Rosario, Santa Fe, Argentina [3] ICREA, Barcelona Spain

Protein language models such as ProGen2 can generate artificial protein sequences resembling natural ones by learning patterns from large-scale sequence databases. However, generating sequences tailored to a specific protein family requires fine-tuning ProGen2 on family-specific data. Beta-lactamases are bacterial enzymes responsible for resistance to beta-lactam antibiotics and represent an attractive case study for family-specific protein generation, given the well-characterized active site of the TEM-1 beta-lactamase. This work aimed to fine-tune ProGen2 with a beta-lactamase-specific dataset and evaluate the structural and functional plausibility of the generated sequences. A multiple sequence alignment of beta-lactamase homologs retrieved from the Pfam database was used to fine-tune ProGen2 (small, medium and base pre-trained models) for different numbers of epochs. New sequences were generated from the fine-tuned models by varying the nucleus-sampling probability (p) and sampling temperature (t) parameters. Generated sequences were folded using ESMFold, and their predicted local distance difference test (pLDDT) scores were obtained. Secondary structure content was calculated with DSSP (Define Secondary Structure of Proteins). Structural similarity to the wild-type TEM-1 beta-lactamase was assessed using TM-align (RMSD, TM-score). Conservation of active site residues (Ser70, Lys73, Ser130, Glu166, Lys234) was evaluated and frustration values were obtained. The base ProGen2 model fine-tuned for 70 epochs showed the best convergence and was used to generate up to 10,000 artificial sequences. Structural alignment with the wild-type sequence yielded TM-scores up to 0.87 and RMSD values as low as 1.48 Å, indicating high structural similarity. Active site residues, particularly Ser70 and Lys73, were frequently conserved among the top-scoring generated sequences and displayed frustration values consistent with the wild-type enzyme. Fine-tuning ProGen2 with a family-specific dataset enables the generation of artificial beta-lactamases that largely conserve key active site residues required for enzymatic function. These preliminary results support the feasibility of this approach for family-specific protein design, while further studies, including molecular dynamics simulations, are needed to confirm functional activity.

DRAFT – UNDER REVIEW

POSTER #4

TRACK: Artificial Intelligence and Systems Biology

The Generalization Gap in Melanoma AI: Cross-Dataset Reliability and the Challenge of Underrepresented Latin American Data

Source: 05_posters/P004.tex · Excel row 5

Virgens GS[1], Iarussi E[2], Rodrigues T[1], Amaral DT[1]

[1] Universidade Federal do ABC (UFABC), Santo André, SP, Brazil [2] Universidad Torcuato Di Tella, Conicet, Argentina

Melanoma prognosis depends on early detection, and deep learning has shown promising performance for skin lesion classification. However, available datasets represent limited populations and clinical settings, raising concerns about model reliability in underrepresented populations, including Brazil and Latin America. Images from HAM10000, PADUFES-20, and PH2 were standardized using the PreSpot preprocessing pipeline. ResNet50, InceptionV3, and EfficientNet-B0 were integrated into an ensemble combining adaptive probability fusion, class weighting, label smoothing, and test-time augmentation. Across 20 independent runs, performance was assessed for discrimination, reproducibility, probability calibration using Expected Calibration Error (ECE), and bidirectional external validation between the international HAM10000 and Brazilian PADUFES-20 datasets. Internally, the ensemble achieved an accuracy of 0.867±0.011, weighted F1-score of 0.883±0.007, and macro AUROC of 0.955±0.003, outperforming individual architectures ($p < 0.005$). In the multi-dataset configuration, melanoma sensitivity reached 0.96. In contrast, external validation revealed severe and asymmetric performance degradation. Models developed on HAM10000 achieved a macro AUROC of 0.740 on PADUFES-20 (accuracy 0.382), whereas models trained on PADUFES-20 and tested on HAM10000 reached an AUROC of 0.393 (accuracy 0.193). Probability calibration also deteriorated under domain shift, with mean ECE increasing from 0.085±0.016 internally to 0.237 and 0.441 externally. These findings show that strong internal performance and reproducibility do not ensure cross-dataset generalizability or reliable probability estimates. The pronounced degradation when transferring models between international and Brazilian datasets highlights the risks of deploying melanoma AI systems in poorly represented populations. Robust clinical translation requires rigorous external validation and larger, diverse, multicenter datasets, particularly from Brazil and other underrepresented Latin American populations.

DRAFT – UNDER REVIEW

POSTER #5

TRACK: Artificial Intelligence and Systems Biology

emb2bind: a deep learning binding predictor for IDRs based on protein language model and energy embeddings

Source: 05_posters/P005.tex · Excel row 6

Duarte SA[1], Bugnon LA[1], Erdős G[2], Milone DH[1], Stegmayer G[1]

[1] Research Institute for Signals, Systems and Computational Intelligence, sinc(i), FICH-UNL-CONICET, Santa Fe, Argentina [2] MTA-ELTE Momentum Bioinformatics Research Group, Department of Biochemistry, Eotvos Lorand University, Budapest, Hungary

Intrinsically Disordered Regions (IDRs) are protein regions that lack a fixed structure and can play key roles in signaling, regulation, and molecular recognition. Many of them carry out their function by binding partner molecules, so locating these binding regions is therefore essential to understand IDR functions and its link to disease. However, this task remains a challenge. On the one hand, deep learning (DL) models using protein language models (pLMs) embeddings have appeared for IDRs predictions, with outstanding results. On the other hand, the IUPred predictor can provide an energy embedding, that is, an energy-like quantity can be assigned to each residue based on the contacts it makes with other residues in its 3D structure. In this work we present emb2bind: a DL model that combines both pLM embeddings and energy embeddings for predicting binding regions. Its architecture comprises a convolutional layer, residual networks, an adaptive max pooling layer, a dropout layer, and a fully connected classification layer for the output. Training was regularized with early stopping and weight decay. Binding regions were obtained from DisProt data release 2025_12, retrieving the corresponding UniProt sequences and discarding those longer than 4,000 amino acids. Data redundancy was controlled with CD-HIT clustering at 40% sequence identity, and 10% of the resulting clusters were held out as an independent test set, so that no test sequence exceeds 40% identity with any training sequence, preventing data leakage. The remaining development set was partitioned by 10-fold cross-validation under the same constraint, and the resulting models were combined into an ensemble by soft voting. On the independent test set, the emb2bind ensemble reached AUC=0.806 and APS=0.459, showing that combining pLM and energy representations is an effective strategy for protein binding prediction. The ensemble has been submitted to the Critical Assessment of Protein Intrinsic Disorder (CAID4) challenge, and its official results are expected by the end of the current year.

DRAFT – UNDER REVIEW

POSTER #9

TRACK: Artificial Intelligence and Systems Biology

Repositioning approved drugs as α-Synuclein aggregation inhibitors through ligand-based virtual screening

Source: 05_posters/P009.tex · Excel row 7

Alberca LN[1,2], Bárcenas López O[3,4], Talevi A[1,2], Ventura Zamora S[3,4]

[1] Laboratorio de Investigación y Desarrollo de Bioactivos (LIDeB), Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Argentina. [2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. [3] Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, España. [4] Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra, España.

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder worldwide, characterized by progressive dopaminergic neuron loss and consequent motor impairment. No disease-modifying therapy currently exists, and available treatments only manage symptoms with variable efficacy over time. α-Synuclein (α-syn), an intrinsically disordered protein abundantly expressed in the brain, is central to PD pathogenesis through its pathological aggregation in dopaminergic neurons. Consequently, inhibition of α-syn aggregation constitutes a promising disease-modifying strategy. Computational approaches have proven valuable for accelerating drug discovery, reducing the cost and time required to identify novel bioactive scaffolds.

Here, we developed and validated ligand-based classification models for virtual screening aimed at identifying novel inhibitors of α-syn aggregation. A curated dataset of 119 active and 93 inactive compounds was compiled from the literature and partitioned, via a clustering-based strategy, into a training set, two validation sets, and a test set; validation sets were enriched with decoys generated using LUDe1. Using in-house Python scripts combining feature bagging with forward stepwise selection, 3000 linear classifiers based on conformation-independent molecular descriptors were generated. Top-performing classifiers were combined into meta-classifiers and assessed through retrospective virtual screening. The best resulting ensemble model achieved excellent discrimination and early-recognition performance in the validation sets (AUC-ROC $\approx$ 0.98; BEDROC(α=100) $\approx$ 0.84). This ensemble was applied to screen DrugBank 5.1.11, identifying 21 approved drugs as candidates for repurposing as α-syn aggregation inhibitors.

Additionally, we implemented a freely accessible web application that lets users upload molecule sets and obtain aggregation-inhibition predictions (https://a-syn-testing.streamlit.app/).

These results show that our ensemble modeling strategy reliably identifies α-syn aggregation inhibitors, supporting its application in drug repositioning. As a follow-up to this screening, we plan to experimentally validate the top-ranked candidates to confirm their therapeutic potential.

1Alberca LN, Prada Gori DN, Fallico MJ, Fassio AV, Talevi A, Bellera CL (2025) LIDEB's useful decoys (LUDe): a freely available decoy-generation tool. Benchmarking and Scope. Artif Intell Life Sci 7:100129.

DRAFT - UNDER REVIEW

POSTER #10 **TRACK:** Artificial Intelligence and Systems Biology

Characterizing Intrinsically Disordered Protein Regions Using ESM-2 Attention Patterns

Source: 05_posters/P010.tex · Excel row 8

Wolter J[1], Schumbera E[1], Sprang M[2], Andrade M[1]

[1] Johannes Gutenberg-Universität Mainz [2] University Medical Center Mainz

Characterizing Intrinsically Disordered Protein Regions Using ESM-2 Attention Patterns

Intrinsically disordered regions (IDRs) are functionally important protein segments that do not adopt a single stable three-dimensional structure. Their high structural flexibility enables multiple roles in signaling, molecular recognition, regulation, and biomolecular condensation, but also makes them difficult to characterize computationally. Protein language models such as ESM-2 learn contextual sequence representations free of structural supervision and have been shown to represent information related to protein structure, function, and intrinsic disorder.

This project will test whether intrinsic disorder is reflected in the self-attention patterns of pretrained ESM-2 and whether different types of IDRs produce distinct attention signatures. We will obtain experimentally supported IDR annotations primarily from DisProt and compare them with ordered reference regions where available. We will analyze ESM-2 attention matrices across layers and heads using quantitative measures such as attention entropy, local versus long-range attention, within-region attention, and information exchange across ordered–disordered boundaries.

Beyond a binary comparison of ordered and disordered regions, we will use region-level attention signatures to investigate heterogeneity among IDRs. We will apply dimensionality reduction and unsupervised clustering to determine whether IDRs group by similarities in their attention profiles. Resulting clusters will then be tested for enrichment of biological functions, pathways, and cellular processes using resources such as Gene Ontology and Reactome. This will address whether ESM-2 captures functionally meaningful organization within the disordered proteome rather than only general sequence properties associated with disorder.

Potential confounding factors such as amino-acid composition, region length, and sequence position will be considered through matched comparisons where feasible. If the pretrained model yields weak or inconsistent attention patterns, fine-tuning for disorder prediction will provide a complementary strategy and allow comparison of attention before and after task-specific training. The same framework could later be extended to additional protein language models designed to assess whether observed patterns generalize beyond ESM-2.

DRAFT – UNDER REVIEW

POSTER #11 **TRACK:** Artificial Intelligence and Systems Biology

Deep Learning Recurrent Architecture for Intrinsically Disordered Protein Classification

Source: 05_posters/P011.tex · Excel row 9

Gamba JC[1,2], Stegmayer G[1], Gerard M[1]

[1] sinc(i) Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional–UNL/CONICET [2] Laboratorio de Tecnología Inmunológica, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral

Intrinsically disordered proteins (IDPs) challenge the structure-function paradigm by driving essential cellular processes through dynamic conformational ensembles. The difficulty of their experimental characterization makes computational prediction essential. However, this task is hindered by sequence length variability and severe class imbalance, as disordered regions typically represent a minority of residues. Thus, high-fidelity prediction may require capturing both the proximal residue environment and the global macro-architectural context.

This work aims to propose a deep learning (DL) based-model that integrates representations from ESM-2 pre-trained protein language models to classify intrinsically disordered residues in protein sequences, together with a two branch recurrent architecture that explicitly takes into account the local and global context of each amino acid.

The designed architecture, built upon high-dimensional evolutionary representations from ESM-2, bifurcates into two processing streams: a local branch analyzing 21-residue windows via Gated Recurrent Units (GRUs) to capture short-range dependencies, and a global branch processing full-length padded sequences using a bidirectional GRU and dilated 1D convolution to extract macro-scale context. Latent representations are fused and projected through a multi-layer perceptron (MLP) for residue-level binary classification, trained under a dynamic window-based undersampling strategy that enforces a 1:1 ratio of ordered to disordered residues per batch to overcome class imbalance.

The model was trained and evaluated on a single random split of 1,600 DisProt proteins (1,200 train, 200 validation, 200 test), using a fixed set of hyperparameters. This first trial yielded at training an AUC-ROC of 0.95, an accuracy of 0.88, and an average F1-score of 0.85. On the validation set an AUC-ROC of 0.76, an accuracy of 0.73, and an average F1-score of 0.67. Furthermore, on the independent test set, it achieved an AUC-ROC of 0.80, an accuracy of 0.76, and an average F1-score of 0.70.

Integrating ESM-2 embeddings with a compact dual GRU/Conv1D architecture represents a promising first approach for IDP prediction. This relatively small model successfully utilizes recurrent networks to extract both the local and global context of the protein sequence. Future work will focus on expanding the database and optimizing hyperparameters to mitigate overfitting and enhance the model's predictive performance.

DRAFT – UNDER REVIEW

POSTER #12

TRACK: Artificial Intelligence and Systems Biology

Incorporating Chemical Probing Data into Deep Learning for RNA Secondary Structure Prediction

Source: 05_posters/P012.tex · Excel row 10

Zablocki LI[1], Bugnon LA[1], Milone DH[1]

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

RNA molecules control many biological processes, including gene regulation and disease. To understand how an RNA works, we need its secondary structure: the pattern of base pairs that folds the sequence into a functional shape. Predicting this structure from sequence is a long-standing problem in computational biology, important for basic research, medicine, and biotechnology.

Two families of methods exist. Classic thermodynamic methods estimate the structure with the lowest folding energy, using dynamic programming. More recently, deep learning models learn to predict structure directly from labeled sequences. Both have limits. Thermodynamic methods rely on simplified energy models and struggle with long-range interactions. Deep learning models need labeled data, which is scarce: few RNA families have experimentally verified structures. When we test these models on RNA families not seen in training, accuracy drops sharply, mainly because the models memorize patterns from known families instead of learning general folding rules. This lack of generalization is the main obstacle to reliable structure prediction today.

Chemical probing offers a way forward. In these experiments, a chemical reagent reacts with unpaired nucleotides, and reverse transcription reveals this as a per-nucleotide signal. Classic thermodynamic methods already use this signal to constrain the folding search, improving predictions. Our proposal is to provide the same benefit to deep learning models, using probing data as additional supervision for training or extra input.

We are testing strategies to combine probing with sequence. One adds a second prediction head, so the model learns to predict structure and probing reactivity together. Another adds probing values as an extra channel alongside the sequence representation, which can be a one-hot encoding or a dense embedding from an RNA language model. A key obstacle is data availability: no public resource links chemical probing profiles to verified structures. To solve this, we built a curated dataset that aligns probing signals from experimental databases with known structures, with quality control and cross-validation. We use this dataset to train and evaluate our models. We will present in detail how this dataset was built, and show prediction results from the different deep learning strategies we tested.

DRAFT - UNDER REVIEW

POSTER #13

TRACK: Artificial Intelligence and Systems Biology

Identifying inflammatory Programmed Cell Death associated features to predict clinical outcome in patients with diffuse gliomas

Source: 05_posters/P013.tex · Excel row 11

Gonzalez Abad A[1], Vilches J[1], Cairo M[1], Gómez Bergna S[1], Ferrelli ML[1], Amorós Morales LC[1], Pidre ML[1]

[1] Laboratorio de Virología Molecular (LVM), Instituto de Biotecnología y Biología Molecular (IBBM-CONICET), Facultad de Cs. Exactas, Universidad Nacional de La Plata (UNLP)

Gliomas are primary tumors of the nervous system derived from glial cells. Among them, Glioblastoma Multiforme (GBM) is the most aggressive brain tumor in adults, with a median survival of approximately 15 months despite standard treatments. To expand therapeutic alternatives, we aimed to characterize the role of pro-inflammatory Programmed Cell Death (PCD) pathways as potential therapeutic targets for the development of gene therapy strategies.

As a first step, we performed computational analyses of patient biopsy datasets to identify candidate genes. Gene expression analyses of components of the PCD pathways of interest were conducted using data from diffuse gliomas and GBM available in The Cancer Genome Atlas (TCGA), comprising clinical, genomic, and transcriptomic information from 702 patients. Python-based analyses revealed significant differences in the expression of the selected genes when patients were stratified according to their IDH mutational status, a clinically relevant diagnostic and prognostic molecular marker in gliomas.

We then developed a survival prediction model using the Random Survival Forest (RSF) machine learning algorithm, which estimates cumulative risk and survival probability. Gene expression levels were used as input variables. Hyperparameter optimization was performed using GridSearch and K-fold cross-validation, training the model on 75% of the samples and reserving 25% for subsequent testing.

The model showed good predictive performance, achieving a C-index of 0.79. Time-dependent AUCs were also calculated for each input variable, with values close to or above 0.8 for most genes at different follow-up times. The model's mean cumulative AUC was 0.85 and exceeded 0.9 between 200 and 1600 days. All analyzed genes contributed relatively evenly to model predictions, with CASP3 showing the highest mean importance (0.038).

These findings suggest that the expression of genes involved in pro-inflammatory PCD pathways is relevant for risk stratification in patients with diffuse gliomas and GBM, supporting their further investigation as potential candidates for future therapeutic strategies.

DRAFT - UNDER REVIEW

POSTER #14

TRACK: Artificial Intelligence and Systems Biology

Héritas Second Brain: An Expert-Curated Knowledge Graph for Clinical Genomic Interpretation

Source: 05_posters/P014.tex · Excel row 12

Pulicari E[1], Zambrano M[1], Marriscurrena J[1], Barbieri M[1], Itzcovich T[1], Castañon A[1], Vidoz M[1], Papazoglu M[1], Brun B[1], Vazquez M[1]

[1] Héritas S.A., Laboratory of Genomics and Precision Medicine, Rosario, Argentina (a CONICET technology-based company).

Clinical genomic interpretation requires integrating evolving evidence with specialist experience. We developed Héritas Second Brain to support expert analysts in exome analysis, cancer genomics and oncohematology through a shared, evolving knowledge graph (KG).

The KG organizes knowledge into interconnected thematic nodes with source provenance, version metadata and links to related topics. Incremental ingestion updates changed sources and incorporates curated analyst interactions. Node prioritization combines graph centrality, access frequency and links to approved outputs, with a relevance bonus and retention safeguards for analyst-curated knowledge. Agents retrieve information and prepare drafts that analysts review and correct; curated interactions and approved outputs feed subsequent KG updates. Personally identifiable information (PII) is protected through structured-field exclusion, automated text redaction and recurring audits. Feedback content is scrubbed before model processing and KG incorporation. Integrity checks identify knowledge gaps and contradictions.

The institutional KG comprised 259 nodes and 129 links, with 1,454 distinct source references. Thirty-nine analyst-curated conversations had been incorporated into nine thematic nodes. A retrospective audit of 11 recorded exome draft reviews (10 cases) found five approvals without edits, five after editing and one draft held for further review. These preliminary workflow outcomes do not establish diagnostic accuracy.

Héritas Second Brain connects specialist curation with persistent, traceable institutional memory. Initial results document knowledge consolidation and human review of agent-assisted analyses. The analyst-KG feedback loop supports controlled knowledge reuse; larger, prospectively evaluated cohorts are needed to assess effects on interpretation accuracy and turnaround time.

DRAFT – UNDER REVIEW

POSTER #15

TRACK: Artificial Intelligence and Systems Biology

A three-dimensional video-tracking protocol for the assessment of neurobehavioral alterations in adult zebrafish (Danio rerio)

Source: 05_posters/P015.tex · Excel row 13

Zanghelini M[1], Porcaro A[1], Casablanca M[1], Muñoz C[1], Paravani E[1], Bianchi M[1]

[1] Laboratorio de Química Ambiental - Facultad de Ingeniería - Universidad Nacional de Entre Ríos

Zebrafish (Danio rerio) is widely used in behavioral neurotoxicology, but the lack of standardized, low-cost protocols limits comparability across laboratories. This work aims to develop and validate a protocol to assess neurobehavioral changes induced by environmental chemicals in adult zebrafish. A 20 × 20 × 20 cm tank with LED lighting and a 45° mirror placed above it allowed simultaneous recording of the top view (X, Y coordinates) and the front view (Z coordinate) with a single camera. Tank wall configuration and lighting intensity were optimized. Fish were sexed by external morphology and weighed to form homogeneous groups of five same-sex individuals, which were recorded for 10 min after a 10 min acclimation period. Tracking was performed with idtracker.ai, an open-source software based on deep learning, after the classic idTracker showed frequent identity swaps. A Python script specifically developed for this purpose merges both views by matching the Y coordinate of each fish through a correspondence score, converts pixels to centimeters using the tank edges as reference, and computes the following endpoints: dispersion radius (group cohesion), time spent in each vertical third of the tank and thigmotaxis (anxiety-like behavior), total distance and mean speed (locomotor activity), and tortuosity index (motor coordination). Interactive three-dimensional (3D) trajectory plots are also generated. Lighting, rather than wall transparency, was the main determinant of stress-related behavior: strong light induced pecking at the tank walls, whereas dim light caused bottom-dwelling and inactivity. An intermediate condition (3400 lux at the tank front and 7500 lux at the water surface) was adopted as reference. Idtracker.ai maintained individual identities robustly and produced tagged videos for tracking quality control, and the script enabled the reconstruction of individual 3D trajectories. Baseline behavior according to sex and body size, as well as protocol reproducibility, is currently being characterized. The proposed setup and analysis pipeline provide a low-cost tool based on open-source software for the quantitative 3D assessment of zebrafish behavior, designed to be reproducible across experiments. Next steps include establishing baseline behavioral patterns and obtaining dose–response curves for environmentally relevant concentrations of at least one xenobiotic.

DRAFT – UNDER REVIEW

POSTER #16 **TRACK:** Artificial Intelligence and Systems Biology

Phthalate exposure and obesity in adults: a small but robust association, and three leakage mechanisms that inflate it.

Source: 05_posters/P016.tex · Excel row 14

Farneti MB[2,3], Ceschin DG[1,2,3]

[1] Hospital Privado Universitario de Córdoba, Córdoba, Argentina [2] Instituto Universitario de Ciencias Biomédicas de Córdoba (IUCBC), Córdoba, Argentina [3] Centro de Investigación en Medicina Traslacional "Severo R. Amuchástegui" (CIMETSA), U.A. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Córdoba, Argentina

Background. Machine learning applied to the National Health and Nutrition Examination Survey (NHANES) is widely used to test whether phthalates, hypothesised metabolic disruptors, are linked to obesity. Urinary phthalate metabolites are measured only in a one-third laboratory subsample, a deliberate data gap that invites analytic choices, such as imputing the exposure, capable of manufacturing the association being tested.

Methods. We analysed ten NHANES cycles (1999–2018) rebuilt from public CDC files. Obesity was defined as body mass index (BMI) $\geq$30 kg/m^2. Associations were estimated by survey-weighted logistic regression. We trained regularised logistic regression and gradient-boosted tree models with 5-fold cross-validation, evaluated by area under the ROC curve (AUC) with bootstrap confidence intervals (CI), against permutation and demographics-only negative controls. No exposure was imputed and body-composition variables were excluded. We then quantified three leakage mechanisms and audited 210 published NHANES obesity machine-learning studies.

Results. Among 16,035 adults (representing 207.7 million), three of five metabolites were associated with obesity after full adjustment: mono-benzyl phthalate (MBzP) odds ratio (OR) 1.098 (95% CI 1.048–1.149), mono-2-ethylhexyl phthalate (MEHP) 0.857 (0.823–0.893) and mono-isononyl phthalate (MiNP) 0.823 (0.775–0.873). Exposure added ΔAUC (change in AUC) of +0.016 (95% CI +0.010 to +0.023) over demographics. Three mechanisms inflate this small effect: tautological body-composition predictors (ΔAUC +0.345); imputation of exposure (AUC 0.894 in imputed vs 0.567 in measured rows, visible only with non-linear models); and proxy-mediated leakage, where excluding the outcome from imputation but retaining waist circumference (Spearman ρ = 0.948 with BMI) yielded imputed values correlating with BMI at $|\rho| >$ 0.86 versus <0.15 when measured, which a linear model largely misses. Of 210 audited studies, 14.3% mentioned survey design in their abstracts, 2.9% imputation, and none any leakage check.

Conclusions. Phthalate exposure is modestly associated with obesity in adults, with an effect small enough that subsample selection determines its detectability. The same data structure makes artefactual effects easy to produce. Excluding the outcome from imputation is insufficient when a strong proxy remains; exposure variables missing by design should not be imputed.

DRAFT – UNDER REVIEW

POSTER #17

TRACK: Artificial Intelligence and Systems Biology

Probing a genomic language model for off-target, pseudogene and chimeric reads in COI eDNA metabarcoding

Source: 05_posters/P017.tex · Excel row 15

Kaplan M[1], Falduto N[1], Colangelo G[1], Martí M[1]

[1] IQUIBICEN, Universidad de Buenos Aires

Environmental DNA (eDNA) metabarcoding of the cytochrome c oxidase subunit I (COI) gene depends on a set of specialised tools to decide whether a read is genuine: alignment to reference databases to recognise off-target amplicons, reading-frame checks to flag pseudogenes, and reference-based detectors to find chimeras. Each answers a single question and requires curated references. Genomic foundation models offer a different route: trained on raw sequence without labels, they may encode the same biology implicitly. Here we asked whether Evo 2, a 7-billion-parameter genomic language model, already stores the information these classical tools rely on.

We scored 140,951 unique COI reads from two eDNA libraries (Brazilian soil and abyssal sediment), plus simulated chimeras of known junction, and extracted three levels of model output: perplexity (the exponential of the mean negative log-likelihood), per-position base probabilities, and the mean internal embedding. Classical methods provided the reference answers: alignment against eukaryotic and bacterial databases labelled reads that had passed every standard COI filter, and an in-frame stop codon marked pseudogene-like reads. We then tested whether simple linear read-outs of Evo 2 outputs recovered these answers on held-out reads, summarised as the area under the receiver operating characteristic curve (AUROC).

Evo2 outputs recovered the alignment-based origin of 8,573 bacterial-like reads concealed among eukaryotic COI with AUROC 0.99-1.00, and identified pseudogene-like reads with AUROC 0.84-0.87. Perplexity alone missed both (0.58-0.67 and 0.59). Without access to the parent sequences, Evo2 also detected chimeras formed by distant parents (AUROC 0.83) and placed their junction with a median error of 13 bp, against 51 bp expected by chance; chimeras of closely related parents remained undetectable.

A genomic foundation model therefore encodes, position by position, much of the biological knowledge that classical tools apply to eDNA reads: gene origin, reading-frame integrity and chimeric structure. It is invisible in aggregate scores such as perplexity but accessible in the model's internal representations. Showing that it is present is the prerequisite for reference-free and label-free quality control of eDNA data built on foundation models.

DRAFT - UNDER REVIEW

POSTER #18 **TRACK:** Artificial Intelligence and Systems Biology

AI-Based Extraction of Structured Lifespan Data from Scientific Literature

Source: 05_posters/P018.tex · Excel row 16

Rafikova E[1,2], Orden PD[2], Nemirovich-Danchenko N[2], Duhart JM[1], Buslje CM[1], Ceriani MF[1]

[1] Fundación Instituto Leloir, Av. Patricias Argentinas 435, C1405BWE, Buenos Aires, Argentina [2] Open Longevity non-profit organization, 15260 Ventura Blvd, Suite 2230, Sherman Oaks, CA 91403, United States

Introduction. We aim to identify genes with the strongest evidence for affecting lifespan, predict understudied candidates and potentially synergistic gene combinations, and experimentally validate them. The project builds on Open Genes (Rafikova et al., NAR 2024), a manually curated human gene database annotating lifespan experiments with up to 35 parameters, including lifespan, cohort size, sex, strain, housing and genetic manipulation. These data will support rule-based gene ranking by evidence quality, quantity, consistency and reproducibility. Because manual curation cannot keep pace with the literature, we developed an AI-based extractor to keep the database up to date.

Methods. The extractor follows Open Genes curation guidelines and controlled vocabularies. Gemini 3.6 Flash identifies experiment-control comparisons from the PDF and extracts shared experimental information. GPT-6 Sol with code execution locates survival curves, calibrates axes from pixels, and reads lifespan at 90%, 50% and 10% survival and other figure data such as cohort size and sex. Gemini extracts 51 fields per experiment, with textual evidence and confidence scores; supplementary text is provided as context, and unanswered fields are re-queried. Finally, Python checks detect unsupported values, recalculate lifespan changes, and flag inclusion-rule violations. Experiments were matched to Open Genes records and compared across 49 fields; mismatches were curator-annotated.

Results. In 20 mouse-based papers with 46 eligible experiments, the extractor identified 43 (93.5%); all experiment-control pairs satisfied the one-altered-gene and unaltered-control criteria. Correct values were obtained for 86.1% (1,815/2,107) of fields. Overall pipeline performance was reproducible across independent paper samples, with accuracy improving from ~80% to 86% and processing time decreasing from 5.5 to 1.5 hours after pipeline refinement. Accuracy varied from 73% to 98% across papers, with poorer performance mainly associated with supplementary-only data, likely due to their variable, non-standard formats.

Conclusions. The extractor reliably identifies eligible experiments and recovers most data from the main text. Error analysis suggests achievable accuracy of 91–92% with current models. After further refinement and validating it on a larger sample, new data will be collected and manually curated before release by reviewing all low-confidence fields and re-checking a random 10% of papers.

DRAFT - UNDER REVIEW

POSTER #20

TRACK: Artificial Intelligence and Systems Biology

Functional characterization of drug repurposing candidates through integrated analysis of molecular targets and biological pathways

Source: 05_posters/P020.tex · Excel row 17

Gomez M[1], Parizzia A[1], Rossi M[1], Cassol I[1]

[1] Facultad de Ingeniería, Universidad Austral, Pilar, Buenos Aires, Argentina.

Introduction: Transcriptomic approaches can identify and rank candidates for drug repurposing. However, rankings based on expression similarity alone may overlook differences in mechanism of action and pharmacological plausibility. This study builds on candidates identified in a previous analysis of breast cancer cells following ubiquitin-specific peptidase 19 (USP19) knockout and integrates complementary evidence to refine their prioritization.

Objective: To develop an integrative workflow for characterizing these candidates through their molecular targets, biological pathways, and experimental context to support their prioritization for future validation.

Methods: Perturbation metadata, including dose, exposure time, and cell line, are linked to candidate identities and target annotations using source-specific mapping. Chemical compounds are queried against ChEMBL and Guide to Pharmacology (GtoPdb), whereas ligands are mapped using OmniPath and GtoPdb. Retrieved protein targets are harmonized by UniProt accession and linked to encoding genes through HGNC identifiers and approved gene symbols, preserving annotation provenance and mapping uncertainty. Target–pathway relationships from KEGG are integrated into a candidate–protein–gene–pathway network. Within- and across-candidate analyses identify candidate-specific targets, shared pathways, and recurrent biological processes. Experimental conditions and pharmacological evidence inform interpretation of these associations.

Results: The study included 37 candidates: 15 chemical compounds, 18 ligands, and 4 antibodies. Target mapping yielded 24 compound–protein and 59 ligand–protein pairs (83 pairs; 75 proteins corresponding to 75 target genes). Pharmacological records from ChEMBL, GtoPdb, and BindingDB were retrieved for 49 pairs (23 compound–protein; 26 ligand–protein); 34 had no retrieved records (1 compound–protein; 33 ligand–protein). Non-retrieval does not imply absence of pharmacological information. Retrieved records are pending review; antibody analysis is in progress. Of these 75 genes, 70 were associated with at least one KEGG pathway, yielding 185 unique pathways, 130 shared by at least two candidates.

Conclusions: This workflow extends transcriptomic candidate identification toward interpretable functional profiles, supporting mechanistic hypotheses and prioritization for experimental validation. Target and pathway annotations provide supporting evidence, not evidence of therapeutic efficacy.

DRAFT – UNDER REVIEW

POSTER #21 **TRACK:** Artificial Intelligence and Systems Biology

Segmentation metrics for the evaluation of Pfam domain annotation

Source: 05_posters/P021.tex · Excel row 18

Vitale R[1], Milone D H[1], Stegmayer G[1]

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

Protein domains are the structural and functional units of proteins, and databases such as Pfam group them into families. Computational methods for domain annotation are usually evaluated as classifiers, using metrics such as accuracy, precision, recall and F1-score. However, when a full-length protein sequence is analyzed, predicting the correct family is not enough: the method must also locate where each domain starts and ends. Classification metrics do not capture boundary errors, fragmented or merged domains, nor spurious detections, which limits detailed comparisons between methods. Thus, proper segmentation metrics for evaluating annotation methods are needed.

In this work we propose, inspired by metrics from another field, a novel segmentation setting evaluation for proteins, where amino acids are the minimal "units" of evaluation and domains are "events" to detect. At the residue level, a confusion matrix is extended to identify residues that are inserted or deleted, residues that fragment a domain or merge two domains, and residues that reflect boundary imprecision, either by extending (overfill) or shortening (underfill) the domain. At the domain level, reference domains are counted as correctly detected, deleted, fragmented or merged, while predicted domains may be counted as inserted. Since these metrics only assess detection, we also propose family-level classification metrics for domains. A predicted domain is labeled as correct if its family matches that of the reference domain, or as family exchanged otherwise. We applied this framework to compare a Deep Learning (DL) model combining protein language model embeddings with convolutional neural networks versus profile hidden Markov models and BLAST-based annotations on the Pfam v32 database.

Segmentation metrics revealed that several DL models reached high classification scores at the cost of inserting domains absent in reference. DL models also fragmented domains more often than HMMs, which deleted more domains. Family exchanges were rare across methods, indicating that most errors arise from detection.

By incorporating segmentation metrics and separating detection from family classification, the proposed metrics provide a more detailed view of model performance than standard classification metrics alone. They revealed strengths and weaknesses that remained hidden in standard evaluations and supported a more informed choice for methods development and selection in each annotation task.

DRAFT - UNDER REVIEW

POSTER #22

TRACK: Artificial Intelligence and Systems Biology

Modeling the occurrence of toxigenic Aspergillus carbonarius in grapes using regularized regression and neural networks

Source: 05_posters/P022.tex · Excel row 19

Cendoya E[1], Ludueña L[2], **Chiotta ML**[1]

[1] Instituto de Investigación en Micología y Micotoxicología (IMICO), CONICET–Universidad Nacional de Río Cuarto, Río Cuarto, Córdoba, Argentina. [2] Instituto de Investigaciones Agrobiotecnológicas (INIAB), CONICET–Universidad Nacional de Río Cuarto, Río Cuarto, Córdoba, Argentina.

