Proceedings X-Meeting eXperience 2020

Editor: AB^{3}C

November 2020

Conference Program

1 — Organizing Committee

AB3C President : Ney Lemke (UNESP)

AB3C Vice President : Marcelo Brandão (Unicamp)

AB3C Secretaries :

- Fabrício Martins Lopes (UTFPR)
- Vasco Ariston de Carvalho Azevedo (UFMG)

AB3C Financial Department :

- Nicole Scherer (INCA)
- Arthur Gruber (USP)

Poster Session Organizers :

- Nicole Scherer (INCA)
- Arthur Gruber (USP)
- André Yoshiaki Kashiwabara (UTFPR)

Highlight Track Organizer :

- Arthur Gruber (USP)

Thesis and Dissertation Award :

- Fabrício Martins Lopes (UTFPR)
- Ney Lemke (Unesp)

2 — Introduction

The Brazilian Association of Bioinformatics and Computational Biology (AB3C) is a scientific society funded in July 12th 2004. Since its creation, AB3C has been responsible for the annual conference entitled "X-Meeting" which is the main Bioinformatics and Computation Biology event in Brazil. This year its 14th edition occurred online in a different format called X-meeting eXperience from November 9th to November 10th. All the posters in this edition were presented as short videos. That could be acessed from this proceeding.

In this edition the Board of Directors of AB3C acknowledged the relevant contributions of Professor Miguel Ortega (UFMG) to both the Association and for the Field of Bioinformatics at Brazil.

We also had some Special Interests Groups (SIGs):

- Escola Parananense de Bioinformática
- Bioinformatics Core Managers

and two schools:

- Python for Bioinformatics
- Train the trainer (with EMBL support)

Bioinformatics is now a strategic area for Brazil and all Latin America and, therefore, it is also strategic to the development of Science, Technology and Economy. The X-Meeting is a Brazilian event with international reach which has an average of 200 participants. The Conference is an opportunity for students, researchers and companies to interact and difuse knowledge. The AB3C has been a pioneer society in the field of Bioinformatics in Brazil and we have a history of ten past very productive meetings. This year we succesfully adapted most of the traditional sections to online versions.

3 — Database and Software Development

”

A database of viral taxon-specific profile HMMs for the detection and classification of viral sequences

Wendel Hime Lima Castro, Arthur Gruber, Liliane Santana Oliveira Kashiwabara

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ, UNIVERSIDADE DE SÃO PAULO

Abstract

Viruses are the most abundant and diverse biological entities on nature, characterized by high rates of replication and evolution. The highly variable nature of their genomes, composed of either single- or double-stranded RNA or DNA molecules, implies that different enzymes are required for replication. No single gene is shared by all viruses and, therefore, there are no universal markers for viral phylogenetic analysis and taxonomic classification. Thus, different viral markers must be established to study distinct taxonomic groups. Profile HMMs are statistical models that incorporate the diversity of a set of primary sequences and constitute a very sensitive approach for detecting known and emerging viruses. Publicly available resources of viral profile HMMs are based on orthologous clusters, which may include sequences from different taxonomic groups, implying that models can often detect a wide and unpredictable range of taxa. We have recently developed TABAJARA, a program for the rational design of profile HMMs. TABAJARA uses a multiple sequence alignment (MSA) as input and is able to find short blocks that are either (1) conserved across all sequences or (2) discriminative for two specific groups of sequences. The program then runs validation tests against the training set and automatically define cutoff scores customized for each developed model. Alternatively, full-length sequences are used for model construction and highly specific profile HMMs may be obtained when using the assigned cut-off scores. This approach has been successfully used to develop models for some taxonomically specific viral groups, including Microviridae phages and viruses of the genus Flavivirus. Given the utmost importance of developing bioinformatic resources and tools for novel virus detection and classification, we decided to extend our approach to all viral taxa represented with protein sequences on the NCBI's Identical Protein Groups (IPG) database. We implemented a Python pipeline to perform the following steps: (1) automatic sequence retrieval from IPG database, according to a pre-defined list of queries; (2) data organization and storage; (3) multiple sequence alignment for each query; (4) profile HMM construction; (5) model validation; and (6) report generation. We obtained 20.749 models, comprising prokaryotic and eukaryotic viruses. Relational searches can be performed on a web front end according to the taxonomic name or ID, protein name, type of model (short or full-length) and host (prokaryotic or eukaryotic), among other parameters. The selected models can then be downloaded. We expect this database will become an important resource for the scientific community.

Funding:

Link to Video:

,,

PLANT CO-EXPRESSION ANNOTATION RESOURCE 2.0: A WEB TOOL FOR THE ASSOCIATION OF PROTEINS OF UNKNOWN FUNCTION TO ABIOTIC STRESSES IN PLANTS

Maurício de Alvarenga Mudadu, Adhemar Zerlotini Neto, Marcos José Andrade Viana

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

The development of genetically modified (GM) crops includes the discovery of candidate genes through bioinformatics analysis, using genomic data, gene expression, among others. Proteins of unknown function (PUFs) are interesting targets for pipelines of GM crops due to the novelty associated and also to avoid copyright protections. One method to infer the possible function of PUFs is to relate them to factors of interest, such as abiotic stresses, using orthology and coexpression networks, applying the guilt by association approach. The objective of this work is the discovery of PUFs involved in responses to abiotic stresses in plants. For this, we analyzed and processed the genomic data of 67 plant species, including 5 important species tolerant to abiotic stresses. Diamond and InterproScan were used to discover PUFs in all of these species. RNA-seq data related to abiotic stress was downloaded from NCBI / GEO and used as inputs to the LSTrAP software to build coexpression networks and clusters whose members are most likely related to the molecular mechanisms associated with abiotic stress in plants. Ortholog groups were created with Orthofinder using all proteins as input. So, we searched for PUFs associated with these groups of ortholog and simultaneously with some gene coexpression cluster related to abiotic stress. With that, we stored in a database provided by the ax software (https://github.com/lmb-embrapa/machado) 2, 136, 336 genes and 2, 714, 161 mRNA, together with their translated proteins. We recovered 78, 416 PUFs with Diamond and Interproscan analyzes, created 91, 172 groups of orthologists and 1, 975 coexpression clusters. We developed a protocol to search for PUF annotations and retrieve PUFs that belong to an ortholog group that also contains proteins whose mRNA belongs to a coexpression cluster related to abiotic stresses. After running this protocol, we recovered 4, 673 PUFs. We conducted a literature search on the proteins that belong to the orthologs groups, for all the PUFs that belong to the species Pearl millet, Populos simonii, Oropethium thomaeum e Boea hygrometrica, all known to be tolerant to abiotic stress (517 PUFs). We found studies related to abiotic stresses, on average, for 67.5% of PUFs. A webserver https://www.machado.cnptia.embrapa.br/plantannot2 is freely available and provides indexed queries of PUFs possibly associated with abiotic stresses. We hope that Plantannot2 will be useful for researchers trying to obtain genes related to responses to abiotic stresses for the production of new GM crops tolerant to the risks of climate change.

Funding:

Link to Video:

,,,

ToxAnalyzer: A tool for Comparative Toxicogenomics Database (CTD)™ data analysis and visualization

Diego César Batista Mariano, Lucianna Helene Silva dos Santos, Carlos Alberto Tagliati, Daniel Ribeiro Rodrigues

UNIVERSIDADE FEDERAL DE MINAS GERAIS, IOC/Fiocruz

Abstract

Toxicology is a field of science that has undergone changes in recent years. Using modern sequencing technologies, it went from an exclusive animal-based science to a field of data-based decision making. Many databases store important data about toxicology experiments, which can be useful to understand the mechanisms concerning a chemical product toxicity. One of these databases is the "Comparative Toxicogenomics Database" (CTD). CTD provides information about diseases, genes, compounds and their role in toxicity. However, the volume of data from a simple search is often overwhelming to be manually and individually analyzed. The tools to analyze CTD's data require previous installation and the knowledge of writing and manipulating programming scripts. Hence, we developed ToxAnalyzer, a web application to help users evaluate CTD toxicogenomic data from a chosen compound. Our application withdraws data from CTD's servers and uses a set of Python scripts to process and display important information about any chemical compound found in CTD. Using the chemical name, a user can gather information about the publications that report the compound, the organisms that are involved, gene interactions, and other helpful plots. Our tool provides a helpful and quick overview of the complete compound data available in CTD to aid a study hypothesis. If users need more information about an interaction between a specific gene and a chemical compound, PubMed ID links are available to access the original publications. Therefore, ToxAnalyzer is a user-friendly tool, since it can be accessed from anywhere, with any device with internet access, requiring absolutely no programming knowledge. ToxAnalyzer is available on https://toxanalyzer.com/.

Funding:

Link to Video:

,,,

A machine-learning scoring function for protein-ligand molecular docking

Oscar Emilio Arrúa Arce, Andrej Aderhold, Adriano Velasque Werhli, Karina dos Santos Machado

Universidade Federal do Rio Grande - FURG, UNIVERSIDADE FEDERAL DO RIO GRANDE

Abstract

In the field of drug design, scoring functions are useful for predicting the binding affinity of protein-ligand complexes. Machine learning approaches are showing promising performance as a result of the increasing amount of data regarding biochemical and biophysical processes, obtained from previous experiments. In this work we propose a machine learning based scoring function for protein-ligand molecular docking. This scoring function was developed according to related works, where: from protein-ligand complexes (training set) were obtained features of proteins, ligands and interactions that are considered as attributes; machine learning methods are to use to train models, including feature selection techniques and hyperparameters optimization; and test sets that are used to evaluate the proposed scoring functions models. As training set, we combine the PDBbind 2016 refined and general sets, CSAR-NRC HiQ and Decoys CSAR-NRC HiQ. As attributes we considered AutoDock Vina score and geometrical, SFCscore, solvent-accessible surface area, DeltaVinaRF20, protein primary and secondary structure and Vina features. We also considered specific software to generate features as PaDEL Descriptor, NNScore 2.0 and RDKit. Random Forest and Gaussian Process were compared as machine learning methods, in addition to LASSO to calculate the weights of the attribute's importance and GridSearchCV as a technique to hyperparameters optimization. Thus, the proposed scoring function was evaluated using the CASF-2016 benchmark, based on Scoring, Ranking, Docking and Screening Power. As a result, for CASF-2016 evaluation, the proposed scoring function achieved good results, comparable to the best scoring functions. As Scoring Power, we obtained 0.81 that corresponds to the Pearson correlation coefficient between predicted affinities and experimental measured affinities. For Ranking Power, the proposed scoring function achieves a Spearman correlation coefficient of 0.66 between the ranks based on the predicted affinities values and the experimentally ones. For the Docking Power, the proposed scoring function obtained 86% success rate in identifying the top best-scored ligand binding pose below 2 root-mean-square deviation from the native pose (and 83.8% without native poses). Finally, for Forward Screening Power, the proposed scoring function has a got 26.5% success rate to identifying potential small-molecule ligands for a chosen target protein at the top 1% level (better than all the scoring functions compared in CASF-2016) while for Reverse Screening Power achieve a 18.5% success rate in identifying potential target proteins for a bioactive small-molecule compound at the top 1% level.

Funding: CAPES, CNPq

Link to Video:

,

Machine Learning models applied to the subtypes classification of Acute Myeloid Leukemia and Myelodysplastic Syndrome

Marcelo Mendes Brandão, Fábio Malta de Sá Patroni

UNIVERSIDADE ESTADUAL DE CAMPINAS

Abstract

Myeloid Malignancies are clonal diseases of hematopoietic or progenitor stem cells. Among the five main types are Acute Myeloid Leukemia (AML) and Myelodysplastic Syndrome (MDS). Leukemias are one of the most common cancers in Brazil. In 2018, they accounted for approximately 3% of new cancer cases. MDS are considered nowadays the most common class of acquired medullary failure syndromes in adults. They are also considered the most prevalent hematologic malignancies. There is a risk of transforming AML in approximately 1 out of 3 patients. Acute Myeloid Leukemia is a potentially fatal disease, common in children and adults, that can lead to death if left untreated. The MDS occurs predominantly in older male patients, with an average age of diagnosis of approximately 70 years. Here we are applying an automated method for classification of AML and MDS into their subtypes using a machine learning politomic classifier. Two models will be created, one for each disease, which are trained with various clinical tests to predict accurate classification results. The models will be written in Python, using two popular frameworks: Scikit-Learn and Tensorflow. Input data for training and testing the models will come from three major public databases: the GDC Data Portal, the GDC Legacy Archive, and the NCBI GEO. The validation of the models will be performed with data from patients of the UNICAMP Blood Center. The classification process will be taken one step further in the research field. The approach proposed here can be used as a tool to help pathologists.

Funding: This work was supported by grants from CNPq (870370/1997-9).

Link to Video:

4 — Education and Outreach

,,,,,,,,

LEAGUE OF BRAZILIAN BIOINFORMATICS: CHALLENGES TO PROMOTE SCIENTIFIC TRAINING

Lucas Miguel de Carvalho, Raquel Riyuzo de Almeida Franco, Elvira C Alves Horacio, Maira Rodrigues de Camargo Neves, Nilson Coimbra, Mayla Abrahim Costa, Flavia Figueira Aburjaile, Sheila Tiemi Nagamatsu, Neli Jose da Fonseca Junior

UNIVERSIDADE ESTADUAL DE CAMPINAS, FIOCRUZ - IOC, RSG BRAZIL, EMBL-EBI, UNIVERSIDADE FEDERAL DE MINAS GERAIS, CTBE/CNPEM, IOC/Fiocruz, UNIVERSIDADE DE SÃO PAULO

Abstract

The scientific training to become a bioinformatician includes multidisciplinary abilities such as biology, mathematics, statistics, biochemistry, and computer science. Besides, it requires the development of soft-skills such as teamwork, scientific communication, resilience, critical thinking, and research. In order to improve and promote the ongoing training of the Brazilian bioinformatics community, we organize a national competition, with the main goal to develop human resources and abilities in computational biology at the international level. The competition framework was designed in three phases, and the competitors had to organize themselves in groups (until 2-3 participants). The first phase was a one-day challenge with 60 multiple choice questions of Biology, Computer Science, and Bioinformatics. In the second phase, they were challenged to solve 5five programming questions in 5 days. While the third phase included the development of an original project evaluated during the 15th X-meeting. The first edition of the Brazilian League of Bioinformatics (LBB) counted with 168 competitors and 59 groups. We reached 76% out of 26 Brazilian States. The participants were majority men (67.2%) from the southeast of Brazil (around 55%). Also, they were distributed into undergraduate students (14.4%), graduate students (12.6% master and 16.8%, Ph.D.), and professionals in the field. The first phase selected 46 teams to proceed in the competition, while the second phase selected the three top-performing teams. The Brazilian League of Bioinformatics included mostly multi-diverse groups, however, the finals were composed only of men. During the competition we were able to stimulate teamwork in the main areas of bioinformatics, with the engagement of all research-level competitors. Furthermore, we identified opportunities to deliver and offer better training to the community and we intend to apply the acquired experience in the second edition of the LBB, which will occur in 2021.

Funding:

Link to Video:

5 — Genes and Genomics

,,,

A survey of bacterial and archaeal genomes reveals novel casposon elements and hosts

Liliane Santana Oliveira Kashiwabara, Bas Dutilh, Arthur Gruber, Táriky Meirelles Rocha

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ, UNIVERSIDADE DE SÃO PAULO

Abstract

Casposons comprise a superfamily of self-synthesizing DNA transposons containing cas1, a gene for endonuclease Cas1, a key enzyme of the adaptive immunity CRISPR-Cas system. All casposons also harbor polb, a type B DNA polymerase gene. Some other genes are also found but are not shared by all elements. Casposons are typically flanked by terminal inverted repeats (TIRs) and target site duplications (TSDs) and these features can be used to predict their boundaries within the host's genome. Four casposon families have been characterized so far. Family 1 is composed of elements of archaeal hosts, presenting a protein-primed PolB, an enzyme closely related to polymerases of archaeal viruses. Elements of the other families were found in Archaea (Families 2 and 4) and Bacteria (Family 3). Phylogenetic analyses suggest that casposons originated CRISPR-Cas systems. To better understand the evolutive role of casposons in the emergence of adaptive immunity in prokaryotes, we decided to perform a survey on PATRIC, a public repository of assembled genomes. First, we used TABAJARA, a program developed by our group, to construct sets of profile HMMs derived from Cas1 and PolB sequences. All models were validated against bona fide datasets of Cas1 and PolB sequences derived from casposons or other sources. Profile HMMs specific to casposon elements were used together with e-Finder, a generic tool that use the models to detect and extract multigene elements from assembled genomes. All elements were annotated with EGene2 and the functional annotation was curated on Artemis. TIRs and TSDs were detected with UGENE program on regions flanking the predicted casposon elements. The survey revealed 136 elements, 90 in archaea and 46 in bacteria. From this set, 44 elements did not present the flanking repeats, with 39 of them showing truncated sequences in at least one end. In total, 92 full-length casposons were found with the expected flanking repeats. A phylogenetic analysis of the elements confirmed their monophyletic character in regard to CRISPRs. Based on the phylogenies of both Cas1 and PolB sequences, we found evidence that some casposons of Family 2 may in fact constitute a novel family. More detailed genome structure information and higher taxa sampling with be necessary to confirm this result. Another interesting finding was the presence of a full Type II CRISPR embedded within casposons of Hyphomonadaceae bacteria. Finally, new hosts were also identified, expanding the current knowledge on the occurrence of casposon elements.