The occurrence of Aspergillus carbonarius in grapes poses a food safety concern, as toxigenic strains can produce ochratoxin A (OTA), a mycotoxin that can be transferred to wine and has nephrotoxic, immunotoxic, genotoxic, and neurotoxic effects, among others. This study aimed to identify climatic variables associated with the incidence of toxigenic A. carbonarius and to evaluate statistical and machine-learning approaches to model and predict its incidence across different wine-growing regions of Argentina. Data collected over five growing seasons from vineyards located in La Rioja, San Juan, Mendoza, Río Negro, and Neuquén were analyzed, considering temperature, relative humidity, and precipitation recorded at different grapevine phenological stages, as well as their cumulative values. Multiple linear regression and regularized Ridge and Lasso regression models were applied, and a neural network was fitted. The multiple linear regression model explained 89.8% of the observed variability (R^2 = 0.898), although multicollinearity among predictor variables was detected. Regularization improved predictive performance, with test-set Mean Squared Error (MSE) values of 21.00 for Ridge and 20.29 for Lasso. In addition, Lasso enabled the selection of a subset of climatic variables associated with fungal incidence. The neural network showed agreement between observed and predicted values, highlighting its potential to capture complex relationships among the variables analyzed. Overall, these results demonstrate the potential of regularized regression and neural networks as complementary approaches for modeling the incidence of toxigenic A. carbonarius and provide a basis for the development and validation of predictive tools to assess the risk of OTA contamination in grapes.

DRAFT – UNDER REVIEW

POSTER #23 **TRACK:** Artificial Intelligence and Systems Biology

AUTISM SPECTRUM DISORDER CLASSIFICATION USING GRAPH NEURAL NETWORKS FROM GUT MICROBIOTA PROFILES

Source: 05_posters/P023.tex · Excel row 20

Leal N[1], Namour J[1], Molina MG[1], Villena J[2], Albarracin L[2]

[1] Laboratorio de Computación Científica, Depto de Ciencias de la Computación, Facultad de Ciencias Exactas y Tecnología, Universidad Nacional de Tucumán. [2] Laboratorio de Inmunología Respiratoria, Escuela de Posgrado de Agricultura, Universidad de Tohoku

AUTISM SPECTRUM DISORDER CLASSIFICATION USING GRAPH NEURAL NETWORKS FROM GUT MICROBIOTA PROFILES

Authors List

Leal N.[1]; Namour J.[1]; Molina MG.[1]; Villena J.[2]; Albarracin L.[2]

Affiliation List

[1] Laboratorio de Computación Científica, Depto de Ciencias de la Computación, Facultad de Ciencias Exactas y Tecnología, Universidad Nacional de Tucumán

[2] Laboratorio de Inmunología Respiratoria, Escuela de Posgrado de Agricultura, Universidad de Tohoku

The gut-brain axis can define the emotions, decisions and actions of a human being. The microbiome is an influential part in our lives. Conventional machine learning methods represent metagenomic profiles as a matrix, leaving behind the relationships and valuable characterization data. Graph Neural Networks (GNN) offer an alternative by representing patients and taxa as interconnected entities to learn directly from this structure. This study uses GNNs for the classification of Autism Spectrum Disorder (ASD) and neurotypical controls using gut microbiome profiles, identifying the different taxa with more relevance in the neurodevelopmental disorder. We used public datasets that have 123 prepubescent patients from China selected by age gap, origin, and drug exposure. Samples were previously analyzed via Illumina sequencing of 16S rRNA V4 region amplicons. Data were processed using QIIME 2 to obtain genus-level taxonomic abundance profiles. A bipartite patient-taxon graph was constructed, where nodes represent patients and taxa, and edges represent taxonomic abundances. Graph Attention Network (GAT) was implemented over this structure to classify individuals. The dataset was split into training and test sets 80:20 using node masking. The model achieved an overall accuracy of 0.84 on the test set. For neurotypical controls, precision was 0.88, recall 0.70, and F1-score 0.78, for ASD, values were 0.82, 0.93, and 0.88, respectively. Analysis via GNNExplainer allowed ranking taxa contributions and identified a subset of nodes with high influence on classification, revealing that the most influential for ASD are: Micrococcaceae, Lachnospiraceae_UCG-003 and Bacillus. Representing microbiota profiles via bipartite graphs combined with attention neural networks is a promising strategy for classifying neurodevelopmental phenotypes. Incorporating explainable AI methods identifies the taxonomic features supporting predictions, enhancing the biological interpretability of the model.

DRAFT - UNDER REVIEW

POSTER #24 **TRACK:** Artificial Intelligence and Systems Biology

Machine learning in species distribution models: A case study between a cactophilic drosophilid and its hosts

Source: 05_posters/P024.tex · Excel row 21

Villarreal RG, Soto IM, Soto EM

Instituto de Ecología, Genética y Evolución de Buenos Aires (IEGEBA – CONICET). Departamento de Ecología, Genética y Evolución. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. Buenos Aires, Argentina

Species distribution models are widely used in spatial ecology to predict the occurrence and habitat suitability of species, and machine-learning algorithms have become increasingly standard tools for this purpose. These models relate species presence data to environmental predictors, typically abiotic, although data on other species can also be included as biotic predictors. Furthermore, predictions from different algorithms can be combined into an ensemble, reducing overfitting and variability while maintaining or improving predictive performance.

In this study we applied three types of machine learning algorithms (MaxEnt, Random Forest [RF] and Support Vector Machines [SVM]) to predict the habitat suitability of the cactophilic species Drosophila antonietae and evaluate the effect of including the distribution of its breeding and feeding cacti hosts (Opuntia monacantha and Cereus hildmannianus). We used 12 abiotic variables (CHELSA 2.0), that were summarized through a principal component analysis (PCA), and four models with the same number of predictors were generated for each algorithm: four components; three components and one cactus species (for each species); and two components with both cactus species. In addition, an ensemble integrating its four models was built for each algorithm in order to compare overall performance among algorithms.

When comparing classification metrics, the inclusion of biotic variables partially improved predictive performance with MaxEnt, whereas with RF and SVM the indices were slightly higher when using abiotic variables only. Ensemble performance was similar among algorithms, with minimal differences and no clearly superior algorithm.

Overall, the inclusion of cactus distributions has different effects depending on the algorithms used to estimate the distribution of D. antonietae. Only when using MaxEnt would the use of cacti distributions improve the estimation. Further studies analyzing the effect of using biotic variables with multiple machine learning algorithms are needed to understand their influence on the species studied.

DRAFT – UNDER REVIEW

POSTER #25

TRACK: Artificial Intelligence and Systems Biology

Simplifications and approximations in molecular systems biology: lessons from the toggle switch

Source: 05_posters/P025.tex · Excel row 22

Alejandro Barton[1], Nara Guisoni[1,2], Luis Diambra[1]

[1] CENEXA, Universidad Nacional de La Plata, CONICET, La Plata, Argentina [2] Instituto de Tecnología, Universidad Argentina de la Empresa, Buenos Aires, Argentina

Background: The toggle switch — a two-gene mutual repression circuit — is a canonical bistable genetic network, extensively modeled using quasi-steady-state (QSS) or Hill-function frameworks that assume promoter states equilibrate instantaneously. These models capture steady-state structure but discard all promoter kinetic information. Cooperative repressor binding can occur via mechanistically distinct routes: a Recruitment Mechanism (RM), where successive binding enhances association rates, and a Stabilization Mechanism (SM), where it reduces dissociation rates. Both yield the same equilibrium constants and are indistinguishable at steady state. Here we introduce a detailed kinetic model for both mechanisms and show that their transient dynamics differ substantially.

Methods: We build a two-level model hierarchy based on Ordinary Differential Equations (ODE). Model I (with 10 ODEs) explicitly tracks all promoter occupancy states for four mechanism combinations (RM-RM, SM-SM, RM-SM, SM-RM). Model II (with 4 ODEs) applies a regulatory function derived via QSS approximation (RQSS), collapsing all combinations of model I onto the same dynamics. We analyze three perturbation protocols from the saddle point. For the asymmetric RM-SM case, we derive an effective regulatory function (Reff) via singular perturbation theory.

Results: We note that promoter relaxation times τprom differ markedly between the two mechanisms: at low repressor concentration, τpromSM is ~34-fold greater that versus τpromRM, converging at high concentrations. In the model I with RM-SM combination we found a kinetic separatrix distinct from the QSS prediction: nearby initial conditions commit to opposite fates based on kinetics alone. Further, we derive an effective regulatory function Reff that captures the kinetic bias analytically, retaining a 4-ODE structure. The promoter of the gene with SM lags behind the increase in the concentration of its repressor during transients, increasing its activity by up to ~35%, enough to redirect fate.

Conclusions: Cooperative binding mechanisms identity are invisible to equilibrium models but decisive during transients. Reff provides a compact, mechanism-sensitive correction to the QSS approximation without additional state variables. Promoter kinetics encode information that RQSS and Hill models systematically discard, with functional consequences for bistable circuit behavior.

DRAFT - UNDER REVIEW

POSTER #27

TRACK: Artificial Intelligence and Systems Biology

A Machine Learning-Derived Glycosignature Associated with Glioblastoma Stem-like Cell Subtypes and Patient Survival

Source: 05_posters/P027.tex · Excel row 23

Castillo JO[1,2], Ferreira GM[1,2], Segatori VI[1,2], Gabri MR[1,2]

[1] Centro de Oncología Molecular y Traslacional, Universidad Nacional de Quilmes [2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

Glioblastoma (GBM) is the most common and aggressive primary brain tumor in adults, characterized by a median overall survival of 15 months despite standard-of-care treatment. This aggressiveness is largely attributed to glioblastoma stem cells (GSCs), which can undergo polarization toward Proneural (PRO) and Mesenchymal (MES) states, with the latter being associated with enhanced invasiveness and therapeutic resistance. Aberrant glycosylation, particularly N- and O-glycans, are well studied to contribute to tumor progression but its specific role within GSCs remains poorly characterized. The objective of this study was to identify a glycosylation signature associated with GSC PRO/MES polarization and patient survival, aiming to uncover novel biomarkers and therapeutic targets. To this end, transcriptomic and clinical data from TCGA (discovery cohort) and CGGA (independent validation cohort) were analyzed, and PRO/MES polarization was quantified as a continuous score using singscore. A two-axis feature selection pipeline combining Boruta (phenotype axis) and Random Survival Forest (survival axis) was applied to 124 KEGG glycosylation-related genes. Then, candidate signatures were selected using repeated 5-fold cross-validation in TCGA against pre-selected performance thresholds, resulting in an 8-gene glycosignature dominated by O-glycosylation-related glycosyltransferases. In addition, developed models trained in TCGA predicted the PRO/MES phenotype using the independent CGGA cohort (R^2 = 0.625, AUC = 0.872). In relation to patient survival, the 8-gene glycosignature could significantly stratify patients in low and high-risk in CGGA cohort (HR = 1.54, p = 0.018). Taken together, these findings present for the first time, an 8-gene glycosylation signature associated with PRO/MES polarization and patient survival, highlight the role of O-glycosylation in GSC biology and lay the groundwork for investigating O-glycosylation as potential therapeutic targets in GBM.

DRAFT - UNDER REVIEW

POSTER #28

TRACK: Artificial Intelligence and Systems Biology

Inference-Time Toxicity Mitigation in Protein Language Models

Source: 05_posters/P028.tex · Excel row 24

Fernández Burda M[1,2], Aranguri S[3], Arcuschin Moreno I[2], Ferrante E[1,4]

[1] Laboratory of Applied Artificial Intelligence (LIAA), Institute of Computer Sciences (ICC), CONICET – Universidad de Buenos Aires, Argentina [2] AI Safety Argentina (AISAR), Argentina [3] Goodfire [4] APOLO Biotech

Protein language models (PLMs) are becoming practical tools for de novo protein design, yet their dual-use potential raises safety concerns. We show that domain adaptation to specific taxonomic groups can elicit toxic protein generation, even when toxicity is not the training objective: the base ProGen2 model generates virtually no toxic sequences, but finetuning on Arthropoda, Arachnida, Gastropoda or Lepidosauria proteins yields predicted toxicity rates of 10–65%. To address this, we adapt Logit Diff Amplification (LDA) as an inference-time control mechanism for PLMs. LDA modifies token probabilities by amplifying the logit difference between a baseline model and a toxicity-finetuned model, requiring no retraining. Across the four taxonomic groups, LDA consistently reduces the predicted toxicity rate (measured with ToxDL2) below the taxon-finetuned baseline, by 8.0 to 29.9 percentage points at the optimal amplification strength (largest in Gastropoda, smallest in Arthropoda). We evaluate quality using Fréchet ESM Distance (FED) and predicted foldability (pLDDT). At the optimal setting, the change in FED remains small (−0.30 to +0.03) and the change in pLDDT ranges from −6.95 to +1.59, indicating that LDA maintains distributional similarity to natural proteins and structural viability, unlike activation-based steering methods, which tend to degrade sequence properties. Our results demonstrate that LDA provides a practical safety knob for protein generators, mitigating elicited toxicity while retaining generative quality.

DRAFT – UNDER REVIEW

POSTER #29 **TRACK:** Artificial Intelligence and Systems Biology

Gut Microbiome as a Source of Biomarkers in Neurodegenerative and Neurodevelopmental Disorders: A Machine Learning and Explainable AI Approach

Source: 05_posters/P029.tex · Excel row 25

Barbara S[1], Peralta R[1], Cherniz A[1]

[1] Universidad Nacional de Entre Ríos, Facultad de Ingeniería

Background: Alterations in the gut microbiota have been associated with the pathophysiology of neurological disorders through the gut-brain axis. Identifying microbial biomarkers for Parkinson's disease (PD), Alzheimer's disease (AD), and Autism Spectrum Disorder (ASD) remains challenging due to the high dimensionality and compositional nature of microbiome data. Machine learning (ML) models combined with Explainable Artificial Intelligence (XAI) offer a powerful framework to address this challenge.

Methodology: 16S ribosomal RNA sequencing data were obtained from 19 public projects in the National Center for Biotechnology Information (NCBI), comprising 1,147 samples for PD (656 cases / 491 controls), 562 for AD (367 / 195), and 487 for ASD (294 / 193). Genus-level microbial profiles were generated using Deblur within QIIME 2, with SILVA 138 as reference. Differential abundance analysis was performed using Linear Discriminant Analysis Effect Size (LEfSe; LDA score ≥ 3.0, $p < 0.05$). Six decision-tree classifiers: Decision Tree (DT), Random Forest (RF), Extra Trees (ET), XGBoost, LightGBM, and CatBoost, were trained with PyCaret using 5-fold stratified cross-validation repeated 10 times. Feature importance was assessed using SHapley Additive exPlanations (SHAP) values averaged across repetitions.

Principal Results: CatBoost achieved the highest discriminative performance for PD (AUC = 0.759 ± 0.030) and AD (AUC = 0.777 ± 0.049), while Extra Trees led for ASD (AUC = 0.720 ± 0.052). Decision Tree served as the lowest-performing baseline. Integration of LEfSe and SHAP identified convergent biomarker candidates: UBA1819 and Akkermansia (enriched in PD), Phascolarctobacterium (enriched in AD, top SHAP rank across all 6 models), and Collinsella (enriched in ASD, top-5 SHAP in all 6 models). Faecalibacterium and Lachnospiraceae members were consistently depleted in PD, consistent with reduced short-chain fatty acid production in the gut-brain axis.

Conclusions: Decision-tree ensemble models combined with SHAP explainability effectively identify microbial biomarker candidates across three neurological conditions. Convergence between LEfSe and SHAP approaches strengthens the robustness of identified genera, particularly Phascolarctobacterium for AD and Collinsella for ASD, which represent novel candidates warranting further validation.

DRAFT – UNDER REVIEW

POSTER #31

TRACK: Artificial Intelligence and Systems Biology

From protein embeddings to complex networks: multiscale organization of molecular recognition

Source: 05_posters/P031.tex · Excel row 26

Giordano GJ[1,2,3], Yanovsky N[1,2,3], Chernomoretz A[1,3,4]

[1] Fundación Instituto Leloir, Buenos Aires, Argentina [2] IIBBA-CONICET, Buenos Aires, Argentina [3] Departamento de Física, FCEN, UBA, Buenos Aires, Argentina [4] INFINA (UBA/CONICET), Buenos Aires, Argentina

Protein–ligand binding sites are local microenvironments that mediate molecular recognition and may remain conserved even between globally dissimilar proteins. In this work, we represent binding sites from BioLiP2 using embeddings derived from the ProtT5 protein language model and construct similarity networks based on their proximity in embedding space. We explore the resulting networks across multiple scales, from their global organization to communities of related binding sites and local patterns within individual proteins. Our results show that network organization reflects ligand and ion specificity, embedding-based communities capture recognition patterns that are not fully explained by conventional classifications based on sequence identity or InterPro annotations. At the intra-protein level, the network further distinguishes fine-grained recognition environments within individual binding pockets. Together, these results suggest that protein language model embeddings and complex networks provide a meaningful multiscale representation of molecular recognition, with potential applications to the study of drug–target relationships and ligand transfer.

DRAFT – UNDER REVIEW

POSTER #32

TRACK: Artificial Intelligence and Systems Biology

Integrating Multi-Omics and Cross-Species Knowledge Graphs to Identify Heat-Responsive lncRNA Regulatory Networks

Source: 05_posters/P032.tex · Excel row 27

Olszanowski E[1], Fernández Burda M[2,3], Zavallo D[2], Ariel F[1], Ferrante E[2,3], Rivarola M[1,2]

[1] Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE), CONICET-Universidad de Buenos Aires, Buenos Aires, C1428EHA, Argentina. [2] APOLO Biotech, Santa Fe de la Vera Cruz, Santa Fe, Argentina. [3] Laboratorio de Inteligencia Artificial Aplicada (LIAA), Instituto de Ciencias de la Computación (ICC), CONICET-Universidad de Buenos Aires, Buenos Aires, C1428EHA, Argentina.

Plant responses to heat stress are governed by complex regulatory networks where long non-coding RNAs (lncRNAs) play essential roles in thermotolerance. Mechanistically, lncRNAs alter chromatin architecture to dynamically regulate heat-responsive genes and mediate transcriptional memory, enabling faster defenses during recurring heat waves. Although heat-responsive lncRNAs are typically identified via transcriptomics, multi-omics integration remains limited. Here, we propose using heterogeneous multi-omics knowledge graphs to uncover heat-responsive non-coding elements, using tomato (Solanum lycopersicum) as a proof-of-concept case study before expanding to other key crops.

While model species like Arabidopsis thaliana possess rich functional annotation networks, crop species often lack comparable systematic knowledge frameworks. To address this, our graph topology builds upon previously curated molecular regulation networks in Arabidopsis, transferring cross-species conservation via orthology mapping with tomato. To enrich this framework, we integrated heat-stress transcriptomic profiles (RNA-seq), chromatin accessibility data (ATAC-seq), transcription factor binding sites (DAP-seq), and functional annotations from Gene Ontology (GO).

The constructed database consists of ~150,000 heterogeneous nodes—including protein coding genes (Arabidopsis and tomato), ~40,000 tomato lncRNAs as well biotic and abiotic stress factors (e.g., heat, drought)—alongside ~1,200,000 heterogeneous edges encompassing ~70,000 orthology links, ~160,000 GO annotations, as well as multiple molecular regulating interactions such as binding and transcription factor regulation. Notably, DAP-seq integration identifies HSFA1a binding targets for 3,065 protein-coding genes and 2,855 lncRNAs in tomato.

The database provides a foundational blueprint for multi-omics, cross-species network analysis. As a future step, we plan to implement an artificial intelligence layer combining LLM-assisted Graph Retrieval-Augmented Generation (GraphRAG) for semantic query navigation, alongside AgroNT sequence embeddings to predict novel lncRNA–target gene regulatory links. This platform offers a scalable computational foundation for functional discovery, demonstrating the power of graph-based multi-omics integration to identify heat stress regulators across agricultural species.

DRAFT – UNDER REVIEW

POSTER #34 **TRACK:** Biological Databases and Bioimaging

BioAIrepo: Making AI Resources Discoverable and Reusable Across the Life Sciences

Source: 05_posters/P034.tex · Excel row 28

Ragueneau E[1], Pajic V[1], Serrano J[1], Martinez M[1], Attafi OA[2], Sarkans U[1]

[1] EMBL-EBI, Wellcome Genome Campus, Hinxton, CB10 1SD UK [2] Department of Biomedical Sciences, University of Padova, Padova, 35131 Italy

Artificial intelligence is increasingly embedded in life science research, generating growing numbers of models, datasets and computational resources across domains including genomics, imaging, structural biology and multi-omics. However, these resources are often distributed across different repositories and described using heterogeneous metadata, making it difficult for researchers to identify suitable models, assess their biological applicability, reproduce published results and reuse existing work.

We present BioAIrepo, an EMBL-EBI initiative providing a common framework for describing and linking AI resources across life science domains. At its core is a biology-aware information model that connects datasets, AI models, optimisation procedures, evaluation results and reproducibility information through biological context. Community ontologies will provide a shared vocabulary for describing concepts such as species, tissues, diseases, cell types and experimental modalities, enabling models and datasets to be discovered and compared according to their biological applicability as well as their computational methodology.

BioAIrepo will support both directly hosted and externally referenced resources, while retaining standardised metadata within the platform. The information model builds on community initiatives including the DOME Registry, Croissant, Hugging Face cards and FAIR4ML, and includes a community-driven vocabulary for evolving AI methodologies.

Our long-term vision is to make AI resource publication and reuse simpler: researchers should be able to describe and submit their resources once, while enabling interoperable dissemination across repositories and services. BioAIrepo aims to provide an open, FAIR and biology-aware foundation for sharing, discovering and reproducing AI-driven research across the life sciences.

DRAFT – UNDER REVIEW

POSTER #35

TRACK: Biological Databases and Bioimaging

Glycosylated IDRs: Biological Implications and a Cross-Reference Between DisProt and GlyConnect

Source: 05_posters/P035.tex · Excel row 29

Lassalle FM[1], Cuccovia Warlet FU[1], Nugnes V[2], Cerrudo CS[1], Belaich MN[1], Aspromonte MC[2], Tosatto S[2]

[1] Universidad Nacional de Quilmes [2] University of Padova

Glycosylations are co- or post-translational modifications (PTMs) that can modulate intrinsic functions or provide new functions to proteins. They can alter the three-dimensional structures of proteins, they can expose or hide binding sites or they can alter the occurrence of other PTMs, either in the same residue or in the spatial proximity. They are involved in several biological processes, like protein folding and trafficking, signal transduction, protein-ligand interactions or cell-cell interactions, among others. Some important human diseases, such as cancers and diabetes, present distinct regulation of several glycosylation sites when compared to healthy controls. A good example of the biological implications of a glycosylation site located in an IDR is that of Histone H3's N-terminal region, as Serine 11 is a site of both glycosylation and phosphorylation, which becomes a competition, and glycosylation of S11 induces the trimethylation of Lysine 10 (active chromatin) ahead of the acetylation of K10 (repressed chromatin). Our objective is to increase the annotation of glycosylations in DisProt, a manually curated database for Intrinsically Disordered Proteins and Regions (IDPs/IDRs), as well as to understand the underlying implications of these PTMs happening within IDPs/IDRs. As this project began as a collaboration with the GlyConnect Database from the Swiss Institute of Bioinformatics, the initial search of experimental evidence of glycosylation sites located within IDRs was made through a cross-reference of this database with DisProt. The resulting dataset consisted of 52 entries with 27 unique proteins and 18 unique reference publications, considering that some proteins had multiple publications as references and that one publication can contain evidence of several proteins. Upon manual revision of the entries, which consisted of reading each publication, checking the proteins UniProt ID, sequence, isoform, possible cleavages, among other analysis, 27 entries were deemed as correct, while the remaining 25 entries were flagged for possible errors. These were further analysed to try to correct the errors and make them available for DisProt annotation. From these, 8 were corrected and deemed as useful, while 17 were determined not to be useful. We also carried out cross-references with other glycosylation databases, such as GlyGen and O-GlcNAc, but the manual curation has not yet begun.

DRAFT – UNDER REVIEW

POSTER #37

TRACK: Biological Databases and Bioimaging

Cross-Beta-DB: An Updated Resource for Naturally Occurring Amyloid Fibrils and Benchmarking Prediction Methods

Source: 05_posters/P037.tex · Excel row 30

Ferrelli ML[1,2], Gonay V[1], Kajava A[1]

[1] CRBM, CNRS, Université Montpellier, Montpellier, 34293, France [2] Instituto de Biotecnología y Biología Molecular, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, CONICET, La Plata, Argentina

Proteins, especially those with intrinsically disordered conformations, are frequently associated with amyloid fibril formation. Amyloid fibrils have a characteristic cross-β structure, with polypeptide chains oriented perpendicular to the fibril axis. Our analysis of existing databases revealed that they often mix true cross-β amyloids of naturally occurring amyloids with other aggregates. Since the goal of computational predictors is to identify disease-related and functional amyloids in vivo, a database restricted to amyloidogenic regions of naturally occurring cross-β fibrils is essential. To address this, we recently constructed Cross-Beta DB, a carefully curated database designed to fill this gap, providing a critical resource for benchmarking amyloid predictors (https://bioinfo.crbm.cnrs.fr/). The initial version of Cross-Beta DB was based on data available as of February 2024. Here, we present an updated release that includes an expanded set of entries derived from both the PDB and the literature. In this update, we also refined the protocol for identifying naturally occurring amyloid fibrils. The updated database enabled us to perform a new benchmark of methods for predicting amyloids, the results of which will be presented. Moreover, the expanded Cross-Beta DB is now used to enhance Cross-Beta Predictor, our machine-learning tool for detecting naturally occurring amyloids.

This work demonstrates how the growth of experimental data on cross-β amyloids can continuously improve predictive performance.

DRAFT – UNDER REVIEW

POSTER #38

TRACK: Biological Databases and Bioimaging

In silico evaluation of the role of BIRC6 in the DNA damage response in lung adenocarcinoma

Source: 05_posters/P038.tex · Excel row 31

Vilches J[1], Marchesini A[1], González Abad A[1], Cairo M[1], Gómez Bergna SM[1], Ferrelli ML[1], Amorós Morales LC[1], Pidre ML[1]

[1] Laboratorio de Virología Molecular (LVM), Instituto de Biotecnología y Biología Molecular (IBBM-CONICET), Facultad de Cs. Exactas, Universidad Nacional de La Plata (UNLP), La Plata, Argentina.

Lung cancer is one of the cancer types with the highest incidence and mortality worldwide. Inhibitors of apoptosis proteins (IAPs) constitute a group of negative regulators of apoptosis that are overexpressed in various tumor types and are associated with tumor development, proliferation, and aggressiveness. In particular, BIRC6 is also involved in other cellular processes, such as the DNA damage response (DDR), a key mechanism for tumor survival and resistance to conventional cancer therapies. In this context, studying the relationship between BIRC6 expression and DDR pathways in lung adenocarcinoma is of interest to understand its potential contribution to tumor survival and therapeutic resistance.

To perform this analysis, the TCGA lung adenocarcinoma (LUAD) database and the UCSC-Xena platform were first used to evaluate the expression of BIRC6 and its correlation with genes involved in different DNA repair pathways. Patient samples were stratified according to BIRC6 expression levels using the median expression value as a reference. Expression analyses were performed using Student's t-test to compare high and low BIRC6 expression groups, while gene-gene correlations were evaluated using Pearson's correlation coefficient.

Subsequently, a spatial transcriptomics analysis was carried out using the STOmicsDB database, evaluating the expression of BIRC6 and genes associated with the repair pathways previously analyzed in TCGA, which allowed us to characterize their distribution and behavior within the tissue context.

We observed a significant positive correlation between BIRC6 and genes involved in the DDR, with a marked trend toward homologous recombination repair. Likewise, spatial transcriptomics analysis showed higher expression of these genes in tumor regions with high BIRC6 expression.

The relationship between BIRC6 and the DDR identified in this work opens new perspectives for the development of therapeutic schemes to help overcome resistance to conventional therapies in this type of tumor.

DRAFT - UNDER REVIEW

POSTER #41

TRACK: Biological Databases and Bioimaging

Bioinformatic Characterization of a Cohort of PD-L1-Negative Non-Small Cell Lung Cancer Patients Through Network Analysis

Source: 05_posters/P041.tex · Excel row 32

Blanco L[1], Assirio MR[1], **Cassol I**[1], Rizzo M[2], Amorín R[2]

[1] Facultad de Ingeniería, Universidad Austral [2] Hospital Universitario Austral

Background: Non-small cell lung cancer (NSCLC) accounts for approximately 85% of lung cancer cases and exhibits substantial clinical and molecular heterogeneity. Patients with PD-L1-negative tumors represent a clinically relevant subgroup in which therapeutic decision-making increasingly depends on the integration of multiple molecular alterations. However, these alterations are frequently evaluated individually, potentially overlooking recurrent patterns of co-occurrence. Methods: A multicenter cohort of 458 patients with PD-L1-negative NSCLC from Latin America and Spain was analyzed by integrating molecular alterations identified through polymerase chain reaction (PCR), immunohistochemistry (IHC), and next-generation sequencing (NGS). Demographic, geographic, and therapeutic characteristics were first described. Molecular alterations were subsequently represented as a co-occurrence network, and the Louvain community-detection algorithm was applied to identify groups of recurrently associated alterations. Associations between molecular communities, clinical variables, and overall survival were evaluated using Kaplan–Meier analysis and the log-rank test. Results: Fourteen candidate molecular communities were identified, of which four included at least six patients:(a) KRAS G12C/ BRAF/ STK11/ KEAP1; (b)KRAS G12V/ HER2-ERBB2/ CDKN2A/ KRAS G12C; (c) MET; and (d) TP53/high tumor mutational burden (TMB)/ KRAS G12C. No significant differences in overall survival were detected among the four retained communities (log-rank p =0.46). Conclusions: Network-based integration of molecular alterations identified distinct molecular communities within PD-L1-negative NSCLC and recovered biologically relevant patterns that may be overlooked when alterations are considered independently. Although no differences in overall survival were observed among communities in this exploratory analysis, this approach provides a framework for characterizing the molecular heterogeneity of PD-L1-negative NSCLC beyond single-biomarker assessment, with comparison against public TCGA data providing an independent context for evaluating the identified molecular patterns.

DRAFT – UNDER REVIEW

POSTER #42 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Differences in the phyllosphere and endophyllosphere fungal communities among three Nothofagus species from the Andean-Patagonian forest.

Source: 05_posters/P042.tex · Excel row 33

Cagnani J[1], Vivanco L[2], Soria MA[1]

[1] Cátedra de Microbiología Agrícola. Facultad de Agronomía. Universidad de Buenos Aires. Ciudad de Buenos Aires. Argentina. [2] Cátedra de Ecología. Instituto de Fisiología y Ecología vinculado a la Agricultura (IFEVA), Facultad de Agronomía, Universidad de Buenos Aires, CONICET, Buenos Aires, Argentina.

Patagonian temperate forests are ecosystems of high ecological value, with Nothofagus as a major component of their tree diversity. Leaf-associated fungi influence host functioning and ecosystem processes, yet their diversity and patterns of association with their hosts remain little understood in Nothofagus. Using ITS-region sequencing we characterized and compared fungal communities associated with green leaves of three coexisting Nothofagus species in Lanín National Park, Argentina: N. dombeyi (evergreen), N. obliqua and N. nervosa (deciduous). We hypothesized that host identity is associated with differences in taxonomic and functional composition of leaf-dwelling fungi.

Five sun-exposed green leaves from different individuals of each species were sampled. Fungal DNA was extracted from whole leaves and the ITS2 region was sequenced using an Illumina platform. Sequences were processed in QIIME2. DADA2 was used to infer amplified sequence variants (ASVs), which were taxonomically assigned against UNITE, and analyzed for alpha and beta diversity, differential abundance, and FUNGuild-defined functional guilds. Representative sequences were queried using BLAST against the NCBI GenBank database.

After quality filtering, 12 samples and 182 features were retained. Alpha diversity did not differ significantly among host species, whereas beta diversity showed significant differences in community composition based on Bray-Curtis (R^2 = 0.51, p = 0.002) and UniFrac distances (R^2 = 0.60, p = 0.001). Only 16.7% of features were shared among the three hosts, indicating substantial taxonomic differentiation. However, it was not paralleled in functional differentiation across Nothofagus species. Comparison with the litter community of the same trees revealed marked differences in taxonomic composition and relative abundances. Querying the ASVs representative sequences by BLAST searches showed limited representation in Genbank, highlighting a gap in reference data for the foliar mycobiota of Patagonian Nothofagus.

These results indicate that host identity is associated with strong differences in the taxonomic composition of leaf-associated fungal communities, even when species coexist under the same environmental conditions. In contrast, functional structure appears comparatively conserved, suggesting that taxonomic turnover may occur without equivalent functional differentiation.

DRAFT – UNDER REVIEW

POSTER #43 **TRACK:** Evolutionary, Comparative and Population Genomics

The Evolution of Disorder: From Prokaryotes to Eukaryotes in Universally Conserved Proteins

Source: 05_posters/P043.tex · Excel row 34

Zapata G[1], Erdős G[2], Dosztányi Z[2], Lagares A[1], **Lozano M[1]**

[1] Institute of Biotechnology and Molecular Biology (IBBM), CONICET-UNLP, La Plata, Argentina. [2] Department of Biochemistry, Institute of Biology, Eötvös Loránd University, Budapest, Hungary.

Intrinsically disordered proteins and protein regions (IDPs/IDRs) are characterized by the lack of a stable three-dimensional structure under physiological conditions. Despite the absence of a defined structure, IDPs and IDRs participate in a wide range of biological functions, including molecular recognition, cellular signaling, and regulation of diverse cellular processes. IDRs are widespread across proteomes and occur in all three domains of life (Archaea, Bacteria and Eukarya), although they have been more extensively characterized in eukaryotes. However, their evolutionary dynamics across distantly related organisms remain poorly understood.

In this study, we aimed to evaluate the evolution of IDRs within genes conserved across Bacteria, Archaea, and Eukaryota. To this end, we selected proteins with homologs in all three domains of life from a previously compiled dataset and obtained their amino acid sequences from the EggNOG database. Those proteins were subsequently analyzed through multiple sequence alignments, secondary structure and disorder predictions, and functional domain identification.

Our results indicate that regions showing high degree of sequence conservation across the three domains generally display low disorder scores, whereas more variable regions tend to exhibit higher disorder. High-disorder regions were more frequently located at protein termini and in linker regions between conserved domains. These results are consistent with previous studies showing that differences in protein length and disorder are not primarily associated with changes in the size of folded domains, which remain relatively similar across the three domains of life, but rather with variation in the length and disorder of the linker regions connecting these domains. Overall, our results support the hypothesis that ancestral proteins were predominantly more ordered, and that the emergence and expansion of intrinsically disordered regions may represent an important feature of protein evolution. Further analysis of specific protein groups in a set of representative genomes selected from a previously established phylogenetic dataset will allow us to determine whether these general patterns are consistently observed across different protein families, or if distinct evolutionary patterns emerge in specific groups.

DRAFT - UNDER REVIEW

POSTER #45 **TRACK:** Evolutionary, Comparative and Population Genomics

Low-coverage sequencing as a strategy for biobanking of underrepresented populations.

Source: 05_posters/P045.tex · Excel row 35

Soria JD[1,11], Luisi P[2,11], Figueroa MI[3,11], Del Gobbo N[1,11], Fernández E[1,11], Bravi C[4,11], Bailliet G[5,11], Argüelles CF[6,11], Miretti M[7,11], Dipierri JE[3,11], Alfaro Goméz EL[3,11], García A[2,11], Ramallo V[8,11], Gonzales-José R[8,11], Sánchez-Loria J[10,11], Poklepovich T[10,11], Dopazo H[1,11]

[1] Laboratorio 2PQ. IEGEBA-CONICET. Departamento de Ecología, Genética y Evolución. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. Ciudad Universitaria, Buenos Aires, Argentina. [2] Instituto de Antropología de Córdoba, IDACOR (CONICET-Universidad Nacional de Córdoba), Cordoba, Argentina. [3] Universidad Nacional de Jujuy, INECOA Instituto de Ecorregiones Andinas (CONICET-Universidad Nacional de Jujuy), Jujuy, Argentina. [4] Instituto Multidisciplinario de Biología Celular, Imbice (CONICET, CIC, Universidad Nacional de La Plata), La Plata, Argentina. [5] Instituto Multidisciplinario de Biología Celular, Imbice (CONICET, CIC, Universidad Nacional de La Plata), La Plata, Argentina. [6] Instituto de Biología Subtropical (CONICET-Universidad Nacional de Misiones), Posadas, Argentina. [7] Instituto de Biología Subtropical (CONICET-Universidad Nacional de Misiones), Posadas, Argentina. [8] Instituto Patagónico de Ciencias Sociales y Humanas "Ma. Florencia del Castillo Bernal", Cenpat-Conicet, Puerto Madryn, Argentina. [10] Unidad Operativa Centro Nacional de Genómica y Bioinformática, ANLIS "Dr. Carlos G. Malbrán" [11] Programa de Referencia y Biobanco Genómico de la Población Argentina (PoblAr), Puerto Madryn, Argentina.

PoblAr is Argentina's first national biobank integrating genetic data with dietary, phenotypic, and social information from human populations. As in many Global South countries, Argentina's scientific system operates under financial and infrastructural constraints, further challenged by political turnover that undermines long-term research continuity. Low-coverage sequencing has emerged as a cost-effective strategy for genomic data generation, enabled in part by recent advances in genotype imputation methods and the growing diversity of reference haplotypes. However, its applicability remains uncertain in populations underrepresented in current genomic reference panels, which are dominated by data from the Global North.

We conducted a pilot study to evaluate the accuracy and analytical reliability of genotypes generated from low-coverage whole-genome sequencing followed by imputation (lcWGS) by directly comparing them with high-coverage whole-genome sequencing (hcWGS) from the same 33 individuals. For common variants, the non-reference discordance rate between hcWGS and lcWGS-Imputed was below 5%. Global patterns of genetic diversity inferred using Principal Component Analysis and ADMIXTURE were highly concordant between datasets, with no differences affecting biological interpretation. lcWGS showed reduced resolution for local ancestry inference, particularly for the ancestry Indigenous to the Americas, with a fourfold increase in unassigned alleles. However, assignment discrepancies between datasets remained below 5%. Importantly, demographic models calibrated from local ancestry tracts yielded nearly identical admixture histories for the Argentinean population using both data types.

Altogether, using a paired design in the same individuals, we show that lcWGS delivers robust population genetic inferences in populations underrepresented in current genomic reference panels, establishing it as a scalable and cost-effective strategy for PoblAr and other biobanks in the Global South. Continued expansion of reference panels to include genetic diversity Indigenous to the Americas will further enhance its performance.

DRAFT – UNDER REVIEW

POSTER #46 **TRACK:** Evolutionary, Comparative and Population Genomics

Evaluating translational rate conservation and cotranslational folding in legumes using minMax profiles

Source: 05_posters/P046.tex · Excel row 36

Dallachiesa D[1], Lozano MJ[1]

[1] Instituto de Biotecnología y Biología Molecular (IBBM) - CONICET-CCT La Plata - Universidad Nacional de La Plata

Synonymous codon usage bias (CUB) influences local ribosome elongation rates, generating kinetic dynamics that, in specific instances, are essential for proper cotranslational protein folding. However, whether these codon-derived kinetic profiles reflect conserved evolutionary selection remains a matter of debate.

We evaluated minMax% profiles as proxies for translation speed conservation across ~5,000 single-copy orthogroups from five legume species (Medicago truncatula, Cicer arietinum, Lotus japonicus, Lens culinaris, and Phaseolus vulgaris). For each orthogroup, the pairwise length-normalized Manhattan distance distribution of minMax% profiles was calculated—using amino acid–guided codon alignments to account for insertions and deletions—and compared to a null distribution derived from shuffled codon sequences. Functional enrichment was subsequently assessed by Pfam domain analysis using eggNOG-mapper.

A bimodal p-value distribution was observed, revealing a distinct subset of orthogroups with significantly conserved minMax% profiles. Proteins exhibiting conserved profiles were significantly larger, supporting a selective drive to mitigate misfolding errors and minimize biosynthetic energy costs. Functional analysis demonstrated clear segregation: conserved profiles were enriched in multidomain proteins (e.g., DEAD-box helicases, Cation ATPases, Dicer/RNA processing machinery, and Sec/Exocyst complexes), whereas non-conserved profiles were predominantly associated with transcription factors, LRR-RLK kinases, and stress-response proteins (such as LEA proteins).

Overall, minMax% profiles appear to be selectively conserved in plant evolution, particularly within large multidomain proteins, highlighting the functional importance of cotranslational folding in legumes.

DRAFT – UNDER REVIEW

POSTER #47 **TRACK:** Evolutionary, Comparative and Population Genomics

A city-scale synthetic population with demographic, socioeconomic and genetic structure for the evaluation of forensic genetic genealogy methods

Source: 05_posters/P047.tex · Excel row 37

Amigo MI[1], Marsico FL[2]

[1] Instituto de Cálculo - Universidad de Buenos Aires [2] Facultad de Ciencias Exactas y Naturales - Universidad de Buenos Aires

Background: Forensic genetic genealogy operates on populations whose demographic, socioeconomic and genetic structure is rarely known in full, and evaluating how identification methods behave under realistic structure is difficult, since real individual-level data are sensitive and largely inaccessible for research. We present a city-scale synthetic population that jointly integrates demographic, geographic, socioeconomic and genetic attributes, as a controlled environment in which database composition and subgroup performance can be examined systematically.

Methods: The population is modeled on the City of Buenos Aires. Each individual is assigned a geographic location, socioeconomic condition, sex and age, with the joint distribution of these attributes calibrated to published demographic sources for the city. Kinship is introduced by reproducing the family structure inferred from real DNA data of comparable urban populations, providing a realistic pedigree topology in terms of family size, generational depth and degrees of relatedness. Genotypes are then simulated using SLiM, a population genetics simulator, so that genetic variation is generated consistently with the underlying pedigree topology. The full genealogy is recorded as a tree sequence, yielding exact IBD segments against which detection methods are evaluated.

Results: Every individual, relationship and genotype is known by construction, providing complete ground truth for method evaluation. We compare IBD detection and cM-to-degree assignment, and characterise database search outcomes (detection rate, number of candidate leads and degree-assignment error) across subgroups defined by geographic location, background relatedness and socioeconomic stratum, using relatedness-matched null comparisons. Identification performance is not uniform across subgroups: detection of a usable match tracks the representation of the target's own subgroup in the database, while elevated background relatedness inflates total shared cM and biases degree assignment towards closer relationships, with error concentrated at the more distant degrees most relevant to genealogical search.

Conclusions: Performance in genealogical search depends on who is represented in the database and on background relatedness, so aggregate performance measures may be insufficient to characterise them. The framework is extensible and supports the evaluation of fairness, robustness and reliability of forensic genetic tools.

DRAFT – UNDER REVIEW

POSTER #48 **TRACK:** Evolutionary, Comparative and Population Genomics

How do reptiles survive hydric stress? A comparative review of the genomic signatures of adaptation in desert and marine environments

Source: 05_posters/P048.tex · Excel row 38

Coba-Males MA[1,2], Morando M[3,4], Olave M[1,5,6]

[1] Instituto Argentino de Investigación de Zonas Áridas (IADIZA), Consejo Nacional de Ciencia y Tecnología (CONICET), Mendoza, Argentina [2] Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba, Córdoba, Argentina [3] Grupo de Herpetología Patagónica (GHP), Instituto Patagónico para el Estudio de los Ecosistemas Continentales (IPEEC), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Puerto Madryn, Argentina. [4] Universidad Nacional de la Patagonia San Juan Bosco, Puerto Madryn, Argentina. [5] Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Cuyo, Mendoza, Argentina. [6] Tree of Life Programme, Wellcome Sanger Institute, Hinxton, United Kingdom.

Climate change imposes new selective pressures that, in some cases, challenge biodiversity, but also provides an opportunity to understand and predict species' ability to survive and persist under changing environmental conditions. Reptiles are a group of particular interest, as they evolved from tetrapods that transitioned from aquatic to fully terrestrial environments and subsequently colonized a wide range of habitats, including, through repeated independent reversions, freshwater and marine environments across their tree of life. Moreover, many species have colonized extreme environments such as deserts. Marine environments and deserts impose strong water-related stress, and reptiles that have repeatedly colonized these environments provide an opportunity to study the physiological and molecular strategies underlying adaptation to hydric stress. However, the general understanding of adaptations in this taxonomic group and in these types of environments remains limited compared to other vertebrates. This study presents a comprehensive comparative review of the literature on the genomic adaptations of reptiles inhabiting deserts and marine environments. In an era marked by advances in genomic sequencing, we focused on identifying candidate genes shared among species and lineages inhabiting these distinct water-related stress environments. Based on this, we evaluated whether these shared genes are associated with biological processes mainly related to osmoregulation, water balance, heat stress response, salinity tolerance, stress response, and energy metabolism. These processes may be related to the ability of reptiles to tolerate the environmental stress conditions of their habitats. We have preliminarily identified genes shared across independent studies within each environment (desert and marine), as well as genes shared between them. These shared genes represent an initial pattern of genetic overlap, although such overlap does not necessarily imply a shared genetic basis for adaptation. The observed patterns are consistent with the possibility that convergent adaptations may arise through either common genetic mechanisms or independent genetic routes. In addition, these results will help identify key knowledge gaps and provide a framework for future genomic research on the adaptive evolution of reptiles. This may contribute to a better understanding of the evolutionary mechanisms underlying adaptation to extreme environments.

DRAFT - UNDER REVIEW

POSTER #49 **TRACK:** Evolutionary, Comparative and Population Genomics

Comparative pangenome analysis of Pasteurella multocida identifies conserved genes with potential as vaccine candidates

Source: 05_posters/P049.tex · Excel row 39

Alvarez L[1], Melano Sosa XM[1], Zumárraga M[1], Del Médico Zajac MP[1], Eirin ME[1]

[1] Instituto de Biotecnología, IABiMo, UEDD INTA-CONICET

Pasteurella multocida is a major veterinary pathogen infecting a range of animal hosts and causing substantial economic losses worldwide, while its extensive genomic diversity poses a major challenge to the development of broadly protective vaccines. A comparative pangenomic approach was applied to identify conserved genes with potential as vaccine candidates across diverse isolates. Twenty publicly available P. multocida genome assemblies from diverse hosts and geographic origins were retrieved from the NCBI, EBI/ENA, and PubMLST databases based on sequencing coverage, assembly statistics, and quality.

The selected assemblies were re-annotated using Prokka, and the pangenome was reconstructed with Panaroo. Core-genome phylogeny was inferred using IQ-TREE with 1,000 ultrafast bootstrap replicates, and population structure was assessed by principal component analysis (PCA) of core-genome SNPs. Pangenome openness was evaluated using rarefaction curves and Heaps' law. Gene clusters were translated using Transeq and functionally annotated with eggNOG Mapper for Clusters of Orthologous Groups (COG) classification. Seed genes, including virulence-associated genes, were used to identify related biological pathways and retrieve additional core gene clusters involved in these pathways. R and RStudio were used to integrate annotations and identify functionally relevant gene clusters.

The pangenome comprised 4096 genes, including 1551 core and 2545 accessory genes, and showed a semi-closed structure (Heaps' law $\alpha = 0.75$). PCA of gene presence-absence profiles revealed no clear geographic clustering. Functional annotation of core genes identified 21 COG categories, with category S (function unknown) being the most represented. The pathway-based analysis expanded the initial 12 seed gene clusters, representing non-dominant functional groups, to 617 gene clusters, identifying 605 additional clusters associated with the biological pathways represented by the seed genes. These results indicate that P. multocida maintains a conserved core gene repertoire across diverse hosts and geographic origins despite substantial variation in accessory gene content. The pathway-guided pangenomic approach expanded a small set of virulence-associated seed genes into a broader repertoire of conserved, functionally related targets, providing a framework for genome-wide discovery and prioritization of novel immunogenic candidates for the development of broadly protective vaccines.

DRAFT - UNDER REVIEW

POSTER #50 **TRACK:** Evolutionary, Comparative and Population Genomics

Comparative genomics and genome mining of Antarctic marine Pseudomonas sp. m6p

Source: 05_posters/P050.tex · Excel row 40

Machuca J[1], Moraga F[1], Griffiths-Sanhueza C[1], Puentes MJ[1], Cerda C[2], Sheen A[2], García K[1], Urrutia IM[1], Pérez-Reytor D[1], Olivares-Pacheco J[3], Bastías R[4], Moya-Beltrán A[2], Plaza N[1]

[1] Instituto de Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Santiago, Chile [2] UTEM group [3] Grupo de Resistencia a los Antimicrobianos en Bacterias Patógenas y Ambientales (GRABPA), Instituto de Biología, Facultad de Ciencias, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile
[4] Instituto de Biología, Pontificia Universidad Católica de Valparaíso, Valparaíso, Chile

Closely related Pseudomonas lineages can be difficult to resolve taxonomically, while their biosynthetic repertoires remain incompletely characterized. We combined whole-genome taxonomic analysis with comparative biosynthetic gene cluster (BGC) mining to contextualize the Antarctic marine strain Pseudomonas sp. m6p. GTDB-Tk placed m6p within the genus Aquipseudomonas. FastANI screening of 23,835 external RefSeq assemblies yielded reportable values for 23,789 genomes, of which 22 showed ANI values $\geq$95% relative to m6p. The closest genomes were Pseudomonas sp. 22189 (99.1866%), Pseudomonas sp. VE 196-7 (99.1201%), and P. koreensis PSB00045 (98.4566%). All-vs-all ANIb analysis of m6p and these 22 genomes formed a single operational cluster at the 95% ANI threshold. For biosynthetic comparison, a curated cohort of 104 genomes was analyzed with antiSMASH and DeepBGC, followed by BiG-SCAPE clustering. Strain m6p contributed 25 normalized biosynthetic records distributed across 24 gene-cluster families (GCFs), all recovered across GCF cutoffs from 0.3 to 0.6. MMseqs2 analysis of 19,810 proteins from m6p-containing GCFs identified 847 homolog groups, including 343 groups containing an m6p protein and present in at least 80% of BGCs within their respective GCFs. Among prioritized loci, the hydrogen-cyanide-associated AS07 cluster showed the closest relationship to a characterized MIBiG reference (d=0.0673), with strong conservation of the HcnABC architecture, while the NRPS-associated AS06 cluster also showed high architectural similarity (d=0.0890). In contrast, AS03 and AS12 were more distant from their nearest characterized references (d=0.5038 and 0.6009, respectively). These results place m6p within a closely related but taxonomically heterogeneous genomic lineage and show that its biosynthetic repertoire combines highly conserved and comparatively divergent cluster architectures.

DRAFT – UNDER REVIEW

POSTER #52 **TRACK:** Evolutionary, Comparative and Population Genomics

Genomic characterization of B-genome-specific pack-MULE transposons in the allotetraploid Arachis hypogaea

Source: 05_posters/P052.tex · Excel row 41

Marichich PB[1,2], Seijo JG [1.3], Samoluk SS[1,3]

[1] Instituto de Botánica del Nordeste (UNNE-CONICET), Facultad de Ciencias Agrarias, Corrientes, Argentina. [2] Universidad Nacional del Chaco Austral, Presidencia Roque Sáenz Peña, Chaco, Argentina. [3] Facultad de Ciencias Exactas y Naturales y Agrimensura, Universidad Nacional del Nordeste, Corrientes, Argentina

Pack-transposons are non-autonomous transposable elements that can capture exogenous genomic DNA, facilitating the duplication and relocation of these fragments to new chromosomal positions, a process known as transduplication. These elements have been shown to incorporate partial or complete sequences from multiple genes, potentially forming novel open reading frames (ORFs) that drive the evolution of new gene functions. Previous characterization of the Dof gene family in the allotetraploid A. hypogaea (2n = 4x = 40; AABB genome) suggested that some members may have undergone gene duplication through transduplication. In this context, we used bioinformatic approaches to identify and characterize transposon-related structures potentially involved in these transduplication events. The genomic analysis identified typical MULE transposon structures flanking six Dof genes, all distributed within the B-subgenome, specifically on chromosomes B01, B02, B07, B09, and B10. These elements range from 1,266 to 1,343 bp in length and share highly conserved 203-bp Terminal Inverted Repeats (TIRs) and a 9-bp target site duplication. The distribution pattern of amino acid domains within the Dof sequences was highly conserved among these elements. The upstream TIR provided functional promoter regions, providing a conserved TATA-box alongside variable cis-regulatory motifs associated with stress and hormone responses. Sequence comparisons suggest that the ancestral state of a Dof gene, located on chromosome B05, served as the most probable donor of the coding region for the Dof genes inserted within pack-MULE transposons. Furthermore, transcriptomic data revealed that the transposon-derived Dof genes maintain a tissue-specific expression profile, mainly in reproductive tissues, consistent with their putative donor gene. Finally, synteny analysis between different subspecies of A. hypogaea and its wild diploid progenitors confirmed that these pack-MULEs remain conserved across both subspecies of the allotetraploid (subsp. hypogaea and subsp. fastigiata) and are exclusively B-genome specific—present in the diploid ancestor A. ipaënsis (B genome) but absent in A. duranensis (A genome). The discovery of pack-MULE elements reveals a previously undescribed mechanism contributing to the diversification of regulatory genes in Arachis genus and provides new insights into the evolutionary dynamics of plant transcription factor families.

DRAFT - UNDER REVIEW

POSTER #53 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Genome Sequencing and Functional Characterization of Extrachromosomal Elements in Rhodococcus erythropolis JCM 2893, a Petroleum-Derived Strain Capable of Growth under Oligotrophic Conditions

Source: 05_posters/P053.tex · Excel row 42

Nashmias ME[1], Zubimendi JP[1], Gallicet MA[1], Torres Tejerizo G[2], Alvarez HM[3], Quelas JI[1]

[1] YPF Tecnología (Y-TEC), Químicos y Biológicos de Especialidad, Plataforma de Biología Molecular y Bioinformática, Berisso, Buenos Aires, Argentina. [2] Instituto de Biotecnología y Biología Molecular (IBBM-CONICET), La Plata, Buenos Aires, Argentina. [3] Instituto de Biociencias de la Patagonia (INBIOP), UNPSJB-CONICET, Comodoro Rivadavia, Chubut, Argentina.

Microorganisms capable of growth under oligotrophic and CO_2-enriched conditions represent promising models for studying metabolic adaptation and biotechnological applications. Here, we present the first genome sequencing, assembly, and functional characterization of Rhodococcus erythropolis JCM 2893, a petroleum-derived strain selected for its ability to grow under oligotrophic and CO_2-enriched conditions and for which no genomic information was previously available.

Sequencing was performed using Oxford Nanopore Technologies (N50 = 12.46 kb; coverage ≈181×). De novo assembly resolved the genome architecture at the replicon level, yielding a 7.07 Mb genome composed of a 6.43 Mb circular chromosome, two closed circular plasmids, and four linear genomic elements that could not be circularized during assembly. Genome annotation identified 6,647 coding sequences, 15 rRNA genes, and 53 tRNA genes.

Comparative analyses revealed that plasmids pRerJCM2893a (79.6 kb) and pRerJCM2893b (100.7 kb) share high sequence identity with replicons described in other R. erythropolis strains, while one linear genomic element exhibited >99% nucleotide identity to a linear plasmid from R. erythropolis JCM 2895, supporting its plasmidic nature.

Functional characterization of extrachromosomal replicons revealed enrichment in genes associated with genetic mobility, recombination, transcriptional regulation, toxin-antitoxin systems, and fatty acid metabolism, together with a high proportion of hypothetical proteins. These results indicate that the extrachromosomal compartment constitutes an important reservoir of functions potentially involved in genome plasticity, environmental adaptation, and metabolic versatility. Notably, the presence of lipid metabolism-related functions is consistent with the reported oleaginous potential of the genus Rhodococcus and may contribute to the physiological flexibility of strain JCM 2893 under nutrient-limited conditions.

These genomic features, together with the experimentally observed ability of strain JCM 2893 to grow under oligotrophic and CO_2-enriched conditions, provide a valuable framework for investigating alternative carbon assimilation strategies and lipid-based biotechnological applications. Overall, this work delivers the first genomic resource for R. erythropolis JCM 2893 and expands current knowledge of extrachromosomal genome architecture within the genus Rhodococcus.

DRAFT – UNDER REVIEW

POSTER #54 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Metagenomic profiling of the metal resistome of microbial communities from the Caviahue-Copahue Geothermal System

Source: 05_posters/P054.tex · Excel row 43

Gallicet MA[1,2], Nashmias ME[1], Quelas JI[1], Urbieta MS[2]

[1] YPF Tecnología (Y-TEC), Químicos y Biológicos de Especialidad, Plataforma de Biología Molecular y Bioinformática, Berisso, Buenos Aires, Argentina. [2] Centro de Investigación y Desarrollo en Fermentaciones Industriales (UNLP-CONICET), La Plata, Buenos Aires, Argentina.

Geothermal environments harbor diverse extremophilic microorganisms adapted to high temperatures, extreme pH, and elevated concentrations of metals and metalloids. Despite advances in sequencing and bioinformatics, uncultivated microbiota of such environments remains poorly characterized at the functional level. Shotgun metagenomics provides a culture-independent approach to explore the genetic repertoire of these communities and identify functions potentially associated with microbial responses to metal stress. Precisely, in this work we characterized the inferred metal-resistance potential of six metagenomes from different anoxic sediments of the Caviahue-Copahue Geothermal System (Argentina), using complementary gene-centric and genome-centric approaches.

Metagenomic reads were quality-filtered, assembled, and used to reconstruct metagenome-assembled genomes (MAGs). Protein-coding sequences were predicted with Prodigal and screened by BLASTP against BacMet, an experimentally curated database of metal -and biocide-resistance genes. Resistance profiles were filtered using $\geq$ 70% amino-acid identity, $\geq$ 50 aligned amino acids, and E-value $\leq$ 1e-5.

The size of the inferred resistome varied among samples, with mercury (Hg) and arsenic (As) resistance genes dominating the profiles. Comparison across metagenomes revealed a core resistome comprising six genes: arsA, arsB, merA, merP, merR, and merT, detected in all six communities; together with 21 genes exclusive to three sampling sites.

To explore the taxonomic distribution of resistance functions, genes were assigned to reconstructed MAGs and represented in a bipartite network linking microbial phyla and target metals. As and Hg showed the highest connectivity, while Proteobacteria displayed the broadest range of metal-resistance associations. The presence of core resistance genes, previously reported in mobile genetic elements, is compatible with a potential contribution of horizontal gene transfer to the dissemination of resistance functions, whereas site-specific profiles may reflect local environmental and/or taxonomic differences.

Overall, the results demonstrate how metagenomic approaches can reveal conserved and site-specific components of the inferred metal resistome in an extreme geothermal ecosystem, providing insights into microbial functional capacity to cope with metal stress and supporting the bioprospecting of communities with promising applications in metal bioremediation and biotransformation.

DRAFT - UNDER REVIEW

POSTER #56 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Comparative Evaluation of Pathway Topology Analysis Methods for Reactome, with a Focus on Intrinsically Disordered Proteins

Source: 05_posters/P056.tex · Excel row 44

Algañaras M[1], Gong C[2], Hermjakob H[2], Lozano MJ[3], Bosch A[1], Ragueneau E[2]

[1] Laboratorio "Biofilms Microbianos" CINDEFI-UNLP-CONICET, CCT La Plata, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Argentina [2] European Bioinformatics Institute (EMBL-EBI) European Molecular Biology Laboratory, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [3] IBBM-UNLP-CONICET-CCT La Plata, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, La Plata, Buenos Aires, Argentina.

Background. Over-representation analysis (ORA) and gene-set enrichment analysis (GSEA) remain standard for pathway analysis, but they treat genes as independent units and ignore the structure, directionality, and interaction context of biological networks. This is especially limiting for intrinsically disordered proteins (IDPs), whose function depends on their position and connectivity within networks rather than on fixed structure, information that pathway topology analysis (PTA) can exploit.

Methodology. We systematically screened over twenty PTA methods and, after filtering on license, availability, programming language, graph model, and scoring strategy, adapted a final selection (SPIA, CePa, GANPA, PathIntegrate, GNNenrich) to the Reactome knowledgebase. We assembled a benchmarking set designed for comparison, covering biological diversity (monogenic versus complex/polygenic diseases across oncology, inflammation, and neurodegeneration), tissue diversity (blood, brain, solid organs), technical and sample-size diversity (multiple platforms, small and large n, high-noise sets), and experimental verification with removal of redundant cases. Cystic fibrosis (CF; GSE15568) served as an in-depth case study. Methods were assessed for accuracy (sensitivity against curated and literature-validated pathways; specificity via label permutation), robustness, and efficiency. Key proteins of significant pathways were cross-referenced against DisProt to report raw sequence disorder content, with no fixed cutoff.

Results. PathIntegrate produced the most exhaustive landscape and captured coordinated, hidden-layer changes; SPIA showed high precision for topologically critical signaling cascades; ReactomeGSA (PADOG) provided a broad reference. In CF, disordered proteins were recovered by topology-aware methods, including EIF4G1 (Translation, ~70.6% disorder content; SPIA and PathIntegrate), PAK1 (Rho GTPase effector, ~48.4%; SPIA), STAT6, and NOTCH1 (PathIntegrate). Method-specific detections indicated complementary coverage of the IDP space.

Conclusions. Topology-aware analysis surfaces IDP-centric mechanisms missed by flat enrichment, supporting integration of a PTA method into Reactome to improve molecular-context analysis for disordered protein resources.

DRAFT – UNDER REVIEW

POSTER #57 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Temporal and spatial persistence of a bacterial core across vineyard soil and wine in an emerging wine-producing region

Source: 05_posters/P057.tex · Excel row 45

Rivas GA[1], Guillade AC[1], Semorile LC[1], Delfederico L[1]

[1] Laboratorio de Microbiología Molecular, Departamento de Ciencia y Tecnología, Instituto de Microbiología Básica y Aplicada (IMBA), Universidad Nacional de Quilmes (UNQ), Bernal, Argentina

Vineyard bacterial communities are shaped by environmental conditions, geographic location, and vintage, although a fraction of the microbiota may persist despite this variability. We integrated bacterial community data from two Malbec vineyards in an emerging wine-producing region of Argentina. Soil-associated communities were analyzed in Saldungaray across three consecutive vintages (2017–2019) and in Al-Este in 2017; wine-associated communities were analyzed across the three vintages in Saldungaray. Communities were characterized by Illumina sequencing of the V3–V4 region of the 16S rRNA gene.

A core of 14 bacterial orders persisted in Saldungaray soil across vintages despite contrasting climatic conditions, including prolonged drought and an unseasonal spring frost that strongly affected grape production. The number of non-core orders increased from 3 in 2017 to 5 in 2018 and 9 in 2019. All 14 core orders were also detected in Al-Este, supporting a conserved bacterial assemblage across vineyards. Al-Este additionally harbored 63 orders not represented in the Saldungaray temporal analysis, although most showed low abundance and variable prevalence. Among these, JG30-KF-CM45 and Nitrospirales showed the highest mean relative abundances and were detected in at least half of the Al-Este samples.

Wine displayed a persistent core of nine bacterial orders across vintages, six of which were also members of the soil core, indicating substantial overlap between the temporally conserved soil and wine associated communities. Several members of this shared core belong to groups commonly reported in vineyard soils and plant associated microbiomes, including lineages associated with plant microbe interactions, nutrient cycling, and plant-growth-promoting traits.

Overall, our results support a model in which a conserved bacterial core coexists with a more variable vineyard and vintage associated fraction. Its persistence across contrasting climatic conditions and vineyards, together with its overlap with persistent wine associated bacteria, is consistent with vineyard soil acting as a potential biological reservoir for a fraction of the wine microbiota.

DRAFT - UNDER REVIEW

POSTER #59 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Revisiting historical 16S rRNA datasets: does improved bioinformatic resolution change our view of bacterial diversity in eutrophic lake sediments?

Source: 05_posters/P059.tex · Excel row 46

Martins G[1,2], Peixoto L[1,3]

[1] CEB – Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal. [2] LABBELS – Associate Laboratory, Braga/Guimarães, Portugal. [3] BioData.pt | ELIXIR Portugal, R. da Quinta Grande, 9, 2780-156 Oeiras, Portugal.

Background: Historical 16S rRNA datasets represent a valuable, but often underexploited, source of microbial diversity information. Since their generation, reference databases, taxonomic frameworks, and phylogenetic methods have substantially evolved. Therefore, this study aimed to reanalyse 16S rRNA sequences generated in a 2012 study of bacterial communities in eutrophic Azorean lake sediments (Martins et al., 2021; http://dx.doi.org/10.1080/01490451.2011.6 and determine whether the original taxonomic interpretations remain supported by current bioinformatic approaches.

Methods: Eight unclassified sequences deposited in GenBank were assessed using current reference databases and taxonomic classification approaches. Original taxonomic assignments and sequence similarities reported in 2012 were systematically compared with current results, with particular emphasis on changes in taxonomic resolution and on sequences previously classified as unclassified.

Results: The new BLASTn provided improved taxonomic resolution for all eight sequences compared with the original assessment. Previously unclassified sequences could now be assigned to Bacillariophyta, Planctomycetota, Spirochaetota, Bacillota, Pseudomonadota, Cyanobacteriota, and Bacteroidota/Chlorobiota groups. In addition, the similarity of the closest relatives exceeded 95% identity for all analysed sequences, compared with values as low as 86% in the original analysis. Contrary, using the SILVA Alignment, Classification and Tree Service, two sequences remained unclassified, and identity starts on 87.6%. SILVA retrivied the classification of tested sequences to Cyanobacteriota, Chloroflexota, Spirochaetota, Chlorobiota, and Candidatus Latescibacterota. This classification was quite different from BLASTn, however, could be resolved to more specific taxonomic levels, reflecting the increased coverage and improved annotation of current reference databases.

Conclusions: The reanalysis substantially improved taxonomic resolution, particularly for sequences that were previously unclassified. However, differences between current classification approaches indicate that improved bioinformatic resolution does not necessarily lead to a single, unambiguous taxonomic interpretation. These results demonstrate the value of revisiting legacy 16S rRNA datasets, potentially extracting additional ecological and taxonomic information without generating new sequencing data.

DRAFT – UNDER REVIEW

POSTER #60 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

DIFFERENTIAL EXPRESSION OF INFLAMMASOME RELATED GENES IN COLORECTAL CANCER AND ITS STRATIFICATION BY MICROSATELLITE INSTABILITY STATUS

Source: 05_posters/P060.tex · Excel row 47

Arias M[1,2], Gutierrez A[3], Muñoz C[4], Larronde C[5,6,7], Viscarra T[3], Buchegger K[2,5,6]

[1] Biotechnology Undergraduate Program, Faculty of Agricultural Science and the Environment, Universidad de La Frontera, Temuco, Chile. [2] Tumor Biology Laboratory, Department of Basic Sciences, Faculty of Medicine, Universidad de La Frontera, Temuco, Chile. [3] Faculty of Health Sciences, Institute of Biomedical Sciences, Universidad Autónoma de Chile, Temuco, Chile. [4] Kinesiology Undergraduate Program, Faculty of Medicine, Universidad de La Frontera, Temuco, Chile [5] Laboratory of Integrative Biology, Center of Excellence in Translational Medicine, Scientific and Technological Bioresource Nucleus (CEMT–BIOREN), Universidad de La Frontera, Temuco, Chile. [6] Millenium Institute of Immunology and Immunotherapy, Santiago, Chile. [7] Doctoral Program in Sciences, with a concentration in Applied Cell and Molecular Biology, Universidad de La Frontera, Temuco, Chile.

Background: Chronic inflammation plays a critical role in colorectal cancer (CRC) tumorigenesis and in shaping the tumor microenvironment. Inflammasomes are innate immune sensors that trigger inflammatory cascades and pyroptosis. However, the expression of their components in human CRC, and its relation to genomic subtypes such as microsatellite instability (MSI), remains poorly characterized. We evaluated a curated 39-gene inflammasome panel in CRC according to tumor-normal and MSI status.

Methodology: In silico analysis of TCGA-COAD/READ RNA-seq data using limma-voom (TMM normalization). Tumor versus normal expression was assessed in a paired intra-patient model (43 tumor/adjacent-normal pairs) and, as a complementary design, in 440 tumors versus 44 unpaired normals. Tumors were stratified by MSI status (331 MSS, 69 MSI-L, 38 MSI-H) with adjustment for tumor sidedness. Overall survival (421 patients, 83 events) was explored with Cox models and expression x MSI interaction tests. P-values were corrected by Benjamini-Hochberg FDR.

Results: In paired analysis, 12 of 39 genes were significantly altered under transcriptome-wide FDR < 0.05 and |log2FC| > 1. In the unpaired design, including upregulated GSDMC (+2.53), IL1A (+1.98) and NOD2 (+1.55) and downregulated CASP5 (-2.36), TXNIP (-1.45) and NAIP (-1.10). MSI status defined a distinct profile: 28 of 39 genes differed between MSI-H and MSS tumors (24 upregulated in MSI-H), led by the DNA sensors AIM2 (log2FC +1.62, FDR 9.9×10^{-8}) and IFI16 (+1.16, FDR 7.4×10^{-10}), both robust to adjustment for tumor sidedness. MSI-L tumors did not differ from MSS in any gene, and adjacent normal mucosa did not differ by MSI status. In preliminary survival analyses, high CASP1 expression was associated with better survival in MSS/MSI-L tumors (HR 0.69, 95% CI 0.55-0.86, FDR 0.037), with an opposite trend in MSI-H tumors (HR 3.31, 10 events; interaction FDR 0.17).