Funding:

Link to Video:

,,,,,

GenTreat: Computational pipeline to perform automated hybrid assembly

Gislenne da Silva Moia, Mônica Silva de Oliveira, Pablo Henrique Caracciolo Gomes de Sá, Jorianne Thyeska Castro Alves, Adonney Allan de Oliveira Veras, Victória Cardoso dos Santos

UNIVERSIDADE FEDERAL DO PARÁ, UNIVERSIDADE FEDERAL DO PARÁ

Abstract

Since the development of the NGS (Next-Generation Sequencing) technologies, the sequencing process of genomes has become faster, with low cost and data generation with high accuracy. To handle a large amount of data and perform different analyzes, for example, assembly, many software have been developed. Among the approaches used to assemble genomes, there is the de novo assembly that performs the task without using a reference organism. Over the years, numerous tools to perform this task were developed, for example, Velvet, Mira, Spades, Abyss, Allpaths. Despite the accuracy of these tools, it is still necessary to perform several assembly rounds, aiming for the best result, it is also possible to use a hybrid strategy where results from different assemblers are used in a joining process to improve the final result. GenTreat is a computational pipeline, with a friendly graphical interface, that performs automated hybrid assembly of prokaryotic genomes, where the assembly result obtained using SPAdes and MegaHit softwares are used as inputs for CISA contigs integrator, the final result is ordered with the Ragoo software and submitted to assessment through of QUAST software, the pipeline accepts as inputs single-end and paired-end reads. To demonstrate the pipeline efficiency, it was executed assemblies using the softwares: Spades, Megahit, Velvet, and the GenTreat, the results showed that the assemblies obtained through of the pipeline are less fragmented, show greater genomic features from analyzing with QUAST, N50 with more value than the others, contigs with lengths greater than 200 base pairs and with the number of bases closest to the expected size of the genome. Therefore, it turns out that the pipeline is a viable alternative to perform hybrid assemblies, in addition to being a tool that accomplishes an automated task without the need for the user to make use of complex and extensive command lines.

Funding:

Link to Video:

,,,,,,,,

COMPUTATIONAL PIPELINE FOR THE IDENTIFICATION OF HLA SOMATIC MUTATIONS: CAN SNV/INDELS EXPLAIN IMMUNOTHERAPY FAILURE IN NON-MUSCLE INVASIVE BLADDER CANCER?

Giulia W. Frigugliett, Diogo A Bastos, Vitor Rezende da Costa Aguiar, Fabiana Bettoni, Diogo Meyer, Erick da Cruz Castelli, Anamaria A. Camargo, Cibele Masotti, Ramon Torreglosa do Carmo

Instituto Sírio-Libanês de Ensino e Pesquisa, HSL Hospital Sírio libanês

Abstract

The treatment for non-muscle invasive bladder cancer (NMIBC) is the complete transurethral resection of the tumor, followed by adjuvant immunotherapy with intravesical BCG (attenuated Bacillus Calmette-Guérin) instillations in high-risk cases of recurrence or progression. BCG immunotherapy significantly decreases the risk of disease recurrence through the stimulation of anti-tumoral immune response. A fraction of patients does not respond to BCG: 30-40% relapse and 10-25% progress to muscle-invasive forms. There are no predictive biomarkers of BCG response in clinical practice, but, as we learn from immunotherapy with immune checkpoint inhibitors, a number of molecular biomarkers have been related to the antigen presentation mechanism. Human Leukocyte Antigen (HLA) class-I molecules present neoantigens at the tumor cell surface and anti-tumoral response may occur. Therefore, HLA somatic mutations (SNVs/INDELs), large deletions or transcriptional silencing may also occur as tumor immune evasion mechanisms. We hypothesized that somatic mutations in HLA impair the tumor antigen presentation and the BCG-promoted anti-tumoral cytotoxic immune response. We aim at quantifying the frequency of this immune evasion mechanism occurring in the context of BCG resistance using NMIBC tumor-only exome sequencing data. Due to high polymorphism, the detection of somatic INDELs and SNVs in HLA genes is a challenge, as unmapped or low quality read mapping interfere in variant calling. We approached this issue through a local pipeline with a multi-referenced genomic alignment (using not a single reference genome, but all HLA alleles described in public databases and Brazilian population allelic frequencies) to evaluate 6 Class-I genes (HLA-A, HLA-B, HLA-C, HLA-E, HLA-F, and HLA-G) across 34 primary tumors of BCG-treated patients (16 responsive and 18 unresponsive). We identified 19 mutations (14 SNVs and 5 INDELs) in 13 patients, mostly in HLA-B and HLA-F. Exons 2 and 3 had 52% of the detected mutations (at the antigen-presentation pocket), and exon 4 had 37%, a domain of HLA-TCR recognition. Tumors from BCG-unresponsive individuals (n=6) did not present significantly more somatic mutations as compared to BCG-responsive (n=5; Fisher's exact test p=1). Also, we did not observe a significant correlation between HLA mutation status and relapse-free survival (Log-rank test p=0.94) or tumor mutational burden status (Fisher's exact test p=0.71), an expected result due to our unpowered sample. This is the first evaluation of HLA somatic variation in the context of BCG response, and we are refining our pipeline in empowered public genomic datasets.

””

Complexity Analysis of Algorithms: A case study about Bioinformatics Tools

Luan Ribeiro Siqueira, Victória Cardoso dos Santos, Mônica Silva de Oliveira, Adonney Allan de Oliveira Veras, Gislenne da Silva Moia

UNIVERSIDADE FEDERAL DO PARÁ

Abstract

The diverse analysis performed by omics sciences, driven by the reduction of costs in the DNA sequencing and reduction of the total time to carry out this process, resulted in an exponential increase in the deposit of all this information in public databases, for example, the National Center for Biotechnology Information - NCBI. The volume of data produced by these sequencing platforms demanded the development of algorithms capable of performing the most varied analysis, such as remotion of redundancy in raw reads from the sequencing process. However, it is worth mentioning that the existence of these various tools which performs this task with proven accuracy through their scientific publications, they do not analyze criteria related to the algorithmic complexity involved in their development. Therefore, this work demonstrated an analysis of algorithmic complexity, through empirical analysis already described in the literature, this analysis was performed with sixteen raw reads datasets with sizes ranging from 900 thousand to 12 million, they were obtained in the NCBI database in the Sequence Read Archive (SRA) format, they were converted to the FASTQ standard through the fastq-dump tool, the selected tools were: MarDRe, NGSReadsTreatment, ParDRe, FastUniq, and BioSeqZip. The analysis was performed on the R statistic platform, by using the GuessCompx package using the processing time of all datasets required by each tool as input, the models created were submitted to the glm adjustment function, in order to identify the function that indicates the complexity observed in each model. To this end, seven Big-O notations were observed:O(n), O(log(n)), O(n^2), O(n^3), O(1), O(n-log-n) and O(2^n). With the analysis of the results plotted graphically, it can be concluded that the NGSReadsTreatment tool obtained the least complexity in the processing of the datasets used in this analysis, presenting a linear complexity behavior, which leads us to infer that for datasets with high volume, this tool shows an interesting alternative to performing data processing.

Funding:
Link to Video:

,,,

DNA methylation and functional annotation of CpG-SNPs in the TIMP3 locus associated with TIMP3 levels and preeclampsia

Daniela Alves Pereira, Natalia Duarte Linhares, Marcelo Rizzatti Luizon, Juliana de Oliveira Cruz

Ufmg- Universidade Federal de Minas Gerais, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

Preeclampsia (PE) is defined by hypertension after 20 weeks of gestation. Matrix metalloproteinases are endopeptidases involved in the extracellular matrix remodeling and trophoblast invasion during placentation, and their activities are regulated by endogenous tissue inhibitors of metalloproteinases (TIMPs). DNA methylation (DNAm) is an essential epigenetic mark and changes in DNAm are involved in the pathogenesis of PE. The promoter of TIMP3 gene was found to be hypomethylated, and the TIMP3 gene upregulated in placentas from PE, which confirmed its relevance in the etiology of PE. Notably, the correlation of SNPs located at CpG sites (CpG-SNPs) with allele-specific methylation in PE is unknown. In this study, we searched for CpG islands in the TIMP3 locus and focused on CpG-SNPs putatively associated with differential DNAm within active regulatory elements associated with TIMP3 levels and/or with PE. We selected SNPs in the TIMP3 locus in association studies with PE and in the Catalog of Published GWAS. The CpG islands were predicted using software MethPrimer and the pairwise linkage disequilibrium (LD) among the selected SNPs was calculated using Haploview. In silico characterization of DNAm and regulatory elements nearby TIMP3 were performed in the UCSC Genome Browser. We selected 13 SNPs in the TIMP3 locus. This region includes seven CpG islands, two in the TIMP3 promoter region. The TIMP3 locus contain 12 regulatory elements. Notably, a 5.8kb promoter/enhancer segment targets TIMP3 and the other two genes, and overlaps with the CpG-SNPs rs140495494, rs58394026, and rs5749511, and the SNP rs9619311. ChIP-seq data for the histone mark H3K4me3 confirmed the presence of a potential promoter element in this region. Moreover, 11 out of the 13 selected SNPs are considered as CpG-SNPs. Nine out of the 11 were CpG-SNPs in the presence of the major allele, and two in the presence of the minor allele. In both cases, the CpG sites are lost in the presence of the alternative allele. We found higher pairwise LD between the SNPs rs2097326 and rs2413151, and the rs9619311 and rs135025 in the European population. The in silico analysis of DNAm showed that the TIMP3 promoter region was methylated. Specifically, the cg15004938 and cg07972276 sites interrogate the CpG-SNPs rs140495494 and rs5749511, respectively, which are methylated. Our findings suggest that CpG-SNPs may affect the epigenetic control of TIMP3 expression, and may help to guide functional studies to elucidate the clinical role of TIMP3 expression and TIMP3 levels in PE and other diseases, including cancer.

Funding:

Link to Video:

””

Resistome profile of Acinetobacter baumannii

Roselane Gonçalves dos Santos, Flavia Figueira Aburjaile, Vasco A de C Azevedo, Francielly Rodrigues da Costa, Diego Lucas Neres Rodrigues

INAPG- França, IOC/Fiocruz, UFMG, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

Acinetobacter baumannii is an important nosocomial pathogen. This Gram-negative bacterium causes several diseases, such as pneumonia, bacteremia, meningitis, osteomyelitis, and erysipelas. It is also a pathogen highly known for its resistance to antimicrobials and its ability to survive in intensive care units assisted ventilation devices. Its characteristics make this pathogen an essential model for studies of resistance to antimicrobials. Besides, in 2017 the World Health Organization announced that A. baumannii was a priority due to their exacerbated resistance to antimicrobials, mainly of the class of β-lactams. In this context, this work objective is to present the predicted resistance gene repertoire of A. baumannii. For this purpose, the public genomes of 206 strains of this species were selected and evaluated by Comprehensive Antibiotic Resistance, Antibacterial Biocide and Metal Resistance genes databases. These data are curated according to classes of antimicrobials. As main results, we obtained a robust resistome composed of 131 genes related to the enzymatic inactivation of the antimicrobial compound and 26 genes encoding putative efflux pumps. We emphasize that the highlighted genes adeK, adeJ, adeI, adeF, adeG, adeL, adeN, abeM, and ampC were identified in all A. baumannii strains. The drug resistance in hospital environments is associated with AmpC β-lactamases in this pathogen, requiring intensive monitoring. On average, each strain showed 26 resistance genes, except for the SDF strain, which presented 12 genes, and the AYE and 2008S11-069 strains that have 38 resistance genes. In conclusion, this pathogen can be used as a good model of bacterial resistance for directing future studies aimed at therapeutic targets.

Funding:

Link to Video:

””

Analysis of UVA induced mutagenesis in translesion synthesis-deficient human cells

Nathalia Quintero Ruiz, Natália Cestari Moreno, Tiago Antonio de Souza, Carlos Frederico Martins Menck, Camila Corradi

UNIVERSIDADE DE SÃO PAULO

Abstract

DNA is constantly subject to endogenous and exogenous agents that cause numerous types of lesions, interfering with its transcription and replication. UVA radiation, responsible for more than 95% of the solar ultraviolet radiation that reaches the Earth's surface, may cause direct and indirect DNA damage, as pyrimidine cyclobutane dimers (CPDs) and nitrogenous base oxidation. Cells have several mechanisms capable of correcting these problems, such as Nucleotide Excision Repair, in addition to pathways that bypasses replication blockage caused by CPDs and oxidized bases, as Translesion Synthesis (TLS). Deficiencies in POLH/XPV gene, which codes for an important protein that acts in TLS, DNA polymerase ? (pol eta), culminates in a rare, autosomal recessive syndrome, Xeroderma Pigmentosum variant (XP-V). Pol eta-deficiency causes an increased frequency of skin cancer in XP-V patients because of their reduced ability to replicate sunlight-induced DNA damage. Decrease of pyrimidine dimers repair induced by UVA light indicates that this process may have a secondary and highly relevant effect on mutagenesis. Furthermore, XP-V cells demonstrated a higher mutational occurrence even in the absence of sunlight exposure, indicating that endogenous oxidative stress increase may be related to internal tumors in those patients. Hence, we analyzed the whole exome sequence of XP-V fibroblasts irradiated with UVA light, treated or not with the antioxidant N-acetylcysteine (NAC). Preliminary results indicate that NAC promoted partial protection of XP-V cells after UVA irradiation, considering we observed an important decrease in total number of exclusive point mutations in exons and splicing regions. NAC treatment also reduced C>A transversions, C>T transitions and CC>TT tandem mutation, indicating that it promoted a protective effect in oxidatively induced DNA damage and it reduced mutations caused by UVA irradiation at pyrimidine dimers, suggesting an important role of oxidative stress in cells that have a decreased capacity to remove DNA damage in the absence of pol eta.

Funding:

Link to Video:

,

Mesoscopic evaluation of DNA mismatches in PCR primers for SARS-CoV-2 detection

Gerald Weber, Pâmella Miranda de Moura

UNIVERSIDADE FEDERAL DE MINAS GERAIS, UFMG - Departamento de Física

Abstract

The pandemic of COVID-19 brought the necessity to a large testing in population. As a golden standard for molecular tests, techniques based in PCR (Polymerase Chain Reaction) has been used to detect SARS-CoV-2 virus, such as RT-PCR (reverse transcription PCR). For the amplification of viral target is used a set of primers to hybrid with. These oligos are single strands DNA sequences of 18-20 bases in length and are designed for sense and antissense direction. An important parameter to obtain a good performance of primers is the melting temperature which is related to the efficiency of primer to hybridise to the DNA molecule. Primers are designed to bind complementarily to DNA, however, it may be include single mismatches in the hybridisation. Mismatch presence can influence in the stabilisation of DNA molecule. In the case of PCR process, one or more mismatches can change the melting temperature of primers and may be interfere in the amplification of DNA molecule. Focusing in how mismatches may impact the detection of SARS-CoV-2 by PCR techniques, we collected and analysed 19 PCR primers sets to verify the behaviour of their melting temperatures in presence of up to three consecutive mismatches. We aligned the primers sets with 21665 genomes of SARS-CoV-2 and applied the Peyrard-Bishop mesoscopic model to obtain the melting temperatures for the resulted alignments. We compared the calculated melting temperatures for mismatch and perfect alignments. Furthermore, we collected genomes of SARS-CoV and other five coronaviruses to be our control group performancing the same workflow. In addition, we collected some data that can contribute to an optimization of primers sets for PCR diagnostic method for SARS-CoV-2. Our results indicate numerous instances where the mismatch presence does not destabilize the primers ensuring their detection capacity.

Funding:

Link to Video:

”

A PUTATIVE ENHANCER REGION ACTIVATED BY METFORMIN OVERLAPS WITH SNPS ASSOCIATED WITH VISTAFIN/NAMPT LEVELS

Lídia Lana Ferreira Coura, Marcelo Rizzatti Luizon, Daniela Alves Pereira

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

Nicotinamide phosphoribosyltransferase (NAMPT) is an adipocytokine with a potential to be a predictive biomarker or a therapeutical target for several diseases, such as nonalcoholic fatty liver disease, type 2 diabetes mellitus, and obesity. Notably, NAMPT was shown to be activated by Metformin, which is the first-line therapy for type 2 diabetes, and is also used as a treatment for other diseases. The single nucleotide polymorphism (SNP) rs1319501 is located in the promoter region of NAMPT gene, and it was found to be associated with plasma NAMPT levels. Notably, the SNPs rs9770242 and rs61330082, which are located 1, 500bp upstream from the transcription start site of NAMPT gene, were in high linkage disequilibrium with the rs1319501 in the European (CEU) population. Moreover, rs61330082 was associated with visfatin/NAMPT levels and adverse cardiometabolic parameters in a cohort of severely obese children. However, whether these noncoding SNPs overlap with active regulatory elements, such as enhancers, is unknown. Therefore, we searched for metformin-responsive regulatory elements in the NAMPT locus, and linked SNPs within them which may be associated with NAMPT levels. First, we examined publicly available ChIP-seq data for active (H3K27ac) and silenced (H3K27me3) histone marks on human hepatocytes treated with metformin, GeneHancer to identify active regulatory elements (enhancers and promoters), and several cis-regulatory elements assignment tools from the Encyclopedia of DNA Elements (ENCODE) to identify enhancers around the NAMPT locus. Next, we performed the functional annotation of noncoding SNPs located in the NAMPT locus using the Genotype-Tissue Expression (GTEx) project data for SNPs linked to NAMPT expression. The SNPs rs1319501, rs9770242 and rs61330082 overlap with a metformin-responsive region enriched for the active histone mark H3K27ac upon metformin treatment, which is located nearby an enhancer element according to GeneHancer (GH07J106288). According to GTEx, the SNPs rs1319501, rs9770242 and rs61330082 are eQTLs for NAMPT expression in the heart tissue. These data support that noncoding variation within a metformin-activated enhancer may increase NAMPT gene expression. However, further studies are needed to reveal whether increased NAMPT expression may represent a beneficial effect. To understand the regulation of NAMPT expression is crucial to reveal its biological functions and the variations under physiological and pathophysiological contexts, which could help to define NAMPT as a biomarker of disease prognosis, a predictive or a pharmacogenetic biomarker. Our study highlights noncoding NAMPT SNPs for further functional studies, which could help to predict NAMPT levels in patients with type 2 diabetes mellitus treated with Metformin.