Conclusions: Inflammasome expression in CRC is sensor-specific rather than pathway-wide, and MSI-H tumors show a distinct, immune-activated profile with induction of the DNA sensors AIM2 and IFI16. CASP1 emerges as a preliminary, hypothesis-generating prognostic candidate that requires validation in independent cohorts and at the protein level.

DRAFT – UNDER REVIEW

POSTER #61 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Pipeline for the taxonomic and functional analysis of plant growth-promoting bacteria from long-read sequencing data

Source: 05_posters/P061.tex · Excel row 48

Andereggen V[1], Massicot A[1], Figuerola ELM[1,2]

[1] Instituto de Biociencias, Biotecnología y Biología Traslacional (iB3), Universidad de Buenos Aires, Argentina.
[2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET).

Plant growth-promoting bacteria represent a sustainable alternative for reducing dependence on chemical fertilizers and contributing to tolerance against biotic and abiotic stresses. However, identifying genes associated with agronomically relevant functions requires running multiple tools and subsequently integrating their results. To automate this process, a Nextflow-orchestrated workflow was developed that enables the analysis of long read sequencing data from isolated bacterial genomes and metagenomic samples. The workflow includes read quality control, assembly with Flye, genome quality assessment with CheckM, taxonomic classification with GTDB-Tk, and functional annotation using Prodigal, KofamScan for KEGG Orthology (KO) assignment, and HMMER hmmscan for Pfam domain identification.

In addition, a Python script was implemented that compares the obtained annotations against a user-configurable database containing genes of interest defined by KO and Pfam identifiers. The system filters the matches and generates summary tables and graphical visualizations. The results are presented in a single Excel file that summarizes the presence of genes associated with biological nitrogen fixation, phosphorus metabolism, phytohormone production, siderophores, and other functions, along with completeness metrics by functional category. The pipeline was evaluated with genomes of bacterial strains isolated from rice (Oryza sativa) roots and validated through in vitro characterization of the plant growth promoting activities of the isolates. The workflow integrated taxonomic and functional information in a single run, making it a useful tool for characterizing plant growth-promoting bacteria and selecting candidate strains for agricultural application.

DRAFT – UNDER REVIEW

POSTER #62 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Cross-domain metabarcoding analytics: an integrated bioinformatics pipeline for bacterial and fungal communities in sugarcane soils

Source: 05_posters/P062.tex · Excel row 49

Ortiz Rocca L[1], Arbuco D[1], Montecchia MS[1,2], Chalco Vera J[3], Acreche MM[4], Correa OS[1], Soria MA[1,2]

[1] Cátedra de Microbiología Agrícola, Facultad de Agronomía, Universidad de Buenos Aires. Av. San Martín 4453 C.A.B.A., Argentina. [2] Instituto de Biociencias Agrícolas y Ambientales, INBA (UBA-CONICET). Av. San Martín 4453 C.A.B.A., Argentina. [3] Instituto de Investigación Animal del Chaco Semiárido (IIACS), Centro de Investigaciones Agropecuarias (CIAP), Instituto Nacional de Tecnología Agropecuaria (INTA)-CONICET, Chañar Pozo s/n, (T4113), Leales, Tucumán, Argentina. [4] INTA Estación Experimental Agropecuaria Salta-CONICET. Ruta Nac. 68, Km. 172, CP 4403, Cerrillos, Salta, Argentina.

Soil microbiomes sustain critical biogeochemical cycles in agricultural ecosystems, yet dual-domain metagenomic characterizations under subtropical crop management remain scarce. In this study, we implemented a comprehensive metabarcoding pipeline to evaluate the structural and topological responses of soil bacterial and fungal communities in long-term experimental sugarcane plots located in Famaillá, Tucumán, Argentina, assessing the interactive effects of sampling season between summer and autumn alongside tillage and harvest management in a factorial design. Dual-marker amplicon sequencing targeting the 16S rRNA V3–V4 and fungal internal transcribed spacer (ITS) regions was processed in QIIME 2, where quality trimming was performed using Cutadapt and single-nucleotide Amplicon Sequence Variants were resolved through DADA2. Taxonomic assignments were conducted against the SILVA and UNITE databases, and feature tables were rarefied to 41,010 reads yielding 40,739 prokaryotic variants dominated by Acidobacteriota, Actinomycetota, and Pseudomonadota, whereas fungal features were rarefied to 165,205 reads yielding 17,622 variants dominated by Ascomycota and Basidiomycota. Downstream bioinformatic workflows implemented in R integrated alpha and beta diversity modeling, multifactorial permutation multivariate analysis of variance on Bray-Curtis dissimilarity at the genus level, Procrustes superimposition, paired matrix correlation via Mantel tests, and sparse inverse covariance network inference using SPIEC-EASI to resolve direct ecological associations. Bacterial alpha diversity responded strongly to seasonal transitions across richness, Shannon, Simpson, and Faith phylogenetic diversity metrics, whereas fungal alpha diversity remained stable. At the community level, while both domains exhibited pronounced seasonal restructuring, fungal assemblages responded significantly to harvest residue retention and tillage disturbance. Multivariate Procrustes rotation demonstrated coordinated global community restructuring between both kingdoms across the field gradient, whereas co-occurrence network architectures revealed contrasting organizational complexity across domains. This integrative workflow illustrates the power of multi-marker bioinformatic pipelines to delineate kingdom-specific ecological sensitivities in subtropical agricultural soils.

DRAFT – UNDER REVIEW

POSTER #63

TRACK: Genomics, Transcriptomics, Epigenomics, Metagenomics

Cake-cutting applied to genomic fragments with repeated sequences

Source: 05_posters/P063.tex · Excel row 50

Fernández I[1], Marenco J[2], Tetzlaff T[1]

[1] Universidad Nacional de General Sarmiento [2] Universidad Torcuato Di Tella

We use the squeezing technique within an evolutionary algorithm to solve the discrete cake-cutting problem, applying it to genomic strands with repeated sequences. For k sequences of interest, the procedure automatically finds k portions that satisfy fair division criteria, ensuring that each genome portion is specialized in one of the repeated sequences. We explain the implications in this context of theoretical results regarding uniqueness, algorithmic complexity, and the existence of solutions. We apply our procedures to real sequences to illustrate their use.

DRAFT – UNDER REVIEW

POSTER #64 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Genome-wide AP-1-directed enhancer demethylation, but a stableLRP1 locus, marks monocyte-to-macrophage differentiation: ananopore methylome analysis

Source: 05_posters/P064.tex · Excel row 51

Giordano LJ[1,2,3], Casale R[1,2,3], Chiabrando GA[1,2,3], Ceschin DG[1,2,3]

[1] Hospital Privado Universitario de Córdoba, Córdoba, Argentina [2] Instituto Universitario de Ciencias Biomédicas de Córdoba (IUCBC), Córdoba, Argentina[3] [3] Centro de Investigación en Medicina Traslacional "Severo R. Amuchástegui" (CIMETSA), U.A. Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Córdoba, Argentina

Background. Low-density lipoprotein receptor-related protein 1 (LRP1) is a protective receptor inatherosclerosis, dynamically regulated during monocyte-to-macrophage differentiation (MMD).

Whether locus-specific DNA methylation drives this regulation, and whether differentiation triggers

broader methylome remodeling, remained untested at nucleotide resolution.

Methods. We performed direct long-read nanopore sequencing of 5-methylcytosine (5mC) and5-hydroxymethylcytosine (5hmC) in a THP-1 model of monocytes, resting macrophages, andLPS-activated macrophages (n=7). Coverage differed across samples (3.4-6.9x); all comparisonsused hypergeometric depth-matching to remove this confound, which otherwise dominates naivestatistics and principal component analysis. A standard region-level dispersion estimator failed aninternal null-permutation control and was replaced by a binomial test on pooled counts with anempirical, permutation-calibrated false discovery rate, applied genome-wide and to LRP1 viaequivalence testing, motif regression, single-molecule analysis, and allele-specific phasing.

Results. The global methylome does not reorganize during MMD (Δ<0.4 points), and LPS producesno detectable change in any compartment. Differentiation instead drives demethylationconcentrated in distal enhancers and repeats (Δ=-0.585 points, z=-30.8 vs permutation), while CpGislands remain unaffected. Motif regression across 76,580 distal enhancers identified AP-1 as thedominant predictor (-1.87 points, z=-18.0), with monotonic dose-response, confirmedindependently by a genome-wide, anticorrelated gain in 5hmC (r=-0.48 with Δ5mC) tracking AP-1density. At LRP1, no regulatory element showed significant methylation change (equivalencebound $\leq$10.2 points), no allele-specific methylation was detected, and the promoter CpG island'sstability ranked at the 83rd percentile among matched CpG islands, indicating a genericcompartment property rather than a locus-specific feature.

Conclusions. MMD triggers a genome-wide, AP-1-directed enhancer demethylation program thatspares CpG-island promoters, including LRP1's own. LRP1 regulation during differentiation is thuscompatible with permissive, histone-driven chromatin remodeling rather than DNAmethylation-based control. This framework offers a template for rigorous nanopore methylomeanalysis under realistic, moderate-coverage designs.

DRAFT – UNDER REVIEW

POSTER #65 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Genomic variation in Argentine isolates of Cercospora kikuchii: contrasting patterns among virulence, fungicide resistance, and phylogenetic genes

Source: 05_posters/P065.tex · Excel row 52

Chirino L[1], Soria M[1], Sautua F[1], Carmona M[1]

[1] Facultad de Agronomía, Universidad de Buenos Aires

Cercospora kikuchii is an important pathogen of soybean, causing Cercospora leaf blight and purple seed stain. Its control relies largely on fungicides, and populations with reduced sensitivity to multiple active ingredients have been reported. However, genomic information on Argentine isolates remains limited. This study aimed to characterize genetic variation in Argentine C. kikuchii isolates by analyzing polymorphisms in genes associated with virulence, fungicide resistance, and phylogenetic inference.

We analyzed genomic data from 85 C. kikuchii isolates collected from soybean crops in different regions of Argentina. A filtered variant dataset generated from comparisons between a C. kikuchii homologous genome and the reference genome was used as the starting point. Genes of interest were identified and mapped to the reference genome using BLAST, and corresponding variants were extracted with bcftools. Genetic distances were calculated for three gene groups, phylogenetic markers, effector/virulence associated genes, and fungicide resistance associated genes. Neighbor-Joining trees were constructed from the resulting distance matrices, and Partitioning Around Medoids (PAM) clustering was used to identify genetic groups based on average silhouette width.

The three gene groups showed contrasting patterns of genetic structure. The phylogenetic gene set was partitioned into six clusters (silhouette width = 0.65), while the effector gene set produced seven clusters (0.64), indicating relatively well-defined groupings. In contrast, fungicide resistance associated genes produced eight clusters with lower separation (0.34). Comparison between clusterings using the adjusted Rand index (ARI) showed the highest agreement between effector and fungicide resistance associated genes (ARI = 0.812), meaning that genes that paired in one category also paired in the other, followed by effector and phylogenetic genes (0.554), and phylogenetic and fungicide resistance associated genes (0.473).

Overall, these results suggest that the three functional categories capture partially shared but distinct patterns of genetic structure among isolates, potentially reflecting differences in evolutionary constraints and selective pressures acting on genes with distinct biological functions. This genomic characterization provides a basis for further investigation of the evolution, epidemiology, and fungicide resistance of C. kikuchii populations in Argentina.

DRAFT – UNDER REVIEW

POSTER #66 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Uncovering Peripheral Biomarkers of Celiac Disease via Integrative Transcriptomic Analysis

Source: 05_posters/P066.tex · Excel row 53

Maggio JF[1,2,3,4], Tolosa G[2], De Marzi MC[1,2,3]

[1] Grupo de Investigaciones Básicas y Aplicadas en Inmunología y Bioactivos (GIBAIB), Instituto de Ecología y Desarrollo Sustentable (INEDES) (Universidad Nacional de Luján – CONICET), Argentina [2] Universidad Nacional de Luján, Departamento de Ciencias Básicas, Argentina [3] FOCIS Center of Excellence - Vaccines and Immuno Therapies against Infections and Cancers (VITIC) [4] Comisión de Investigaciones Científicas (CIC)

Celiac disease (CeD) is an immune-mediated disorder driven by related prolamin proteins known as gluten, found in wheat, rye, and barley, the prevalence of which is increasing worldwide. Diagnosing CeD is complex due to its clinical variability and the heterogeneous response to a gluten-free diet, highlighting the need for novel immunological biomarkers.

In a previous study, we selected 34 genes using differential expression (DE) analysis of three RNA-Seq datasets derived from duodenal biopsies of celiac patients versus controls. Based on these results, we aimed to evaluate the performance of these genes in distinguishing celiac patients from controls, specifically those on a gluten-free diet, using a microarray dataset of PBMCs. Furthermore, we compared this performance against datasets involving blood and PBMCs from conditions with symptoms similar to CeD, such as ulcerative colitis (UC) and irritable bowel syndrome (IBS). To do this, we trained various Random Forest models using the R package caret to assess the importance of the variables, utilizing expression matrices downloaded from the NCBI GEO database that had been quantile-normalized, gene-wise centered, and log-transformed. We used the Gini values obtained from the trained models to compare the genes' ability to discriminate between patients and controls. Additionally, we performed differential expression analysis using the R package DESeq2 and compared the results for the selected genes. Applying a 30% relative importance cutoff, we identified a 14-gene signature capable of distinguishing celiac patients from controls in blood samples; genes showing statistically significant overexpression in the DE analysis included GBP1, CD24, DLGAP5, and PFKFB4, while ABCC2, ACE, and SLC15A1 showed significant underexpression. This signature differed for the other diseases in terms of both differential expression and Gini values.

The molecular signature obtained from blood samples effectively distinguished CeD patients from controls. Furthermore, the results indicate that, potentially, it can serve to differentiate patients with various diseases presenting symptoms similar to CeD.

DRAFT – UNDER REVIEW

POSTER #67 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Improving Microbiome Beta Diversity Analysis through a Parametric Model of Abundance Gradients

Source: 05_posters/P067.tex · Excel row 54

Harvey N[1], Lucero Cáceres JF[1], Farall A[1], **Cassol I**[1]

[1] Facultad de Ingeniería, Universidad Austral, Pilar, Buenos Aires, Argentina

Introduction: A β-diversity distance collapses the comparison between two samples into a single value, obscuring which components of their difference determine it. Sample pairs with similar Bray–Curtis distances may differ in the number of taxa involved, the distribution of abundance changes across taxa, and their direction, since conventional formulations based on $|\Delta|$ discard sign. Objective: To decompose pairwise abundance differences into interpretable parameters and determine which components are captured by commonly used β-diversity distances. Methodology: For each pair of samples expressed as proportions, $\Delta j = p1j - p2j$ was divided using $\epsilon = 10^{-4}$ into positive and negative tails and a neutral zone. Each tail was fitted with a power law $|\Delta_{(r)}| \approx C \cdot r^{-\alpha}$, yielding 13 parameters describing tail size, scale and shape, neutral taxa, and redistributed abundance. Thirteen public human gut microbiome cohorts were evaluated using eight non-redundant distances, and results were pooled using random-effects meta-analysis. Results: Across cohorts, the eight distances were primarily associated with two components: the magnitude of abundance change and the number of taxa involved, which were essentially uncorrelated ($\rho \approx 0.01$). The total magnitude of abundance change strongly predicted Bray_edgeR (r = 0.96) and, after scaling by sequencing depth, Manhattan_raw (r = 0.99). The number of taxa present in only one sample predicted Manhattan_log (r = 0.96), Raup (r = 0.85), and Mountford (r = 0.79). These associations are consistent with the mathematical structure of each distance rather than indicating distinct biological signals. None of the eight distances captured tail shape or the directionality of abundance changes. Discussion: β-diversity distances are not interchangeable summaries of community differences: the evaluated metrics primarily represented magnitude and the number of taxa involved while discarding tail asymmetry and shape, information retained by the parametric model. Metric selection should therefore reflect which components of Δj are relevant to the biological question. Ongoing work evaluates additional incidence-based parameters and robustness to sequencing depth, rarefaction, and rare-taxon filtering.

DRAFT – UNDER REVIEW

POSTER #68 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Systematic Profiling of Expressed HRR Variants: An RNA-seq Analysis Pipeline Applied to Prostate Cancer Cohorts

Source: 05_posters/P068.tex · Excel row 55

Viappiani L[1,2], Korniansky E[1,3], Cayol F[4], Ruiz MS[1,2], Cotignola JH[1,2], Gueron G[1,2], Bizzotto JA[1,2,3]

[1] CONICET - Universidad de Buenos Aires, Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN), Buenos Aires, Argentina [2] Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Química Biológica, Laboratorio de Inflamación y Cáncer, Buenos Aires, Argentina [3] Universidad Argentina de la Empresa (UADE), Instituto de Tecnología (INTEC), Buenos Aires C1073AAO, Argentina [4] Hospital Italiano de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina

PARP inhibitors are approved for some prostate cancers with homologous recombination repair (HRR) gene alterations, so accurate detection of these mutations is essential for patient selection. This is usually based on DNA sequencing, which is unavailable for most public cohorts, limiting comparisons of mutational landscapes across studies. Public RNA-seq datasets are abundant, but extracting clinically relevant variants from them remains technically challenging. We developed a reproducible pipeline to identify HRR mutations from public RNA-seq data and used it to characterize the molecular landscape of prostate cancer.

Raw reads from 13 public datasets (872 unique samples from 1063 sequencing runs; normal tissue, localized tumors and metastases) were downloaded from NCBI SRA. Expressed variants were called with RNAmut; custom Python scripts converted SNVs, indels and internal tandem duplications into GRCh38 VCF files, which were normalized and annotated with Ensembl VEP and OncoKB. A variant allele frequency (VAF) filter was evaluated to reduce low-confidence calls, and clinical metadata were integrated for downstream analyses.

The proportion of patients with at least one Oncogenic/Likely Oncogenic variant varied widely across cohorts: it was lowest (7%) in a tumor and matched normal tissue cohort, and exceeded 90% in both an metastatic castration-resistant prostate cancer (mCRPC) cohort and a cohort of diagnostic biopsies spanning non-metastatic to polymetastatic disease. Multiple correspondence analysis of 338 pre-treatment primary tumors showed that three radical prostatectomy cohorts had similar profiles with no cohort-driven separation, supporting their integration. In contrast, a cohort of paired pre- and post-androgen deprivation therapy samples was more heterogeneous, with higher mutation burden, and clustered closer to the mCRPC cohort. A VAF > 0.2 filter markedly increased specificity: in the mCRPC cohort, patients with oncogenic variants fell from 95.1% to 25.4%, still well above localized disease. Across all cohorts, PARP1 emerged as the most frequently altered gene, detected in 32% of samples, frequently co-occurring with MSH2 (11.6% of the total cohort, representing 36% of PARP1-mutated cases) and ATM (11.2%).

This scalable pipeline enables systematic profiling of HRR mutations in public RNA-seq data lacking matched DNA sequencing, providing a basis to relate expressed mutations, genomic alterations and clinical outcomes in prostate cancer.

DRAFT - UNDER REVIEW

POSTER #69 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Inferring spatial gene expression in the non-model giant kidney worm Dioctophyme renale by extrapolation from Brugia malayi spatial transcriptomics

Source: 05_posters/P069.tex · Excel row 56

Khatchikian VL[1], Lopez Malizia A[2], Franchini G[3], Kamenetzky L[1]

[1] Institute of Biosciences Biotechnology and Translational Biology (iB3), Department of Physiology and Molecular and Cellular Biology, University of Buenos Aires (UBA), Argentina. [2] Josep Carreras Leukaemia Research Institute, Barcelona, Spain. [3] Institute of Biochemical Research of La Plata "Prof. Dr. Rodolfo R. Brenner" (INIBIOLP-CONICET), National University of La Plata (UNLP), Argentina.

Nematodes are a highly diverse group of helminths affecting human health, livestock and agriculture, yet advanced omics technologies have been adopted late in parasitic helminths. Our group is part of the Dioctophyme renale genome project, which has produced the genome, transcriptome and proteome of this parasite, the giant kidney worm able to exceed one metre and destroy its host's kidney. To deepen its expression profiling, we assess the feasibility of spatial transcriptomics (ST) in this non-model organism. As no direct spatial data exist for D. renale, we hypothesise that spatial gene expression patterns are conserved with phylogenetically related parasitic nematodes and can be inferred by orthology analysis of their spatial transcriptomic data.

Surveying all public parasitic-helminth spatial-omics repositories, we found only two peer-reviewed genome-scale datasets, of which a single one came from a parasitic nematode. Accordingly, we selected Brugia malayi as reference, using a publicly available ST dataset of three adult female sections (563 spots). Following the same parameters as the reference paper for data normalization, integration, and batch correction, we processed the data using the R package Seurat. We then applied two independent clustering methods: graph-based clustering (Louvain), as used in the original study, and k-means, with k evaluated from 2 to 16 via elbow plot and silhouette score.

Louvain recovered four spatial clusters, each with significant marker genes (adjusted $p<0.05$, log2 fold-change>0.5; 191 markers total). Compared to k-means at k=4 (selected based on Louvain's output and the elbow plot) and k=5 (chosen for the silhouette analysis), marker-gene overlap between methods exceeded chance (p down to ~1e-62). The two partitions assigned up to 67% of spots concordantly and recovered up to 59% of each other's marker genes, and chance-corrected agreement increased with resolution (adjusted Rand index 0.24 at k=4 to 0.29 at k=5), with k=5 additionally recovering a small cluster missed at k=4.

Ongoing work includes annotating these clusters to anatomical regions through orthology to Caenorhabditis elegans and reference expression atlases, and will then identify D. renale orthologs and project their spatial signatures in silico to predict the tissue localization of genes of interest. This transferable framework will expand knowledge of D. renale functional biology and apply to other non-model nematodes lacking direct spatial data.

DRAFT – UNDER REVIEW

POSTER #70 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Microbiome Composition and Predicted Functions Across Sequencing Platforms

Source: 05_posters/P070.tex · Excel row 57

Stricker A[1], Bravo F[1], Mangano A[1], Fernández MF[1]

[1] Servicio de Microbiología, Hospital de Pediatría "J. P Garrahan", Argentina.

Different sequencing technologies can influence microbiome characterization and downstream analyses. Illumina MiSeq and Oxford Nanopore MinION were compared for microbiome profiling in pediatric fecal samples and in a microbial standard of known composition, focusing on taxonomic profiles, microbial association networks, and predicted functional profiles.

A total of eight fecal samples from pediatric patients undergoing hematopoietic stem cell transplantation (HSCT), collected before fecal microbiota transplantation (PRE-FMT) and seven days after transplantation (FMT+7), were analyzed on both platforms. Illumina data were processed with Nextflow Ampliseq (DADA2 and SILVA) and MinION data with nanoASV (minimap2 and SILVA). Downstream analysis included microbial association networks using NetCoMi and functional inference using PICRUSt2. Four ZymoBIOMICS standard samples were included to assess platform-associated differences under controlled conditions. Agreement between profiles was assessed using Bray-Curtis dissimilarity and Spearman correlation; lower Bray-Curtis values and higher Spearman correlations indicated closer agreement between profiles.

In clinical samples, differences increased with taxonomic resolution, with mean Bray-Curtis dissimilarities of 0.69 at genus level and 0.86 at species level. Network structure showed partial conservation: 56% of associations were shared between platforms at PRE-TMF and 44% at TMF+7, with complete agreement in shared associations. Functional profiles were considerably more similar, with a mean Bray-Curtis dissimilarity of 0.18 and a mean Spearman correlation of 0.92 across matched patient samples. In the ZymoBIOMICS standard, MinION showed closer agreement with the theoretical taxonomic composition, particularly at species level, while Illumina remained slightly closer at genus level. Functional results remained highly concordant, with 344 metabolic pathways shared (91.0% overlap) and Spearman correlations above 0.96.

Overall, platform-dependent differences persisted in clinical taxonomic profiles, requiring caution in cross-platform comparisons. MinION nevertheless showed substantial agreement in downstream analyses and closer agreement with the expected composition in the controlled standard. These results support its potential as an alternative for microbiome profiling, while highlighting the need for platform-specific validation.

DRAFT – UNDER REVIEW

POSTER #71 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Transcriptomic dynamics of the peripheral immune response to FMDV vaccination in cattle

Source: 05_posters/P071.tex · Excel row 58

Netzel P[1], Gupta S[2], Capozzo AV[3], Gupta D[2], Freire T[1]

[1] Laboratorio de Inmunomodulación y Vacunas, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay. [2] Translational Bioinformatics Group, International Centre for Genetic Engineering and Biotechnology (ICGEB), Jawaharlal Nehru University, New Delhi, India. [3] Centro de Altos Estudios en Ciencias Humanas y de la Salud (CAECIHS). Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)- Universidad Abierta Interamericana, Buenos Aires, Argentina.

Foot-and-mouth disease (FMD) is a highly contagious viral disease affecting livestock and is caused by FMD virus (FMDV). Vaccination is essential for FMD control, although variability in immune responses to inactivated vaccines has been described. The objective of this study was to characterize the transcriptomic response in peripheral blood mononuclear cells (PBMCs) associated with FMDV vaccination in cattle. Bulk RNA-seq was performed on PBMCs collected from calves vaccinated on days 0 and 140 with an inactivated oil-adjuvanted FMDV vaccine containing A24/Cruzeiro and O1/Campos strains. Samples were collected on days 0, 60, 140 and 170 (relative to the first vaccination). Differentially expressed genes (DEGs) were identified, and their functional enrichment was evaluated using over-representation analysis (ORA) and Gene Set Enrichment Analysis (GSEA).

Longitudinal analysis identified a sustained transcriptional response, with IL2RA (immune signaling), CREB3L1 (transcriptional regulation) and SLCO4A1 (membrane transport) consistently upregulated at T60, T140 and T170 compared with baseline (T0). At T60, uniquely upregulated genes included IL12B, CCL3, NOS2, SERPINB2 and TNFSF15, consistent with inflammatory and innate immune activation. GSEA showed positive enrichment of TNFA signaling via NF-κB, IL2-STAT5 signaling, T-cell differentiation and interferon-related antiviral responses. At T140, TNFA signaling remained enriched, while oxidative phosphorylation, electron transport and cell-cycle processes were negatively enriched, indicating a distinct pre-booster state. The T170 vs. T140 comparison showed the strongest transcriptional shift, with positive enrichment of oxidative phosphorylation, B-cell receptor signaling, myeloid immunity, and antigen processing and presentation, alongside negative enrichment of TNFA signaling, P53, apoptosis and hypoxia. Blood transcriptional module (BTM) analysis through GSEA supported this post-booster reorganization, showing increased B-cell/humoral, myeloid and antigen-presentation programs, and reduced T-cell signaling.

Overall, primary vaccination elicited inflammatory and T-cell-associated transcriptional responses, whereas booster vaccination induced complementary immunometabolic remodeling, with enhanced adaptive immune signaling following the booster. Together, these distinct but complementary responses provide a framework for identifying gene expression signatures associated with vaccine-induced protection.

DRAFT – UNDER REVIEW

POSTER #72 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Annotating the secretome of an under-represented parasitic nematode, the giant kidney worm Dioctophyme renale

Source: 05_posters/P072.tex · Excel row 59

Sananez I[1], Inchausti Ravella LC[2], Giorello N[3], Calupiña K[1], Maldonado L[1], Ingravidi M[1], Macchiaroli N[1], Smircich P[2], Franchini G[3], Kamenetzky L[1]

[1] Institute of Biosciences Biotechnology and Translational Biology (iB3), Department of Physiology and Molecular and Cellular Biology University of Buenos Aires (UBA), Argentina. [2] Bioinformatics Laboratory, Department of Genomics, Clemente Estable Biological Research Institute (IIBCE), Montevideo, Uruguay [3] Biochemistry Research Institute of La Plata "Professor Doctor Rodolfo R. Brenner" (INIBIOLP), Faculty of Veterinary Sciences, Universidad Nacional de La Plata (UNLP), Argentina

Parasitic nematodes outside the model species are strongly under-represented in sequence and structure databases, which limit functional annotation: helminth genomes encode between 10,000 and 20,000 protein-coding genes, about half of them functionally unknown, and their closest sequence relative is often another uncharacterized protein from a neglected species (Palevich et al., 2018). This matters for secreted proteins, which mediate the interaction with the host and are the source of candidate antigens and drug targets.

From the 16,629 proteins of the parasitic nematode Dioctophyme renale, we identified in silico 2,449 excretory/secretory proteins (ESPs). We screened this set against a curated database of 66 experimentally characterized helminth immunomodulatory proteins (helminth-immuno-DB) by BLASTP (identity>30%, coverage>60%). Next, the resulting immunomodulatory ESPs (Drenale-immuno-ESPs) were annotated by using sequence, structure and experimental evidence as follows: a) protein domain and functional family assignment from sequence, b) structures predicted with ColabFold (AlphaFold2) and searched against AFDB-SwissProt v6 and PDB100 using Foldseek, and c) proteomic analysis of adult worm tissues by mass spectrometry (MS/MS).

We identified 93 Drenale-immuno-ESPs matching helminth-immuno-DB proteins, which map to fourteen protein families, dominated by peptidases and peptidase inhibitors, with cathepsin B-like proteins and serpins accounting for half (n=46). Structural analyses supported the sequence-based annotation of 89 Drenale-immuno-ESPs through concordant Pfam domain assignments, with no discordant case; the remaining four did not meet the quality criteria. Notably, 40 Drenale-immuno-ESPs were supported by worm proteomic data, unevenly across families, including 15 of 30 cathepsin B-like proteins, 9 of 16 serpins, all 3 cathepsin L, and none of the 11 acetylcholinesterases.

This is the first survey of putative immunomodulatory effectors in D. renale, a Clade I nematode. Structural comparison reproduced the family assignment throughout, while proteomic detection reached fewer than half. Candidates supported by more than one line of evidence, and detected in the proteome, are where follow-up as antigens or therapeutic targets should start. Moreover, 37% of predicted D. renale ESPs lack any Pfam domain assignment: extending this annotation to the whole secretome is the next step, and where it should contribute most.

DRAFT – UNDER REVIEW

POSTER #73 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Unveiling the structure of bacterial communities in groundwater using long-read metagenomic sequencing.

Source: 05_posters/P073.tex · Excel row 60

Massicot A[1], Figuerola E[1,2]

[1] iB3- Instituto de Biociencias, Biotecnología y Biología Traslacional - DFBMC - FCEN - UBA - CABA - Argentina.
[2] CONICET - CABA - Argentina.

Nitrate contamination in groundwater is a global environmental concern that particularly affects the Pampas region in Argentina, compromising the Pampeano and Puelche aquifers (key drinking water sources for Buenos Aires Province). High nitrate intake could cause methemoglobinemia in vulnerable populations and has been linked to nitrosamine formation.

Among available industrial treatments, biological denitrification (BD) stands out for converting nitrate into harmless nitrogen gas, offering better water recovery, lower environmental impact, and reduced operating costs compared to ion exchange or reverse osmosis. In a previous study (Dotto et al., 2021), we optimized conditions for a pilot BD system and observed an acclimation period of 100 hydraulic retention times, representing significant resource costs. Bioaugmentation with selected strains emerges as a strategy to shorten this period.

To identify suitable candidates, we applied a top-down approach to characterize natural denitrifying communities from laboratory-scale reactors using long-read Oxford Nanopore Technologies (ONT) sequencing.

Our results describe taxonomic distribution and relative abundance during the acclimation period. We found a positive correlation between the Pseudomonas genus and nitrate concentration over time, while Acidovorax and other genera from the Comamonadaceae family remained stable. Hydrogenophaga and Methyloversatilis abundance increased with nitrite concentration. In particular, Methyloversatilis reached its maximum abundance at 50 hours.

Regarding functional potential, narG abundance was positively correlated with nitrate concentration. For nitrite reduction, nirK abundance increased with a delay relative to peak nitrite levels, coinciding with the maximum abundance of Methyloversatilis. norB remained nearly constant and nosZ increased during the first 50 hours before stabilizing. By integrating these results with genomic sequencing of isolates from laboratory reactors, we identified Pseudomonas and Acidovorax as strong candidates for bioaugmentation.

DRAFT – UNDER REVIEW

POSTER #74 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Bioinformatic Identification and Experimental Validation of microRNAs Associated with Rejuvenation in Ilex paraguariensis

Source: 05_posters/P074.tex · Excel row 61

Duarte MJ[1,2], Acevedo RM [1,2,3], Ortiz NL[1,2], Alvarez MY [1,2], Sansberro PA[1,2,3]

[1] Universidad Nacional del Nordeste, Facultad de Ciencias Agrarias. [2] IBONE, Laboratorio de Biotecnología Aplicada y Genómica Funcional. [3] CONICET.

In plants, the transition from the juvenile to the adult phase involves coordinated molecular and developmental changes in which conserved microRNAs (miRNAs) play key regulatory roles. This process is particularly relevant in woody species in which maturation reduces the capacity for vegetative propagation. In this study, we investigated miRNA profiles associated with rejuvenation in Ilex paraguariensis, a species of economic and ecological importance, using small RNA sequencing and bioinformatic approaches.

Small RNA libraries from juvenile, adult, and serially rooted plants were processed through a quality-control and miRNA discovery pipeline. Raw reads were evaluated with FastQC, adapter and low-quality sequences were removed using Trimmomatic and Cutadapt, and reads outside the 18–34 nt range were discarded. Filtered sequences were mapped to the I. paraguariensis reference genome with Bowtie to assess their genomic origin. Redundant reads were collapsed into unique sequences and reads corresponding to rRNAs, tRNAs, small nuclear and small nucleolar RNAs were removed. Known miRNAs were identified by comparison with miRBase v22.1, whereas unmapped candidates were analyzed with miRDeep-P, integrating genome alignment, precursor extraction, secondary-structure prediction with RNAfold, and plant miRNA structural and expression criteria. Potential miRNA targets were predicted using psRNATarget, followed by Gene Ontology enrichment analysis with BiNGO in Cytoscape.

For differential expression analysis, filtered reads were mapped against known and predicted miRNAs and analyzed with DESeq2. Low-abundance families were excluded, counts were normalized, and differentially expressed miRNAs were identified using fold-change and statistical significance thresholds. Selected candidates were subsequently validated by RT-qPCR using the $\Delta\Delta$Cq method and IpRTF as reference gene.

The analysis identified conserved miRNA expression patterns associated with developmental phase, including higher miR156/miR157/miR528 abundance in juvenile material and predominant miR172 expression in adult plants. These results support the involvement of miRNA-mediated regulation in the molecular changes associated with rejuvenation and provide a computational framework for identifying candidate regulatory miRNAs in woody plants with limited clonal propagation capacity.

DRAFT - UNDER REVIEW

POSTER #75 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

High-resolution single-nucleus transcriptomics reveals striatal molecular remodeling during the off-L-DOPA state

Source: 05_posters/P075.tex · Excel row 62

Martínez BF[1], Schuster CD[3], Li CL[2], Martí MA[3], Ferrario JE[1], Cenci MA[2], Bordone MP[1]

[1] Instituto de Biociencias, Biotecnología y Biología Traslacional (iB3), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, CABA, Argentina. [2] Department of Experimental Medical, Science, Wallenberg Neuroscience Center, Lund University, Lund, Sweden [3] IQUIBICEN, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, CABA, Argentina.

Parkinson's disease (PD) is characterized by progressive dopaminergic degeneration, leading to motor impairment and profound alterations in striatal circuitry. Although L-DOPA remains the most effective treatment for motor symptoms, chronic administration frequently results in L-DOPA-induced dyskinesia (LID). Transcriptomic studies of LID have mainly focused on the striatum during the dyskinetic response, while the molecular landscape of the off-L-DOPA state, when drug effects decline and parkinsonian features re-emerge, remains largely unexplored.

Here, we used single-nucleus RNA sequencing (snRNA-seq) and integrative computational analyses to investigate striatal transcriptional remodeling during the off-L-DOPA state. Hemiparkinsonism was induced by unilateral 6-hydroxydopamine (6-OHDA) lesion, followed by chronic saline or L-DOPA administration. Striatal samples were collected 24 h after the last injection and processed for snRNA-seq. After quality control and preprocessing, major neuronal and non-neuronal populations were identified, followed by higher-resolution characterization of spiny projection neurons (SPNs). D1-SPN and D2-SPN subtypes were resolved using transcriptional markers associated with dopamine receptor expression, striatal compartmentalization, and spatial distribution. Differential expression and functional enrichment analyses were then used to characterize condition- and cell type-specific molecular changes. In parallel, CellChat was applied to infer ligand–receptor interactions and investigate the reorganization of intercellular communication across striatal populations.

This integrative analysis revealed distinct transcriptional responses to dopaminergic denervation and chronic L-DOPA treatment across SPN populations, together with changes in the predicted organization of striatal cell–cell communication. Higher-resolution SPN characterization further revealed transcriptional heterogeneity beyond conventional D1-SPN and D2-SPN classifications.

Overall, our study provides a computationally resolved view of the striatal molecular landscape during the off-L-DOPA state. By integrating cell-type and subtype-specific transcriptomics with functional and intercellular communication analyses, this work expands the characterization of an understudied phase of L-DOPA treatment and provides a framework for investigating molecular adaptations associated with PD and LID.

DRAFT – UNDER REVIEW

POSTER #76 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Plastic-degrading potential in the South Atlantic plastisphere: taxonomy and polymer targets

Source: 05_posters/P076.tex · Excel row 63

Pomme P[1], Hozbor MC[2], Peressutti S[2], Di Mauro R[2,3], Corti-Monzon G[3,4], **Soria M**[1]

[1] Cátedra de Microbiología Agrícola. Facultad de Agronomía, INBA-CONICET. Universidad de Buenos Aires. Argentina [2] Instituto Nacional de Investigación y Desarrollo Pesquero (INIDEP), Mar del Plata, Argentina. [3] Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET), Buenos Aires, Argentina. [4] Instituto de Ciencia y Tecnología de Alimentos y Ambiente, INCITAA, Universidad Nacional de Mar del Plata, Argentina.

Marine microplastics (MPs) are rapidly colonized by biofilms whose members may encode enzymes able to attack synthetic polymers, yet the taxonomic identity and polymer specificity of that potential remain poorly resolved in the South Atlantic. We surveyed the plastisphere of MP particles hand-picked from surface waters of the Argentine-Uruguayan Common Fishing Zone (34-40 S), collected with a continuous onboard filtration system in late spring. Illumina shotgun reads were quality controlled with FastQC, multiqc, adapter-related reads were removed and low-quality ends trimmed with FastP. Contigs were assembled with MEGAHIT, genes predicted with Prodigal, contig abundances estimated with CoverM, and classified with MMseqs2; predicted proteins were screened against a combined PAZy-PlasticDB reference to retrive putative plastic-active enzymes.

Contigs carrying at least one candidate gene spanned 284 genus-level lineages and were dominated by Gamma- and Alphaproteobacteria: Luminiphilus (8.8% of coverage-normalised abundance), Pelagibacter (3.3%), Pseudoalteromonas (2.1%), Pseudothioglobus (1.8%) and Salinivivens (1.6%); the ten leading lineages summed to 34.5% and the top fifty to 71%. Taxa previously linked to hydrocarbon or polymer turnover were recurrent but never dominant (Alloalcanivorax 1.0%, Thalassospira 1.0%, Cobetia 0.9%, Hyphomonas 0.7%, Marinobacter 0.6%), and archaea (Poseidoniales, Marine Group II, Nitrosopumilus) contributed 3.4%. Notably, 52% of the abundance belonged to uncultured lineages without formal names, indicating a largely uncharacterized reservoir.

Hits mapped to 21 polymer targets, with hydrolysable esters and polyethers prevailing: PHA (31.5%), PEG (20.6%), PLA (polylactic acid, 10.7%), PET (polyethylene terephthalate, 10.3%), polyamide (7.5%) and polyurethane (5.3%) accounted for 85.9% of the signal. Ester- and amide-bond polymers together reached ~73%, whereas enzymes targeting carbon-backbone polymers (PE, LDPE, PS, PP) were scarce (4.4% combined), consistent with their recalcitrance. The PHA signal probably reflects native turnover of microbial storage polyesters rather than a response to pollution.

These preliminary results point to a broad, taxonomically dispersed and phylogenetically novel hydrolytic repertoire biased towards ester-bond polymers.

DRAFT – UNDER REVIEW

POSTER #77

TRACK: Genomics, Transcriptomics, Epigenomics, Metagenomics

Multi-Omic Characterization of Hepatocellular Carcinoma across Liver Fibrosis Stages: A Comparison of Ishak 0 and Ishak

Source: 05_posters/P077.tex · Excel row 64

Del Savio V[1], Nadel M[1], Cassol I[1]

[1] Universidad Austral

Introduction: Hepatocellular carcinoma (HCC) is the most common primary liver cancer and frequently develops in the context of chronic liver disease and hepatic fibrosis. Different stages of fibrosis may be associated with distinct molecular alterations, contributing to the biological heterogeneity of HCC. Characterizing these alterations across multiple molecular layers may provide insight into the molecular features associated with different fibrosis stages. Objective: To characterize molecular differences between HCC patients without fibrosis and those with established cirrhosis by comparing patients classified as Ishak 0 and Ishak 6 across multiple omic layers. Methodology: Clinical and molecular data from The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC) cohort are being analyzed. Patients are classified according to the Ishak fibrosis score into two groups: Ishak 0, corresponding to no fibrosis, and Ishak 6, corresponding to established cirrhosis. Molecular differences between groups are being investigated using RNA sequencing, somatic mutation, and DNA methylation data. Each molecular layer is subjected to quality control, preprocessing, and statistical analyses appropriate to the data type. Differentially expressed genes, molecular alterations, and enriched biological pathways are being evaluated to characterize differences between the two fibrosis groups. The analysis is currently ongoing, and the integration of results across molecular layers is in progress.

DRAFT – UNDER REVIEW

POSTER #78 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

BAL-comp: an automated workflow for Lactic Acid Bacteria comparative genomic analysis

Source: 05_posters/P078.tex · Excel row 65

Córdoba JM[1], Urquiza Martinez MP[1], Núñez-Tomás E[2], Marín-García P[2], Saavedra L[1], Hebert EM[1]

[1] Centro de Referencia para Lactobacilos (CERELA) - CONICET [2] mgviz.org

Introduction. Lactic acid bacteria (LAB) are a diverse group of microorganisms widely used in food fermentation and increasingly recognized for their health-promoting properties and biotechnological potential. Advances in next-generation sequencing (NGS) technologies have enabled the rapid and cost-effective generation of large-scale genomic datasets, creating new research opportunities based on publicly available genome sequences. Consequently, there is a growing need for computational tools that facilitate and orchestrate comparative genomics analysis, and help to explore the genetic diversity and evolutionary relationships among LAB species and subspecies. However, integrating multiple bioinformatic tools and analytical steps can hinder reproducibility, scalability, and data standardization. To address these challenges, we developed an automated workflow for producing LAB de novo genome assemblies from short reads and comparative genomic analysis. Methodology. The workflow was implemented in Snakemake within a modular framework using reproducible software environments, configurable parameters, version control, and automated reports. The pipeline includes read quality assessment, de novo assembly, assembly evaluation, taxonomic analysis, genome annotation, average nucleotide identity (ANI) analysis, phylogenetic inference, and identification of genomic features associated with strain safety and biotechnological potential. As a proof of concept, the workflow was applied to publicly available genomes and genomes obtained from strains of interest sequenced within our laboratory. Results. The automated implementation consolidated all analytical stages into a unified framework, reducing manual processing and improving consistency and reproducibility across datasets. The workflow also produces standardized comparative reports summarizing assembly statistics, genome quality indicators, taxonomic assignments, phylogenetic relationships, ANI values, and functional genomic features, facilitating strain comparison. The modular design facilitates the incorporation of additional analytical components and larger genomic datasets. Conclusions. This workflow provides a reproducible and scalable pipeline for comparative genomics of LAB, facilitating the identification of genetic traits associated with safety, adaptation, and biotechnological potential, while supporting Findable, Accessible, Interoperable, and Reusable (FAIR) practices in microbial genome research.

DRAFT – UNDER REVIEW

POSTER #79 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Same genes, different crowd: host community turnover underlies seasonal shifts in the sugarcane soil resistome

Source: 05_posters/P079.tex · Excel row 66

Amago Prato L[1], Ortiz Rocca L[1], Montecchia M[1,2], Chalco Vera J[3], Acreche M[4], Correa O[1], Soria M[1,2]

[1] Cátedra de Microbiología Agrícola, Facultad de Agronomía, Universidad de Buenos Aires. Av. San Martín 4453 C.A.B.A., Argentina [2] Instituto de Biociencias Agrícolas y Ambientales, INBA (UBA-CONICET). Av. San Martín 4453 C.A.B.A., Argentina [3] Instituto de Investigación Animal del Chaco Semiárido (IIACS), Centro de Investigaciones Agropecuarias (CIAP), Instituto Nacional de Tecnología Agropecuaria (INTA)-CONICET, Chañar Pozo s/n, (T4113), Leales, Tucumán, Argentina [4] INTA Estación Experimental Agropecuaria Salta-CONICET. Ruta Nac. 68, Km. 172, CP 4403, Cerrillos, Salta, Argentina

Antimicrobial resistance causes over one million deaths annually, and addressing it under One Health requires understanding environmental reservoirs of antibiotic resistance genes (ARGs). Soil is the most diverse reservoir and the evolutionary source of determinants acquired by clinical pathogens through horizontal gene transfer, yet subtropical South American agroecosystems remain understudied. We characterized the potential resistome of sugarcane soils at INTA Famaillá (Tucumán, Argentina). This site, under soybean-sugarcane rotation since 2012, has no record of bactericide use, lies 5.6 km from the nearest town and has no major watercourses or reported livestock operations nearby. Four metagenomes were sequenced per sampling (December and March).

ARGs were identified with DeepARG, a deep-learning classifier suited to divergent environmental genes, at the level of reads (normalized per 16S rRNA gene copy) and of proteins predicted with Prodigal from MEGAHIT-assembled contigs ($\geq$1500 bp). Contigs were classified with MMseqs2 against the Genome Taxonomy Database, quantified with CoverM and screened for plasmid origin with Platon.

Total ARG abundance rose from 3.4 (December) to 3.7 (March) copies per 16S rRNA copy and was dominated by multidrug efflux genes (80% and 74%, respectively). The increase was concentrated in glycopeptide (VanRS system), tetracycline and rifamycin resistance genes, whose read-based abundance increased 1.5- to 2-fold. At the contig level, Actinomycetota accounted for 93% of the increase in ARG abundance, while ARG density within each phylum remained stable. In parallel, Actinomycetota increased from 6% to 24% of the assembled community, driven by the same families carrying the additional ARGs (Gaiellaceae, Solirubrobacteraceae, Nocardioidaceae, Streptomycetaceae), which include natural producers of these antibiotics. Community composition changed more than the ARG class profile (Bray-Curtis 0.19–0.22 vs. 0.15). No ARG-bearing contig showed robust plasmid evidence (0/301 in December; 4/757 in March) or was assigned to clinically relevant taxa such as Enterobacteriaceae or Enterococcus.

These results point to an intrinsic, predominantly chromosomal resistome whose seasonal variation reflects shifts in host community composition rather than anthropogenic selective pressure. They provide a first baseline for environmental ARG surveillance in sugarcane agroecosystems of Northwestern Argentina.

DRAFT - UNDER REVIEW

POSTER #80 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Shotgun Metagenomic Profiling of Pelagic and Benthic Microbial Communities Across the Argentine-Uruguayan Common Fishing Zone

Source: 05_posters/P080.tex · Excel row 67

Palermo C[1], **Derosa B**[1], Soria MA[1,2], Hozbor C[3], Peressutti S[3], Jurquiza V[3], Ortiz Rocca L[1]

[1] Cátedra de Microbiología Agrícola, Facultad de Agronomía, Universidad de Buenos Aires. Av. San Martín 4453 C.A.B.A., Argentina [2] Instituto de Biociencias Agrícolas y Ambientales, INBA (UBA-CONICET). Av. San Martín 4453 C.A.B.A., Argentina [3] Instituto Nacional de Investigación y Desarrollo Pesquero (INIDEP). Mar del Plata, Argentina.
* Both authors contributed equally.

Marine microorganisms play an essential role in biogeochemical cycles, with community composition and diversity varying across pelagic and benthic environmental gradients. Analyzing their taxonomic architecture across these transitions requires robust shotgun metagenomic approaches. The Argentine-Uruguayan Common Fishing Zone is a highly productive marine ecosystem characterized by marked environmental heterogeneity. Here, we implemented a read-based shotgun metagenomic workflow to evaluate prokaryotic community structure and diversity across bathymetric and oceanographic gradients. Water and sediment samples were collected at 22 stations along three transects perpendicular to the isobaths, covering Argentine shelf, Río de la Plata estuarine, and Uruguayan sectors from coastal waters to the continental slope. Total DNA paired-end 150 bp Illumina sequencing yielded 44 biological datasets, comprising 27 water column samples spanning depths from 5 to 1200 meters and 17 benthic sediment samples. Reads were quality-filtered with fastp and classified against the Standard database using Kraken2, with abundances re-estimated at genus level via Bracken to evaluate community composition across depth layers, transects, and environments. Within the water column, regional transects accounted for significant variations in alpha diversity, Kruskal-Wallis $p<0.0001$, and beta diversity, PERMANOVA $p=0.0001$, whereas depth was not significant for alpha nor beta diversity. Taxonomic composition at genus level was structured by Synechococcus 18%, Candidatus Pelagibacter 17%, and Pseudoalteromonas 9%. Candidatus Nitrosomarinus reached 29% off Uruguay, Candidatus Pelagibacter reached 91% in AR6 at 115 meters, and Pseudoalteromonas peaked at 34% at 1200 meters in RP12. Conversely, benthic sediment exhibited spatial homogeneity across transects for alpha and beta diversity. Sediment was dominated by Bacillota, averaging 45% versus 1% in water, and Pseudomonadota at 38%. At genus level, pelagic dominants were absent, and sediment was characterized by psychrophilic Psychrobacter averaging 18% and reaching 78% in RP11, alongside spore-forming Bacillota including Fictibacillus, Rossellomorea, and Metabacillus.

Keywords: Shotgun metagenomics; Marine microbiome; Argentine-Uruguayan Common Fishing Zone; Taxonomic profiling.

DRAFT – UNDER REVIEW

POSTER #81 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

DESIGN AND IMPLEMENTATION OF A REPRODUCIBLE HPC PLATFORM FOR AUTOMATED RNA-SEQ DATA ANALYSIS

Source: 05_posters/P081.tex · Excel row 68

Suarez M.[1], Torres Peralta T.[1], Molina MG.[1], Villena J.[2], Albarracin L.[2]

[1] Laboratorio de Computación Científica, Depto de Ciencias de la Computación, Facultad de Ciencias Exactas y Tecnología, Universidad Nacional de Tucumán [2] Laboratorio de Inmunología Respiratoria, Escuela de Posgrado de Agricultura, Universidad de Tohoku

RNA-seq data analysis is a fundamental approach for studying gene expression and identifying transcriptional changes under different biological conditions. However, its processing involves multiple stages, heterogeneous tools, complex configurations, and substantial computational and storage requirements, especially for large datasets.

This project presents the design, implementation, and evaluation of a reproducible computational platform for RNA-seq analysis using high-performance computing (HPC). The workflow covers quality control, preprocessing, ribosomal RNA filtering, alignment or pseudo-alignment, quantification, differential expression analysis, and the generation of quality reports and visualizations. Experimental wet-lab procedures are excluded, as the workflow operates directly on FASTQ sequencing data.The computational workflow is implemented using Nextflow as its workflow management system, with nf-core/rnaseq serving as a reference framework for the organization and integration of established bioinformatics tools. The pipeline adopts a modular and configurable architecture in

which individual computational components are encapsulated using container technologies, enabling controlled software versions and reproducible execution environments.

The platform features input validation, execution management, structured logging, error handling, automated testing, version control, and provenance tracking. These ensure traceability of data, tools, parameters, and computational resources.

Additionally, the project includes a containerized R and Shiny-based application for exploring

analysis results. This GUI allows researchers to visualize differential expression data, principal component analyses, heatmaps, volcano plots, quality-control reports, and associated tables.

Ultimately, this platform integrates workflow engineering, HPC, software engineering,

reproducibility, and data analysis into a unified environment. By separating the execution layer from the result-exploration layer, it provides a reproducible, configurable, and accessible solution for RNA-seq analysis.

DRAFT – UNDER REVIEW

POSTER #82 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Genome-wide quantitative profiling of human core promoters reveals sequence determinants of transcriptional activity and noise

Source: 05_posters/P082.tex · Excel row 69

Iungman M[1,2], Uriostegui-Arcos M[3], Fiszbein A[3,4], Schor IE[1,2]

[1] Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE), Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina. [2] Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina. [3] Department of Biology, Boston University, Boston, MA, USA. [4] Faculty of Computing & Data Sciences, Boston University, Boston, MA, USA.

Sequence-to-expression models assume that promoter activity is largely hard-wired in DNA, but they are trained on endogenous data that confound the promoter with its genomic context, and they ignore other properties of transcription, such as its intrinsic noise (the cell-to-cell variability). We built a single-locus MPRA to measure, for nearly all annotated human core promoters, both the mean and the cell-to-cell variance of expression at one defined genomic site in HEK293T cells. Pooled cells were sorted into six EGFP gates and sequenced. We developed a pipeline that reconstructs single-promoter expression distributions using sampling-effort correction and spike-in standards (12,214 high-confidence promoters; replicate r = 0.86), and we scored noise with a mean-independent AUC-ROC metric. We integrated the measurements with endogenous promoter activity from FANTOM5 CAGE, ChIP-seq for more than 900 DNA-binding proteins (ReMap) and histone marks (ChIP-Atlas), conservation data and sequence-to-expression model predictions (PUFFIN).

Isolated core promoter activity correlated with endogenous usage across tissues, highlighting the information encoded in core promoter sequence, though more strongly for housekeeping than for tissue-specific promoters. It also correlated inversely with PUFFIN-predicted context dependence, linking intrinsic strength to autonomy. Young, transposon-derived core promoters were among the weakest overall, yet mixed models of reporter–endogenous discordance showed that they are still more active in the reporter than at their native locus, whereas TATA-box promoters show the opposite bias. Core promoters overestimated by the reporter are bound by chromatin repressors (SETDB1, NCOR1, L3MBTL2), suggesting that context-dependent silencing restrains them at their endogenous loci. CpG islands and TATA-boxes both increase activity but have opposite effects on noise, in agreement with previous scRNA-seq studies. Low noise was also associated with MLL1/2 binding and H3K4me3 deposition, even in promoters without CpG islands. Finally, alternative promoters of the same gene were no more similar than random pairs, and secondary promoters were typically weaker and noisier than main ones, except for switch-like pairs, where the secondary promoter becomes dominant in some tissues. This resource provides a context-controlled ground truth for benchmarking sequence-to-function models and adds noise as a missing dimension for promoter design.

DRAFT – UNDER REVIEW

POSTER #83 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Sex-hormone crosstalk shapes gut microbiome-immune responses to microbial stimuli

Source: 05_posters/P083.tex · Excel row 70

Ilid M[1,3], Docena GH[1,3], Muglia CI[1,3], Diambra L[2,3]

[1] Instituto de Estudios Inmunológicos y Fisiopatológicos (IIFP), Universidad Nacional de La Plata [2] Centro de Endocrinología Experimental y Aplicada (CENEXA), Universidad Nacional de La Plata [3] CCT-La Plata, CONICET, La Plata, Argentina.

Background: The gut microbiome shapes host immune responses in a manner that varies between individuals and between the sexes, yet most analyses of microbiome-cytokine associations following ex vivo microbial stimulation have not accounted for sex or hormonal status. We previously showed that microbiome associations with basal, unstimulated cytokine production in the 500 Functional Genomics (500FG) cohort are sexually dimorphic and modulated by oral contraceptive (OC) use; whether the same holds for stimulated immune responses remained untested.

Methodology: We reanalyzed the 500FG stimulated-cytokine dataset, comprising responses to lipopolysaccharide, heat-killed Staphylococcus aureus, and a fungal panel (Candida albicans and Aspergillus fumigatus) across three ex vivo systems: whole blood, peripheral blood mononuclear cells, and monocyte-derived macrophages. Gut microbial composition was profiled from stool samples by shotgun metagenomic sequencing, with taxonomic abundances inferred via MetaPhlAn. Cytokine levels were adjusted for age, body mass index, sex, and OC use via ordinary least squares regression, and the residuals were tested for association with microbial abundances using Spearman correlation, with Benjamini-Hochberg FDR correction, separately for men and for OC-using and non-using women.

Results: The gut microbiome-immune axis proved highly context-dependent, reorganized in a stimulus-, cytokine-, and cell-type-specific manner rather than following a single, generalized architecture. Sex-specific interaction networks were pervasive, and a small set of taxa, including Faecalitalea cylindroides and Bacteroides ovatus, emerged as homeostatic anchors linked to anti-inflammatory cytokine profiles across stimuli. Critically, pooling OC users and non-users concealed a substantial fraction of associations: stratifying women by OC use revealed up to seven times as many significant associations as the pooled analysis, showing that OC use reorganizes, rather than simply attenuates, microbiome-cytokine coupling.

Conclusions: Host sex and hormonal status fundamentally alter the rules of engagement between the gut microbiota and immune sentinels following microbial challenge. These findings argue for routine sex- and hormone-based stratification, rather than covariate adjustment alone, in microbiome-immune research, with direct implications for personalized microbiome-targeted therapeutics and mucosal-barrier interventions.

DRAFT – UNDER REVIEW

POSTER #85 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Transcript-anchored mapping and evidence-weighted curation of germline breast cancer predisposition variants in a Panamanian cohort

Source: 05_posters/P085.tex · Excel row 71

De Gracia L[1,2,3], Cedeño J[1], Brunello F[3], Turjanski A[3,4]

[1] Centro Especializado de Genética Médica y Genómica Clínica, Caja de Seguro Social, Panama City, Panama [2] Secretaría Nacional de Ciencia, Tecnología e Innovación (SENACYT), Panama City, Panama [3] Departamento de Química Biológica, FCEN, Universidad de Buenos Aires, Argentina [4] IQUIBICEN, CONICET–Universidad de Buenos Aires, Argentina

Background and aim: No germline BRCA1/BRCA2 data from Panama have been reported. We characterised the molecular spectrum, transcript position and verified classification of breast cancer predisposition variants in a Panamanian cohort.

Methods: Germline DNA from 127 women was sequenced using a 117-gene next-generation sequencing (NGS) panel; reads were aligned to GRCh38, and variants were called and annotated with the Bitgenia B Platform. Recurrent calls were assessed for gnomAD v4.1.1 population frequency, repeat context, and pseudogene homology. Retained variants were normalised to MANE Select transcripts, and BRCA1/BRCA2 variants were mapped to exons and functional domains. Variants were curated in R under the applicable gene-specific ClinGen/ACMG-AMP specifications (ENIGMA for BRCA1/BRCA2, CDH1, PTEN), together with CanVIG-UK guidance and the Tavtigian Bayesian framework.

Results: Among 36 recurrent calls, 30 (83%) were identified as artefacts or benign findings. After curation of the complete dataset, 20 unique variants were retained in 23 women: eight missense, six frameshift, four nonsense, one splice-region variant, and one in-frame deletion. Seventeen women (13.4%) carried pathogenic/likely pathogenic (P/LP) variants, 13 in BRCA1/BRCA2; one further call remains under technical review. Five women carried only variants of uncertain significance (VUS), including BRCA2 NM_000059.4:c.632-3C>G, initially reported as likely pathogenic and sustained as uncertain pending RNA study. Curation also corrected a gnomAD frequency copied from an unrelated record, removed an unsupported rsID, and replaced criteria applied outside the PTEN expert-panel specifications. BRCA1 NM_007294.4:c.5123C>A p.(Ala1708Glu), a founder variant in Colombia and a recurrent allele reported in Mexico, was identified in four unrelated women. Five of nine BRCA1 carriers harboured variants within the

BRCT domain. Among variants evaluated using the point-based Bayesian framework, only BRCA1 c.5123C>A (13 points) and BRCA2 c.9371A>T (12 points) independently reached the pathogenic threshold ($\geq$10 points).

Conclusions: Transcript-anchored mapping and evidence-weighted curation refine the molecular profile of hereditary breast cancer predisposition in Panama while avoiding classification of uncertain or technically unresolved findings as positive. These data provide a population-specific reference for germline BRCA1/BRCA2 variation in Panamanian women with breast cancer.

DRAFT – UNDER REVIEW

POSTER #86 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

METAGENOMIC CHARACTERIZATION OF MICROBIAL COMMUNITIES FROM THE AMBARGASTA AND HUYAMAMPA SALT FLATS IN NORTHERN ARGENTINA

Source: 05_posters/P086.tex · Excel row 72

Zerda Moreira A[1], Kurth D[2], González JM[1]

[1] Instituto de Bionanotecnología del NOA (INBIONATEC), CONICET-UNSE - Santiago del Estero – Santiago del Estero - Argentina [2] Planta Piloto de Procesos Industriales Microbiológicos (PROIMI)-CONICET - San Miguel de Tucumán - Tucumán – Argentina

Saline environments are extreme ecosystems that harbor microbial communities adapted to salinity stress and represent a potential source of genetic and metabolic diversity. In this study, microbial communities from the Ambargasta and Huyamampa salt flats, located in Santiago del Estero, Argentina, were analyzed by integrating read based and genome resolved metagenomics. Five Illumina shotgun metagenomes were analyzed, two from Ambargasta and three from Huyamampa. Read quality was assessed using FastQC, and preprocessing was performed with Trimmomatic. Read based taxonomic profiling was conducted using MetaPhlAn. For the assembly based analysis, each metagenome was independently assembled with MEGAHIT, followed by binning with SemiBin and quality assessment with CheckM. Bins with $\geq$50% completeness and <10% contamination were selected, and GTDB-Tk was used for taxonomic classification and phylogenetic placement.

Read-based taxonomic profiling revealed communities dominated by Proteobacteria, particularly Gammaproteobacteria, including genera associated with saline environments, such as Halomonas and Marinobacter. Taxonomic composition varied among samples, with a distinct representation of Halobacteria in Ambargasta. The assembly-based approach recovered 111 bins meeting these quality criteria, while two additional bins close to the completeness threshold were retained for exploratory analysis. Taxonomic classification and phylogenetic placement of the recovered genomes revealed diverse bacterial and archaeal lineages, including Pseudomonadota (Proteobacteria), Bacteroidota, Halobacteriota, Thermoproteota, and Nanoarchaeota. Concordance between both approaches was observed, with the recovery of MAGs belonging to genera also detected by MetaPhlAn, including Marinobacter and Sphingomonas. Furthermore, several MAGs could not be assigned at lower taxonomic ranks, with some remaining unclassified at the genus or family level, indicating limited representation of closely related genomes in the reference database and suggesting potential taxonomic novelty. These results provide new insights into the microbial diversity of two previously unexplored saline environments in Northern Argentina and highlight the value of integrating complementary metagenomic approaches for their taxonomic, genomic, and phylogenetic characterization while enabling further exploration of their metabolic potential and mechanisms of adaptation to saline conditions.

DRAFT – UNDER REVIEW

POSTER #89 **TRACK:** Proteomics and Protein Structure and Function

Can ML Diffusion Models Capture Intrinsically Disordered Protein Ensembles?

Source: 05_posters/P089.tex · Excel row 73

Kulemeyer G[1], Bugnon L[1], Erdös G[2], Milone DH[1], Stegmayer G[1]

[1] sinc(i) - Research institute for signals, systems and computational intelligence, UNL - CONICET, Santa Fe, Argentina [2] Department of Biochemistry, Institute of Biology, Eötvös Lorand University, Budapest, Hungary.

Intrinsically disordered proteins (IDPs) play key roles in cellular signaling, regulation and molecular recognition. Their conformational flexibility enables diverse interactions, while their dysfunction is associated with cancer and neurodegeneration. Understanding these proteins requires characterizing heterogeneous conformational ensembles rather than a single structure. Diffusion models offer promising tools for sampling protein conformations, with approaches such as idpSAM learning to generate disordered conformations from simulation-based ensembles.

Here, we assess whether a general-purpose generative model can characterize IDP ensembles from an energetic perspective. We build on ProteinEBM, a sequence-conditioned energy-based model trained through denoising score matching relating the diffusion score to the negative gradient of a learned energy function. Unlike models trained directly on IDP ensembles, this formulation is less tied to ensemble-specific data and can support broader applications.

We combined ordered structures from the Protein Data Bank (PDB), disordered regions annotated in DisProt, with their corresponding AlphaFold models, and reference ensembles from the Protein Ensemble Database (PED), with small-angle X-ray scattering (SAXS) data. We evaluated energy per residue across generated conformations per sequence, and analyzed the radius of gyration (Rg), its evolution along reverse diffusion trajectories, and its dependence on sequence length under Flory scaling.

AlphaFold-derived disordered regions showed higher diffusion energies per residue than for ordered structures, with overlapping distributions. Diffusion-based conformations exhibited a logarithmic dependence of Rg on sequence length across ordered proteins, disordered proteins and isolated disordered regions, with greater variability in disordered proteins. These findings suggest that generative models can capture IDP conformational properties and motivate the development of representations linking protein sequence, energetics and conformational behavior.

Our results support the use of model energy to characterize differences between ordered and disordered structures and identify the representation of extended conformations as a key area for improving IDP ensemble generation. This highlights the importance of ensemble-based validation and provides a basis for investigating sampling strategies and model adaptations to better reproduce IDP conformational properties.

DRAFT - UNDER REVIEW

POSTER #91 **TRACK:** Proteomics and Protein Structure and Function

Annotating experimental conditions of amyloid structures to explore polymorph relevance

Source: 05_posters/P091.tex · Excel row 74

Mencucci MV[1], Bárcenas O[2], Ventura S[2]

[1] Centro de Endocrinología Experimental y Aplicada (CONICET- Universidad Nacional de La Plata) [2] Institut de Biotecnologia i Biomedicina (Universitat Autònoma de Barcelona)

Introduction: Amyloid fibrils are highly ordered protein aggregates linked to amyloid diseases. A key feature is amyloid polymorphism, in which the same protein can form distinct structures. Different polymorphs may be associated with distinct disease phenotypes. The Protein Data Bank (PDB) contains amyloid structures, but the experimental conditions under which they were obtained are not systematically curated.

Objective: To systematically annotate experimental conditions of amyloid structures in the PDB, and use this information to explore the potential biological relevance of polymorphs.

Methods: We searched the PDB for amyloid structures ("Protein Fibril", "Amyloid", "Amyloids", "Fibril", "Filament"). Structures were annotated as ex vivo or in vitro. Ex vivo structures were annotated for biological model/disease and tissue; in vitro structures for monomer type/concentration, mutations/modifications, fibrillization solution, pH, temperature, agitation/sonication, and seeding. Polymorphs were inferred by structural-similarity clustering and classified by the physiological relevance of their formation conditions: ex vivo > ex vivo-seeded > physiological-like in vitro > non-physiological in vitro.

Results: The dataset included 554 structures from 50 proteins, with Tau, α-synuclein, and amyloid-β42 being the most represented. Approximately 17% lacked an associated publication or at least one required variable.

Alzheimer's disease and brain tissue were the most represented ex vivo sources. For in vitro structures, some variables showed similar trends across the complete dataset and the three most represented proteins, such as the use of near-physiological pH and temperature, and phosphate buffer. Other variables varied across proteins. For example, the use of seeding was rare for Tau (5.7%) but common for α-synuclein (51.2%) and amyloid-β42 (45.5%).

Several clusters for the three most represented proteins included ex vivo or ex vivo-seeded structures, which may represent more physiologically relevant polymorphs. In addition, some clusters combined multiple in vitro structures with an ex vivo structure associated with a specific amyloid disease, potentially guiding the selection of in vitro conditions to study disease-associated polymorphs.

Conclusions: We provide a systematic annotation of experimental conditions for amyloid structures, supporting polymorph characterization and highlighting the need for standardized annotation and deposition practices.

DRAFT - UNDER REVIEW

POSTER #92 **TRACK:** Proteomics and Protein Structure and Function

A novel structure-driven strategy for Parkinson's disease immunotherapy

Source: 05_posters/P092.tex · Excel row 75

Pesce G[1], Estivill-Alonso M[1], Fornt-Suñé M[1], Bárcenas O[1,2], Anton-Villanueva L[1], Pallarès I[1], Ventura S[1,3]

[1] Institut de Biotecnologia i Biomedicina and Universitat Autònoma de Barcelona [2] Institute of Advanced Chemistry of Catalonia (IQAC) [3] Hospital Universitari Parc Taulí

Parkinson's Disease (PD) is the second most common and fastest-growing neurodegenerative disorder, yet current treatments only address early symptoms and fail to halt or reverse disease progression.

A hallmark of the disease is the aggregation of α-Synuclein (αSyn) into amyloid fibrils that accumulate in the brain as insoluble inclusions known as Lewy bodies, whose presence is closely linked to the progressive loss of dopaminergic neurons. Interfering with αSyn amyloid formation and its associated toxicity represents a promising therapeutic strategy for PD and other synucleinopathies. However, αSyn aggregation is a complex and dynamic process in which different aggregated species coexist, posing a significant challenge in selectively targeting the toxic ones. In this landscape, soluble oligomeric intermediates have recently gained particular attention, as these non-fibrillar species are considered the primary drivers of neurodegeneration and disease progression. In this context, our group has identified an oligomer-specific conformational region whose engagement halts the transition to fibrils and which, if targeted, could provide the basis for novel therapeutic strategies.