Funding:

Link to Video:

”

FUNCTIONAL ANNOTATION OF INTRONIC SNPs OF ARG2 GENE ASSOCIATED WITH FETAL HEMOGLOBIN LEVELS IN PATIENTS WITH SICKLE CELL ANEMIA TREATED WITH HYDROXYUREA

Rahyssa Rodrigues Sales, Marcelo Rizzatti Luizon, Bárbara Lisboa Nogueira

Universidade federal de minas gerais, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

Sickle cell anemia (SCA) a β-hemoglobin disorder. Fetal hemoglobin (HbF) ameliorates the clinical outcomes and severity that are associated to SCA. Hydroxyurea (HU) is the main drug used to treat SCA patients, which improves their clinical course by raising HbF levels. Notably, HU was suggested to act as a nitric oxide (NO) donor in SCA. Recently, HU was shown to modulate NO signalling pathway in red blood cells (RBC), RBC rheology and oxidative stress through its effects on HbF and possibly on NO bioavailability. However, the NO-related effects of HU on RBC physiology and NO signalling pathway are not fully known. While BCL11A and HBS1L-MYB are the major loci regulating HbF levels, other candidate genes were associated with significant changes in HbF levels in SCA patients treated with HU, including two intronic SNPs (rs10483801 and rs10483802) of ARG2 gene. Arginase 2 was described to play a role in the regulation of extra-urea cycle arginine metabolism, in down-regulation of NO synthesis, and also extrahepatic arginase functions to regulate L-arginine bioavailability to nitric oxide synthase (NOS). Therefore, we examined whether these intronic SNPs of ARG2 (rs10483801 and rs10483802) may be linked with the actual functional regulatory elements that may regulated ARG2 expression. Here, we performed the identification of cis-regulatory elements at ARG2 locus using several assignment tools, including The ENCyclopedia Of DNA Elements (ENCODE) ChIP-seq data for the active histone mark H3K27ac, the ENCODE registry ofcandidate cis-regulatory elements (cCREs) using SCREEN (https://screen.encodeproject.org/). Next, we performed the functional annotation of these intronic ARG2 SNPs using the Genotype-Tissue Expression (GTEx, www.gtexportal.org/home/) project and the RegulomeDB (https://regulomedb.org/). Notably, rs10483801 and rs10483802 SNPs are located 400 bp distant in the last intron of ARG2 and overlap with H3K27ac peaks for three ENCODE cell lines, namely K562, NHEK and NHLF. Moreover, they are linked to transcription factors and are located next to a region with proximal enhancer-like signature identified by the ENCODE registry of cCREs. These data support the presence of an enhancer element in the last intron of ARG2. Notably, in SCA hemolysis results in the release and activation of arginase, an enzyme that reciprocally regulates NO synthase activity and thus, NO production. Considering that more than half of the patients with SCA present endothelial dysfunction caused by a local decrease of NO bioavailability, the knowledge of genetic variants which may increase the NO levels constitutes an important alternative to reduce the complications of SCA.

Funding:

Link to Video:

”

The genomic signature of Fungi at the gene- and pathway-levels

Francisco Pereira Lobo, Alison Menezes, Aline Soares dos Reis

UNIVERSIDADE FEDERAL DE MINAS GERAIS, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

Fungi comprise a highly diverse eukaryotic taxon that plays important ecological and biotechnological roles. They can also be a threat due to their pathogenicity: fungal toxicity and drug resistance are key concerns in medicine and agriculture. Therefore, the molecular and functional characterization of fungi is both noteworthy and needed to achieve a better understanding of their modus operandi and to detect possible new targets for molecular intervention. KEGG database describes groups of homologous genes shared across genomes as Kegg Orthology groups (KO). We use the KEGG API to obtain all KO groups for each complete eukaryotic genome available, which were divided into fungi (F, 129 genomes) and non-fungi (NF, 406 genomes). To select KO groups enriched in F, we used the following strategy: 1) genome bootstrapping in both groups (100 bootstraps); 2) For each bootstrap, perform the Fisher's test, followed by FDR correction, to evaluate if a KO was more observed in F than in NF; 3) a KO is considered overrepresented in F if FDR-corrected q-values are < 0.05 in 95% of the bootstraps and absent in NF. From the set of 13962 KOs found in at least one genome, 495 (3.5%) are significantly enriched in Fungi. Among the 50 most enriched KOs, we found enzymes of fungal activity as decomposers and also of metabolic activities of commercial interest. Interestingly, we also found several enzymes already targeted by antifungals, indicating our strategy also detects known druggable proteins. We also found several interesting candidates for future research that play major roles in fungal biology, such as transcriptional regulators and components of cell wall processes. We detected enriched pathways in F by computing the ratio of enriched KOs in a pathway by the total count of KOs in the same pathway. Among the pathways with the highest ratio of KOs enriched in fungal genomes we found both fungi-specific processes and fungal-specific modules within conserved eukaryotic pathways, such as DNA repair, protein synthesis and fatty acid metabolism. Mapping of KOs into specific pathways will also allow us to search for chokepoints, defined as steps that either uniquely consumes a specific substrate or uniquely produces a specific product, therefore comprising interesting targets for systemic interventions in fungal metabolism. Together, our comparative genomics analysis provides a gene- and pathway-level signature of fungal biology.

Funding:

Link to Video:

,,,,,,

In silico strategies to identify IP3 receptors and their function in activation of the plasma membrane H(+)-ATPase in Saccharomyces cerevisiae cells

Sayonara Quaresma Alves Campos, Cleberson Marcelo dos Santos, Rodrigo Bentes Kato, Aureliano Claret da Cunha, ROGELIO LOPES BRANDAO, Izinara Rosse, Raquel D. Penido

UNIVERSIDADE FEDERAL DE MINAS GERAIS, UNIVERSIDADE FEDERAL DE OURO PRETO, UFOP- UNIVERSIDADE FEDERAL DE OURO PRETO, UFOP

Abstract

Saccharomyces cerevisiae is a fungal specie commonly known as yeast. It is widely used in several products such as breads, biofuels, enzymes production and mainly for alcoholic beverages. Moreover, it is a model organism, having a well characterized cellular organization, short generation time that facilitates its replication, a small genome with about 6000 genes and orthologous pathways of intracellular signaling to higher eukaryotes. The plasma membrane H(+)-ATPase enzyme is essential for the physiology of this fungi and for the fermentation process. It is known that in the addition of glucose, the H(+)-ATPase may be involved in the intracellular calcium signaling, but this pathway has not been elucidated yet. Evidence indicates that a relationship between Inositol 1, 4, 5-trisphosphate (IP3) and the calcium channel Yvc1p is mediated by a receptor, that are related to the enzyme activations and calcium signaling in yeast. However, these receptors have not been identified in S. cerevisiae. In this context, this study, aimed to identify IP3 receptors in S. cerevisiae using in silico strategies. We've searched for similarity between the sequences of IP3 receptors of other species already deposited in GenBank. Therefore, the gene and protein sequences of IP3 receptors of type 1, 2 and 3 from 729 species deposited were selected using the search terms Inositol 1, 4, 5-Trisphosphate receptor type 1, type 2, type 3 and IP3 receptor. The gene sequences were aligned against the genome of S. cerevisiae, strain S288C, using the BLASTn. Additionally, a multiple alignment of these sequences, at gene and protein level, was performed using the Muscle algorithm implemented in Molecular Evolutionary Genetics Analysis (MEGA) tool, to verify the presence of regions that were conserved throughout the evolutionary process in IP3 receptors. No similar sequences to the receptors of any of the other analyzed species was observed in the yeast genome. However, it was possible to identify from the multiple alignment, several conserved regions that in some of the individuals are conserved domains. It was also observed that in species of fungi in which IP3 receptors have already been described, the conserved patterns differ from the other species analyzed. The results indicate that the IP3 receptors in fungi species may be exclusive, this is the reason that lead us to believe that, if one confirm the presence of this receptor in S. cerevisiae, probably will be also with an exclusive sequence. Consequently, it is still necessary to define what is the standard conserved in the IP3 receptors in fungi species and if it is conserved in S. cerevisiae.

Funding:

Link to Video:

,,,

Characterization of bacteriocins in Xanthomonas citri

Juan Carlos Ariute, Flavia Figueira Aburjaile, Ana Maria Benko-Iseppon, Milena Gomes Cabral

IOC/Fiocruz, UNIVERSIDADE FEDERAL DE PERNAMBUCO

Abstract

The resistance against antimicrobial components shows up as a growing concern worldwide and makes treating infectious diseases spread in humans, animals, and plants impracticable. In crops, Xanthomonas citri is a Gram-negative bacterium that infects plants within the Citrus genus, responsible for developing citrus canker and causing considerable economic losses. However, this microorganism produces some secondary metabolites, such as bacteriocins, that appear as an alternative to elucidate the problem since they can inactivate or kill microorganism targets. Thereby, this study aims to identify and characterize potential bacteriocins produced by X. citri. We have evaluated in silico 78 complete genomes in GenBank/NCBI database. The RAST server automatically annotated these genomes. After that, the annotated sequences were submitted to BAGEL4 and antiSMASH to identify genomic regions containing potential bacteriocins. Results obtained from BAGEL4 showed 76 genomes presenting the zoocin A, which was characterized as a bacteriocin-like inhibitory peptide of class III that attaches with the peptidoglycan region of some bacterias causing its lysis. The only ones that haven't shown zoocin A were Xanthomonas citri pv. mangiferaeindicae XC01 that has revealed a rhodonadin and a microcin and Xanthomonas citri pv. phaseoli fuscans CFBP7767 that hasn't shown any area of interest. Both of them have a lasso structure, which gives them more stability. Furthermore, the results obtained from antiSMASH revealed 22 bacteriocins and 55 lasso peptides in those genomes, except for three: Xanthomonas citri pv. aurantifolii str. 1566, Xanthomonas citri pv. aurantifolii FDC 1561 and Xanthomonas citri pv. vignicola CFBP7113. Therefore, we suggest that X. citri can produce metabolites with antimicrobial activity that could be used for industrial applications and to characterize this bacteriocin repertoire in this species.

Funding:

Link to Video:

,,,,,,,,

TEN COMPLETE MITOCHONDRIAL GENOMES OF GYMNOCHARACINI (STETHAPRIONINAE, CHARACIFORMES)

Iuri Batista da Silva, Rubens Pasa, Fabiano Menegidio, Igor Henrique Rodrigues-Oliveira, Dinaíza Abadia Rocha-Reis, John Seymour (Pat) Heslop-Harrison, Trude Schwarzacher, Karine Frehner Kavalco, Matheus Lewi Cruz Bonaccorsi de Campos

UNIVERSIDADE FEDERAL DE MINAS GERAIS, UNIVERSIDADE DE MOGI DAS CRUZES

Abstract

Stethaprioninae is a subfamily of characiform fish that comprises small animals popularly known as tetras. Some species of the genus, such as Astyanax, share several common features that difficult their recognition, leading to efforts to identify diagnostic characteristics or molecular signatures for the group. In an attempt to contribute to these efforts, we are presenting eight new complete mitogenomes of species/cytotypes from Neotropical Ecozone belonging to the Astyanax and Psalidodon genus: A. aeneus, A. altiparanae, P. fasciatus (from two basins), A. lacustris, P. rivularis (two cytotypes) and P. rioparanaibano. Total genomic DNA was extracted from liver and heart samples of six species. The Whole Genome Sequencing from these species was performed in a Novaseq 6000. We assembled the mitogenomes from raw reads on Novoplasty v3.7 in a parallel cluster computer using the mitogenome of P. paranae from GenBank as seed. We annotated the obtained sequences on MitoAnnotator (MitoFish). In the Galaxy platform, we accessed the quality of raw reads (using FastQC) and filtered with Fastp tool. For broader comparisons, we also assembled the mitogenome of two species with raw reads available on European Nucleotide Archive: P. fasciatus from Upper Parana river basin and A. aeneus from Mexico. We perform comparative genomics analysis by BLAST comparison of all Astyanax/Psalidodon mitochondrial genomes against a reference (P. paranae) generated by Blast Ring Image Generator. Our results have shown that all mitogenomes content and gene order were identical, with 13 protein-coding genes (PCGs), 22 tRNA genes and two rRNA genes, following an expected order according to already described Characiformes mitogenomes. All PCGs and tRNAs are on the heavy chain, except the Nd6 gene and eight tRNAs. The length of mitochondrial sequence range from 16, 626bp in P. fasciatus to 16, 812bp in P. rivularis. The average length of D-loop was 1, 061bp. Deepening the knowledge about the D-loop, can play a fundamental role in understanding the evolutionary history of the Astyanax and Psalidodon genera. In this work, we observed that the size variation between different Astyanax/Psalidodon mitogenomes occurs mainly due to the extension of the D-loop. In conclusion, our methodology used in the reconstruction of the mitochondrial genome proved to be satisfactory and able to access the length of this type of genome, plus the composition and nature of the D-loop, solving possible gaps in previous methodologies.

Funding:

Link to Video:

,,,

Metagenomic analysis of the enzyme a-Galactosidase in two soil samples

Sthephanye Katherine Ferreira Gomes, Maricy Raquel Lindenbah Bonfá, Rodrigo Matheus Pereira, Jéssica Patricia De Sá Souza

FUNDAÇÃO UNIVERSIDADE FEDERAL DA GRANDE DOURADOS, UFGD - UNIVERSIDADE FEDERAL DA GRANDE DOURADOS

Abstract

Alpha-Galactosidase is a glycoside hydrolase enzyme, it is present in many plants, and its performance is in biotechnological applications and several industrial fields, such as the food industry. Metagenomics has the ability to access DNA from microbial populations in different environments, allowing the identification of non-cultivable microorganisms. This research aimed to analyze the occurrence of genes that encode the enzyme a-galactosidase, identify the microorganisms that produce thes enzyme, andanalyze the statistical differences in the occurrence of thes enzyme in two soil samples: Native Forest and agricultural management of Conventional Planting. The soil samples were collected and made available by the company Embrapa Agropecuária Oeste located in the city of Dourados - MS. DNA extraction was performed using the DNA SPIN KIT. DNA sequencing was performed using Illumina technology, assessing its quality by using the FastQC program, filtering the low-quality ones through the Prinseq-lite program. With the identification of ORFs obtained by the FragGeneScan program, sequence identification was performed using BLAST 2. Comparisons were made with a local database built from a-galactosidase enzymes obtained from the NCBI Identical Protein Group. The comparison of the microbial data of the samples was performed using the MEGAN 6 program and the statistical analysis were performed using the STAMP program. The locally built a-galactosidase enzyme database had in store a total of 162, 552 enzymes. The native forest soil sample carried a total of 46, 430 ORFs of which 11, 184 belonged to the enzyme a-galactosidase, the agricultural management soil sample of Conventional Planting had 71, 762 ORFs and 1, 427 of the enzyme a-galactosidase were obtained. The comparison of soil samples indicated that the phyla with the greatest a-galactosidase enzyme representativeness were Actinobacteria, Deinococcus-Thermus, Ascomycota, Firmicutes and Acidobacteria, and the genera were Thermus, Streptomyces, Talaromyces, Bacillus and unclassified Acidobacteria. Statistical analyzes show that the phylum Deinococcus-Thermus has a statistically significant difference occurring more in the soil of Conventional Planting, while in the phylum Acidobacteria it has a statistical difference occurring more in the Native Forest and in the genus Talaromyces, while the genus Thermus has a statistical difference occurring more at Conventional Planting.

Funding:

Link to Video:

,,,,,,,

How much primer choice affects the perceived biodiversity? A case study of bat diet according to three COI gene regions.

Renato Renison Moreira Oliveira, Enrico Bernard, Ronnie Alves, Eder Soares Pires, Angélico Fortunato Asenjo Flores, Gisele Nunes Lopes, Guilherme Oliveira, Marcele Laux

ITV - Instituto Tecnológico Vale, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

With the advent of the metabarcoding strategy, the study of diet components and niche overlap has reached an unprecedented taxonomic resolution, especially concerning the Arthropoda phylum. The cytochrome c oxidase subunit I (COI), widely adopted as marker for Metazoa, presents distinct variation levels at distinct gene regions, making the primer choice essential to amplification success of targeted taxa. The bat diet recovered from guano eDNA amplified by three distinct COI primer pairs was compared between each primer pair dataset in terms of taxa recovery, Arthropoda taxonomic coverage, and community structure response to the spatial component according to diversity indexes and beta-diversity patterns. The guano samples were collected in eight caves located in Amazon biome and one cave sampled along one year in Caatinga biome, totalizing 13 sample units. The P23 short primer pair (130bp) recovered mostly Chiroptera reads (56%), presented the lowest Arthropoda recovery, but the most even proportion of Arthropoda orders. The P34 primer pair, which amplifies the longest and more variable amplicons (370bp), recovered the largest proportion of Arthropoda reads (57%), 73% of it recruited by Diptera and Lepidoptera orders. The P56 primer pair, amplifying a 350bp COI amplicon originally described as the most conserved region, showed the narrowest taxonomic coverage, with over 95% of Arthropoda reads assigned to Lepidoptera. Both P34 and P56 long amplicons recovered the largest portion of reads from Caatinga cave, showed better Arthropoda recovery and resolution than P23, and presented a nested beta-diversity structure, probably as a result from the narrow taxonomic coverage and the time-series sampling design. Even so, the three primer pair datasets shared similar compositional responses to spatial scale, with most part of the variance explained by sample units, followed by regions and biomes. Since there was a spatial-ecological shared response between the datasets, mostly between P34 and P56, their Arthropoda community of reads could be complementary, potentially improving the niche breadth detected. The amplicon variability observed was substantially different from described by the literature, reinforcing the critical importance of local ecological and biological attributes in primer choice and expected outcome.

Funding:

Link to Video:

,,,

Mining of 141, 456 high-quality human exomes and genomes reveals the presence of 10, 909 putative immunoglobulin heavy chain IGHV variants

Tiago Mendes, Liza Figueiredo Felicori Vilela, Lucas Alves de Melo Pontes, Fábio Martins

UNIVERSIDADE FEDERAL DE MINAS GERAIS, Ufv, UFMG

Abstract

The correct identification of alleles can assist in the study of several human diseases associated with the antibody repertoire and in the development of new therapies using antibody engineering techniques. The advent of next-generation sequencing of human genomes and antibody repertoires enabled the development of several tools for the mapping and identification of new immunoglobulin (Ig) alleles. Some of these tools use 1, 000 Genomes (G1K) data for new Ig alleles discovery. However, genome data from G1K present low coverage and variant call problems. For these reasons, we used in this work, data from the Genome Aggregation Database (gnomAD), the largest high-quality catalogue of variation from 125, 748 exomes and 15, 708 human genomes. The methodology developed in this work identified 10, 909 putative immunoglobulin heavy chain variable region gene (IGHV) alleles, in which 10, 828 of them are new and 2, 024 appear at least in 6 different alleles. IGHV2-70 was the IGHV gene segment with the largest number of variants described. The majority of the variants were found in the framework 3 and most of them are missense. Interestingly, a large number of variants were found to be population exclusive. A database integrated with a web platform was created (YGL-DB) to store and make accessible the likely new variants found. This database is the largest human putative IGHV alleles repository to date. This available data can help the scientific community to validate new IGHV variants through the design of new primers (specific or not to a given population) or even to validate new variants found from AIRR-seq data.

Funding:

Link to Video:

”

Elucidating the multiple genetic lineages using SNP data in livestock

Marcos Vinicius Gualberto Barbosa da Silva, Saulo Moraes Villela, Luiz Afonso Glatzl Junior

UNIVERSIDADE FEDERAL DE JUIZ DE FORA

Abstract

Differences observed among populations of the same species regarding a trait can be attributed to differences in the environments in which they are found or to genetic differences among them. When seeking greater efficiency in animal production, there are two paths. The first one is to make improvements in general management, and the second one, slower and with a permanent and cumulative character, is a genetic improvement, carried out through the selection of desirable phenotypes and mating. In traditional animal breeding, the additive genetic value derived from the phenotype is estimated. With the advancement of Biotechnology and Genomics, the use of molecular markers such as SNP (Single Nucleotide Polymorphism) has resulted in more accurate selection methodologies. Relationship presupposes similarity of genotypes and its measurement is fundamental in the correct use of the selection and mating tools. The main purpose of models based on machine learning is to obtain generic conclusions through a particular dataset. With supervised learning, classifiers can be generated using data previously labeled for adjustment. As an alternative to traditional animal relationship identification techniques, this work aims to develop a lineage classifier based only on the animal's molecular markers. The dataset used in this study is composed of the genotype of 14, 242 Gir Leiteiro animals, allocated to the GGP Indicus SNP Chip, considering only markers from autosomal chromosomes, resulting in 33, 336 SNPs. Initially, quality control was performed using filters for minor allele frequency and call rate, resulting in 1, 380 markers. The five lineages with the largest number of individuals genotyped were identified, resulting in 1, 061 animals. In model development, each SNP was coded using one-hot-encoding and the dataset was split into training and test sets in the proportion of 70% and 30%. The training set was balanced through oversampling, being subdivided into 5 parts to perform a cross-validation. Four different machine learning models were adjusted, k-Nearest Neighbors (k-NN), Multilayer Perceptron (MLP), Support Vector Machine (SVM), and Random Forest (RF). The metrics used in the model selection were the accuracy and standard deviation of the test sets, and the area under the ROC curve. The best performance was obtained through SVM. From the developed work, it was possible to verify the importance of evaluating new models in animal breeding problems, which helps the development of methodologies with less computational effort and greater accuracy results.

Funding:

Link to Video:

,,,,,,,,,

Role of functional polymorphisms in the p53 pathway genes and the genetic susceptibility to Congenital ZIKV Syndrome

Eduarda Sgarioni, Igor Araujo Vieira, Ana Cláudia P Terças-Tretell, Marcial Francis Galera, Maria Denise Fernandes Carvalho de Andrade, Lavinia Schuler-Faccini, Fernanda Sales Luiz Vianna, Patrícia Ashton-Prolla, Lucas R Fraga, Julia do Amaral Gomes

Unichristus - Centro Universitário Christus, UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Abstract

Congenital ZIKV Syndrome (CZS) occurs in up to 42% of individuals exposed to ZIKV prenatally. Deregulation in gene expression and protein levels of components of the p53 signaling pathway, such as p53 and MDM2, due to ZIKV infection has been reported. Here, we evaluate functional polymorphisms in genes of the p53 signaling pathway as risk factors to CZS. This study was approved by the Ethics Committees of all participating hospitals. Forty children born with CZS (case group) and forty-eight children exposed to ZIKV, but born without congenital anomalies (control group) were included in this study. Case group was recruited in five Brazilian research and/or assistance centers in North (n=4), Northeast (n=21), Midwest (n=14) and South (n=1) regions of Brazil. Control group was recruited specially from Midwest (n=46), but also from North (n=1) and South (n=1) regions. Sociodemographic and pregnancy characteristics were obtained from questionnaires. Clinical data were obtained from chart review and consultation done by physicians. Blood samples were collected from participants for the DNA extraction. Genotyping was performed using the TaqMan© Genotyping Assay method in the Step One PlusTM Real-Time PCR System. Gestational and sociodemographic information as well as the genotypic and allelic frequencies of functional polymorphisms in TP53, MDM2, MIR605 and LIF genes were compared between the two groups. Quantitative variables were compared between groups by Student's t test or Mann–Whitney U test and categorical variables were compared by chi-squared test or Fisher's Exact Test. Hardy-Weinberg equilibrium was tested for all polymorphisms. A p-value <0.05 was considered statistically significant. We found children with CZS exposed predominantly in the first trimester and controls in the third trimester (p<0.001). Moreover, children with CZS were predominantly from families with a lower socioeconomic level (p=0.008). We did not find a statistically significant association between the investigated polymorphisms and development of CZS; however, we found an association of the TP53 rs1042522, which is associated with a more potent p53-induced apoptosis, and the phenotype lissencephaly in children with CZS (p=0.007). Our findings suggest that the ZIKV infection in the first trimester of pregnancy as well as the socioeconomic level may be environmental risk factors to CZS as well as the TP53 rs1042522 could be a possible genetic risk factor for the development of lissencephaly in children with CZS.

Funding: CNPq; INAGEMP; FIPE-HCPA

Link to Video:

””

Large runs of homozigozity regions in Arapaima gigas (Pirarucu) suggest the existence of genomic regions in single-copy and a sex-related dosage-compensation mechanism.

Tetsu Sakamoto, Jorge Estefano Santana De Souza, SIDNEY EMANUEL B DOS SANTOS, José Miguel Ortega, Renata Lilian Dantas Cavalcante

UNIVERSIDADE FEDERAL DE MINAS GERAIS, UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE, UNIVERSIDADE FEDERAL DO PARÁ

Abstract

Arapaima gigas, known as Pirarucu, is the largest freshwater bony fish in the world that dwells in the Amazon Basin. It is an Osteoglossiformes with great potential for aquaculture because of its large size, fast growing rate, and quality of flesh. Albeit these advantages, Pirarucu's reproductive management in captivity is limited, in particular, due to the lack of external sexual dimorphism, which makes sex identification difficult in this species. Here we used genomic data from Pirarucu to seek molecular markers which could assist in their sex identification and in unraveling the molecular mechanisms involved in sex differentiation. Previous studies on genome comparison of both sexes identified male specific regions, suggesting the XY system for Pirarucu. However, because of its small size, the topic is still in debate. In this study, we aimed to search for single-copy genomic regions in one sex of Pirarucu, which could be associated with the sex determination mechanism, by identifying sex-specific runs of homozygosity (ROH) regions. We used genomic data from six adults, three males and three females available at NCBI under the Bioproject IDs PRJEB22808 and PRJNA540910. BWA (v. 0.7.12) and Samtools (v. 1.7) were used for the genome indexing, mapping, and alignment steps of reads against the reference genome of Pirarucu ID GCA_900497675.1 deposited also in NCBI. Bedtools genomecov (v. 2.24) was used to analyze the total sample coverage. For calling variants, Varscan (v. 2.4.0) was used. The results were processed using R (v. 3.6.0) and the graphics were generated with ggplot2 library (v.3.2.1). To identify ROH regions, we verified the occurrence of heterozygous sites along the genome of each sample and selected those scaffolds that presented (1) the average of sites in heterozygosis = to 5 for one sex and (2) the size of the scaffold being = to 500.000 bp. We identified 22 scaffolds with large ROH in one sex and with the antagonistic scenario in the other one, suggesting that these regions are in single-copy in one sex and being affected by the dosage compensation mechanism in the opposite sex. These scaffolds are promissory candidates of a molecular marker for sex differentiation in Pirarucu, although analyses with more samples and molecular tests are needed. Furthermore, identifying ROH for comparative genomic analysis demonstrated to be an interesting strategy to assist us in the unraveling of the sex determination system in Pirarucu.

Funding:

Link to Video:

””

Unraveling potential probiotic features in the genome of Lactobacillus paragasseri

Flavia Figueira Aburjaile, Rodrigo Profeta Silveira Santos, Marcus Vinicius Canário Viana, Vasco Ariston de Carvalho Azevedo, Rodrigo Dias de Oliveira Carvalho

IOC/Fiocruz, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

The vaginal microbiota is dominated by lactobacilli, which exerts important health-promoting effects and prevention of bacterial vaginosis (BV). Lactobacillus paragasseri is a sister taxon related to L. gasseri and has been recently described in 2018 as a new species. Therefore, its role in the vaginal environment must be investigated. Moreover, previous studies suggest the vaginal ecosystem as a potential source of probiotic strains to be used in the treatment of other dysbiosis related diseases such as gastrointestinal disorders. Currently, comparative genomics studies have been shown as a promising tool in the prediction of the possible mechanisms related to their beneficial activity. In this work two L. paragasseri strains, CRI16 and CRI18 were recovered from healthy women of reproductive age. The strains were isolated from healthy subjects while CRI22 was isolated from a BV patient. The genomes were assembled with SPAdes and annotated via RAST. Eight genome sequences of L. paragasseri were obtained from NCBI and included in the comparative analysis. Genes involved in protective mechanisms, such as bacteriocins were predicted using BAGEL. The metabolic pathway of other health-promoting features such as vitamins and short-chain fatty acids and the adaptation to gastrointestinal stress conditions were characterized via PATRIC. The presence of genetic mobile elements such as plasmids and phages was evaluated in the strains to evaluate their safety for human use. In this context, antibiotic resistance genes and virulence factors were investigated using CARD and VFDB databases respectively. Our results revealed several beneficial features such as the presence of three and six bacteriocin genes in CRI16 and CRI18 respectively. Metabolic pathway analysis indicates both strains are able to produce vitamins K2, B1 and B9. Furthermore, they also seem to be able to produce Lactate and to degrade biogenic amines, including putrescine and spermidine although they may produce cadaverine as well. Interestingly, the strains present enzymes genes involved in bile salts deconjugation, suggesting adaptation to gastrointestinal conditions. Regarding safety issues, one intact prophage sequence was predicted in both strains and only CRI16 harbors a plasmid of 40kb, although no antibiotic resistance or virulence gene was found. Our study represents the first step in the characterization of the L. paragasseri strains CRI16 and CRI18 suggesting them as potential candidates for probiotic use.

Funding: CAPES

Link to Video:

,

Predicting melting temperatures and deriving hydrogen bonds using the Peyrard-Bishop model for LNA+DNA:DNA sequences

Gerald Weber, Izabela Ferreira

UFMG - UNIVERSIDADE FEDERAL DE MINAS GERAISn, UFMG - Departamento de Física

Abstract

Locked nucleic acids are nucleic acids modified by introducing a 2'-O-, 4'-C methylene bridge. This modification induces a conformation change in the backbone, locking the ribose ring in a C3-endo conformation. Thus inducing a favorable entropic variation in its vicinity increasing the overall helix stability. It has been shown to improve mismatch discrimination, compatibility, and specificity toward complementary DNA and RNA strands. Several applications have been described in nucleic acid-based therapeutic strategies both in vitro and in vivo. Even so, there are not many accurate temperature prediction methods applicable to LNA probes. Such temperature predictions are important, as an example, for probe design and PCR applications since both rely on the melting temperature. For instance, it is not yet fully established how much of the improvement in affinity and specificity is actually due to stacking interactions or hydrogen bonds. Here, we use the Peyrard-Bishop mesoscopic model that was successfully used for describing DNA, RNA, and more recently DNA/RNA hybrids, to characterize the thermodynamic properties of LNA. We use existing melting temperatures of LNA/DNA hybrids, to extract model parameters which can be interpreted in terms of hydrogen bonding and stacking interaction. Our results show a considerable increase in the hydrogen bonds of the modified nucleotides and also an instability on some stacking potentials, which is relatable to the destabilizing effect in some LNA modified probes.

Funding: capes, cnpq

Link to Video:

6 — Phylogeny and Evolution

,,,,,,,,

AMONG GENERA: A PHYLOGENETIC INFERENCE WITH MITOGENOMES OF NINE SPECIES OF STETHAPRIONINAE (CHARACIDAE, CHARACIFORMES)

Matheus Lewi Cruz Bonaccorsi de Campos, Fabiano Menegidio, Rubens Pasa, Igor Henrique Rodrigues-Oliveira, Dinaíza Abadia Rocha-Reis, John Seymour (Pat) Heslop-Harrison, Trude Schwarzacher, Karine Frehner Kavalco, Iuri Batista da Silva

UNIVERSIDADE DE MOGI DAS CRUZES, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

The Stethaprioninae is a species-rich subfamily that comprises small freshwater fishes with wide distribution in the Neotropical Ecozone. Most of the diversity of the group comes from Astyanax, a polyphyletic and high diverse genus with around 170 species. But with great diversity comes numerous taxonomic problems and even with wide studies with the group, many phylogenetic relationships remain unsolved. Given this scenario, we aimed to provide a better understanding of the relationships among the three genera through a mitogenomic phylogenetic inference. To achieve this goal, we sampled a total of 12 mitogenomes from four species of Astyanax, four of Psalidodon, one from Deuterodon giton and one from Brycon nattereri, which we assigned as an outgroup. We removed all rRNA, tRNA and the control region D-loop, leaving only the 13 protein-coding genes (PCGs) which we aligned one by one using the ClustalW algorithm. Then, we calculated the pairwise distances and conducted a Maximum Likelihood (ML) analysis with bootstrap as branch support value on MEGA X. Next, we concatenated the alignment on SequenceMatrix to partitioned and attribute the best evolutionary nucleotide model for each gene with Partition Finder 2.1. Ultimately, we performed the Bayesian Inference (BI) on MrBayes 3.2.7 with 4 independent Markov chains, with 10 million generations which 25% of them were discarded at the end of the analyses and used Tracer 1.7 software to verify the effective sample size (ESS) and strand convergence. We obtained phylograms from ML and BI analysis that show strong correlations with the genetic distance, with solid bootstrap and posterior probabilities in most of the branches, respectively. Showing a consistent clade subdivided in 2 monophyletic groups: Psalidodon and Astyanax, with Deuterodon as sister group of them. The topologies obtained leads to similar conclusions seen in previous works with Astyanax, reinforcing the Psalidodon status and the early divergence of the Deuterodon from the other genera. Hereby, we conclude that mitogenomes poses as a great tool to the phylogenetic inference among genera.