Nanobodies are small single-domain antibodies capable of binding antigens with high specificity and affinity. Their compact size and exceptional stability make them well-suited for a wide range of therapeutic, diagnostic and research applications. Here, we propose a novel immunotherapeutic approach based on structurally informed rational design to generate a series of nanobodies targeting the conformational region that orchestrates the oligomer-to-fibril conversion. Our results indicate that these nanobodies bind selectively to oligomeric αSyn species and that this binding inhibits their aggregation in vitro, modulating aggregation kinetics and supporting the therapeutic potential of this approach for PD.

DRAFT – UNDER REVIEW

POSTER #94 **TRACK:** Proteomics and Protein Structure and Function

Diverse sequences, equivalent ensembles: Understanding the molecular grammar underlying conformational buffering in IDRs

Source: 05_posters/P094.tex · Excel row 76

Nuñez PF[1], Díaz A[2], Glavina J[1], García Alvarez H[1], Vranken W[2], Chemes LB[1]

[1] Instituto de Investigaciones Biotecnológicas (IIBio), Universidad Nacional de San Martín (UNSAM) [2] Bio2Byte, Interuniversity Institute of Bioinformatics in Brussels (IB)2, Université libre de Bruxelles (ULB)-Vrije Universiteit Brussel (VUB)

Intrinsically disordered regions (IDRs) lack a stable 3D structure but evolve under selective pressure to preserve their function. Studies on the E1A viral oncoprotein identified a compensatory mechanism termed "conformational buffering" which conserves linker's end-to-end distances through compensatory changes in sequence length and composition. This mechanism explains how highly diverged IDRs preserve functions and proteome-wide studies suggest it may operate across many IDRs.

To assess the generality of conformational buffering across IDRs, we selected a dataset of 1,137 human IDRs and their orthologs that exhibit features of conformational buffering (~52,000 sequences in total). We developed a pipeline to integrate multiple sequence-based computational tools including localCIDER (sequence properties), ALBATROSS (ensemble structural features) and Bio2Byte's B2BTools (biophysical propensities including protein dynamics, early folding events, and disorder) as well as quantitative biophysical parameters (Rg, Re, and ν) extracted from CALVADOS simulations.

Building upon this foundation, our current work aims to explore the molecular grammar governing this compensatory mechanism. Preliminary unsupervised k-means clustering based on sequence features (e.g., NCPR, hydrophobicity, FCR) partitions the IDR space into distinct sub-populations characterized by their charge composition/patterning: positive, negative and near-neutral clusters. Interestingly, these divergent clusters exhibit remarkably similar compaction (ν) distributions and different compensatory mechanisms where proline and hydrophobic content play a major role in sequences with a lower fraction of charged residues, suggesting that diverse sequence-level strategies can achieve equivalent structural ensembles.

To further investigate the molecular basis of this phenomenon, we will cross-reference these sequence-based classifications with Bio2Byte biophysical predictions. While these predictors operate per-residue, we implemented the b2btoolstats package to aggregate them into per-IDR fractional metrics and statistical descriptors. By analyzing these global fractional profiles in representative orthologous families, we will explore whether IDRs maintain specific biophysical propensities. Overall, this framework aims to contribute to a broader understanding of the biophysical constraints that influence IDR evolution.

DRAFT UNDER REVIEW

POSTER #95 **TRACK:** Proteomics and Protein Structure and Function

Grafting Site Modulates Intramolecular Inhibition in Monomeric Domain-Linker-Motif Constructs

Source: 05_posters/P095.tex · Excel row 77

Kramar JM[1], Garrone NA[1], Nuñez PF[1], Clerc I[3], Suarez CF[1], Glavina J[1], Rodriguez S[2], Cortes J[3], Bernadó P[2], Chemes LB[1]

[1] Universidad Nacional de San Martin [2] CBS-CNRS [3] LAAS-CNRS

Intramolecular inhibition is often mediated by motif-domain interactions within the same protein chain. Here, we developed a model system to study intramolecular inhibition based on the p107 pocket protein. The p107 pocket domain contains binding sites for short linear motifs (SLiMs), including the LxCxE and E2F motifs, which bind the LxCxE and E2F clefts, respectively. We developed monomeric p107 variants designed to specifically occlude either cleft through intramolecular fusion of an LxCxE- or E2F-containing SLiM using flexible linkers. This model system provides an experimental tool to investigate how linker length and grafting sites regulate intramolecular interactions. We first generated a p107 construct fused to a flexible linker and a high-affinity LxCxE SLiM from HPV16 E7 (p107L-Linker-E7). While effective concentration (Ceff) calculations predicted strong intramolecular inhibition, pull-down assays revealed incomplete LxCxE-pocket occlusion. Structural analysis suggested that the grafting site influenced SLiM positioning. We therefore generated a p107S-Linker-E7 construct harboring a new grafting site and a shorter linker. Ceff calculations predicted a threefold increase upon linker shortening and a 300-fold increase when accounting for the change in anchoring site. Experimentally, p107S-Linker-E7 completely occluded the LxCxE cleft while retaining E2F cleft accessibility. The strategy was also tested using an E2F2-derived SLiM, revealing an opposite dependence on grafting site and highlighting the critical role of grafting position on pocket occlusion. Orthogonal validation by SAXS, ITC, and simulations performed using a stochastic all atom conformational sampling method supported these findings. SAXS showed that p107L-Linker-E2F2 had a smaller radius of gyration than p107L-Linker-E7, consistent with stronger intramolecular engagement of the E2F2 SLiM for this construct. In contrast, p107L-Linker-E7 displayed a larger radius of gyration and required dimeric species for fitting, indicating inefficient positioning of the E7 SLiM within the LxCxE cleft. Overall, our results establish a rational strategy for generating monomeric p107 variants that specifically occlude interaction clefts, and provide clues for understanding how linker length and grafting sites regulate intramolecular inhibition.

DRAFT – UNDER REVIEW

POSTER #96 **TRACK:** Proteomics and Protein Structure and Function

Glycosylation across IDRs and non-IDR regions: patterns and functional associations

Source: 05_posters/P096.tex · Excel row 78

Cuccovia Warlet FU[1], Lassalle FM[1], Nugnes MV[2], Cerrudo CS[1], Belaich MN[1], Aspromonte MC[2], Tosatto S[2]

[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. [2] Department of Biomedical Sciences, University of Padova, Padova 35131, Italy

Protein glycosylation and intrinsic disorder are ubiquitous features of proteomes, yet their relationship remains unexplored. Glycosylation regulates protein stability, interactions and signalling, whereas intrinsically disordered regions (IDRs) and proteins (IDPs) confer flexibility and multifunctionality. Whether both features converge and shape functions remains unclear. Here, we analysed this by integrating glycosylation data from key resources like GlyGen, O-GlcNAc, Eukaryotic Linear Motif (ELM) and GlyConnect with IDPs/IDRs data extracted from DisProt, the reference database for validated IDRs. We systematically mapped glycosylation sites to IDRs and non-IDRs for DisProt proteins and proteins absent from DisProt. GlyConnect sites were manually reviewed to correct misannotated positions. We compared N- and O-linked glycosylation distribution, glycan diversity, recognition by ELM motifs, and Gene Ontology (GO) association using UniProt annotations. Statistical comparison was performed using Fisher's exact and Mann–Whitney U tests. Our analysis highlighted that O-linked glycosylation was preferentially associated with IDRs, whereas N-linked glycosylation was more frequent in non-IDR regions. Glycans in IDRs were larger than those in non-IDR regions, with a significant difference for N-linked sites. All sites were recognised by ELM motifs; however, O-linked sites required only a serine or threonine at the modified position. Further analysis of sequence, physicochemical and structural features revealed no pattern accounting for these glycosylations. Functional associations differed according to the structural context of glycosylation, with proteins glycosylated within and outside IDRs showing the highest functional diversity. IDR/IDPs glycosylation was mainly associated with programmed cell death and negative regulation of cellular processes (GO:0012501, GO:0048523), whereas non-IDRs/IDPs glycosylation was associated with metabolic regulation and stress responses (GO:0009893, GO:0006950). Proteins glycosylated in both contexts combined these functions and were additionally linked to developmental processes (GO:0050793, GO:0033554). DisProt proteins without glycosylation showed weaker associations, focused on response to radiation and post-translational modification. Glycosylation outside DisProt showed no functional association. Collectively, these results indicate that glycosylation within IDRs could be a key determinant of protein multifunctionality.

DRAFT – UNDER REVIEW

POSTER #97 **TRACK:** Proteomics and Protein Structure and Function

Insights into the allosteric modulation of IDO1: molecular dynamics simulations reveal the binding mode of a novel inhibitor

Source: 05_posters/P097.tex · Excel row 79

Chinestrad P[1], Cabrera M[1], Zinni A[1], Blanco R[1], Lorenzano Menna P[1]

[1] Laboratorio de Farmacología Molecular, Universidad Nacional de Quilmes

Indoleamine 2,3-dioxygenase 1 (IDO1) is a key immunosuppressive enzyme in the tumour microenvironment. Its AGGSAG loop transitions between ligand-free (LF, inactive) and ligand-bound (LB, active) conformations, opening up allosteric, LF-exclusive pockets as a novel way to target the enzyme. We previously identified C26, a drug-like compound highly specific to

an allosteric pocket present in inactive conformations. Here we analysed the interaction between C26 and IDO1 and its impact on the AGGSAG loop.

Two 200 ns molecular dynamics simulations were run for IDO1-C26 and ligand-free IDO1. According to molecular mechanics Poisson-Boltzmann surface area estimates, C26 remained bound throughout the simulation, with a favourable mean binding energy of −29.96 ± 6.43 kcal/mol, showing no dissociation along the trajectory. Between 80 and 120 ns, the ligand rearranged within the pocket while retaining contacts with 11 of 13 DUCK pocket residues in over 40% of frames, as determined by ligand root-mean-square deviation (RMSD), minimum distance to the protein and residue contact occupancy. Over the same window, IDO1 underwent

a defined transition along the first principal component (PC1), driven mainly by pocket residues 14-24, which also showed the largest per-residue Root Mean Square Fluctuation (RMSF) increase. Moreover, the variation of PC1 correlated strongly with ligand RMSD (r = 0.88), indicating a coupled process. Finally, AGGSAG backbone conformations were projected onto the axis joining the LF (6AZUB) and LB (4PK5A) references, defining an advance and a perpendicular deviation. In ligand-free IDO1, the AGGSAG loop advanced 2.12 ± 0.44 Å along the 2.82 Å LF-to-LB axis (76%). Meanwhile, in the IDO1-C26 simulation, the loop advanced only 0.06 ± 0.46 Å (2%) and never exceeded 1.2 Å, with a larger perpendicular deviation (2.70 vs 1.88 Å). The complex also recovered LF-exclusive AGGSAG loop contacts nearly absent without ligand (Leu124, Val125, Cys129; occupancy 0.90-1.00 vs 0.03-0.08) and lost every LB-exclusive contact (Gln271, Phe291, Met299).

These results show that C26 establishes a durable interaction with the DUCK pocket, preventing the spontaneous drift towards the active conformation observed in the apo enzyme.

Overall, C26 not only inhibits IDO1 activity but also stabilises the enzyme in a conformation distinct from those reported for known inhibitors. Together, these findings position LF-exclusive

DRAFT - UNDER REVIEW

pockets as a promising new avenue for targeting IDO1.

DRAFT - UNDER REVIEW

POSTER #98 **TRACK:** Proteomics and Protein Structure and Function

AmyloTools: a Python package and web application for exploration and quantitative characterization of amyloid fibril structures.

Source: 05_posters/P098.tex · Excel row 80

Garcia-Messina J[1], Navarro AM[1], Marino-Buslje C [1].

[1] Fundación Instituto Leloir. Buenos Aires, Argentina

Amyloid fibrils are protein aggregates in which copies of protein segments stack into layers that repeat along a helical axis to form protofilaments, which can further associate into fibrils. They are strongly associated with neurodegenerative diseases and systemic amyloidoses, though the same architecture also occurs as a functional assembly (e.g in bacterial biofilms).

As of September 2026, more than 800 fibril structures have been deposited in the Protein Data Bank, and the number continues to grow. Yet systematic comparison remains limited because coordinate files do not explicitly encode fibril organization. Geometric and energetic descriptors that depend on this organization must therefore be reconstructed from the coordinates for each entry, which makes analyses difficult to automate and standardize.

To address this problem and make amyloid fibril structure analysis more accessible, we define the structural hierarchy as consisting of four levels: chain, layer, protofilament, and filament. We then developed a computational toolkit that automatically reconstructs this hierarchy directly from atomic coordinates. The toolkit identifies layers as chains arranged in planes orthogonal to the fibril axis, protofilaments as layers related by common helical twist and rise, and filaments as the complete assemblies. This approach operates independently of chain labels, symmetry annotations, or deposition metadata, enabling consistent analysis across PDB entries.

Users can select chains, layers, and protofilaments directly, extend resolved chains to desired n-mers to reconstruct fibrils, or reduce structures to selected subsets. The toolkit integrates the computation of descriptors commonly reported in the literature, such as subunit twist and rise, layer planarity, fibril diameter, ordered-core length, buried surface area, and solvation free energy, which otherwise require combining separate tools and custom scripts.

We implemented this toolkit as AmyloTools, with both programmatic and interactive interfaces. The amylotools package is distributed through PyPI and provides a command-line interface and Python API in which each descriptor is one command or function call, while the web application exposes the same operations without requiring local installation. The web application will be available at https://amylotools.leloir.org.ar.

DRAFT - UNDER REVIEW

POSTER #99

TRACK: Proteomics and Protein Structure and Function

Analysis of Antiplatelet Activity through Protein Disulfide Isomerase (PDI) Inhibition by Peptides

Source: 05_posters/P099.tex · Excel row 81

Rodríguez-Cabello F[1], Méndez D[1], Ulzurrun E[2], Teitgen AE[2,3], Guzmán F[4], Trostchansky A[5], Carrasco B[6], Campillo NE[2], Palomo I[1], Fuentes E[1]

[1] Centro de Investigación en Trombosis y Envejecimiento Saludable, Universidad de Talca [2] Centro de Investigaciones Margarita Salas (CSIC), España [3] Instituto de Ciencia de Materiales de Madrid (CSIC), España [4] Núcleo Biotecnología Curauma, Pontificia Universidad Católica de Valparaíso, Chile [5] Departamento de Bioquímica and Centro de Investigaciones Biomédicas, Universidad de la República, Uruguay [6] Centro de Estudios en Alimentos Procesados (CEAP), Chile

Protein disulfide isomerase (PDI) is an enzyme that catalyzes the formation and reorganization of disulfide bonds, playing a critical role in protein folding. As its dysregulation is associated with various diseases and its inhibition exerts antiplatelet effects, PDI has emerged as a promising therapeutic target for regulating blood coagulation. In this study, we evaluated the PDI inhibitory potential of computer-designed peptides through a combined approach of molecular docking simulations and functional experimental assays. Initially, peptide sequences were generated combinatorially and assessed for structural compatibility with PDI's active site cysteines. Molecular docking revealed favorable binding energies and the formation of stable disulfide bonds, suggesting a strong potential for enzymatic inhibition. To validate these computational predictions, the leading peptides were subjected to biological evaluation. Our experimental results demonstrated that the selected peptides significantly inhibit platelet aggregation and reduce platelet activation markers as assessed by flow cytometry. Furthermore, the peptides successfully impaired clot retraction and exhibited protective antithrombotic efficacy in a thrombosis model. In conclusion, we demonstrate that these computer-designed peptides effectively interact with PDI and interfere with its enzymatic activity, translating into potent antiplatelet and antithrombotic effects. These findings highlight the potential of targeted peptide-based PDI inhibitors as novel therapeutic agents for cardiovascular diseases.

DRAFT – UNDER REVIEW

POSTER #100 **TRACK:** Proteomics and Protein Structure and Function

Computational identification of structurally uncharacterized gamete fusogens in fungi

Source: 05_posters/P100.tex · Excel row 82

Manzano J[1], Aguilar P[1]

[1] Laboratorio de Biología Celular de Membranas, Instituto de Fisiología Biología Molecular y Neurociencias (IFIBYNE, UBA-CONICET).

Biological membrane fusion is mediated by specialized proteins known as fusogens. While viral and intracellular SNARE-mediated mechanisms are well characterized, cell-cell fusion remains less understood despite its fundamental role in fertilization (gamete fusion) and tissue development (e.g., placenta and muscle). The identification of a superfamily of proteins named FUSEXINS (FUSion proteins essential for Sexual reproduction and EXoplasmic merger of plasma membranes) such as HAP2/GCS1, a protein that mediates gamete fusion, has provided a crucial framework for studying these processes. The presence of HAP2/GCS1 across most eukaryotic groups supports the hypothesis that this fusogen existed in the last eukaryotic common ancestor (LECA). However, its apparent absence in lineages such as fungi and vertebrates suggests that either it evolved as highly divergent orthologs that escape current computational detection, and/or it has been replaced by alternative fusogenic proteins. Uncovering these missing fusogens is crucial, as fundamental cell-fusion pathways are often evolutionarily conserved, making fungal insights potentially translatable to vertebrate gamete fusion. Because structural homology-based searches failed to identify fusogen candidates on fungal and vertebrata genomes, we aimed to identify small, intrinsically disordered proteins such as viral FAST fusogens. We constructed a comprehensive small ORFs (sORF) database by running a custom-mining Python script on 133 genomes that encompass the diversity of the fungal kingdom. This script systematically extracts all possible sORFs ranging from 90 to 600 nucleotides by scanning from each stop codon back to its respective start codon, agnostic to any transcriptional or translational evidence. After filtering to retain conserved sequences, we subjected this database to both feature-based and Hidden Markov Model-based screening pipelines, ultimately aiming to deliver a prioritized list of fungal candidate genes with a high probability of acting as novel fusogens.

DRAFT - UNDER REVIEW

POSTER #101 **TRACK:** Proteomics and Protein Structure and Function

Beyond Annotation: A Curated Proteomic and Bioinformatic Approach to Map the Small Protein Universe of Bordetella pertussis

Source: 05_posters/P101.tex · Excel row 83

Gonzalo Z[1], Bottero D[1], Hozbor D[1], **Mauricio L**[1]

[1] Instituto de Biotecnología y Biología Molecular, Universidad Nacional de La Plata, La Plata, Argentina

Small open reading frames (smORFs) encode functional small peptides (smORF-encoded polypeptides, SEPs) involved in a wide range of cellular processes across both eukaryotic and prokaryotic organisms. Due to their small size, the identification of smORFs and their products is challenging. Obtaining reliable bioinformatic predictions is difficult, as it can be problematic to distinguish smORfs encoding functional proteins from those arising by chance or as products of proteolytic degradation. Likewise, conventional laboratory techniques also overlook these small proteins, further complicating their detection.

In this study, we analyzed proteomic datasets previously generated in our laboratory, together with proteomic and transcriptomic datasets retrieved from public repositories, to identify and characterize small proteins in Bordetella pertussis, the causative agent of the human respiratory disease named whooping cough or pertussis. We conducted a comprehensive comparative analysis of proteomic data derived from B. pertussis outer membrane vesicles (OMVs) and whole-cell preparations, with the aim of mapping the distribution of SEPs across these fractions. In parallel, we systematically benchmarked data-independent (DIA) against data-dependent (DDA) acquisition proteomic strategies, assessing their relative suitability and performance for SEP detection. Separately, we evaluated three bioinformatic tools for the prediction of smORFs (ProsmorfPred, MipepID and smorfer) to identify novel previously unannotated SEPs.

Our results show that DIA approaches are better suited for SEP identification, even in the absence of peptide enrichment methods, as they enable the detection of all SEPs identified in whole-cell and OMV DDA proteomes. Evaluation of the prediction tools showed that they generated a large number of candidate sequences, of which only a very small percentage was experimentally supported. Nevertheless, six novel SEPs were supported by both proteomic and transcriptomic evidence, while four additional candidates were identified but their classification as SEPs remains uncertain.

Analysis of proteomic and transcriptomic datasets under two different phenotypic phases of the bacteria: virulent and avirulent phase, revealed differential expression of several SEPs between physiological states. Overall, these findings expand the known SEP repertoire of B. pertussis and support their relevance to bacterial physiology.

DRAFT - UNDER REVIEW

POSTER #102

TRACK: Proteomics and Protein Structure and Function

Impact of mutations at evolutionary conserved sites on the conformational dynamics of tandem-repeat proteins

Source: 05_posters/P102.tex · Excel row 84

Chiappinelli M[1], Saldaño[1], Fernandez-Alberti[1]

[1] Universidad Nacional de Quilmes

Tandem-repeat proteins (TRPs) are composed of arrays of repeating structural units that assemble into extended, superhelical, or horseshoe-shaped architectures stabilized mainly by short-range interactions. Their distinctive sequence–structure–dynamics–function relationships have been extensively studied to understand the molecular principles that distinguish them from globular proteins. Here, we investigate how mutations affect the conformational mechanics of PR65, the HEAT-repeat scaffold of protein phosphatase 2A (PP2A), which acts as an elastic connector between its catalytic and regulatory subunits. Using computational approaches to characterize collective conformational changes, we found that the impact of mutations on the dynamics associated with PR65 binding to the catalytic subunit correlates with evolutionary conservation. Our results also reveal a common pattern among repeat units in the way mutations perturb collective dynamics, while highlighting differences associated with the specific functional role of individual units. Thus, although mutations in different repeats share a common influence on the global dynamics of the protein, their local contribution to function introduces additional variability in their effects. Interestingly, these relationships are not observed for the subsequent conformational changes occurring during binding of the PR65–catalytic subunit complex to the regulatory subunit. These findings highlight both shared and unit-specific features in how mutations modulate the conformational dynamics of tandem-repeat proteins. More broadly, they contribute to understanding the interplay between evolutionary conservation, conformational dynamics, and function in modular protein architectures.

DRAFT - UNDER REVIEW

POSTER #104 **TRACK:** Proteomics and Protein Structure and Function

Detection of Bacteroides fragilis Division II (cfiA+) by Mass-Shift Analysis of MALDI-TOF MS Spectra and Machine Learning Classification

Source: 05_posters/P104.tex · Excel row 85

Bravo F[1], Otero S[1], Cejas D[2], Viale D[1], Mangano A[1], Fernández MF[1]

[1] Department of Microbiology, Hospital de Pediatría Prof. Dr. Juan P. Garrahan, Buenos Aires, Argentina.
[2] Department of Microbiology, Immunology and Biotechnology, Faculty of Pharmacy and Biochemistry, University of Buenos Aires (FFyB-UBA), Buenos Aires, Argentina.

Background: Bacteroides fragilis Division II isolates carry cfiA, encoding a class B metallo-β-lactamase associated with carbapenem resistance. Although cfiA is often silent or poorly expressed, insertion sequences upstream of the gene may activate its expression, leading to high-level carbapenem resistance and potential therapeutic failure. MALDI-TOF MS identifies B. fragilis but does not discriminate Division I (cfiA−) from Division II (cfiA+), requiring molecular confirmation. This study aimed to develop and evaluate a MALDI-TOF MS mass-shift classifier combined with machine learning for the detection of cfiA+ isolates.

Methods: Fourteen B. fragilis isolates were characterized by cfiA PCR and classified as cfiA+ or cfiA−. Fresh subcultures were processed in the same MALDI-TOF MS run to standardize spectral acquisition. From the resulting .rep files, five marker regions within ~4688–5282 Da were analyzed, identifying 9–23 Da mass shifts associated with cfiA status. For each marker, the observed mass was compared with previously reported reference masses for Divisions I and II. Peak-distance features were used to train a machine-learning Random Forest classifier, evaluated by leave-one-strain-out cross-validation (LOSO-CV), with all replicates from each isolate kept within the same fold.

Results: Among the 14 B. fragilis isolates, 5 were cfiA+ and 9 were cfiA− by PCR. At least four of the five spectral markers were detected in all cases, while the Random Forest model achieved 100% correct classification. The model achieved 100% sensitivity (95% Wilson CI: 56.6–100%), 100% specificity (70.1–100%), and 100% accuracy (78.5–100%). The ~4700 Da and ~4820 Da markers showed the highest relative importance in the model (33.8% and 32.2%, respectively).

Conclusions: Mass-shift analysis of routine MALDI-TOF MS spectra discriminated Division II (cfiA+) isolates with complete agreement with PCR. This strategy could serve as a screening method to select isolates requiring molecular confirmation, reducing reagent use and hands-on time. These preliminary results are limited by the small sample size (cfiA+, n=5) and require validation in a larger cohort.

DRAFT – UNDER REVIEW

POSTER #105 **TRACK:** Proteomics and Protein Structure and Function

Beyond the nucleolus: Integrative interactome and transcriptome analysis identifies MAGEB2 as a ribosome-associated regulator of p53 under nucleolar stress

Source: 05_posters/P105.tex · Excel row 86

Vidal C[1], Escalada M[1], Suberbordes M[1], Silva M[1], Soubeyran P[2], Monte M[1]

[1] Laboratorio de Oncología Molecular, DQB, FCEN-UBA. IQUIBICEN, UBA-CONICET, Buenos Aires, Argentina.
[2] Centre de Recherche en Cancérologie de Marseille (CRCM), Université and Institut Paoli-Calmettes, Marseille, France.

MAGE-I proteins are cancer-testis antigens aberrantly expressed in tumors with aggressive clinical behavior. MAGEB2 is a pro-proliferative protein that participates in rRNA transcription and, unlike most MAGE-I members, localizes to the nucleolus but relocalizes to the nucleoplasm upon nucleolar stress. This stress-induced relocalization led us to explore its molecular functions and relationship with nucleolar stress signaling by integrating interactomic and transcriptomic approaches. The experiments were conducted in doxycycline-inducible MAGEB2 HCT116 cells under nucleolar stress induced by BMH-21 and actinomycin D, respectively. Network analysis of MAGEB2 interactors using MCL clustering identified functional clusters enriched in cytoplasmic translation, ribosome assembly, and protein ubiquitination. These associations were further supported by overlap with published MAGEB2 interactomes and reciprocal identification of the MAGEB2 mouse ortholog, Mageb4, in a murine ribosome-associated interactome. In addition, previous studies have shown that MAGEB2 binds non-polyadenylated RNA, further supporting a role extending beyond the nucleolus and an association with cytoplasmic ribosomes. Given the role of ribosome-associated proteins in nucleolar stress signaling toward p53 and the ability of several MAGE-I proteins to modulate p53 via proteasome-dependent mechanisms, we investigated whether MAGEB2 modulates this response. Under stress, MAGEB2 lost its interaction with PICT1 while maintaining its interaction with RPL5. These findings support a model in which MAGEB2 impairs the assembly of the inhibitory RPL5–RPL11–5S rRNA–MDM2 complex, potentially preserving MDM2 activity toward p53 and attenuating p53 activation. Over-representation analysis and gene set enrichment analysis (GSEA) of RNA-seq data under nucleolar stress identified pathways related to p53 stabilization and regulation of its activity among the most significantly altered. RT-qPCR validation in MAGEB2-WT and -KO cells confirmed reduced induction of p53 targets under stress, and cycloheximide chase assays showed impaired p53 stabilization that was largely restored by proteasome inhibition. Together, these results identify MAGEB2 as a ribosome-associated protein that attenuates nucleolar stress-induced p53 activation, revealing a potential mechanism of tumor adaptation.

DRAFT – UNDER REVIEW

POSTER #106 **TRACK:** Proteomics and Protein Structure and Function

Proportional phylogenetic assignment of shared peptides for accurate microbial biomass estimation in quantitative metaproteomics

Source: 05_posters/P106.tex · Excel row 87

Fabre Barbero N[1], Käll L[2], Rosano G[3]

[1] Faculty of Exact and Natural Sciences, University of Buenos Aires, CABA, Argentina. [2] Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden. [3] Institute of Molecular and Cellular Biology of Rosario (IBR-CONICET-UNR), Santa Fe, Argentina.

Background: While metaproteomics effectively estimates microbial biomass, peptide redundancy from genomic homology complicates precise taxonomic assignment. Common mitigations like the lowest common ancestor (LCA) algorithm often yield incomplete lineages. Therefore, systematically benchmarking quantification strategies with defined mock communities is essential to ensure accurate and reliable biomass estimates.

We aimed to develop and implement a proportional mixture algorithm (proportional mixture model) in Python that assigns the intensity of shared peptides across the phylogenetic topology, addressing the biases of rigid LCA and unique-peptide methods.

Methodology: Twenty-four LC-MS/MS (liquid chromatography–tandem mass spectrometry) runs from mock communities of known composition were processed with quantms (1% false discovery rate), and peptides were mapped to the NCBI taxonomy. We implemented a proportional mixture model in which unique peptides anchor species abundances in each run and the intensity of shared peptides is redistributed among descendant taxa. We benchmarked it against unique-peptide and LCA-based quantification across the three experimental conditions (Constant, Protein, Uneven), using the added protein mass as ground truth.

Principal results: At the species level, the log-scale Pearson correlation coefficient (r) with the added protein mass increased from 0.46–0.56 (unique peptides and rigid LCA, which are identical at the leaves) to 0.84–0.93 with the proportional mixture model across the three mixing designs, with the largest gain in the Uneven design (0.46 $\rightarrow$ 0.93). Recovery of fold changes between conditions was comparable across methods (r = 0.74–0.94).

Conclusions: Proportional redistribution of shared peptides mitigates the quantitative distortions of rigid LCA while retaining all identified peptides, markedly improving species-level composition in a defined mock community. The gain was specific to absolute composition and not to between-condition fold changes. Future work should evaluate it on more complex communities, and integrate it with Bayesian error-propagation models such as Triqler, which currently discards shared peptides.

DRAFT - UNDER REVIEW

POSTER #107 **TRACK:** Proteomics and Protein Structure and Function

Conservation of the catalytic core and active-site diversification in class A flavin-dependent monooxygenases

Source: 05_posters/P107.tex · Excel row 88

Castillo F[1], Mascotti ML[1]

[1] Instituto de Histología y Embriología de Mendoza. IHEM-CONICET

Class A flavin-dependent monooxygenases (FMO A) are a widely distributed enzyme family. They share a conserved domain architecture composed of a Rossmann fold (dinucleotide binding) and a substrate-binding domain; and mainly catalyze the hydroxylation of aromatic compounds using FAD and NAD(P)H [1]. They are involved in the degradation of phenolic compounds, antibiotic biosynthesis, and essential pathways such as ubiquinone synthesis. Their mechanism is well characterized in model enzymes such as p-hydroxybenzoate hydroxylase (PHBH; EC 1.14.13.2) [2]. However, the sequence determinants governing their different reactivities and substrate specificities remain unclear. In this work, we explored the functional diversification of FMO A from an evolutionary perspective.Using a dataset of previously characterized enzymes, homologous sequences were collected across the three domains of life: Bacteria, Archaea and Eukarya. Family membership was assigned based on the conserved motifs; GxGxxG, associated with FAD binding, and D-G-x_5-R and GDx_{10}-G, linked to the active site [2]. Notably, sequences from Archaea represented only 2% of the total number of collected ones. Phylogenetic analysis allowed the identification of seven clades with high statistical support (SH-aLRT/UFboot: 95/92–100/96), six of which include experimentally characterized enzymes. Their taxonomic distribution was mostly mixed: clades I and V contain representatives of all three domains, whereas clades II and III are exclusively bacterial, and clade VI is mainly composed by eukaryotic sequences. Clade VII, comprises both eukaryotic and bacterial species. Based on this clustering, chemical reactivity and substrate preference were further analyzed. Structural analysis through residue conservation mapping allowed the identification of clade-specific substitutions, many of them located within the substrate-binding site. These substitutions suggest a reorganization of the active-site periphery associated with the functional diversification of each clade.

Our results indicate that the common ancestor of FMO A functionally diversified by preserving essential catalytic elements while modifying peripheral residues in each lineage. The differential conservation of substrate-interacting residues suggests an evolutionary history linked to the substrate/product pair transformed by each group.

DRAFT - UNDER REVIEW

POSTER #108 **TRACK:** Proteomics and Protein Structure and Function

Same motif, different partners: an evolutionary-structural view of human sHSP hetero-oligomers

Source: 05_posters/P108.tex · Excel row 89

Racigh V[1,2], Fornasari MS[1], Rodriguez-Sawicki L[1,3]

[1] Structural Bioinformatics Group, Department of Science and Technology, National University of Quilmes. [2] Center for the Study of Organic Compounds, Faculty of Exact Sciences, National University of La Plata [3] Institute of Biochemical Research of La Plata, CONICET, Faculty of Medical Sciences, National University of La Plata

Background

Small heat shock proteins (sHSPs) are ATP-independent chaperones composed of a conserved alpha-crystallin domain (ACD) flanked by disordered N- and C-terminal regions. ACD dimers are the building blocks of larger oligomers and present a central and two lateral grooves which bind conserved and distal segments of the N-terminal region (NTR), respectively. Following gene duplication, oligomeric paralogs are expected to evolve selective interactions. Human HSPB1 (B1), HSPB5 (B5) and HSPB6 (B6) co-expressed in several tissues, can form hetero-oligomers, and their chaperone activity depends on both paralog composition and client identity. As NTR-ACD interactions may contribute to hetero-oligomerization, we investigated shared and partner-specific features of these interactions.

Methods

Per-site evolutionary rates were estimated with Rate4Site using vertebrate ortholog datasets (242 sequences each, from the same species set). Homo- and heterodimeric ACD models bearing distal and conserved NTR peptides were generated with AlphaFold2-Multimer (500 conformers per model). Secondary structure was assessed with DSSP, 3D contact-density maps were computed, and groove occupancy and intermolecular interactions were analyzed with ProLIF.

Results

ACD residues showed evolutionary rates below the sequence average. Distal and conserved NTR linear motifs also showed below-average rates, indicating selective constraint within a dynamic region. Distal peptides preferentially occupied lateral grooves, whereas conserved peptides populated the central groove, consistent with experimental evidence. The β2 strand bordering the central groove was the main interaction surface for conserved peptides. β2-strand formation was less prevalent in B6 than in B1 and B5 ACD chains, coinciding with deeper conserved-peptide localization in B6-containing models. Across all combinations, the conserved motif, particularly its F-x-Q-x-F core, dominated conserved-peptide interactions with the central groove; the relative contributions of motif residues and interactions involving other peptide residues varied with the partner.

Conclusions

B1, B5 and B6 share evolutionarily constrained NTR-ACD interaction elements and a common interaction pattern across homo- and heterodimeric contexts, while exhibiting partner-dependent specificities. This shared framework may contribute to their compatibility in hetero-oligomeric assemblies.

DRAFT – UNDER REVIEW

POSTER #109 **TRACK:** Proteomics and Protein Structure and Function

Disordered tails, conserved contacts: a proteome-wide search for non-catalytic autoregulation

Source: 05_posters/P109.tex · Excel row 90

Bravo FNE[1], Rodriguez Sawicki L[1,2], Fornasari S[1]

[1] Structural Bioinformatics Group (SBG), Department of Science and Technology, National University of Quilmes, Argentina. [2] Institute of Biochemical Research of La Plata (INIBIOLP), CONICET, Faculty of Medical Sciences, National University of La Plata, Argentina.