Funding:

Link to Video:

””

Effects of Alignment Characteristics on Distance-based Tree Reconstruction

Roberto Tadeu Raittz, Jeroniza Nunes Marchaukoski, Dieval Guizelini, ALEXANDRE GORI DE CASTILHO, Guilherme Taborda Ribas

UNIVERSIDADE FEDERAL DO PARANÁ

Abstract

The analysis and comparison of nucleotide and amino acid sequences are the cornerstones of computational biology and bioinformatics studies. Trees are important tools for comparative studies between gene sequences, proteins and genomes. As well as in studies of phylogeny, protein homology and taxonomy based on molecular characteristics. The constructions of the trees are directly affected by the multiple alignment algorithms, by the measures of distances between the sequences and in the research ambiguity of the interpretation of the distance values, observed in the comparison of the sequences composition and the evolutionary events. In this work, we revisited classical hierarchical clustering methods, applied to nucleotide and amino acid sequences, to measure and identify the effects on the resulting dendrogram when one varies the size and number of aligned sequences. And, also, to verify whether any of those methods can reproduce the reference tree - or at least close for some alignment variation. The Average-linkage is widely studied and has known problems of consistency in phylogenetic reconstruction, especially when the distance matrix is not time corrected with evolutive models. However, with the nowadays softwares for simulating theoretical alignments and trees, it is possible drawing massive experiments to verify how the input data can affect the final tree obtained, and whether there are variations that can identify ranges where distance and linkage methods can be consistent. In the completed stages, 840 data sets were simulated and, from different clustering methods, 5, 880 trees were generated. The data sets were obtained by the Cartesian combination between size and sequence quantity. Where sizes ranged from 10 to 10, 000 bases, in nucleotide and amino acid sequences and; sequence amounts ranged from 25 to 100 sequences. We used the geodesic distance method to do trees comparisons. Among the methods evaluated for nucleotide sequences, the best methods are: single, complete, average and weighted. While for amino acid sequences, the highlighted methods are: median and centroid. We conclude that the larger the sequence size, the greater the consensus of the trees produced and, fairly often, the average-linkage method is not the most suitable for the reconstruction trees when compared to the other methods covered in this study. And, at this point in the research, none of these clustering methods reproduces the reference tree.

Funding: Capes

Link to Video:

,,,,,,

First-ever described Virome of the Amazonian Lake Bolonha: contributions to the understanding of water-related public health concerns

Bruna Verônica Azevedo Gois, Kenny da Costa Pinheiro, Andressa de Oliveira Aragão, Ana Lidia Queiroz Cavalcante, adriana ribeiro carneiro folador, Rommel Thiago Jucá Ramos, Wylerson Guimarães Nogueira

UFPA - UNIVERSIDADE FEDERAL DO PARÁ/GEAM, Universidade Federal do Pará -UFPA, UNIVERSIDADE FEDERAL DE MINAS GERAIS, UNIVERSIDADE ESTADUAL DA PARAÍBA, UNIVERSIDADE FEDERAL DO PARÁ, UNIVERSIDADE DO ESTADO DO PARÁ

Abstract

The availability of safe water supplies and adequate sanitation is vital to protect populational health and is one of the basic human rights, according to the World Health Organization. Despite the importance of water supplies, the understanding of the ecology of freshwater viruses still lacks an in-depth description of its diversity, as well as the microbiological interactions that occur in these environments. The Amazonian Lake Bologna, from Belém, capital of the Brazilian State of Pará, is a key source of water that supplies the city and all of its metropolitan region, yet it remains unexplored regarding the contents of its virome and viral diversity composition. Therefore, this work's main aim is to clarify in terms of taxonomic diversity the species of DNA viruses that are present in this lake, especially bacteriophages and cyanophages, since they can act both as transducers of resistance genes and reporters of water quality for human consumption. For this work, we used the metagenomic sequencing data generated by Alves et al. (2020), and we analyzed it at the taxonomic level using the tools Kraken2, Bracken, and Pavian. Later, the data was assembled using Genome Detective, which performs the assembly of viruses. The results observed in this work suggest the existence of a widely diverse viral community and an established microbial phage regulated dynamics in the Lake Bolonha. This work is the first-ever to describe the virome of Lake Bolonha using a metagenomic approach based on high-throughput sequencing, as it contributes to the understanding of water-related public health concerns regarding the spreading of antibiotic resistance genes and population control of native bacteria and cyanobacteria.

Funding: CAPES; FAPESPA; FAPEMIG

Link to Video:

,,,,,,,

Assembly, annotation and gene editing of the genome of the PH8 strain of Leishmania amazonensis with focus on multigene families encoding virulence factors

Carlos Rodolpho Ferreira Brasil, João Luís Reis Cunha, Anderson Coqueiro-dos-Santos, Viviane Grazielle da Silva, Daniella Bartholomeu, Ana P. S. M. Fernandes, Santuza Maria Ribeiro Teixeira, Wanessa Moreira Goes

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

Leishmania (Leishmania) amazonensis is one of the etiological agents of cutaneous leishmaniasis, a disease that has 21, 000 cases / year in Brazil. Different molecules of the parasite have already been studied as they play a crucial role in the establishment of infection in the mammalian host and contribute to the pathogenesis of leishmaniasis. Among the most studied virulence factors are multigene protein-coding families such as amastins, GP63 metalloproteases and A2 proteins. To deepen the understanding of the role of these virulence factors, it is essential to obtain well-assembled and annotated complete genomes of different isolates of L. amazonensis and also to be able to manipulate the genes that encode these factors using the CRISPR-Cas9 technology. Here, we report the sequencing and chromosome level assembly of the PH8 strain of L. amazonenis using a strategy based on the combination of long PacBio sequencing reads, short Illumina reads and synteny data with the Leishmania mexicana genome. The initial contigs were generated using only the PacBio reads and the Canu assembler. The scaffolding step was performed using SSPACE and gap filling with GapFiller, based on short paired-end reads Illumina. Finally, the assembly was polished using Pilon, and the scaffolds were ordered based on the chromosomes of L. mexicana using Abacas. The final assembly, composed of 34 chromosomes and 44 small scaffolds not incorporated in the 34 chromosomes, represents a genome of 32 Mb. The annotation of the L. amazonensis genome was transferred from the annotation of 8225 genes present in the genome of L. mexicana. Of these, 7072 are protein coding genes, among which 82 encode tRNAs, 13 rRNAs and 270 ncRNAs. In addition, genes belonging to the families of amastins and metalloproteins GP63 were characterized. Amastins and GP63 were identified based on homology with proteins from other Leishmania spp and T. cruzi, adding up to 33 and 5 genes, respectively in L. amazonensis. The alignment of amastin genes, using MUSCLE and phylogeny analyzes, using the neighbor-joining algorithm with 1000 bootstrappings resulted in groupings corresponding to the four sub-classes of amastins known as alpha (a), beta (β), gamma (?) and delta (d) amastins. Analysis of the sequences encoding GP63 showed the conservation of important domains, such as HExxH and SRYD, which are important for protein structure and binding to macrophage surface receptors, respectively. Finally, we tested a CRISPR-Cas9 protocol to generate knockout cell lines of the Miltefosine Transporter gene (TM) of L. amazonensis as a proof of concept. Expression of Streptococcus pyogenes Cas9 (SpCas9) in L. amazonenses promastigotes was achieved after transfection with pLDCN, an episomal vector. Alternatively, Staphylococcus aureus Cas9 (SaCas9) was also expressed in Escherichia coli. Promastigotes were

7 — Proteins and Proteomics

,,,,,,

The extracellular vesicles produced by the probiotic Propionibacterium freudenreichii in different conditions share a core proteome that includes immunomodulatory proteins

Brenda Silva Rosa da Luz, Fillipe Luiz Rosa do Carmo, Gwenaël Jan, Yves Le Loir, Vasco A de C Azevedo, Eric Guedon, Vinícius de Rezende Rodovalho

INAPG- França, UNIVERSIDADE FEDERAL DE MINAS GERAIS, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

The probiotic properties of several organisms have been consistently associated to surface or secreted factors, including specific proteins that mediate interactions with the host. One of such organisms is Propionibacterium freudenreichii, a Gram-positive dairy bacterium that has been long used in the production of cheese, vitamin B12 and organic acids. We recently described that the extracellular vesicles (EVs) produced by this bacterium contain proteins that participate on their health-promoting roles. EVs are spherical nanostructures, delimited by a lipid bilayer and containing several molecules in their interior. In the case of P. freudenreichii, EVs contained surface-layer proteins previously associated to immunomodulation, such as surface-layer protein B (SlpB) and E (SlpE). In addition, EVs exerted an anti-inflammatory effect in vitro, with the mitigation of NF-?B activation and IL-8 release by intestinal epithelial cells. In our first study, EVs were purified by size-exclusion chromatography (SEC) from cultures in milk ultrafiltrate (UF). However, it remained unclear if EVs characteristics and content would be similar in other growth or purification conditions. Therefore, we analyzed EVs derived from P. freudenreichii cultures in yeast extract-lactate (YEL), in addition to UF medium. Moreover, we employed ultracentrifugation (UC) in a sucrose density gradient as a purification method, in addition to SEC. Proteomics characterization was performed by NanoLC-ESI-MS/MS analysis. We found that, considering all 4 conditions, EVs share a core proteome of 308 proteins, that included the immunomodulatory proteins SlpB and SlpE. The detection of immunomodulatory proteins in EVs obtained from different conditions opens up the possibility of EVs yield and activity optimization, while retaining or improving immunomodulatory properties. Moreover, functional enrichment analysis showed that the core proteome was mainly associated to energy and carbon metabolism, ribosomal structure and biogenesis, quorum sensing, protein export and peptidoglycan biosynthesis. This indicates that some functional roles are conserved among different conditions of EVs obtention. The availability of this core proteome will allow a comprehensive analysis of the vesicular content of EVs produced by this probiotic strain and the identification of general traits of the proteins that are exported inside EVs. Overall, this study contributes with a robust proteomic characterization of EVs produced by a relevant probiotic strain of P. freudenreichii.

Funding:

Link to Video:

””

ANTIGENIC VARIABILITY OF DENGUE VIRUS NON-STRUCTURAL PROTEIN 1 AND ITS IMPACT ON THE DEVELOPMENT OF DIAGNOSTIC MODELS

RICARDO LEMES GONÇALVES, Breno De Mello Silva, ANA CLARA GOMES DE SOUZA, Túlio César Rodrigues Leite, UBIRATAN DA SILVA BATISTA

UNIVERSIDADE FEDERAL DE OURO PRETO, UFOP - UNIVERSIDADE FEDERAL DE OURO PRETO

Abstract

According to the Pan American Health Organization, over 2 million cases of human infections in the Americas were caused by the Dengue virus (DENV) in 2020 alone with more than 800 deaths. The circulation of the four DENV serotypes (1-4) is still a challenge for the development of efficient diagnoses. The available tests target the structural proteins (E, M, and C) and non-structural protein 1 (NS1). However, global variations in the primary sequences of these proteins are countless and hinder the development of effective vaccines and diagnostic tests. Among these proteins, NS1 has been increasingly using as a target for new diagnostic methods and vaccines, albeit very variable. It is secreted into the extracellular medium in a hexameric form and can be detected at high levels in the bloodstream turning it available to the immune system. The present work mapped the variability of circulating NS1 protein main domains in the American continent and evaluated its impact on antigenic regions through in-silico approaches. 132 coding sequences of Dengue virus NS1 protein (39 from DENV1, 33 from DENV2, 34 from DENV3, and 26 DENV4) from South, Central, and North Americas were analyzed. These sequences were aligned using the MEGA X software through the Muscle algorithm. Next, the variant amino acid (aa) residues were mapped in the exposed wing and β-ladder domains. This analysis identified 78 residues of aa variants in the NS1 protein for DENV whose 34.61% of residues show a prevalence higher than 10%. The Bepipred - 2.0, NetCTL, and NetMHCII tools were used to map antigenic regions. Eighty linear B cell, cytotoxic T, and T - Helper epitopes were predicted for the 4 serotypes. 65.6% of the 32 predicted B cell epitopes, 33.3% of the 18 cytotoxic T cell epitopes, and 43.3% of the 30 T-helper epitopes showed amino acid residue variations. Although there is little information in the literature on how these variations may impact the accuracy of tests on the market, several NS1 variations occur in predicted antigenic regions. Such a situation can be greatly underestimated due to the small amount of sequencing carried out in underdeveloped tropical countries. Together, these findings suggest that further studies are needed to better understand the distribution and influence of NS1 variability in DENV diagnostic tests and raise a red flag about NS1-based tests to other Flaviviruses.

Funding: UFOP, CNPQ e CAPES

Link to Video:

Classification of amino acid residue pairs using GMM and EM Algorithm

Higor Coimbra Amorim

CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS

Abstract

RID (Residue Interaction Database) is a system built to propose site-directed mutations on 3D protein structures using PDB files. One of the steps of RID process is to indicate which amino acid residue pairs of a protein are able to receive a mutation and then, all of the candidates are classified according to their atomic structure similarities. On RID, these classifications are based on a score produced by the overlap of all candidates atomic structures using LSQKAB. However, for a large dataset of PDB files, for example a dataset with about 16000 elements, as the one used in this research, the overlap made by LSQKAB can be a very large time consuming process. One of the proposals made to replace LSQKAB was the use of an atom to atom distance matrix to replace the residue pairs' PDB files and the use of K-Means clustering algorithm to replace the score overlap classification. Results showed that K-Means was a viable method to cluster residue pairs and could be used inside the RID system, as part of the structural classification process. Following the good results obtained by K-Means, this research proposes the use of a more flexible method of clustering: Gaussian Mixture Models (GMMs) used within the Expectation-Maximization (EM) algorithm. The first results showed that GMMs can also be a viable clustering algorithm and a candidate to replace LSQKAB. The clusters' overall biological similarity for a number of 500 and 750 clusters on the GMM are 4.73%, 0.1% higher, respectively, than the ones obtained by K-Means, based on a dataset of 16383 PDB files of amino acid residue pairs.

Funding:

Link to Video:

,,,,,,,,

IMMUNOINFORMATICS APPROACHS TO DESIGN A CHIMERIC MULTI-EPITOPE VACCINE AGAINST MYCOPLASMA PNEUMONIAE

Vasco A de C Azevedo, Arun Kumar Jaiswal, Helioswilton Sales de Campos, Rodrigo Bentes Kato, Stephane Fraga de Oliveira Tosta, Siomar de Castro Soares, sandeep tiwari, Marta Giovanetti, Thaís Cristina Vilela Rodrigues

UFMG - UNIVERSIDADE FEDERAL DE MINAS GERAIS/ UP- Universidade do Porto, UNIVERSIDADE FEDERAL DE MINAS GERAIS, INAPG- França

Abstract

Pneumonia is a serious health problem with global effects, being the death cause of over one million people annually. Among the main microorganisms responsible by pneumonia, Mycoplasma pneumoniae is one of the most common, with a significant increase in the last years. The vaccines are fundamental in diseases prevention besides to considerably avoid the need of health services and funding resources. In this way, the proposal of the present study is to construct multi-epitope vaccine against M. pneumoniae through immunoinformatics approach. Multi-epitope vaccines are constructed by epitopes properly selected to induce targeted immune responses and avoid adverse reactions. Therefore, seven reported vaccine candidates were selected based on the reverse vaccinology approach from one of our published research article and three vaccine candidates through a literature search. Afterwards the search for MHCI, MHCII and B epitopes were performed. Furthermore, the overlapping epitopes, capable to induce both humoral and cellular responses were identified. Those epitopes were filtered according to their immunogenicity, and population coverage. The epitopes with best features were joined with classical peptide linkers and the heat-labile enterotoxin from Escherichia coli as adjuvant, then, the structure of the vaccine was predicted. The vaccine was considered physically stable, non-toxic, non-allergen, not significantly similar to human proteome and with appropriate antigenic and immunogenic properties. The molecular docking of the vaccine with the Toll-Like Receptor 2 (TLR2) was performed and the dynamic simulation will be executed to ensure the affinity and stability between these complexes. In silico cloning was tested in an expression vector with positive results. In addition, the immune simulation for vaccine efficacy was also tested with promising findings. Through immunoinformatic approaches we constructed an effective multi-epitope vaccine candidate, that with further tests could contribute to prevention of pneumonia in a massive scale. Besides that, the study assists to better understanding of the immune mechanisms regarding M. pneumoniae infections and its interaction with the host.

Funding:

Link to Video:

’’’

Which substilisin/kexin like proprotein convertases PCSKs might be or not be implicated in SARS-Cov-2 cell to cell infection?

José Miguel Ortega, Lucianna Helene Silva dos Santos, Lucas Bleicher, Arthur Pereira Fonseca

UNIVERSIDADE FEDERAL DE MINAS GERAIS, IOC/Fiocruz, UFMG, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection became pandemic and has boosted research on possible treatments and medicines. Apparently, the virus has undergone a series of genetic mutations, allowing it to efficiently infect humans. One of these mutations is the insertion of a cleavage site on its spike protein, which is expected to be target of at least one protein convertase from the PCSK (substilisin/kexin like proprotein convertase) subtype: Furin, also known as PCSK3. When it is cleaved the spike is divided into two subunits, which increases significantly the virus cell to cell infection. In total there are nine PCSKs with varying tissue distributions. These enzymes contain a catalytic triad, Asp-His-Ser, and an oxyanion hole, Asn, in the peptidase domain. We set out to investigate which PCSKs may act cleaving the spike protein. We searched the coronavirus spike protein loop bearing the putative cleavage site and prepared it in pdb format. Then we docked this peptide to the PCSKs, using CABSdock, asking if they would bind it at the catalytic domain. Later we ran this complex in a molecular dynamic for 100 ns using NAMD. Hitherto, results for five convertases indicate preferential cleavage by a subset and not all enzymes. Furin and PCSK2 showed similar results, with the peptide bound in the catalytic site and contacting the catalytic triad and oxyanion hole. Preliminary results (30 ns) indicate that PCSK9 may also cleavage the spike loop. In contrast, in PCSK1 and PCSK5 complexes the peptide disconnected from the active site and moved around freely, although managing to attain an initial correct docking, indicating wispy efficiency on the spike cleavage. MBTPS1 (PCSK8) is the most distant in the family and a modeled structure was obtained by threading with C-I-Tasser; the enzyme was discarded because the model was unable to dock the peptide correctly in the catalytic domain. Dinamics analysis for PCSKs 4, 6 and 7 are currently running and might be reported in the conference. Furin is inhibited by Pirfenidone and PCSK9 by evolocumabe. But as PCSK2 emerged as a new possible enzyme acting on the coronavirus, a search on Drugbank showed that there is no currently inhibitor for it. We then suggest that, once proven in vitro, this convertase should be target of new drugs to suppress its activity.

Funding: Support: CAPES

Link to Video:

,,,

Unrevealing novel potential therapeutic targets in fungi extracellular vesicle proteins: a comparative study between pathogenic and non-pathogenic fungi species

Letícia Graziela Costa Santos, Marcio Rodrigues, Fabio Passetti, Vinicius da Silva Coutinho Parreira

Fiocruz-PR, Fiocruz, Instituto Carlos Chagas, Fiocruz

Abstract

Extracellular vesicles (EV) are double-membrane vesicles associated to intercellular communication. In higher eukaryotes, EVs are regularly used to cell-cell communication from close cells to far tissues. In pathogenic organisms, EVs play a role in pathogen-host interplay. The EVs contain different molecules, including nucleic acids, proteins, polysaccharides, and lipids. Since the discovery of EV production in the fungus Cryptococcus neoformans, the importance of EVs release in its pathogenicity have been elucidated. Shotgun proteomics is the standard strategy to study the proteome composition of a given sample. Therefore, shotgun proteomics has been used to identify fungi EV proteins in many clinical samples. To date, few studies have addressed the proteomic content of EVs from multiple pathogenic fungi species. In this context, orthology analysis allows to identify relations about genes and its proteins in different species through the peptide sequences comparison. Our main objective was to use an orthology approach to compare EV shotgun proteomics data derived from 8 pathogenic and 1 non-pathogenic species, as follows: Aspergillus fumigatus, Candida albicans, Cryptococcus deuterogattii, C. neoformans, Histoplasma capsulatum, Paracoccidioides brasiliensis, Sporothrix brasiliensis, Sporothrix schenckii, and Saccaromyces cerevisiae. Proteins detected through shotgun proteomics were compared using an orthology approach based on Uniprot and FungiDB databases. We integrated data for 11, 433 proteins detected in fungi EVs, resulting in 3, 834 different orthogroups. Proteins with the Hsp70 Pfam domain were clustered in the unique orthogroup (OG6_100083) identified for all fungi species. Such proteins are associated with stress response, survival, and morphological changes of different fungi species. Although no orthogroup limited to pathogenic fungi EV was found, we identified 5 orthogroups exclusive for S. cerevisiae. Using the criteria of at least 6 pathogenic fungi species to define a cluster, we detected the following unique pathogenic orthogroups, which are listed according to the most frequent Pfam annotation: ATPase family associated with various cellular activities (OG6_101915), Nucleoside diphosphate kinase (OG6_100304), Ribosomal S17 (OG6_100832), Core histone H2A/H2B/H3/H4 (OG6_100082), and RNA recognition motif (OG6_100425). Taken together, our data suggest that Hsp70-related proteins play a key role in fungi EVs, regardless the pathogenic status. Using an orthology approach, we identified at least 5 protein domains that may be investigated as novel potential therapeutic targets against pathogenic fungi.

Funding: CAPES, CNPq, and Fiocruz

Link to Video:

8 — RNA and Transcriptomics

”

Metatranscriptomics analysis reveals diverse viral RNA in cutaneous papillomatous lesions and normal tissues of cattle

GERLANE DOS SANTOS BARROS, Marcus Vinicius de Aragão Batista, Adriana de Oliveira Fernandes

Federal University of Sergipe

Abstract

Bovine papillomavirus (BPV) is the pathogen associated with bovine papillomatosis, which mainly affects domestic cattle and acts by forming benign warts in epithelial tissues, as well as malignant lesions. Studies which aim to promote the molecular detection have shown that the same animal could be infected not only by BPV, but also by other viruses, in a way similar to humans. Thus, it is possible that these coinfections may influence the disease progression. Therefore, this study aimed to identify and describe the functions of viral genes in cutaneous papillomatous lesions and normal tissues in cattle through a metatranscriptomic approach. A RNA-seq computational pipeline was used to analyse six libraries of RNA sequences from epithelial tissues from bovines, three of them from cutaneous papillomatous lesions and other three from normal tissues. Sequences were obtained from Gene Expression Omnibus (GEO) database. Trinity was used to assemble reads into contigs and Bowtie2 was used to map the reads with the bovine reference genome. Non-mapped sequences were converted into FASTA format to undergo a BLASTx search against the Swissprot viral protein database. Functional annotation of the expressed viral genes was performed using Blast2GO, KEGG and STRING. In total, 106.740.353 raw paired reads were obtained, which resulted in 183.333 after assembling, mapping and filtering to recover only the non-mapped contigs. We have found a total of 25 viral families, being 25 of them present in the cutaneous papillomatous lesions and 24 in the normal tissues. It was possible to notice that both libraries shared similarities in term of viruses and genes found, and that the most prominent viral families were of clinical interest, such as Poxviridae, Retroviridae and Herpesviridae. The functional annotation revealed that the expressed genes had functions related to viral replication, immune suppression, cell cycle malfunctions and cell growth. In this study we managed to identify several viral families in bovine warts and normal cutaneous tissues with expressed genes related to several cell cycle control activity. This analysis is vital to extend the knowledge regarding the viral diversity in bovine papillomatous lesions and to aid in the compreheension of the clinical implications of theses viruses within the main disease.

Funding:

Link to Video:

,,,,,,,

Environmental plasticity of Staphylococcus aureus extracellular vesicles RNA content

Aurélie Nicolas, Svetlana Chabelskaya, Vinícius de Rezende Rodovalho, Yves Le Loir, Vasco A de C Azevedo, Brice Felden, Eric Guedon, Brenda Silva Rosa da Luz

INAPG- França, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

The role of bacterial extracellular vesicles (EVs) in cell-to-cell signaling have arisen a lot of interest in the past years. These membranous spheres released by many living cells carry various bioactive macromolecules (e.g. proteins, nucleic acids), some of which were shown to be delivered to target cells to perform functional roles. Studies reveled that bacterial EVs charters RNA export, probably for broad communication with surrounding bacteria, as well as with their infected/colonized hosts. However, data in this area is still lacking, mainly in Gram-positive bacteria. Staphylococcus aureus is a serious human and animal pathogen that releases EVs, nevertheless, no studies characterizing the presence of RNA inside its EVs have been reported. Here we address if S. aureus EV cargo comprises RNAs, and if so, which RNAs. A high-throughput RNA sequencing approach was used to evaluate samples of producing S. aureus cells (clinical strain HG003) and its derived EVs under different in vitro conditions: early- and late-stationary phases and in presence or absence of a sub lethal concentration of vancomycin. Results showed no significant difference in the particle yields between the conditions tested, however, EVs released in late-stationary growth phase were approximately 55 % larger than those from early-stationary phases. Various RNAs were identified into the S. aureus EVs, including tRNAs, rRNAs, mRNAs, and sRNAs, which are also present in the EVs producing cells. EVs enclose mRNAs expressing virulence factors genes, ribosomal proteins, transcriptional regulators, and metabolic enzymes. The sRNA RsaC implicated into the oxidative stress adaptation was also detected. The amount and nature of the RNAs detected into the purified EVs was significantly impacted according to the growth phase and the presence/absence of vancomycin, while in a much less extent in the EVs producing cells. Finally, differential RNA abundance was observed among the environmental conditions tested, and between EVs and EVs producing cells, suggesting that not only environment shapes the RNA content, but the packing of these molecules into EVs is a regulated process. This is the first exploratory work characterizing RNA cargo from S. aureus and its derived EVs under different conditions. Since several RNAs present inside EVs are implicated into staphylococcal virulence and survival (RNAIII, Spa, Sbi, Hld, RsaC), our report indicate possible roles of these RNAs on bacterial-host cell communication, virulence, and pathogenesis of S. aureus, paving the way for future functional studies in this area.

Funding:

Link to Video:

,,,,,,

Analysis of expression of xylanases encoded in the genome of rust coffee fungus during different stages of infection

Túlio Morgan, Rafaela Zandonade Ventorim, Renato Lima Senra, Isabel Samila Lima Castro, Eveline Teixeira Caixeta, Tiago Mendes, Júlia Santos Pereira

UNIVERSIDADE FEDERAL DE VIÇOSA, UNIVERSIDADE FEDERAL DE MINAS GERAIS, UNIVERSIDADE FEDERAL DE VIÇOSA, UFV - UNIVERSIDADE FEDERAL DE VIÇOSA campus Viçosa, Ufv

Abstract

Coffee leaf rust is a major disease caused by the fungus Hemileia vastarix that affects many coffee producers around the world. Since H. vastatrix is a biotrophic fungus, its growth and reproduction are totally dependent on the cells of the living host, and because of that, they infect the tissue without causing necrosis. Also, it is known that some fungi, during plant interaction, can express genes involved in the formation of infectious structures as well as synthesize enzymes responsible for the degradation of the host cell wall. Many of them produce enzymatic cocktails capable to degrade cell wall components, which are basically cellulose, hemicelluloses and lignin. The most abundant group of hemicellulases are xylans, which has aroused industrial interest for many applications, such as biobleaching in the pulp and paper industry and as prebiotics in animal nutrition. However, more studies are needed to evaluate interactions between fungi-plant and other factors that can activate fungus pathogenicity. Also, it is desirable to identify active xylanases of commercial interest. Because of that we propose on this work to evaluate if genes that encode xylanases are being expressed during H vastatrix infection. First, the fungal protein was predicted by Augustus ab initio prediction. The program was set for the fungus Puccinia graminis. The functional annotation of the predicted genome was performed by dbcan2 (release 8.0) selecting all CAZymes from the genome. Among the 345 CAZymes found, 162 belongs to the Glycoside Hydrolases (GH) and only 3 are xylanases (GH10). After that, was performed the analysis of RNA-seq data of C. arabica cv. caturra vermelho CIFC 19/1 (Bioproject: PRJNA353185) to 0, 12, 24, 96 hai. Read quality was assessed with FastQC software version 0.11.5 and trimmed with Trimmomatic software version 0.36. Next the "Tuxedo" pipeline was executed using Hemileia vastatrix HvCat (PZQQ00000000.1) as the reference genome – the same used to gene prediction in Augustus. The preliminary results indicate that the 3 xylanases present different expression profile, but are being most expressed in the early stages of the infection: 12 and 24 hai, which corresponds to the phases when the fungus is penetrating the plant. For the next steps of this work we aim to perform gene expression analysis of the 3 xylanases using real-time quantitative PCR and execute activity assays with the xylanases.

Funding:

Link to Video:

,,,

microRNAs, target-genes and pathways related to Gestational Diabetes Mellitus: searching for potential biomarkers

Marcelo Rizzatti Luizon, Izabela Mamede Costa Andrade da Conceição, Graziele Pimentel Coelho Almeida, João Rafael Gonçalves Pereira

UNIVERSIDADE FEDERAL DE MINAS GERAIS, UFMG

Abstract

Gestational Diabetes Mellitus (GDM) is the most common metabolic disease of pregnancy, defined as carbohydrate intolerance first diagnosed during pregnancy that may or may not persist after delivery. In addition, women with GDM are at higher risk for complications such as obesity and cardiovascular disease. Adequate glycemic control has been proven to reduce the risk of complications related to GDM. MicroRNAs are widely recognized post-transcriptional regulators, as they hybridize with complementary sequences in the messenger RNA and silence gene expression by destabilizing or preventing mRNA translation. MicroRNAs have already been proposed to play a role in the pathogenesis in cancer and diabetes, but their effective use as biomarkers is still undefined, due to the ability of a microRNA to have made on a large group of genes. Indeed, there are still not enough data to understand if microRNAs can be promising biomarkers for prediction and monitoring pregnancy complications, such as GDM. Our goal is to identify microRNAs differentially expressed in GDM in literature reviews and their possible target genes and the associated molecular pathways, which could be related to mechanisms present in GDM pathophysiology using bioinformatics tools. From a selection of microRNAs identified in literature reviews on GDM, target genes were identified using Harmonizome (https://maayanlab.cloud/Harmonizome/), a collection of pre-processed datasets used for data mining about gene products and proteins. From the name of each microRNA identified, searches were carried out on Harmonizome to find their respective target genes, which were used to identify associated molecular pathways in the Enrichr, a functional enrichment tool for multiple gene-sets (https://maayanlab.cloud/Enrichr/), and we have focused on the Reactome Database pathways (https://reactome.org/). Pathways with p-value lower than 0.05 were selected for further analysis. From the literature, we selected four microRNAs for their possible effect as central regulators in GDM pathophysiology, namely miR-20a-5p, miR-16-5p, miR-17-5p, mir-29a-3p. In our research, we found 14 regulatory pathways associated with the target genes for these four microRNAs that passed the 0.05 cutoff. Among these pathways, "cell cycle related to PTK6" was the pathway with the highest combined score, which may be related to diabetic cell death. Another important result is "TGF-beta related pathways", since it is one of the main factors which are related to Diabetic Nephropathy. Our results suggest that microRNAs may be potential biomarkers to clinical outcomes and mechanisms related to GDM pathophysiology. However, further functional studies should explore the mechanistic role of these microRNAs in GDM pathophysiology.

Funding:

Link to Video:

,,,

Pathway analysis of target-genes for microRNAs related the pathophysiology of Pre-eclampsia

Izabela Mamede Costa Andrade da Conceição, João Rafael Gonçalves Pereira, Marcelo Rizzatti Luizon, Graziele Pimentel Coelho Almeida

UNIVERSIDADE FEDERAL DE MINAS GERAIS, UFMG

Abstract

Pre-eclampsia (PE) is defined as arterial hypertension (systolic blood pressure above or equal to 140 mmHg and/or diastolic blood pressure above or equal 90 mmHg) identified for the first time after the 20th week of gestation which is frequently associated with proteinuria, among other possible clinical features. PE continues being the major cause of maternal and fetal morbidity and mortality worldwide. However, there are no predictive biomarkers for PE. MicroRNAs have a prominent role in regulating the amount of RNAs and proteins produced in eukaryotic genes. Due to its broad regulatory potential and its presence on peripheral blood, microRNAs have been suggested as possible biomarkers for a variety of human conditions including cardiovascular diseases and cancer. Here we propose a way to search for microRNAs which could act as regulators in pathways related to PE pathophysiology, which could help to find potential biomarkers for PE. From narrative review articles, five microRNAs were selected which could have potential effects on the PE pathophysiology, namely miR-210, miR-126, miR-152-5p, miR-216 and miR-148a. The selected microRNAs were then inserted in HARMONIZOME, a collection of pre-processed datasets, used in data mining of gene products and proteins, in order to search for the relation of those five microRNAs with their possible target-genes. A total of 374 genes were found as targets, which were then submitted to EnrichR, a tool to functional enrich gene sets used to search for pathways which may participate the target-genes. We used Reactome pathways, and a total of 15 pathways with a p-value lower to 0.05 were selected for further analysis. The pathway with the highest combined p-value was "Interleukin, inflammation", which represents a key role on PE pathophysiology. This finding suggests that those five microRNAs could regulate inflammatory genes and enhance the inflammatory response in PE. Noteworthy, "transcription pathways" were also distinctively present, together with "splicing and gene silencing". Interestingly, this finding raises questions about how the mechanisms related to PE may modify the transcriptome. Although there are no studies in the literature regarding alternative splicing in PE, it is possible to take part into the mechanisms related to PE pathology. "Apoptosis pathways" were also present, which correlates to the key role of cell death and hypoxia present on PE. In summary, the microRNAs miR-210, miR-126, miR-152-5p, miR-216 and miR-148a could act as important regulators of mechanisms in PE, and may help to unravel possible biomarkers and targets in PE.

Funding:

Link to Video:

”

miRNA-BASED PROGNOSTIC PREDICTOR FOR OVARIAN TUMORS USING MACHINE LEARNING

Alexandre Chiavegatto Filho, Mariana Boroni, Cristiane Esteves Teixeira

Brazilian National Cancer Institute (INCA)

Abstract

Ovarian cancer is one of the neoplasms with the highest incidence among women worldwide, with significantly high mortality. The vast majority of patients are diagnosed in advanced stages of the disease since early stages present unspecific symptoms and inaccurate diagnoses. Currently, there are not many biomarkers available clinically to help diagnose or predict its prognosis because of contradicting diagnostic accuracy. In this sense, the application of multi-omic integration approaches combined with machine learning techniques is promising, not only to better understand cancer prognosis but to identify effective prognostic biomarkers related to ovarian cancer. Therefore, our goal is to build a predictor of prognosis for patients diagnosed with ovarian cancer, as well as to identify new biomarkers. Patients from diagnosed with ovarian cancer obtained from the database of The Cancer Genome Atlas (TCGA) project were classified into two groups according to their prognosis Patients with < 3 years of survival and Dead status were allocated as the group with poor prognosis and patients with = 3 years were included in the good prognosis group. Based on the miRNA expression data we were able to identify relevant predictors by applying variable selection methods (FCBF, Cox Univariate Regression, and ElasticNet). Subsequently, data were divided into training (70%) and test (30%) data sets. Nine machine learning algorithms (Support Vector Machines, Extreme Gradient Boosting Machine, Gradient Boosting Machine, Stochastic Gradient Boosting Machine, Random Forest, Conditional Random Forest, MLP (Multilayer perceptron), Generalized Linear Models, and Ranger Random Forest) were trained to build the prognosis predictor. Also, we identified the potential target genes for the miRNAs selected as relevant predictors through the analysis of differentially expressed genes. After applying the variable selection methods, 78 miRNAs were selected. Regarding the performance of the models on the test data set, the MLP, an artificial neural network model, presented the best metrics: 0.684 of AUC (Area Under the Curve); 0.740 specificity; 0.612 sensitivity; 0.603 of F1-Score and with an accuracy of around 70%. The miRNA-483, which was observed differentially expressed in more advanced stages of ovarian cancer in other studies, was considered to be an excellent predictor by the MLP model to classify patients in good or poor prognosis. When evaluating the top 1000 differentially expressed genes (655 up- and 345 down-regulated) in ovarian cancer samples compared with the healthy tissue, 196 were identified as potential targets for the 78 relevant miRNAs. Among them, 15 genes are regulated by the miRNA-483 and are associated with the prognosis and survival of patients with ovarian cancer.