Background: Intrinsically disordered regions (IDRs) flanking folded domains are pervasive in the human proteome and can regulate domain function through intramolecular contacts. In enzymes this shows altered catalytic activity; in non-catalytic proteins (chaperones, transcription factors) there is no simple readout, so these mechanisms remain unmapped, despite their likely roles in signaling and gene regulation and the growing number of disease variants found in IDRs. We previously showed that conserved IDR motifs binding the α-crystallin domain of HSPB1 coevolve with it, modulating chaperone activity and oligomerization. Here we asked how widespread this regulation is in the human proteome.

Methods: From 20,431 reviewed human proteins (Swiss-Prot), we selected 3,835 with an IDR (MobiDB-lite) within 60 residues of a folded domain. IDR-domain contacts were identified on AlphaFold2 models (Cβ–Cβ < 8 Å). For proteins with contacts, vertebrate orthologs (OrthoDB) were aligned (MAFFT), a tree was inferred (FastTree) and site-specific rates were estimated (LEISR). Rates at IDR contact and non-contact positions were compared using the Mann-Whitney U test with Benjamini-Hochberg FDR correction ($p < 0.05$). Significant proteins were grouped into families and tested for functional enrichment.

Results: A 92-protein pilot dataset with the IDR-domain-IDR architecture closest to HSPB1 validated the approach: 75% showed IDR–domain contacts, and in 24 the contact positions evolved significantly slower than the rest of the IDR. At full scale, 2,412 of 3,835 proteins had contacts, and in 208 ($p<0.05$) contact positions evolved significantly slower. This is an underestimate, as ortholog retrieval and large-scale rate estimation remain limiting, so the catalog will grow. Family recruitment increased with scale: the pilot's bHLH signal was confirmed and amplified, now including the full HIF/ARNT and HES/HEY networks and joined by bZIP (ATF/CREB/FOS), while SMADs, chromatin remodelers, and neurodegeneration-linked proteins (ATXN3, FUS) emerged as new groups.

Conclusions: The workflow delivers a first map of candidate non-catalytic cis-autoregulation, previously found only case by case. The signal concentrates in dimerizing transcription factors, suggesting flanking-IDR tuning of dimerization and DNA binding is a recurrent strategy. Each candidate includes its contact positions, defining residues to mutate and a structural framework for interpreting disease variants in IDRs.

DRAFT – UNDER REVIEW

POSTER #110 **TRACK:** Proteomics and Protein Structure and Function

In silico Optimization of Artificial Cellulosomes: Simulating Multienzyme Cellulose Hydrolysis via Agent-Based Reaction–Diffusion Dynamics

Source: 05_posters/P110.tex · Excel row 91

Nakasone N[1], **Momoli J**[1,2], Iglesias Rando MR[3], Craig PO[1,2]

[1] Departamento de Química Biológica FCEN, UBA, [2] IQUIBICEN (UBA/CONICET), [3] King's College London

Lignocellulosic biomass represents a major renewable feedstock for sustainable biofuel generation, yet its cost-effective conversion into fermentable sugars is hindered by the low catalytic efficiency and high cost of cellulolytic enzymes. Natural cellulosomes circumvent these barriers by spatially clustering diverse cellulases onto structural scaffolds, boosting degradation through enzymatic synergy. Inspired by this principle, our group designs synthetic cellulosomes based on the decameric Brucella abortus Lumazine Synthase scaffold to co-assemble cellulose-binding modules and cellulases for enhanced catalytic output.

To investigate the physical and kinetic mechanisms governing this synergy, we developed an agent-based reaction–diffusion model to simulate cellulose breakdown by endo- and exoglucanase combinations. To access long simulation timescales without losing key mechanistic details, the model's structural complexity was reduced and kinetic parameters (kon, koff, kcat) were appropriately scaled to capture system dynamics beyond strict diffusion limits.

A systematic parametric exploration comparing clustered assemblies with matching free-enzyme mixtures demonstrated that cellulosome efficiency heavily depends on specific kinetic regimes. The artificial cellulosomes outperformed free enzymes across a broad region of moderate-to-high substrate affinities, yielding a 2- to 3-fold functional gain. Strikingly, optimal enhancements of 4- to 8-fold were achieved in regions characterized by moderate affinity and high association rates, whereas performance dropped below free enzymes under low-affinity conditions.

These functional gains are driven by enhanced binding avidity from localized enzyme clusters and potential substrate channeling effects. In silico experiments featuring targeted mutant variants are currently underway to decouple these distinct contributions. Overall, this model offers a robust predictive strategy to guide the rational design of high-efficiency multienzyme platforms for industrial biomass processing.

Acknowledgment: Supported by grants from ANPCyT, UBACYT, and CONICET.

* Contributed equally

Corresponding author: pocraig@qb.fcen.uba.ar

DRAFT – UNDER REVIEW

POSTER #111 **TRACK:** Proteomics and Protein Structure and Function

Uncovering Drug-Repurposing Opportunities across CCR5 and CCR3: A Structure-Based Strategy toward Chemokine Receptor Modulation in Cancer

Source: 05_posters/P111.tex · Excel row 92

Gutierrez A[1], Reyes ME[1], Mora-Lagos B[1]

[1] Instituto de Ciencias Biomédicas, Universidad Autónoma de Chile

CCR5 and CCR3 are chemokine G protein-coupled receptors that orchestrate leukocyte trafficking and inflammatory signaling, processes increasingly recognized as relevant to tumor–immune interactions and cancer progression. Here, we developed an integrated structure-based drug-repurposing workflow to uncover clinically available compounds with previously unexplored potential to interact with the transmembrane pockets of both receptors.

A library of 2,768 approved or clinically used compounds was prepared considering physiologically plausible protonation/tautomeric states and screened with AutoDock-GPU. CCR5 docking used the experimental maraviroc-bound structure (PDB 4MBS), whereas CCR3 was modeled from AlphaFold using a binding region centered on functionally relevant residues Y41, Y113 and E287. Of the initial library, 2,553 compounds completed docking against each receptor. The top 250 ligands per target were further prioritized using physicochemical descriptors, ADMET predictions, ligand efficiency, systemic exposure, Pareto optimization and chemical clustering. The analysis revealed distinct yet partially overlapping repurposing landscapes across CCR5 and CCR3. Among CCR5 candidates, montelukast emerged as a compelling hit, combining favorable predicted binding with a reproducible non-canonical interaction pattern and an established pharmacological profile in inflammatory signaling. For CCR3, cyclopenthiazide stood out as a high-priority non-steroidal candidate, offering an alternative chemotype less likely to reflect the steroid-driven docking bias observed across the screening. Several bile acids and steroid-like molecules ranked highly for both receptors; however, this recurrent enrichment also exposed the tendency of lipophilic GPCR cavities to favor promiscuous hydrophobic scaffolds, reinforcing the value of prioritizing structurally distinct non-steroidal candidates. Maraviroc showed favorable predicted affinity for CCR5, providing a qualitative internal benchmark. Together, these findings highlight montelukast and cyclopenthiazide as representative leads and support integrated docking, ADMET, exposure and chemical-diversity filters to move beyond raw binding scores. This dual-target strategy provides a framework for experimentally testable drug-repurposing hypotheses at the interface of chemokine signaling, tumor inflammation and cancer therapeutics.

DRAFT – UNDER REVIEW

POSTER #112 **TRACK:** Proteomics and Protein Structure and Function

Methodological Optimization and Molecular Dynamics Validation of the Catalitic Core in Herpes Virus Type 1 DNA Polymeras: Preservation of the Mg(II) Coordination Sphere

Source: 05_posters/P112.tex · Excel row 93

Gerstner A[1,2], Salvatierra KA[1,2]

[1] Facultad de Ciencias Exactas, Químicas y Naturales. Universidad Nacional de Misiones. [2] Laboratorio MADAR BSL2 PLUS. Universidad Nacional de Misiones.

Rigid-receptor docking cannot capture receptor plasticity, explicit water networks, or polarization at metallic centers. In the catalytic subunit (UL30) of Herpes Simplex Virus type 1 (HSV-1) DNA polymerase, activity and ligand binding, such as pyrophosphate analogues, depend strictly on two divalent magnesium ions (Mg(II)) in the palm domain. However, classical force fields often distort divalent metal coordination, requiring tailored molecular dynamics (MD) workflows to ensure active-site stability before screening new inhibitors. To address this limitation, the closed-conformation UL30 structure (PDB ID: 8EXX) was prepared retaining the two catalytic Mg(II) ions and foscarnet (PPF). The receptor was processed with pdb4amber and PDBfixer, capping unresolved loops (>40 residues), while protonation states at physiological pH were assigned using Reduce. PPF was parameterized with GAFF2 and AM1-BCC charges, and the complex was solvated in a truncated octahedral TIP3P box with 150 mM NaCl, implementing the Li-Merz 12-6-4 potential to preserve Mg(II) octahedral coordination. Following geometry validation via CheckMyMetal, systems were simulated in OpenMM through staged minimization, NVT heating (10 to 310 K), and NPT equilibration (1 bar, 310 K). System assembly proceeded without topological errors, and the 12-6-4 parameters preserved canonical octahedral coordination for both ions. Specifically, Mg1095 was coordinated by Asp617 (OD1 at 2.05 Å; OD2 at 2.33 Å), Asp788 (OD2 at 2.20 Å), and four first-shell waters (2.02–2.13 Å), whereas Mg1096 adopted regular six-coordinate geometry via Glu825 (OE1 at 1.98 Å) and five waters (2.06–2.15 Å). Foscarnet remained within the exchange channel without direct contacts (< 2.5 Å), as coordinating waters satisfied both coordination spheres without distorting the active site. Together, these findings demonstrate that this protocol preserves active-site geometry and metal coordination in explicit solvent, validating the MD workflow for closed-conformation HSV-1 DNA polymerase. Ultimately, this validated receptor provides a robust structural reference to evaluate and compare future MD simulations.

DRAFT - UNDER REVIEW

POSTER #114 **TRACK:** Proteomics and Protein Structure and Function

A workflow for structural phylogenetic analysis of divergent protein domains

Source: 05_posters/P114.tex · Excel row 94

González JM[1]

[1] Instituto de Bionanotecnología del NOA, INBIONATEC, CONICET-UNSE. Santiago del Estero.

Tracing the evolutionary origins of ancient protein superfamilies is typically hindered by phylogenetic signal loss due to extreme sequence divergence. At these timescales, multiple sequence alignments (MSAs) suffer from severe inaccuracies, where gaps appear almost arbitrarily across homologous regions, leading to unreliable phylogenies. Instead, the true relatedness between two highly divergent polypeptide sequences can only be established through structural comparison, since 3D structures are far more conserved than linear 1D sequences. While this approach has been historically limited by the availability of relatively few experimental structures in the Protein Data bank, present day AI-driven structure prediction algorithms have fundamentally transformed this scenario, paving the way to a new era of Structural Phylogenetics. Herein, a topology-driven workflow is presented, to trace the evolutionary origins of ancient protein families. As a proof-of-concept, this workflow was deployed to characterize the superfamily of proteins sharing an ATP-grasp domain. The pipeline successfully clustered the superfamily into 34 distinct families, mapping sequence conservation to the 3D geometry of their ATP binding pockets. Structural homology search was accomplished through querying the AFDB50 database with Foldseek, retrieving structurally homologous sequences with domain boundaries defined by secondary structure elements. This approach offers an alternative to customary HMM profile search for domain curation and deep evolutionary analysis. For each family, MSAs were calculated, curated, and used to estimate ML phylogenies, where IQ-Tree was used to obtain representative ATP-grasp domains of each family through ancestral sequence reconstruction. Finally, all ancestors were modeled with AlphaFold3, allowing detailed structural analyses such as all-to-all superposition for template-modeling scoring, and cross-validation of sequence-based (cladistic) and structure-based (phenetic) trees. By anchoring evolutionary histories to physical 3D geometries, this framework offers a robust pipeline for structural analysis of divergent proteins, allowing us to decode the evolution of ancient enzymes long before their linear sequence similarity has faded.

DRAFT – UNDER REVIEW

POSTER #116 **TRACK:** Proteomics and Protein Structure and Function

Backbone scaffold conformation influences the stability of amyloid cross-β structures: an all-atom molecular dynamics study

Source: 05_posters/P116.tex · Excel row 95

Fernández ML[1,2,3], Kajava A[4], Marino-Buslje C[3,5]

[1] Departamento de Física, FCEyN, UBA, Buenos Aires, Argentina [2] CONICET - UBA, Instituto de Física Interdisciplinaria y Aplicada (INFINA), Buenos Aires, Argentina [3] Instituto Tecnológico de Buenos Aires (ITBA), Buenos Aires, Argentina [4] CRBM/CNRS UMR5237, Université de Montpellier, 1919 Route de Mende, 34293 Montpellier, France [5] Fundación Instituto Leloir (FIL)/IIBBA-CONICET. Buenos Aires, Argentina

Amyloid fibrils are a structural hallmark of many neurodegenerative disorders, including Alzheimer, Parkinson, and amyotrophic lateral sclerosis. Fibrils formed by unrelated proteins are built on a common cross-β architecture. Within it, the backbone adopts a wide variety of folds. The intrinsic stability of a fold may determine its efficiency as a seed for fibril elongation. In natural sequences, however, the backbone contribution cannot be separated from that of side-chain packing. We therefore isolate this effect using fibril models consisting of 10-residue poly-alanine chains arranged in four distinct backbone architectures: flat, single-arc (V-shaped), two-arcs same-sense (horseshoe-like), and two-arcs alternating (zig-zag).

Each architecture was simulated by all-atom molecular dynamics in explicit solvent at two sizes of the cross-β stacks, three and ten chains, with 50 independent replicas per condition. Stability was quantified at t50, the time required to lose half of the native inter-chain backbone hydrogen bonds. Backbone RMSD relative to the starting configuration and solvent-accessible surface area (SASA) were followed along each trajectory and also evaluated at t50. All observables yielded consistent results. The t50 analysis established a stability hierarchy: flat < V-shaped ≅ zig-zag < horseshoe-like. While SASA at t50 provided clear discrimination among the scaffolds, recapitulating the stability ranking obtained from the inter-chain hydrogen-bond analysis, RMSD at t50 goes to the same direction, but discriminated less sharply among the scaffolds, indicating that hydrogen-bond loss precedes large-scale structural drift.

The same ordering was recovered in both the three-chain and ten-chain assemblies, indicating that the stability hierarchy is not an artifact of the minimal model. PolyVal fibrils were consistently more stable than their polyAla counterparts, showing that side-chain might also contribute to fibril stability.

Backbone conformation alone, independent of residue side-chain, establishes a stability hierarchy among amyloid scaffolds, with β-arched topologies being more resistant to dissociation than flat β-sheets. These findings provide a basis for future studies of how backbone conformation governs amyloid nucleation and seeding, suggesting that β-arch-containing conformers may be better candidates for nuclei or seeds of fibril growth and offering a criterion for ranking scaffolds in models of nucleation and cross-seeding.

DRAFT - UNDER REVIEW

POSTER #118 **TRACK:** Evolutionary, Comparative and Population Genomics

Experimental warming alters fungal transcriptional signals recovered from Colobanthus quitensis RNA-seq data

Source: 05_posters/P118.tex · Excel row 96

Juan Machuca[1], Olman Gómez-Espinoza[2], María José Contreras[3], **Karla Leal**[3]

[1] Doctorado en Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Temuco 4810101, Chile. [2] Instituto de Ciencias Aplicadas, Facultad de Ingeniería, Centro de Investigación e Innovación, Universidad Autónoma de Chile, Huechuraba 8580000, Chile. [3] Instituto de Ciencias Aplicadas, Facultad de Ingeniería, Universidad Autónoma de Chile, Temuco 4810101, Chile. * Corresponding author: Karla Leal (karla.leal@uautonoma.cl)

Our knowledge of how plant-associated microorganisms respond to experimental warming is quite limited, especially in challenging environments like Antarctica. Plant transcriptomes that are publicly available might contain RNA from these microorganisms, presenting an underutilized opportunity to investigate their ecological and functional responses to environmental changes.We reanalyzed poly(A)-selected RNA-seq from mature aerial tissue of the Antarctic pearlwort Colobanthus quitensis collected after three growing seasons under open-top chamber (OTC) warming or open-area (OA) conditions. After uniform quality and fungal-signal assessment, six libraries (three OTC and three OA) were retained and processed through sequential host subtraction, positive fungal classification, de novo Trinity reconstruction, competitive cross-kingdom DIAMOND/LCA validation, Salmon quantification, DESeq2 differential modeling, and independent InterPro and Gene Ontology (GO) aggregation. OA libraries yielded 8.1-fold more recoverable fungi-only fragments than OTC libraries. The fungi-only assembly comprised 2,618 transcripts representing 1,886 Trinity genes; 972 of 1,217 predicted ORFs met strict fungal classification criteria. Ascomycota dominated the recovered fungal fraction, with Sclerotiniaceae and Botrytis showing the greatest mean representation at family and genus levels, respectively. Transcript-level analysis identified 60 significant features (BH $q<0.05$) representing 50 Trinity genes, with 44 showing greater representation in OA and 16 in OTC. Independent functional models identified 29 significant InterPro entries and 29 GO terms. OA-associated signals showed greater representation of ribosomal, translation, RNA-binding, and ribosome-biogenesis functions, whereas OTC-associated signals included HSP70/HSP90, unfolded-protein binding, ATP-dependent chaperone activity, oxidoreductase functions, and aldehyde-dehydrogenase-related regions. Across 100 depth-matching iterations, all 118 significant transcript and functional signals retained their original direction in at least 80% of iterations. Similarly, these findings indicate that Antarctic transcriptomic archives, initially created for host research, can be analysed with suitable bioinformatic methods to extract further ecological and functional insights into how plant-associated fungi react to environmental changes.

DRAFT – UNDER REVIEW

POSTER #120 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

A Post-Hoc Analytical Framework for the Characterization of Disease Endotypes Applying Biclustering

Source: 05_posters/P120.tex · Excel row 97

Gómez López R[1], Lujan Cecchini R[1,2], Soto AJ[1,2], Ponzoni I[1,2]

[1] Instituto de Ciencias e Ingeniería de la Computación, CONICET–UNS [2] Departamento de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur (UNS)

Biclustering is a machine learning technique designed to detect local patterns within a data matrix. Unlike traditional clustering methods, biclustering simultaneously identifies subsets of rows and columns that exhibit a coherent mathematical pattern. This is particularly relevant for transcriptomic data, where different molecular programs may be active only in specific subgroups of patients and represented only by a subset of genes, potentially remaining undetected by clustering approaches.

In this work, we propose a post-hoc framework designed to identify and characterize disease endotypes through the systematic application of multiple biclustering approaches. The framework is applied to systemic sclerosis as a case study to characterize transcriptomic heterogeneity and identify candidate molecular endotypes. This framework begins with the application of biclustering methods, such as QUBIC-2, FABIA, and ISA, followed by the standardization and integration of their results. Subsequently, the identified biclusters (defined as subsets of genes and samples) are statistically assessed against an empirical null distribution derived from the dataset using the Virtual Error as an internal coherence metric to evaluate their statistical significance, distinguishing non-random patterns from those that may arise by chance.

Next, an integration step is applied across statistically supported biclusters to identify recurrent patient groups. A sample-by-sample similarity matrix is then constructed from patient co-membership across significant biclusters, and hierarchical clustering is applied to detect core patient groups characterized by recurrent co-occurrence. Subsequently, a gene signature is derived for each core patient group based on gene membership across its supporting biclusters. The resulting signatures are characterized through functional enrichment analyses to identify related biological processes and molecular pathways. Finally, each candidate endotype is compared with previously reported disease endotypes to assess biological concordance and explore whether the identified patterns are consistent with known endotypes or reveal mixed disease profiles.

Overall, these complementary analyses provide a systematic basis for the identification and characterization of recurrent molecular subgroups, enabling the identification of shared and distinct molecular features across candidate disease endotypes.

DRAFT - UNDER REVIEW

POSTER #121 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

16S rRNA-based analysis of storage-associated shifts in the relative composition of the raw boar seminal microbiome at 17 °C

Source: 05_posters/P121.tex · Excel row 98

Machuca J[2], Carrasco-Zambrano N[1], Leal K[1], Núñez-Montero K[3], Pezo F[4,5], Contreras MJ[1]

[1] Instituto de Ciencias Aplicadas, Facultad de Ingeniería, Universidad Autónoma de Chile, Temuco, Chile. [2] Doctorado en Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Temuco, Chile. [3] Instituto de Ciencias Aplicadas, Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Santiago, Chile. [4] Facultad de Ciencias Agropecuarias y Medioambiente, Universidad de La Frontera, Temuco, Chile. [5] Centro de Excelencia en Biotecnología Reproductiva (CEBIOR), Facultad de Ciencias Agropecuarias y Medioambiente, Universidad de La Frontera, Temuco, Chile

Liquid storage of boar semen at 15–17 °C is routinely used in artificial insemination, but the temporal changes of the seminal bacterial community under storage remains incompletely characterized. We assessed storage-associated changes in the relative composition and diversity of the raw boar seminal microbiome using 16S rRNA gene sequencing and a reproducible bioinformatic workflow. Twenty-four 16S rRNA gene V3–V4 amplicon libraries from six Landrace boars sampled on days 1, 4, 8, and 12 of storage at 17 °C were processed in QIIME 2 2024.10. Primer removal, denoising, and amplicon sequence variant (ASV) inference were performed with Cutadapt and DADA2, and taxonomy was assigned with a V3–V4-specific Naive Bayes classifier trained on SILVA 138. After taxonomic curation and filtering, 3,229 ASVs were retained. Alpha diversity was evaluated using Chao1, Shannon, Simpson, and observed richness, whereas Bray–Curtis dissimilarity and principal coordinate analysis (PCoA) were used to explore between-sample compositional patterns. Within-sample diversity was highest on day 1 and lower from day 4 onward. Taxonomic profiles changed across storage time points, with later samples showing greater relative representation of genera including Pseudomonas, Escherichia-Shigella, and Enterococcus, although the magnitude of these patterns varied among animals. PCoA further revealed temporal variation in community composition, with later storage time points showing increasingly distinct compositional profiles relative to baseline. Together, these findings indicate storage-associated changes in bacterial diversity and relative taxonomic composition during semen storage at 17 °C. Overall, the analysis describes storage-associated changes in diversity and relative community composition of raw boar semen and provides a reproducible framework for characterizing temporal microbiome patterns under the experimental conditions evaluated.

DRAFT – UNDER REVIEW

POSTER #122

TRACK: Artificial Intelligence and Systems Biology

Emergent Time-Sharing in Competitive Systems with Cross-Inhibition and Slow Self-Adaptation

Source: 05_posters/P122.tex · Excel row 99

Ruiz Bochides EA[1,2], Sanchez de la Vega C[3,4], Ventura AC[1,2]

[1] Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina [2] Instituto de Fisiología, Biología Molecular y Neurociencias (IFIBYNE), CONICET–Universidad de Buenos Aires, Buenos Aires, Argentina [3] Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina [4] Instituto de Cálculo (IC), CONICET–Universidad de Buenos Aires, Buenos Aires, Argentina

Time-sharing, understood as the alternating use of a shared resource, represents a potential mechanism of temporal organization in competitive systems. However, the mechanisms underlying its emergence in biological systems remain incompletely understood. In cellular contexts, multiple processes may compete for limited resources, such as enzymes, ribosomes, or regulatory molecules, raising the question of how their utilization is dynamically organized.

In this work, we propose a minimal dynamical system in which time-sharing emerges without centralized control. We consider a nonlinear model consisting of two activity variables coupled through mutual inhibition, each associated with a slow, activity-dependent self-adaptation variable representing the accumulation of inhibitory factors. Using a system of coupled differential equations, we show that the interplay between strong mutual inhibition and slow adaptive negative feedback gives rise to robust antiphase oscillations. Consequently, the alternating dominance of the activity variables can be interpreted as an emergent form of temporal resource partitioning, resulting in a time-sharing regime.

Our results show that spontaneous alternation between competing activities can emerge in systems with memory, without requiring an explicit coordination mechanism or external periodic forcing. This framework identifies a minimal dynamical mechanism linking competition, adaptive regulation, and temporal organization in biological systems.

DRAFT – UNDER REVIEW

POSTER #123 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

An Integrated Bioinformatics Framework for ORF Characterization and Visualization

Source: 05_posters/P123.tex · Excel row 100

Cirignoli C[1], **Sierra C**[1,2], Romano S[1], Osipovitch M[1]

[1] ZS, Research & Development Excellence, Discovery [2] Universidad de Buenos Aires

Recent advances in ribosome profiling (Ribo-seq) and mass spectrometry-based proteomics have revealed a complex landscape of translated open reading frames (ORFs), including many outside canonical protein-coding annotations. However, integrating translation evidence, peptide support, and functional annotations remains challenging due to fragmented analytical workflows and visualization tools.

We developed an integrated bioinformatics framework and interactive R Shiny platform for the exploration and characterization of translated ORFs. The framework combines Ribo-seq-derived ORFs, peptide back-mapping, sequence information, and functional annotations into standardized genomic representations, harmonizing outputs from multiple analytical pipelines into a unified data model. ORF predictions are enriched with genomic and transcriptomic context, amino acid sequences, peptide evidence, and annotation metadata, enabling comprehensive downstream analysis.

The centerpiece of the framework is an interactive R Shiny application that serves as a platform for data exploration, interpretation, and prioritization of translated ORFs. By integrating multiple evidence layers within a shared genomic context, the application allows users to visualize ORF structures, Ribo-seq coverage profiles, peptide mappings, transcript annotations, sequence features, and functional information through a single interface. Researchers can efficiently query, filter, and investigate candidate ORFs, eliminating the need to manually reconcile results across multiple tools.

Application of the framework enabled rapid assessment of translated ORFs and streamlined evaluation of supporting evidence from diverse experimental sources. The unified view of translational activity, peptide validation, and annotation data improved interpretability, facilitated quality assessment, and accelerated candidate prioritization.

This framework bridges analytical pipelines and interactive analysis, providing a scalable foundation for integrating additional transcriptomic, proteomic, and annotation datasets while supporting systematic investigation of the translated genome.

DRAFT – UNDER REVIEW

POSTER #124 **TRACK:** Evolutionary, Comparative and Population Genomics

Comparative genomic analysis of budded virion production in baculoviruses

Source: 05_posters/P124.tex · Excel row 101

Ledesma FT[1], Cuccovia Warlet FU[1], Simonin AJ[1], Lassalle FM[1], Maciel Alonso G[1], Motta FL[1], Cerrudo CS[1], Belaich MN[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.

Baculoviruses comprise a diverse group of insect viruses with substantial genomic and phenotypic diversity. Although the production of budded virions (BV) is a central component of the baculoviral replication cycle, some baculoviral lineages lack this infectious form. Whether this reflects ancestral absence followed by acquisition or secondary loss of genes involved in BV formation remains unclear. We explored this question by comparing genes associated with baculoviral replication and virion formation across representative viral lineages, evaluating gene conservation, functional domain architecture, genomic organization and codon usage.

As an initial case study, we analysed the major occlusion body protein (MOBP), polyhedrin-granulin, a highly abundant structural component of occlusion bodies. Comparative codon usage analysis revealed substantial variation among MOBP genes, with several sequences showing limited correspondence with the codon usage of their respective viral genomes. PCA based on codon usage revealed distinct patterns among MOBP genes. Notably, the AcMNPV MOBP gene grouped in PCA space with CuniNPV, the most phylogenetically divergent lineage among those analysed, rather than with those from closely related baculoviruses, suggesting a possible contribution of host-associated codon usage patterns to the evolutionary history of this highly expressed viral protein.

These results provide preliminary evidence that gene composition, protein domain architecture and codon usage may contribute to the diversification of baculoviral replication strategies. Comparative analysis of conserved and lineage-specific features may help distinguish patterns consistent with gene gain or loss and provide insight into the evolutionary diversification of BV production and baculovirus–host adaptation.

DRAFT - UNDER REVIEW

POSTER #125 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Complete genome assembly of Argentinean Creole bovine using long reads

Source: 05_posters/P125.tex · Excel row 102

Guizzo D[1,2], Irazoqui JM[1,2], Santiago GM [1,2], Eberhardt MF [1,2], Arroyo P [3], Amadio AF[1,2]

[1] Universidad Nacional de Rafaela (UNRaf) [2] Instituto de Investigaciones de la Cadena Láctea (IDICAL) CONICET-INTA [3] Instituto de Innovación para la Producción Agropecuaria y el Desarrollo Sostenible (IPADS) CONICET-INTA

The Argentinean Creole is a cattle breed with origins tracing back to the Spanish conquest in the 16th century. Its introduction into various Latin American countries led to rapid expansion, driven by favorable climatic and nutritional conditions. Adapted to the diverse environmental conditions of Latin America, Creole cattle possess genetic diversity that distinguishes them from other commercial breeds. This work reports the generation of the first complete reference genome sequence for the Creole bovine using long read sequencing, and was achieved as part of the International Bovine Pangenome Consortium.

In this context, Oxford Nanopore long read technology was employed to sequence the complete genome of a single male individual. High molecular weight DNA was extracted to construct genomic libraries, which were then sequenced using a PromethION P2solo platform. The resulting sequences provided approximately 45X coverage of the ~2.8 Gbp genome. Various assembly strategies enabled the reconstruction of the complete Creole bovine genome for the first time, including the assembly of several chromosomes without gaps. Some assemblers, provided haplotype specific assemblies, something that was impossible a few years ago. Comparison of the assembly we obtained with reference genomes of other breed was done, showing excellent completeness results and, sometimes, longer chromosomes containing telomeres.

To compare genotyping results obtained via short read sequencing, we compared genotyping results derived from mapping 17 short read sequenced Creole cattle against two references: the standard bovine reference (ARS2.0) typically used in breeding programs and the assembly generated in this study. Using an appropriate reference (matching the specific breed) enables more accurate genotype analysis, yielding better mapping and genotype assignment rates. This principle underpins the International Bovine Pangenome Consortium, which aims to sequence all cattle breeds worldwide; our contribution will facilitate the creation of pangenomes that represent the genetic diversity of each breed more accurately than a single linear reference sequence.

DRAFT – UNDER REVIEW

POSTER #126 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Inferring microRNA expression patterns and regulatory circuits at single-cell resolution through pri-miRNA detection

Source: 05_posters/P126.tex · Excel row 103

Alarcón Pastor D[1,2], Cavagnaro Rago D[1,2], González PM[1,2], Aguirre PA[1,2], **Fededa JP**[1,2]

[1] Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín (UNSAM) – Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina. [2] Escuela de Bio y Nanotecnologías (EByN), Universidad Nacional de San Martín, Argentina.

DRAFT · UNDER REVIEW

Detecting microRNAs at single-cell resolution is still a methodological challenge. Current methods for detecting mature microRNAs are still not compatible with commercial single-cell RNA sequencing (scRNA-seq) platforms, limiting the study of microRNA expression and regulatory relationships in complex tissues made up of many cell types. Much of the regulation of microRNA expression happens at the transcriptional level: Detecting immature, polyadenylated microRNA precursors (pri-miRNAs) could therefore serve as a proxy for mature microRNA expression.

Here, we developed a pipeline that detects pri-miRNAs in scRNA-seq datasets generated with poly(A) capture technologies (10x Genomics Chromium 3'). For the alignment strategy, we generated curated GTF annotations of pri-miRNA genes in a parallel alignment strategy, so coding transcripts and pri-miRNAs are quantified from the same cells. Using this pipeline in published developing brain scRNA-seq datasets, we identified pri-miRNA expression in silico and characterized the distribution of pri-miRNA species across brain cell types and regions.

To validate pri-miRNAs as a proxy for mature microRNAs, we analyzed their expression across developmental time and brain regions and found that the expression patterns of several pri-miRNAs matched those reported in the literature for the corresponding mature microRNAs.

By combining pri-miRNA and coding-transcript expression, we explored candidate microRNA–target relationships across cell types and found that the number of target sites for different microRNAs in a transcript's 3' UTR is critical for the strength of the regulatory relation. Finally, we explored microRNA families containing members with identical mature sequences that are processed from different pri-miRNAs transcribed at different loci, identifying the locus of origin of these microRNAs in specific cell types and tissues.

By taking advantage of single-cell resolution analysis, this strategy identifies novel candidate microRNA-mediated regulatory circuits and expands the ways these molecules can be studied at the single-cell level.

POSTER #127

TRACK: Artificial Intelligence and Systems Biology

Consistency and Stability of XAI Explanations to Understand QSAR Models trained for Predicting Developmental Toxicity

Source: 05_posters/P127.tex · Excel row 104

Rambo AR[1], Ponzoni[2], **Cravero F**[3]

[1] Instituto de Investigación, Desarrollo e Innovación en Informática, FCEQyN-UNaM, Misiones, CP 3300, Argentina, alirambo@fceqyn.unam.edu.ar [2] Instituto de Ciencias e Ingeniería de la Computación, UNS-CONICET, Bahía Blanca, CP 8000, Argentina, ip@cs.uns.edu.ar [3] Departamento de Informática e Inteligencia Artificial, UCU, Montevideo, CP 11600, Uruguay, fiorella.cravero@ucu.edu.uy

Different QSAR (Quantitative Structure–Activity Relationship) models trained on the same data (identical sampling and validation strategies) may yield divergent predictions for the same molecule. Each machine learning technique introduces its own inductive bias that determines how it organizes chemical space and, consequently, how it predicts. One way to detect these biases is through the use of Explainable artificial intelligence (XAI) methods such as LIME and SHAP, which assign feature importance values. However, the stability and consistency of these explanations play a key role when it comes to trusting them. This study presents a comparative analysis of explanations obtained with LIME and SHAP across three QSAR models for predicting developmental toxicity which have different inductive bias. Developmental toxicity models were trained on two different datasets, D1 and D2, using 27 RDKit molecular descriptors. Three QSAR models were trained using Support Vector Machine (SVM) with kernel RBF, Gaussian Naive Bayes (NB) and k-nearest-neighbors (KNN), and were explained with KernelSHAP and LIME under an identical computational budget. Explainer performances were quantified using Rank Consistency (RC), Importance Stability (IS), and GES = ½(RC + IS). Regarding the results, each inductive bias is associated with a coherent explainer, and this mapping is reproduced across both datasets: NB's conditional independence aligns with Shapley values (SHAP: GES 0.89–0.91), whereas SVM's global non-linearity and KNN's locality are better captured by LIME's local approximation (GES 0.84–0.93). SHAP showed high consistency for NB based models, with GES values of 0.89 and 0.91 for D1 and D2 respectively. LIME showed high consistency for SVM, with GES values of 0.86 and 0.93 for D1 and D2 respectively, and for KNN, with values of 0.90 and 0.84 for D1 and D2 respectively. A divergent prediction for clindamycin is analyzed in detail, where SVM and NB-based models predict wrongly and KNN-based models predict correctly. The explainers clearly reflect the biases of the three models: NB concentrates importance on heterocyclic fragments, and KNN assigns greater importance to continuous molecular connectivity descriptors over structural fragments; for SVM, the SHAP importance values are unstable for this case (IS=0.41). In conclusion, these results suggest that comparative XAI analysis can provide complementary information about models behavior beyond predictive performance alone.