Funding:

Link to Video:

”

Insights into plant adaptations to occupy a challenging Amazonian habitat

Cecílio Frois Caldeira, Guilherme Oliveira, Mariana Dias

ITV - Instituto Tecnológico Vale, UFMG/ITV

Abstract

Canga substrates are a well-known example of a harsh environment for plants. Canga is the name of Brazilian ferruginous field formations, and it refers to the ecosystems associated with superficial iron crusts typical of Minas Gerais and Pará. This ecosystem is marked by high temperature and UV radiation, acidic and nutrient-depleted soils (especially phosphorus, magnesium, and calcium), and high metal concentrations (such as iron and manganese), all challenging for the establishment of plants. In the present study, we generated the first transcriptome data from two native Fabaceae species widely distributed in the canga outcrops in the State of Pará, Parkia platycephala Benth. and Stryphnodendron pulcherrimum (Willd.) Hochr, to understand the adaptive genetic mechanisms behind the establishment of these plants in the canga environment. Transcriptomics were carried out from leaves of plants growing in canga and forest substrates collected at the Carajás Mineral Province, Pará, Brazil. A combination of methods was used to recover complete and accurate transcriptomes. We achieved over 95% of the complete single-copy genes of eudicotyledon orthologs with BUSCO evaluation. Both species had more down-regulated genes in plants growing under canga substrate compared to the forest substrate: 391 up-regulated and 723 down-regulated for P. platycephala and 264 up-regulated and 574 down-regulated for S. pulcherrimum. The gene ontology enrichment analysis showed that the enriched differentially expressed genes (DEGs) of P. platycephala are associated with response to abiotic stimulus with the up-regulation of thiazole and thiamine (vitamin B1) biosynthetic process. For S. pulcherrimum, the DEGs are mainly associated with the up-regulation of guard cell differentiation and stomatal complex development. However, the plants' shared 64 up-regulated and 163 down-regulated genes, mainly associated with the rhythmic process and polysaccharide catabolic process. The KEGG enrichment pathway analysis revealed also enriched DEGs involved in the plants' circadian rhythm regulation, biosynthesis of secondary metabolites, and starch and sucrose metabolism. The differences in gene expression between P. platycephala and S. pulcherimum in canga substrates were lower than between plants in forest substrates, suggesting a more conservative strategy when they grow in canga. Cross-species differential expression analysis was conducted using single-copy orthologues shared between species. Samples clustered by species indicated the difference in the adaptive strategies of each species. Looking at the DEGs between canga and forest, 204 gene orthologs were found to be DE between substrates. Our results reveal some insights over the adaptive convergence in the canga environment and suggested different strategies between species.

Funding:

Link to Video:

,,,,,

De novo transcriptomes of endangered Amazonian bats and the detection of a rich source of virus information

Santelmo Vasconcelos, Leonardo Trevelin, Mariane Ribeiro, Renato Renison Moreira Oliveira, Guilherme Oliveira, Mariana Dias

ITV - Instituto Tecnológico Vale, UNIVERSIDADE FEDERAL DE MINAS GERAIS, UFMG/ITV, ITV - Instituto Tecnológico Vale

Abstract

Bats are the second richest mammalian order, present a wide diversity of feeding behaviors, and play a crucial role in maintaining the ecosystem balance. They provide essential ecological services, such as pest control, pollination, and seed dispersion. Bats are also known as a frequent virus reservoir. The occurrences of multiple viral diseases like Ebola, SARS, and, more recently, Covid-19 have motivated the scientific community to look deeper into the relationships of bats and infectious agents and discover potentially pathogenic viruses. In this study, we used RNA-sequencing to reveal both the expression of genes associated with bats' innate immunity and the detection of active viruses. We performed a de novo transcriptome assembly and annotation of three regionally endangered bat species from Serra dos Carajás, Pará, Brazil: Furipterus horrens (Furipteridae), Lonchorhina aurita (Phyllostomidae), and Natalus macrourus (Natalidae). Total RNA was isolated from several organs of one specimen per species. Transcriptomes were assembled with Trinity. After removing redundant contigs, functional annotation was performed using Trinotate. We recovered from the assembled transcriptomes over 90% of the mammalian orthologues, higher than previously published bat transcriptomes. Highly expressed genes recovered for all three species are associated with a wide range of cellular activities, including more than 900 genes related to the immune response. Besides, we could identify virus sequences from tissue reads of the three analyzed bats. We customized a bat genome database with representatives of all three bat families and aligned the reads from all tissues of each species to remove host sequences. The unmapped reads were de novo assembled with Trinity, and Kaiju was used for the initial taxonomic classification of the generated transcripts, which were further validated with BLAST searches against complete virus reference sequences. We identified 32 virus families in the tissues of F. horrens, and 27 in L. aurita and N. macrourus. Eighteen virus families were shared among the species. The families with a higher number of representatives in the three species were Retroviridae, Herpesviridae, and Poxviridae. Viruses primarily associated with human infections, such as the human herpesvirus, human endogenous retrovirus, and human papillomavirus, were found in the three species. We did not observe coronavirus in the Amazonian bats. Also, high diversity in the expression pattern in the different tissues was observed for all bat species.

Funding:

Link to Video:

’’’

Evaluation of lncRNAs as potential prognostic biomarkers in metastatic melanoma using machine learning

Cristiane Esteves Teixeira, Nicole de Miranda Scherer, Mariana Boroni, Guilherme Duarte de Moraes

Instituto Nacional de Câncer, INCA - Instituto Nacional de Câncer, Brazilian National Cancer Institute (INCA)

Abstract

Melanoma, despite having approximately 5% of incidence among skin neoplasms, is the most lethal of them. Although immunotherapies involving immune checkpoint blockade such as CTLA-4 and PD-1 have shown promise, there is a great variation in response to treatment, mainly related to genomic heterogeneity and immune infiltrates in patients. Therefore, prognostic biomarkers play a critical role in understanding disease progression and optimizing treatment and patient survival. Long non-coding RNAs (LncRNA) have already been described as prognostic biomarkers of several types of cancer, including cutaneous melanomas. In this sense, our work aims to investigate the role of lncRNAs as prognostic biomarkers specifically in melanoma patients who have metastasized, and to develop a model for predicting patient survival based on the expression profile of lncRNAs. We use machine learning (ML) techniques such as supervised machine learning algorithms on clinical-pathological data and gene expression of lncRNA from samples of patients diagnosed with metastatic melanoma obtained from the database TCGA (The Cancer Genome Atlas). A cohort of 348 patients was selected after filtering clinical data for metastatic samples only. Samples were divided into training and testing in a 4:1 ratio. We tested each of the 13, 954 lncRNA found in the samples using univariate cox regression to associate their level of expression with the overall survival of patients, resulting in 618 genes with $p < 0.05$. Ongoing analyses include the development of predictive models using three different regression algorithms (LASSO, Elastic Net and Ridge) in order to find a relevant lncRNA signature to classify patients according to the prognosis. The formula generated by the algorithms will be applied to calculate the risk of each patient and divide the cohort in low and high risk groups by the median. In addition, we will perform pathway enrichment analysis and generate a clinical nomogram.

Funding:

Link to Video:

’’’’

Modeling alelle-specific expression in complex polyploids

Agnelo Furtado, Antonio Augusto Franco Garcia, Robert James Henry, Gabriel Rodrigues Alves Margarido, Fernando Henrique Correr

Escola Superior de Agricultura "Luiz de Queiroz" / Universidade de São Paulo (ESALQ/USP)

Abstract

Allele-specific expression (ASE) represents the difference in the magnitude of expression between haplotypes of the same gene. Assessing ASE relies on identifying polymorphisms from genotypic data, whose allele expression levels are measured by high-throughput methods. Allelic imbalance occurs if the ratio of expression between two alleles shows deviations from their expected equivalent expression. However, this is not straightforward for polyploids, especially autopolyploids, as knowledge about the dosage of each allele is required for accurate estimation of ASE. This is the case for the genomically complex Saccharum species, characterized by high levels of ploidy and aneuploidy. Two species in this genus were the basis for developing sugarcane cultivars, which are interspecific hybrids. We propose a model to test for allelic imbalance in Saccharum that can be easily expanded to other polyploids. Our study is the first approach to assess ASE in a complex polyploid system using estimated allelic dosages. First, we identified SNPs and estimated allele dosages in a panel of Saccharum and other closely-related accessions. Then, we quantified the expression of each allele using sequenced libraries from leaves of six genotypes. To test for ASE in the i-th SNP of the k-th genotype, our null hypothesis was that the proportion of the reference allele from RNA counts (?ik) was equal to ratio of the dosage of this allele (Pik) in the genome. Our model followed a Beta-Binomial distribution in which the a priori distribution of ?ik was modeled by a Beta distribution using as parameters the dosage of each allele. To obtain the a posteriori distribution we used the Bayesian Markov chain Monte Carlo procedure, calling an ASE SNP if Pik was outside the high density interval of ?ik. We found that genes showing ASE were common in Saccharum, with highest frequencies in sugarcane hybrids. Genes with ASE were related to a broad range of processes, mostly associated to the general metabolism, organelles, responses to stress and responses to stimuli. Although many processes were specifically associated to particular genotypes, we found that conserved Liliopsida orthologs were significantly enriched with genes showing ASE. However, there was no significant enrichment among orthologs of Saccharinae or Saccharum. We then hypothesize that monocot core genes show ASE to preserve essential functions. These results provide evidence of the ASE importance in the evolution of Saccharum, justifying the maintenance of higher expression levels of some beneficial alleles.

Funding:

Link to Video:

”

Interactive Meta-analysis Framework for Biomarker Identification in Breast Cancer subtypes.

Milton Yutaka Nishiyama Junior, Ana Claudia Oliveira Carreira, Nathan de Oliveira Nunes

Instituto Butantan, Instituto Butantan

Abstract

Breast cancer is the most diagnosed type of cancer among women in the world, classified into 4 subtypes (HER2, LumA, LumB, Triple negative-TNBC), with TNBC being the most aggressive and lethal. We have developed an interactive meta-analysis framework to identify new common and exclusive molecular markers and signaling pathways in each study of breast cancer subtypes and in non-tumoral (MCF10A) and tumoral (Hs578T) cell lines with alteration of the expression of the CD90/Thy-1 gene. The interactive framework based on the shiny R package provides an intuitive and easy approach to compare multiple studies using gene expressions profiles (RNA-seq) in different conditions and able to parallelize the high-throughput data analysis. This framework is based on the steps and methods: M1: Load of data, followed by inter (RLE / TMM / UQ) and Intra (ComBat) sample normalization. M2: Classification of samples in subtypes/clusters. M3: Differentially expressed genes (DEGs) identification (edgeR), M4: Qualitative (ReactomePA) and Quantitative (FGSEA) pathway enrichment analysis, M5: Relevance and significance for genes (Ranking Product (RP)) and Pathways (Fisher / Stouffer methods) and M6: Connectivity between studies (TF-IDF and average precision). There are two meta-analysis approaches, one based on gene expression by M3 and p-value from enriched pathways by M4, followed by the meta-analysis methods in M5, and establishing the similarity matrix in M6, to measure the connectivity and average precision between the studies. The best normalization methods were established to correct bias and errors within and between samples. The classification approaches allowed us to cluster the samples in a biologically meaningful way. The DEGs or enrichment analysis provided genes/pathways information for further downstream meta-analysis. The methods of RP for the genes and Fisher for the pathways increased the identification in more than 80% of the TNBC canonical genes and pathways and confirmed the expected associations between the studies. Important and already described genes in the literature such as CGA, PLUNC and SMR3B have been identified in TNBC by our approach. Among the highlighted pathways, we found ECM and GPCR as one of most relevant between all breast cancer subtypes. We have developed a framework that can be applied to any set of studies, integrating a robust set of statistical methods and bioinformatics approaches in an user-friendly visualization tool aiming to identify the common, exclusive and promising genes and pathways for our studies (MCF10A/CD90+ and Hs578T/shCD90) compared to Breast cancer subtypes.

Funding: CNPQ

Link to Video:

,

Classification of mRNA and ncRNA sequences: a study based on complex networks and filter by exclusivity

Fabrício Martins Lopes, Murilo Montanini Breve

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ, Universidade Tecnológica Federal do Paraná (UTFPR)

Abstract

The large volume of biological data generated in recent decades and the need to analyze them to produce consistent discoveries has led to the development of a new area of knowledge, Bioinformatics. This is an interdisciplinary area that presents several important computational challenges related to Biology, one of them is the need to distinguish mRNAs and ncRNAs effectively. The correct identification of these RNA sequences is important due to the existence of thousands of non-coding RNAs, whose function and meaning are not yet known, as well as the challenge of understanding their genetic expression and possible regulatory action. This work adopts the complex network theory, which is being successfully used in many problems and, in several contexts. Thus, the proposed method consists of generating a complex network for each sequence of RNA, and then applying a filter in order to select the most exclusive edges of each class, increasing the distinction between the networks. For each filtered network, some topological measurements were extracted that are organized in a database, and thus used to classify each sequence in mRNA or ncRNA. For this purpose, the Random Forest classification method was adopted in the R project. Experiments were carried out to evaluate the proposed method considering a data set with six different species and comparing its acertivity with relevant methods in the literature such as CPC1, CPC2 and PLEK. The results indicated a high distinction between the RNA classes due to the filtering of the exclusive edges, which allowed the proposed method to reach average rates of accuracy higher than 98% in the mRNA and ncRNA classification considering all the adopted species. Finally, the proposed method presented less variations in its results when compared to the competing methods, indicating its adequacy for the classification of the RNA sequences. The application and a more in-depth study of this method can lead to a better understanding of non-coding RNA structure and as results its function.

Funding: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)
Link to Video:

,,,,,,

A wide tRNA-derived small RNA annotation in Astatotilapia latifasciata with B chromosome presence or absence.

Adauto Lima Cardoso, Rafael Takahiro Nakajima, Bruno Evaristo de Almeida Fantinatti, Rafael Luiz Buogo Coan, Cesar Martins, Jordana Inácio Nascimento Oliveira, Luiz Augusto Bovolenta

IBB - UNESP Botucatu, UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO, UNESP - Instituto de Biociências - Botucatu/SP, UNESP - Universidade Estadual Paulista

Abstract

Considered in the past years just as transfer RNAs (tRNAs) degradation products, tRNA-derived small RNAs (tRFs) are now considered an emerging class of small non-coding RNAs. tRFs are commonly associated with stress conditions and other functions are related to ribosome regulation, cell proliferation and differentiation, transposable elements control and coevolution of host and its associated microbiota, apart from several diseases. In turn, supernumerary B chromosomes (B), dispensable extra chromosomes found in several species of eukaryotes are a source of transcription of several RNAs. Some questions remain open and one of them is the Bs impact over gene regulation. To evaluate this, we developed a pipeline to the annotation of tRFs in Astatotilapia latifasciata fishes with the B chromosomes presence or absence (B+ and B-, respectively). We searched by homology among 4, 051, 654 tRNA sequences and B+/B- genome assemblies. All homologue sequences were submitted to tRNAScan-SE. Conserved genomic blocks between B+ and B- assemblies were identified using Mugsy. We obtained RNA-seq data from small RNA libraries extracted of encephalon, gonad, muscle and gill of male and female individuals with B+ or B-. All libraries contained biological replicates (N=3 encephalon and gonad; N=2 muscle and gill). We filtered the data by quality using FastQC and UEA sRNA workbench. Filtered reads were aligned to B+ tRNA sequences using Shortstack. tRFs sequence and read count were selected based on the secondary structure. TMM normalization was applied using the edgeR package. We applied the Generalized Linear Model, Quasi-Likelihood F-test and Shannon's entropy for pairwise comparison. Putative tRF-mRNA binding sites were predicted by homology. We annotated 1120 and 1007 tRNA genes in B+ and B- assemblies, a gene copies gain of 10% (113). In general, 5'-tiRs and tRFs-5c are the most impacted corresponding 31% (20) of differentially expressed among B+ and B-. We also observed that some tRFs may play different roles among tissues or sex. As an example, in the gill, the ala-tRNA-GluTTC-B-271-5tiR and ala-tRNA-GluTTC-B-271-tRF-5c play a putative contradictory role for male (up-regulated) and female (down-regulated) with B. We also observed variations in the entropy, B+ samples show a distribution more significantly extend than individuals B-. In specific tRF types, 5 and 3tiRs are more variable with B+ in the encephalon. In conclusion, our pipeline was capable of annotating and quantifying putative tRFs among B presence and absence individuals and indicated an expression variability gain correlated with B presence.