DRAFT – UNDER REVIEW

POSTER #130

TRACK: Biological Databases and Bioimaging

PHENOTYPING OF FUNGAL INFECTIONS IN TOMATO PLANTS USING MULTISPECTRAL IMAGING

Source: 05_posters/P130.tex · Excel row 105

Muschietti B[1], Ferrante E[1,2], Ariel F[2,3]

[1] Laboratorio de Inteligencia artificial Aplicada - Instituto de Ciencias de la Computación - Facultad de Ciencias Exactas y Naturales - UBA / CONICET [2] APOLO Biotech [3] Laboratorio de Epigenética y ARNs no codificantes - Instituto de Fisiología, Biología Molecular y Neurociencias - Facultad de Ciencias Exactas y Naturales - UBA / CONICET

Early detection of plant diseases is essential for crop management. To this end, multispectral technology has emerged as a powerful tool for detecting subtle changes in leaf physiology. Due to intracellular phenomena resulting from infection, variations in reflectance across the electromagnetic spectrum serve as an early indicator of lesion development, such as necrotic spots. Based on this, we set out to explore the scope and results of this methodology under laboratory conditions with the goal of eventually transferring it to open-field conditions. To do so, it was necessary to develop a comprehensive system which included: i) the design of a non-conventional experimental setup using commercial materials, ii) the generation of a method for processing raw images to obtain leaf segments, iii) the selection of features calculated from spectral values, and iv) the implementation of machine learning models such as random forest (RF) and support vector machine (SVM) to classify infected leaves. Using this complete procedure, we achieved an F1-score of 85 % and a recall of nearly 70 % for infected leaves —even those with nearly imperceptible lesions— using a canopy-level perspective.

DRAFT – UNDER REVIEW

POSTER #131 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Generation and evaluation of a bioinformatics pipeline for the detection of viral sequences in insects using public NGS data

Source: 05_posters/P131.tex · Excel row 106

Maciel Alonso G[1], Lassalle FM[1], Cuccovia Warlet FU[1], Ledesma FT[1], Simonin JA[1], Motta LF[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.

Biological pest control using entomopathogens represents a sustainable alternative or complement to chemical insecticides. Among these agents, baculoviruses are widely used, but viral diversity in insects is extensive and its biopesticidal potential remains largely unexplored. Therefore, characterizing viromes associated with pest insects is crucial for identifying novel viruses with potential applications in pest management. Genome Next-Generation Sequencing (NGS) bioprojects generate large amounts of data containing reads that do not map to the host and can be exploited to detect viruses. By applying a metagenomic approach to these unmapped reads, it is possible to recover complete/partial viral genomes (of environmental contaminants, microbiome-associated viromes, persistent infections, endogenous elements) associated with cells, organisms, or tissues.

In this work, a bioinformatics pipeline was developed to identify and assemble putative viral reads from public NGS datasets, which was validated for the pest Spodoptera frugiperda (CM028954 as reference) and associated bioprojects (ERR1212036: cells; SRR29409529: larvae). Briefly, the total raw reads for the project were downloaded and mapped against the insect reference genome using Bowtie2 to recover only unmapped reads. These reads were assembled with metaSPAdes, and the resulting contigs were clustered and converted into scaffolds using MaxBin2. BLAST was used to search viral databases for identification and taxonomic assignment. After mapping, 59.56% and 35.24% of non-insect reads were recovered from cells and larvae, respectively. Finally, the de novo assembly of these reads resulted in 1,700,000 and 1,022,824 contigs. In both datasets, most (84%) of the detected viral genomes belonged to Telodnaviria, Heunggongvirae, and Orthornavirae, while 10% could not be classified. As expected for the case study, both samples showed lower proportions of Polydnaviridae and Nudiviridae. Additionally, Nucleocytoviricota was detected in SRR29409529, whereas Bamfordvirae and Anelloviridae were detected in ERR1212036.

This approach enables the reuse of public datasets to characterize viromes associated with cells and to discover previously unreported viruses in hosts of interest. The open-source tools employed make the pipeline adaptable to different organisms, sequencing depths, and assembly qualities, facilitating its scalability, reproducibility, and adjustment to available datasets.

DRAFT - UNDER REVIEW

POSTER #133 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Prophage profile of Staphylococcus aureus isolates from persistent infections.

Source: 05_posters/P133.tex · Excel row 107

Dotto C[1], Muñoz B[1], Quiroga MP[1], Buzzola F[1].

[1] Instituto de Investigaciones en Microbiología y Parasitología Médica (UBA-CONICET)

Staphylococcus aureus is an opportunistic pathogen that can cause severe infections with high mortality. Its ability to form biofilms hinders eradication and contributes to the persistence of infections. The success of this pathogen is due to its arsenal of virulence factors combined with antibiotic resistance and its genomic architecture. Phages integrated into the bacterial genome, or prophages, can contribute to the virulence of S. aureus by altering the structure of its genes and by carrying genes associated with its pathogenicity. In this study, 16 S. aureus isolates from persistent infections in our country were analyzed to determine the presence of prophages, their gene content, and biofilm-forming capacity. DNA was extracted and sequenced using Illumina MiSeq technology. The resulting fragments were assembled using SPAdes, assembly quality was assessed using Quast, and the genomes were annotated using Prokka. Thirty-seven prophage regions were identified using PHASTEST and classified as intact (22), questionable (2), or incomplete (13). All isolates possessed at least one prophage. Two isolates contained a single intact prophage in their genomes. Regarding virulence genes carried by prophags, the presence of leukocidins, superantigens, hemolysin, immune evasion proteins, and pathogenicity islands was determined. The presence of a β-lactamase gene was determined in one of the intact prophags. Biofilm formation was determined spectrophotometrically using the crystal violet method. The results showed that 6.6% were classified as high (HF), 13.3% as medium (MF), 46.06% as low (LF), and 33.33% as non-biofilm forming (NF). It is noteworthy that AF isolates possessed prophages that did not contain virulence genes, while these genes were found in prophages from BF or NF isolates. Overall, the characterization of the prophages allowed the identification of a diversity of genetic elements potentially associated with S. aureus virulence, although the presence of genes related to biofilm formation was not determined. Integrating genomic analyses with phenotypic assays and studying a larger number of isolates will allow for the evaluation of the potential of prophage profiles as genomic indicators of characteristics relevant to infection persistence.

DRAFT – UNDER REVIEW

POSTER #134 **TRACK:** Proteomics and Protein Structure and Function

Identification of Novel Agricultural Antifungal Candidates Targeting Cytochrome b through Virtual Screening, Molecular Modeling and Resistance-Mutation Analysis.

Source: 05_posters/P134.tex · Excel row 108

Miranda ML[1], Otero J[2,3], Richmond V[4,5]

[1] Departamento de Química Biológica, FCEN-UBA [2] IFIBA (CONICET-UBA) [3] Departamento de Física, FCEN-UBA
[4] UMYMFOR (CONICET-UBA) [5] Departamento de Química Orgánica, FCEN-UBA

Fungal diseases represent a major threat to global agricultural productivity, and the increasing occurrence of resistance to available fungicides highlights the need for new modes of action and novel active compounds. This study aimed to identify potential agricultural antifungal agents targeting cytochrome b, a key component of the mitochondrial electron transport chain and a well-established fungicide target.

A computational workflow integrating homology modeling, virtual screening, molecular docking, and binding-pocket characterization through Molecular Dynamics (MD) simulations was developed to identify promising cytochrome b inhibitors. Structural models of wild-type and resistance-associated mutant cytochrome b proteins were generated and analyzed to investigate changes in binding-site architecture and physicochemical properties. Large chemical libraries were screened using molecular docking to prioritize compounds with favorable binding affinities and interaction profiles across different protein variants.

Results revealed a predominantly hydrophobic binding cavity embedded within the transmembrane region of cytochrome b. Selected compounds established key hydrophobic interactions and displayed binding modes consistent with favorable calculated binding energies. Comparative analyses of wild-type and mutant proteins showed that resistance-associated amino acid substitutions can alter pocket accessibility and interaction networks. Notably, several candidate compounds maintained strong interactions across multiple cytochrome b variants, suggesting reduced susceptibility to resistance-related structural changes. These findings provided a molecular basis for prioritizing compounds with potential broad-spectrum efficacy against susceptible and resistant fungal strains.

Overall, the integration of structural modeling, resistance-mutation analysis, virtual screening, and MD-based binding-site characterization have proved to be an effective approach for the discovery of novel cytochrome b-targeting antifungal candidates. These findings provide a framework for the rational development of fungicides with improved potential to overcome emerging resistance mechanisms.

DRAFT – UNDER REVIEW

POSTER #136 **TRACK:** Proteomics and Protein Structure and Function

A modular workflow for residue-level analysis of peptide–protein interactions from molecular dynamics trajectories

Source: 05_posters/P136.tex · Excel row 109

Pineda M[1], Aguirre M[1], Maidana E[1], Balatti GE[1]

[1] Instituto Universitario para el Desarrollo Productivo y Tecnológico Empresarial de la Argentina (IUDPT), Ciudad Autónoma de Buenos Aires, Argentina.

Interpreting molecular dynamics trajectories of biomolecular complexes requires distinguishing persistent interfacial contacts from transient proximity. We present the design of a modular workflow for residue-level analysis of peptide–protein interactions, developed as an undergraduate research project. The workflow organizes four complementary tasks: calculation of inter-residue distances and contact persistence using a configurable cutoff, geometric characterization of selected contacts, integration of Coulomb and Lennard-Jones interaction energies when available, and prioritization of residues based on the combined evidence. A 0.6 nm distance threshold is used as an initial operational definition of contact, while time-resolved outputs are intended to reveal changes in the interface across a trajectory. The resulting analysis plan specifies the input data, intermediate measurements, and residue-level outputs required for application to existing molecular dynamics simulations. This framework provides a defined basis for implementing and evaluating a reproducible analysis tool. Its prospective application to peptide–protein systems may help select interfacial residues for subsequent computational or experimental study.

DRAFT – UNDER REVIEW

POSTER #138 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Identification of the heavy metal ATPases (HMAs) genes in Yerba Mate

Source: 05_posters/P138.tex · Excel row 110

Dellamea CV[1,2], Duarte MJ[1,2], Alvarez MY[1,2], Angelina EL[3,4], Acevedo RM[1,2], Sansberro PA[1,2]

[1] Facultad de Ciencias Agrarias - Universidad Nacional del Nordeste (UNNE) [2] Instituto de Botánica del Nordeste (IBONE) - UNNE - CONICET [3] Facultad de Ciencias Exactas, Naturales y Agrimensura - Universidad Nacional del Nordeste (UNNE) [4] Instituto de Química Básica y Aplicada del Nordeste Argentino (IQUIBA-NEA) - UNNE - CONICET

Yerba Mate (Ilex paraguariensis) is commercially cultivated in southern Brazil, southeastern Paraguay, and northeastern Argentina. The commercial product consists of processed leaves and young stems. The increasing contamination of agricultural soils with heavy metals is a concern for the international community due to the associated toxic effects on human health. HMA (heavy metal-transporting ATPase) transporters are proteins with translocase activity that couple ATP hydrolysis to the transmembrane transport of Zn^{2+}/Cd^{2+}; essentially, they function as pumps that load metals into the xylem, phloem, or apoplast for redistribution to other plant parts (such as leaves and shoots) as part of a redistribution strategy. HMA proteins are encoded by a gene family that has been studied in several species, including Hordeum vulgare, Oryza sativa, Arabidopsis thaliana, and Thlaspi caerulescens. This work aims to identify and structurally characterize the genes encoding the HMA transporter family in Yerba Mate, and to predict the structure of their encoded proteins.

Homologous sequences of HMA genes were selected and characterized based on the genome and leaf and root transcriptomes of yerba mate. The Oryza sativa HMA2 sequence (UniProt accession number: A3BF39) was used as a reference for the search. A phylogenetic tree was constructed using HMA proteins from other. TargetP, WoLF-PSORT, and Plant-mPLoc were used to predict subcellular protein localization. Transmembrane domains and the terminal localization of the identified HMA proteins were predicted using TMHMM and TMPred. The NCBI-CDD tool was used to identify conserved motifs in the yerba mate HMA proteins. Finally, the three-dimensional structure of the proteins was predicted using AlphaFold3.

Five genomic sequences were identified showing 69.8–82% nucleotide similarity and 58.6–75% amino acid similarity to the Oryza sativa HMA2 sequence. Furthermore, the Yerba Mate genomic sequences found phylogenetic relationship with the HMA proteins of A. thaliana and O. sativa. All HMAs predicted protein structures of Yerba Mate exhibit the transmembrane helix segments and characteristic domains of these transporters.

Knowledge of the sequences of these genes in Yerba Mate is the essential first step enabling studies to quantify their expression in various organs and in response to different stimuli, as well as analyses of their subcellular localization, among many other studies.

DRAFT – UNDER REVIEW

POSTER #139 **TRACK:** Proteomics and Protein Structure and Function

Protein disorder in Rhizobium favelukesii LPU83 and its association with the acid stress response

Source: 05_posters/P139.tex · Excel row 111

Iglesias J[1], Pistorio M[1], **Nilsson JF**[1]

[1] IBBM (Instituto de Biotecnología y Biología Molecular), UNLP - CONICET - La Plata - Buenos Aires – Argentina

Intrinsically disordered proteins (IDPs) lack stable secondary and tertiary structure under physiological conditions yet perform essential biological functions. While some IDPs remain constitutively disordered throughout their functional cycle (e.g. entropic chains), others adopt defined conformations upon interaction with their physiological partners, a phenomenon known as induced folding. This structural flexibility confers an enhanced capacity to engage multiple, structurally diverse binding partners. Consequently, IDPs are frequently found at the core of processes that depend on multi-partner interactions, including molecular recognition, macromolecular assembly, cell cycle control, signal transduction and transcriptional regulation. IDPs and intrinsically disordered regions (IDRs) have been linked to gene regulation, signaling and environmental stress responses in diverse organisms. Rhizobium favelukesii LPU83 is an acid-tolerant bacterium capable of establishing symbiosis with alfalfa (Medicago sativa). To elucidate the relevance of IDPs and IDRs in this organism in the context of the acid stress response, the complete proteome of LPU83 (GCA_055533505.1; 7,092 proteins) was analyzed using IUPred3, a disorder predictor that estimates inter-residue interaction energies. Approximately 600 proteins harboring at least one IDR (minimum length 30 amino acids) were identified. These results were cross-referenced with previously published proteomic and transcriptomic data obtained under acid stress conditions. IDR-containing proteins were significantly enriched among differentially expressed genes (DEGs) relative to the rest of the proteome, positioning IDRs as potentially relevant factors in the acid stress response. Over 50 IDR-containing proteins showed higher expression under acidic conditions. Functional category analysis revealed 10 regulators and signaling proteins with IDR among the DEGs, including methyl-accepting chemotaxis protein (MCP) receptors. Additionally, 29 secreted or membrane-associated IDR proteins were identified among the DEGs. These results constitute the first systematic analysis of protein disorder in R. favelukesii and lay the basis for further investigation into the role of IDRs in the acid stress biology of rhizobia.

DRAFT – UNDER REVIEW

POSTER #140 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Comparative Genomic Analysis of Lentilactobacillus kefiri CIDCA 8348 and CIDCA 83111: Probiotic Potential and Safety Assessment

Source: 05_posters/P140.tex · Excel row 112

Assandri MH[1,2], Escalante G[2], Valeff N[1], Leon LB[4], Malamud M[3], Serradell MA[1,2]

[1] Catedra de Microbiologia, Facultad de Ciencias Exactas, Universidad Nacional de La Plata [2] Instituto de Ciencias de la Salud, Universidad Nacional Arturo Jauretche, Florencio Varela. [3] Institute of Experimental Pathology, Center of Molecular Biology of Inflammation , University of Münster, Münster, Germany. [4] Centro de Investigación y Desarrollo en Fermentaciones Industriales, La Plata

Lentilactobacillus kefiri, commonly found in kefir, has gained interest for its potential probiotic properties. This study compared the genomes of L. kefiri CIDCA 8348 and CIDCA 83111 to identify genetic features potentially associated with probiotic properties and assess genomic safety markers.

Bioinformatic analyses were performed using the Galaxy platform and specialized tools for genome annotation, comparative genomics, and screening of genes associated with host adhesion, bacteriocin production, and S-layer proteins (SLPs). Antimicrobial resistance-associated determinants, virulence factors, and mobile genetic elements were also investigated.

CIDCA 8348 had a genome of 2,453,653 bp, a G+C content of 41.72%, and 2,375 coding sequences (CDSs), whereas CIDCA 83111 had 2,513,687 bp, 41.6% G+C, and 2,434 CDSs. Three CDSs associated with host adhesion were identified in CIDCA 8348, compared with 15 in CIDCA 83111. One gene related to bacteriocin production was identified in CIDCA 8348, compared with four in CIDCA 83111. Both strains contained eight SLP-encoding genes, with sequence differences between them. CIDCA 8348 showed high genomic similarity to other L. kefiri strains, with a maximum average nucleotide identity (ANI) of 99.94%.

No virulence factors were detected in either strain. Both genomes contained annotations associated with antimicrobial resistance, including vanZ-family genes and determinants related to tetracycline resistance and multidrug efflux. Phenotypic testing confirmed vancomycin resistance in CIDCA 8348. No mobile genetic elements or plasmids were detected in CIDCA 83111.

Overall, both strains showed genomic features potentially associated with probiotic properties, with CIDCA 83111 displaying a larger repertoire of candidate genes related to host adhesion and bacteriocin production. Further phenotypic and functional studies are needed to validate these findings and support a comprehensive safety assessment.

DRAFT - UNDER REVIEW

POSTER #142 **TRACK:** Proteomics and Protein Structure and Function

Bioinformatic strategies to detect specific anti-enzyme antibodies of the primary immune response

Source: 05_posters/P142.tex · Excel row 113

Cauerhff A[1]

[1] Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales (UBA-CONICET)

In recent decades, despite its complexity, the IgM format has been considered for therapeutic use in clinical trials and has caught the interest of biopharmaceutical companies. For that reason, our goals include research on the primary humoral immune response towards various antigens (Ags), especially enzymes. We obtained 15 IgM isotype monoclonal antibodies (mAbs) through two immunization protocols using chicken egg lysozyme (HEL) as a model Ag. Our study includes the variable regions VH and VL sequencing, affinity measurement, and the characterization of recognized epitopes (Ep) by these anti-HEL IgM. For epitope mapping, in vitro experiments such as ELISA were carried out, and several bioinformatics strategies as homology modeling and docking experiments. For this purpose, we used online servers such as AlphaFold3 for homology modeling and docking, and AbBuilder 2 to model the Fv fragments of the IgM under study. After validating the resulting structures with the Molprobity server, we refined the AbBuilder 2 structures through energy minimization using the Yasara server, yielding notable improvements, particularly in reducing atomic clashes. We docked the energy-minimized structures with high-resolution HEL structures using servers such as ClusPro and HADDOCK 2.4, yielding 20–29 structures from ClusPro and ten clusters of 40 structures from HADDOCK. Consequently, we needed an effective screening method to identify the most probable models that were also consistent with the ELISA results. In this regard, we aligned the protein sequences of the VH and VL regions of mouse anti-HEL IgM with those of anti-HEL IgG from the same species whose structures had been determined by X-ray crystallography. This method allowed the calculation of the theoretical paratope for each specific IgM, while structural analysis determined the theoretical epitope (TE). To detect IgM specific to the active site of HEL (AS), the same approach was carried out using anti-HEL VHHs from the Camelidae family, whose CDRs interact with the HEL AS. By combining all the strategies mentioned in this summary, it was possible to identify, with greater confidence, at least two IgM likely to recognize the HEL AS.

DRAFT – UNDER REVIEW

POSTER #143 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Identification of Bacterial Species of the Genus Leptospira and Preliminary Pathogenicity Assessment Based on Sequencing Data from Environmental Samples Collected from the Salado, Coronda, and Paraná Rivers and Associated Lagoons in the Santa Fe Metropolitan Area

Source: 05_posters/P143.tex · Excel row 114

Mohamed G[1], Cabrera JM[2], Margenet L[3], Carletti J[4], Previtali MA[4,5], Rueda EC[4,5]

[1] Facultad de Tecnologías e Innovación para el Desarrollo, Universidad Nacional de Rafaela, Rafaela, Santa Fe, Argentina. [2] Instituto de Investigación y Desarrollo en Bioingeniería y Bioinformática, Universidad Nacional de Entre Ríos / Consejo Nacional de Investigaciones Científicas y Técnicas (UNER-CONICET), Oro Verde, Entre Ríos, Argentina. [3] Instituto Nacional de Enfermedades Respiratorias "Dr. Emilio Coni", Santa Fe, Argentina.
[4] Departamento de Ciencias Naturales, Facultad de Humanidades y Ciencias, Universidad Nacional del Litoral
[5] CONICET, CCT-Santa Fe, Santa Fe, Argentina.

Leptospirosis is a zoonotic disease with a significant impact on public health. Certain environmental conditions favor bacterial survival, including warm, humid weather with heavy rainfall, particularly in areas prone to flooding. The city of Santa Fe presents hydrological conditions that favor the persistence and dispersal of Leptospira spp.; however, knowledge of the diversity of species present in its water bodies and their pathogenic potential remains limited. The aim of this study was to identify and characterize species of the genus Leptospira isolated from cultures obtained from freshwater environmental samples collected from the Salado, Coronda, and Paraná rivers and associated lagoons, using bioinformatics tools, and to preliminarily assess their pathogenic potential.

A total of 21 sequencing files were analyzed. The 16S rRNA marker, sequenced using Illumina technology, was used for taxonomic characterization. The workflow included quality control, trimming, sequence assembly, identification using BLASTn, and confirmation through phylogenetic analysis.

The 16S rRNA analysis enabled the identification of sequences predominantly belonging to the genus Leptospira and revealed a predominance of saprophytic lineages. However, the high conservation of this marker hindered discrimination between closely related species. The phylogenetic relationships inferred from 16S rRNA corroborated the assignment of the analyzed sequences to the saprophytic clade.

These results provide evidence of the diversity of Leptospira present in the aquatic environments studied. The bioinformatics approach generated relevant information on the environmental diversity of this genus and provides a foundation for future epidemiological surveillance and environmental monitoring studies in the region.

DRAFT - UNDER REVIEW

POSTER #144 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Contributions to the study of the genetic basis of the dehiscence trait in the forage grass Panicum Coloratum

Source: 05_posters/P144.tex · Excel row 115

Cravero L[1], Irazoqui JM[1,2], Santiago G[1,2], Amadio A[1,2], Tomas MA[1,2], Diaz M[3], Eberhardt MF[1,2]

[1] Instituto de investigación de la Cadena Láctea (IDICAL, INTA-CONICET) [2] Universidad Nacional de Rafaela (UNRaf)
[3] Centro de Recursos Naturales Renovables de la Zona semiárida (CERZOS, CONCIET-UNS)

Seed shattering is an adaptive mechanism that promotes progeny dispersal in wild species; however, it represents a major constraint on harvest efficiency in agricultural systems. Although cereal domestication largely attenuated this trait through the selection of mutations in genes regulating abscission zone formation, it remains a critical bottleneck in many forage grasses. Such is the case for Panicum coloratum L., a C4 perennial grass adapted to semi-arid environments, whose predominantly outcrossing nature and tetraploid background hinder the fixation of seed-retention traits via conventional breeding. To elucidate the genetic architecture underlying this trait, candidate genes previously characterized in grasses (SvLes1 and Sh1) were retrieved, and their homologs were identified across reference genomes of the genus Panicum. Specific primers were designed to amplify these loci from two contrasting P. coloratum populations: SR (TEM-SR1), displaying 26% seed retention at 42 days post-anthesis, and EM (El Mirador), a shattering-susceptible population with only 14% retention over the same timeframe. The resulting amplicons were sequenced using Oxford Nanopore long-read technology; reads were quality-filtered and mapped against the P. miliaceum reference genome using the minimap2 aligner. This framework enabled the mapping of conserved regions, as well as the identification of sequence variants, haplotype structures, and distinct mutational profiles. These findings provide foundational genomic resources for future breeding programs aimed at developing cultivars with improved seed retention.

DRAFT – UNDER REVIEW

POSTER #145 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

in vitro caracterization of the activity of the CASK protein kinase domain

Source: 05_posters/P145.tex · Excel row 116

Vaca Pereira E[3], Espada R[1,2], Román E[3,4]

[1] Universidad de Buenos Aires, Facultad de ciencias exactas y naturales, Departamento de Física [2] Instituto de fisiología, Biología Molecular y Neurociencias, CONICET [3] Departamento de química biológica, FCEN, UBA [4] Instituto de química biológica de FCEN, CONICET

CASK syndrome is a rare genetic disorder, with only around 400 reported cases. It is caused by mutations in the calcium/calmodulin-dependent serine-kinase protein (CASK) gene, which plays an important role in brain development and function. The CASK protein contains a kinase domain whose function can be altered by point mutations. To date, the molecular mechanisms underlying CASK function, as well as the effects of different syndrome-associated mutations on its activity, have not been systematically characterized.

The objective of this project is to investigate the effects of mutations on the activity of the CASK kinase domain, with the aim of establishing a relationship between protein functionality and its stability and dynamic properties. To this end, an enzymatic assay will be developed to evaluate the activity of the kinase domain and, subsequently, that of its mutants, allowing the effects of specific mutations to be quantified.

As part of this project, the kinase domain was expressed in bacteria and subsequently purified. Its enzymatic activity will be assessed using an assay based on the quantification of ATP by nuclear magnetic resonance (NMR). The same assay will be used to evaluate approximately 20 pathogenic variants. For this analysis, missense mutations identified in patients diagnosed with CASK syndrome were selected from the ClinVar database.

The study of these variants will help determine whether the alterations identified in patients affect the activity of the kinase domain and will contribute to a better understanding of how mutations alter the molecular function of CASK.

DRAFT - UNDER REVIEW

POSTER #146 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

Characterization of the Resistome and Mobilome in Microbiota Metagenomes of Wild Penguins from the Strait of Magellan.

Source: 05_posters/P146.tex · Excel row 117

Piekar M[1], Aguilar A[1], Daniel Plaza[2], Mendiondo N[1], Vallejos D[1,3], Ochoa M[4], Souza V[4], Acuña P[5], Centrón D[1,3], **Quiroga MP**[1,3]

[1] Clúster de Bioinformática, Instituto de Investigaciones en Microbiología y Parasitología Médica, Facultad de Medicina, Universidad de Buenos Aires - Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET), Buenos Aires, Argentina. [2] Laboratorio de Investigaciones en Mecanismos de Resistencia a Antibióticos, Instituto de Investigaciones en Microbiología y Parasitología Médica, Facultad de Medicina, Universidad de Buenos Aires - Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET), Buenos Aires, Argentina. [3] University of Oxford, Oxford, United Kingdom. [4] Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, Coyoacán, Ciudad de México, México [5] Centro de Estudios del Cuaternario Fuego-Patagonia y Antárctica (CEQUA), Punta Arenas, Chile

The study of antimicrobial resistance (AMR) in wildlife contributes to understanding its environmental distribution through the One Health approach. Therefore, our objective was to characterize the mobilome and resistome in the microbiota metagenomes of king penguins from the Strait of Magellan. Thirteen metagenomes obtained via shotgun sequencing of feather samples from wild animals were analyzed. The reads were evaluated and processed using FastQC, MultiQC, and Trimmomatic. De novo assembly was performed using MEGAHIT, and the assembly was evaluated with MetaQUAST. Initial taxonomic profiles were determined based on the reads from each sample using MetaPhlAn. This study revealed the presence of Escherichia coli which could be from (One Health lineages) in all analyzed metagenomes, with varying relative abundances. Subsequently, sets of contigs corresponding to potential bacterial genomes (bins) were reconstructed; their taxonomic classification using the GTDB-Tk program enabled the identification and confirmation of the presence of E. coli. The presence of antibiotic resistance genes (ARGs) in the assembled sequences was determined using the RGI program, selecting matches with >60% identity and >70% coverage. Targeted searches for clinically relevant ARGs and integrons were also conducted using BLASTn. This study revealed the presence of sequences associated with resistance to various antibiotic families, including beta-lactams and tetracyclines, as well as efflux pump components in all samples. Additionally, intI1, intI2, and intI3 integron integrase genes were identified in different metagenomic samples. The mobilome analysis was performed using the programs Integron Finder, ISEScan, and PlasmidFinder on four sets of contigs defined by their length and the filters applied to matches with GRA. This analysis revealed a total of 1,281 clusters of attC sites lacking integron-integrases (CALINs), 38 complete integrons, and seven integrases with no detected gene cassettes. The detection of E. coli, AMR-associated sequences, and mobile genetic elements in the microbiota of wild penguins underscores the environmental spread of these determinants. These findings support including wildlife in AMR surveillance under a One Health framework and investigating potential links with environmental and human reservoirs.

DRAFT – UNDER REVIEW

POSTER #147 **TRACK:** Genomics, Transcriptomics, Epigenomics, Metagenomics

GraphPert: Network propagation-aware drug repurposing from transcriptional

Source: 05_posters/P147.tex · Excel row 118

Conti GA[1], Espinoza AA[1], Gomez Chavez JL[1], Perez ER[1], Angelina EL[1]

[1] Lab. Estructura Molecular y Propiedades, IQUIBA-NEA, Universidad Nacional del Nordeste, CONICET

In disease states, the function of specific proteins can be altered, and these anomalies can propagate through the protein–protein interaction (PPI) network. Similarly, the initial effect of a drug may involve a limited number of proteins but propagate progressively through the PPI via physical interactions, reaching other proteins and functional modules relevant to the disease. This implies that a drug does not need to act on the same proteins altered in the disease to be effective: it is enough for it to reach them indirectly, through the propagation of its effect in the PPI. GraphPert employs a graph autoencoder (GAE) pretrained on the human PPI to model this propagation, generating latent representations that are leveraged to screen drugs according to their propagated effect in the network.

The central innovation of the method consists of translating gene expression changes into a modification of the graph without retraining the model. From expression signatures of the Connectivity Map L1000 platform, perturbations are constructed for each condition via an operator that rescales node connectivity according to their expression z-scores. In the BRAF V600E melanoma case study in the A375 line, candidates are ranked by cosine similarity between their latent representations and those corresponding to BRAF silencing, used as a therapeutic reference.

In a retrospective screening with known MAPK pathway inhibitors, both GraphPert and raw transcriptional signature similarity achieved an AUC above 0.80, although raw signatures showed slightly superior overall performance. However, analysis of prioritized compounds showed a relevant difference: while raw signatures mainly recovered compounds whose targets are near the MAPK pathway, GraphPert additionally identified mechanistically diverse compounds, such as HDAC, proteasome, and PARP inhibitors. Although their targets are farther from MAPK, these compounds converge with the reference in functional modules such as ribosomal biogenesis, RNA processing, and chromatin remodeling. Thus, propagating transcriptional signatures using a pretrained GAE complements screening based on direct expression similarity and broadens the identification of repurposing candidates toward alternative mechanisms of action.

DRAFT – UNDER REVIEW

POSTER #148

TRACK: Artificial Intelligence and Systems Biology

Hydra effect emerges from optimal control in structured populations

Source: 05_posters/P148.tex · Excel row 119

Maxwell J[1], Moreno RJ[2], Ventura AC[1], Sánchez de la Vega C[3]

[1] Instituto de Fisiología Biología Molecular y Celular Universidad de Buenos Aires, Departamento de Física
[2] Universidad Austral [3] Instituto de Cálculo Universidad de Buenos Aires

Counterintuitive responses are increasingly recognized as fundamental features of complex biological systems. Among them, the Hydra effect—where increased mortality leads

to higher population size or stability—challenges the conventional view that survival maximization is the primary driver of population success. Despite extensive empirical and theoretical evidence, a unifying mechanistic interpretation of this phenomenon remains elusive.

Here, we propose that the Hydra effect can be understood as the outcome of an underlying optimal control principle. We formulate population dynamics within a non-linear optimal control framework, in which time-dependent mortality acts as a control variable shaping population trajectories. Using Pontryagin's Maximum Principle, we analytically characterize the structure of optimal strategies that maximize population-level objectives over finite time horizons.

Our results reveal that optimal solutions naturally exhibit bang–bang control structures, consisting of abrupt switches between extreme regimes. Within this framework, an increase in mortality emerges as an optimal strategy that enhance long-term population performance, providing a rigorous theoretical basis for the Hydra effect.

DRAFT – UNDER REVIEW

POSTER #149 **TRACK:** Proteomics and Protein Structure and Function

Generating holo structures from apo conformations with AI structure-prediction models: pipeline design and a docking baseline on apo–holopairs

Source: 05_posters/P149.tex · Excel row 120

Pliauzer JC[1], Prieto JM[1], Martí MA[1]

[1] Laboratorio de Bioinformática, Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Background: Ligand binding often requires conformational changes that rigid-receptor docking cannot capture, so predicting holo (ligand-bound) states from apo (unbound) structures remains a central problem in structure-based drug design. Artificial intelligence (AI) structure-prediction models such as AlphaFold and Boltz sample conformational space efficiently, but they were not trained or evaluated for the apo-to-holo transition. We are building a modular pipeline to test whether AI-generated conformations recover holo-competent binding sites.

Methods: Phase 1 curates apo-holo pairs from the Protein Data Bank and superposes each apo structure onto its holo partner, aligning the ligand-binding chain by sequence. Phase 0 sets the baseline: the crystallographic ligand is docked with AutoDock4 into both structures, classically and biased toward the known site with AutoDock Bias. Phase 2 generates 10 unguided conformers per candidate system with Boltz-2 under multiple sequence alignment (MSA) subsampling, and classifies each by its root-mean-square deviation (RMSD) to the apo and holo structures. Phase 3 docks the ligand into the generated conformers and compares them against the Phase 0 baselines. Phase 4 biases the generative model toward holo-like conformations, testing steering strategies informed by docking, molecular dynamics (MD) and other structural priors.

Results (preliminary, Phase 0, 418 pairs): Docking reproduced the crystal pose (ligand RMSD < 2 Å) in 46.3% of holo structures but only 7.6% of apo structures (n = 410). Site-biased docking lowered the median apo ligand RMSD from 8.36 to 6.20 Å (245 systems improved vs 158 worsened; sign test p = 1.7e-5), but did not increase the fraction of correct apo poses and reduced correct holo poses to 37.8%. The holo-apo docking energy difference correlated with ligand size (r = -0.31), so we also report a size-corrected residual. Superposing apo onto holo before placing the docking box was essential: without it, the apo box fell on solvent in 60% of 537 systems. Phase 2 generation is under way, and results through Phase 4 are expected to be presented at the meeting.

Conclusions: Neither rigid nor site-biased docking closes the apo-holo gap, which defines a quantitative target (46.3% vs 7.6% correct poses) that AI-generated conformations must approach. The pipeline and its controls are in place to test whether unguided and biased generation produce holo-competent binding sites.

DRAFT – UNDER REVIEW



a2b2c.org.ar

a2b2c

bioinformatica_ar