Funding: National Council for Scientific and Technological Development (CNPq) [grant number: 167444/2017-4]

Link to Video:

””

Bioinformatics analysis reveals novel long non-coding RNA candidates in Ewing sarcoma

Caroline Brunetto de Farias, André Tesainer Brunetto, Marialva Sinigaglia, Ney Lemke, Rafael Luiz Buogo Coan

UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO, ICI - Instituto do Câncer Infantil

Abstract

Long non-coding RNAs (lncRNAs) are defined as RNA molecules with more than 200 nucleotides, which don't encode proteins. They are present exclusively on the nucleus, cytoplasm or in both. They can interact with cellular components to form RNA-DNA, RNA-RNA and RNA-protein complexes, modulating gene wide expression. LncRNAs are also players in several diseases, including Ewing sarcoma (ES), which is a childhood malignant neoplasm that affects bones and soft tissues. A vital molecular alteration in ES is the translocation between chromosomes 11 and 22, resulting in the fusion protein EWS-FLI1, which acts as a transcription factor, altering genome-wide gene expression. In this study, our goal was to establish a bioinformatics pipeline to determine new lncRNA in ES patient samples. We used Illumina RNA-Seq data from dbGaP (phs000768v2p1) to identify novel lncRNA candidates in 26 ES patient samples consisting of EWS-FLI1 types I, II and III fusions. Raw reads were trimmed and quality filtered with Trimmomatic 0.36, then aligned to the human genome (hg38) with HISAT2 2.1.0. We then performed a new guided assembly on each sample with Stringtie 1.3.4, which was followed by merging the 26 new transcriptomes into a single file. We used the filter module from FEELnc 0.1.1 to exclude transcripts overlapping sense protein-coding genes from Gencode v.30 (150, 140 transcripts) and lncRNA from RNAcentral v.16 (554, 174 transcripts). After filtering, 527 candidate lncRNA remained, which were subjected to two coding potential estimators to computationally evaluate their protein-coding ability. We used FEELnc coding potential module and PLEK 1.2 for this task, only keeping the consensus transcripts found on both programs. A total of 459 transcripts lasted. We quantified the expression of the candidate lncRNA with Salmon v.1.1.0 and made between sample normalization with DESeq2. There are several novel transcripts with various levels of transcription, which may indicate a level of activity in ES. Our next steps include further genomic characterization of candidate lncRNA, plus in vitro and in vivo validation of potential transcripts involved in ES pathology. Although in early steps, our results show the potential of bioinformatics analysis to identify new candidate lncRNA that may be involved in ES biology.

Funding: CNPq 142129/2018-6

Link to Video:

””

Integrative analyses of single-cell and bulk transcriptomics to study the tumor microenvironment in high-grade serous ovarian cancer

Mariana Boroni, MARCO ANTONIO MARQUES PRETTI, Gabriela Rapozo Guimarães, Palloma Porto Almeida, Giovanna Resk Maklouf

INCA - Instituto Nacional de Câncer, Instituto Nacional de Câncer

Abstract

Ovarian cancer (OC) is the major cause of death related to gynecological tumors, with high-grade serous ovarian cancer (HGSOC) being the most common subtype of this disease accounting for the majority of deaths. The elevated mortality rates of HGSOC have been reported to be related to the interconnected signaling networks, cellular composition, and spatio-temporal localization of cells within the tumor microenvironment (TME). In this way, the aim of this project is to characterize the tumor microenvironment, integrating bulk and single-cell transcriptomics of HGSOC and investigate the impact of different subpopulations in patients' outcome. In order to accomplish that, single-cell RNA sequencing data (scRNA-seq) from public datasets were analyzed using the Seurat R package. Quality control, normalization, clustering, and differential gene expression analysis were performed to define the resulting clusters into cell types. Also, a thorough review of gene markers was made to support the annotation of the subpopulations. For a comprehensive assessment of the subpopulation's roles in a larger cohort, we used CIBERSORTx, a deconvolution tool that employs machine-learning to detect the abundance of subpopulations in bulk RNA-seq data based on a single-cell reference signature matrix. Subsequently, Cox Proportional Hazard Models were used with Survival and Survminer R packages to determine which cell type significantly influenced the patients' risk of death. Using the scRNA-seq data of 26, 786 cells derived from 5 patients, we have identified 24 clusters with distinct profiles within malignant, immune, and stromal major populations relevant to HGSOC. The cell types identified in the analysis comprised not only commonly observed subpopulations related to cancer, e.g. T CD4 and T CD8, but also more rare ones, e.g. pericytes and adipogenic fibroblasts. Finally, deconvolution with bulk RNA-Seq data of 454 patients from The Cancer Genome Atlas and International Cancer Genome Consortium databases revealed that cancer-associated fibroblasts, endothelial cells, and malignant cells were the most abundant subpopulations in HGSOC patients. However, regarding the hazard ratio (HR) related to the overall survival of the patients, the enrichment in T CD8 (HR = 0.82, p = 0.002) and T CD4 (HR = 0.85, p = 0.012) signatures were associated with a good prognosis, which corroborates with the high lymphocyte density as a common indicator of better outcomes at different stages of disease in many malignancies, including HGSOC. Our approach to explore the HGSOC TME will provide insights into how intratumoral content besides cancer cells can operate as a prevalent factor in the patient's prognosis, as well as offer potential diagnostic biomarkers and therapeutic targets in future clinical practice.

Funding: MS/INCA, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) e Chan Zuckerberg Initiative (CZI)
Link to Video:

,,,,,

Comparative midgut transcriptome analysis of Helicoverpa armigera feeding on natural conditions

Natalia Faraj Murad, Karina Lucas da Silva Brandão, Fernando Luis Cônsoli, Celso Omoto, Marcelo Mendes Brandão, André Ricardo Oliveira Conson

CBMEG/UNICAMP, UNIVERSIDADE ESTADUAL DE CAMPINAS, UNIVERSIDADE ESTADUAL DE CAMPINAS

Abstract

The cultivation of different annual crops may provide ideal conditions for feeding and survival of lepidopteran pests presenting generalist feeding habits. With occurrence in Brazil since 2013, one of the most important species from Noctuidae family is Helicoverpa armigera. It has a high capacity to tolerate many insecticides. Both private companies and public policy makers seeks to develop an integrated pest management approach to reduce the usage of pesticides and the emergence of populations. In addition, the main enzymes responsible for the digestive process in insects are peptidases involved in the initial digestion of plant proteins. Thus, to develop efficient ways to control pests, it is mandatory first to know which genes are involved in the digestive and detoxifying processes. Helicoverpa armigera individuals feeding on natural conditions were collected in order to characterize differentially expressed transcripts associated with soybean, cotton and bean diets. Total RNA from midgut was extracted and cDNA libraries sequenced (paired-end) using an Illumina Hiseq 2500. A de novo assembly of the short reads using both Mira and Trinity resulted in 145, 284 transcripts (1000 bp N50) and a length of 98.63 Mb. We identified 31 transcripts differentially expressed between dietary conditions. The largest number of differentially expressed transcripts was obtained in the cotton versus soybean contrast, where 26 transcripts were found up-regulated in cotton diet. From these, 11 transcripts were also found to be up-regulated in cotton diet relative to bean. Functional analysis showed that these transcripts are involved in biological processes like proteolysis, electron transport chain and lipid catabolic process. This is the first study of H. armigera transcriptome feeding under natural conditions and assembled transcripts are a powerful resource for future research promoting an improved understanding of the gene regulation of digestive peptidases.

Funding: Processo FAPESP 2018/19461-1

Link to Video:

””

Metformin regulates cells epigenomic landscape leading to decreased proliferation and inflammation in hepatocytes

Lucio Rezende Queiroz, Marcelo Rizzatti Luizon, Paula Silva Matos, Gloria Regina Franco, Izabela Mamede Costa Andrade da Conceição

UFMG, UNIVERSIDADE FEDERAL DE MINAS GERAIS, UFMG - Departamento de Bioquímica e Imunologia, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

Metformin is the first-line oral therapy for type 2 diabetes. It has been approved for use in other diseases, like polycystic ovary syndrome, obesity, and promising clinical trials for cancer. However, metformin's mechanism remains to be entirely elucidated, with its potent anti-aging, anti-carcinogenic, and epigenetic-regulator effects commonly seen but not thoroughly explained. We propose a mechanism for metformin's beneficial effect on inflammation and proliferation, hyper and down expressing transcripts which act as potent epigenetic regulators. We analyzed high-throughput RNA-seq data of primary human hepatocytes with the standard laboratory pipeline. Then, we selected transcripts that acted as epigenomic regulators according to our functional enrichment analysis and queried their translated sequences for the presence of whole domains. From all differentially expressed transcripts (DETs), six were present in epigenetic regulation pathways, four upregulated and two downregulated, all being protein-coding transcripts with their active domains present. The four upregulated DETs belong to the histone lysine demethylase (KDM) subfamily and use a JumonjiC (JmjC) domain that converts a-ketoglutarate (a-KG) to succinate during the demethylation process. High levels of succinate inhibit a-KG conversion, and metformin is known to reduce intracellular succinate levels, leading to increased activity of KDMs. At the gene-level, our candidate KDMs were also up-regulated in hepatocellular carcinoma (HCC), which does not comply with metformin proposed effects of reducing inflammation and proliferation. Articles that linked KDMs super expression to proliferation did not analyze at the isoform-level, and a previous study showed that KDM isoforms that retain the JmjC domain show an anti-carcinogenic effect. Contrastingly, the ones which increase proliferation are short isoforms that lost the JmjC domain. Therefore, according to our data Metformin would upregulate the anti-carcinogenic KDMs. The two other DETs were downregulated isoforms of the Methionine Adenosyltransferase 2A (MAT2A). MAT2A is the leading synthesizer of S-adenosylmethionine (SAM), the main cellular donator of methyl groups. On Liver it is mostly expressed in extra-hepatic tissues, while its paralogue, MAT1A, is present in hepatocytes. A switch in MAT1A:MAT2A ratio in hepatocytes is positively correlated to liver diseases, such as HCC and fibrosis. Metformin downregulating MAT2A leads to more presence of MAT1A, which increases SAM, leading to more methyl groups available. Our findings point towards a robust epigenetic regulatory axis controlled by isoform-specific differential expression induced by metformin. This mechanism leads to new understanding of metformin's role in the hepatic microenvironment and new ways in which those pathways can be targeted in hepatic disorders.

”

Expression Analysis Integration with Inferencing of Gene Regulatory Networks

Fabrício Martins Lopes, David Menotti, Juliana Costa-Silva

UNIVERSIDADE FEDERAL DO PARANÁ, Universidade Tecnológica Federal do Paraná (UTFPR)

Abstract

Application of next generation sequencing technology in cDNA sequencing (RNA-Seq), in transcriptomic studies has become suitable for transcript discovery, depicting mechanisms of gene regulation and differential gene expression analysis. Efforts to understand the complex network of reactions and influences that regulate the functioning of organisms involve an interdisciplinary research, in particular the development of computational models. As advances in generating biological data take place, computing plays a fundamental role in the processing of biological data. In this context, several works were produced with the objective of identifying the interaction networks between genes and their functionalities in the most diverse organisms. This work proposes the development of an interactive software tool analysis of RNA-Seq data. This software tool will create a stream to infer gene regulatory networks, based on differential expression analysis output, guaranteeing the order and a pattern of process. Besides, it will allow the user to choose among different techniques of analysis for each processing step, resulting in a useful tool for RNA-Seq analysis. It has adopted the Python language programming for development in order to make possible its use in an online page and make easy maintenance. The data analysis provided by the proposed approach is composed by three main steps: i) Expression analysis: BaySeq, edgeR, DESeq, DESeq2, EBSeq, limma-voom, SAMSeq, sleuth or consexpression can be used; ii) Network Inference: will use the results of the expression analysis to generate the inference of gene networks with RNA-Seq data, based on the premise that the generated networks are of the small world and scale-free type, considering the relationships indicated in previous studies. iii) Network validation and integration: To evaluate the characteristics of the generated genetic networks, KEGG tool and the ontology of this KO tool will be used. The association of the results with a functional annotation results in a complete analysis, which responds to a need for research in the area. We expect to produce a software tool for RNA-Seq data analysis and interpretation, as well as the spread of these analysis techniques. We consider that access to this tool can facilitate transcriptomic studies and help simplify its use in Bioinformatics training or in classes about RNA-Seq analysis.

Funding: CAPES

Link to Video:

,,,,,

De novo transcriptome assembly and functional annotation of cobia (Rachycentron canadum) liver

Bruno Cavalheiro Araújo, Alexandre Wagner Silva Hilsdorf, Renata G. Moreira, Luiz R. Nunes, Fabiano Bezerra Menegidio, David Aciole Barbosa

UNIVERSIDADE DE MOGI DAS CRUZES

Abstract

Rachycentron canadum (cobia) is considered the marine fish with the highest production potential in Brazil, mainly due to its fast growth, good feedstock conversion rate, and excellent fillet quality. However, the lack of genomic information hinders more comprehensive investigations in this species, a problem that could be minimized by de novo transcriptome assembly, providing data that could contribute to a variety of studies including conservation genetics, selective breeding, reproductive biology, and fish nutrition. In this report, we provide a de novo assembled transcriptome, using material from of R. canadum liver cells, highlighting the main functional annotations for this economically relevant fish. Briefly, 90 cobia juveniles (128.85 $\pm$ 18.43 g), obtained from a commercial hatchery, were randomly distributed in 3 tanks (2.000 L) under controlled temperature (23$\pm$1.5 oC) and photoperiod. Fish were equally hand-fed (twice a day) with commercial marine fish diet, during 6 weeks, anesthetized (4 g benzocaine in 10 ml ethanol), placed into 40 L of seawater, and euthanized by spinal cord section. Liver samples were frozen in liquid nitrogen and stored at -80 oC, until total RNA was extracted and transcriptome sequencing performed, using 150 bp (2 X 75) paired-end strategy, in an Illumina Nextseq sequencer. This analysis resulted in 1, 761, 965, 794 sequences, which were further processed in a Galaxy server (usegalaxy.eu). Libraries were submitted to quality control (FastQC and MultiQC) and low-quality reads(Q<30), adapters and other contaminant sequences were removed (Fastp), providing 1, 652, 319, 304 high-quality reads (93.8% of raw data), which were then used for de novo transcriptome assembly and metrics evaluation, using Trinity. This process resulted in the identification of 101, 789 unigenes and 163, 096 isoforms, with an average length of 1, 617.34 and 950.61 bp respectively. Median sizes (N50) were 7, 843 bp for unigenes and 2, 312 bp for isoforms. In total, 163, 096 transcripts were generated, with 95, 075 (58%) presenting more than 500 bp. Transcriptome completeness was assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO), identifying 81.7% of the universal complete actinopterygii genes (3746/4584, OrthoDB v9) supporting the high quality of our transcriptome assembly.. Eukaryotic Non-Model Transcriptome Annotation Pipeline (EnTAP) functionallly annotated 75, 554 transcripts (RefSeq: 25, 728 [34%]; Nr: 29, 155 [39%], Swiss-Prot: 15, 507 [21%]; EggNOG: 75, 068 [99%]), which were related to 75, 060 Gene Ontology (GO) terms (molecular function: 47, 766 [64%]; biological process: 48, 550 [65%]; cellular component: 34, 492 [46%]) and 24, 100 [32%] Kegg pathway assignments.

Funding: CAPES, CNPq, FAPESP

Link to Video:

”

Molecular characterization of Saccharomyces cerevisiae industrial strain PE-2 under high ethanol conditions in industrial fermenters

Marcelo Mendes Brandão, Flavia Vischi Winck, THIAGO SENA SIMOES

CBMEG/UNICAMP

Abstract

Ethanol production has received great investments in the world scenario, considering that biofuel is a renewable energy source. The production process embraced by the Brazilian industry is the biological one, using Saccharomyces cerevisiae yeast in the fermentation of the must obtained from sugarcane. New technologies are being developed with the aim of increasing the productivity of fermentative cycles, including Very High Gravity Fermentation (VHG), which uses high concentrations of sugars. However, this technique and the other procedures used in bioethanol plants subject yeasts to highly stressful conditions, reducing cell viability. The sugar and alcohol industry mainly use wild strains PE-2 and CAT-1, which are more resistant to the industrial environment. The elucidation of the various stress response processes is of great relevance for the selection and obtaining of the most efficient S. cerevisiae strains. The present work aims the molecular characterization of the PE-2 strain used by the Brazilian industry, analyzing gene expression and identifying the main active metabolic pathways during the fermentation cycle. The fermentative tests of the PE-2 strain of Saccharomyces cerevisiae were carried out under industrial conditions. Gene expression data were obtained at 0, 6, 12 and 18 hours from microarray experiments. The GO enrichment analysis was performed with topGO and GOrilla tools. The coexpression network was obtained using the R language. The analysis of the genes differentially expressed indicated that the expression profiles of the times 6h and 18h are closer to each other, when compared to 12h. In both cases, genes related to sexual reproduction (CDC31, SPS100, etc.) are activated. Within 6 hours, yeast cells are proliferating in an environment containing high levels of sugars and at 18 h, the sugar reserves were consumed and the cells are likely reproducing through spores, because these structures are more resistant to unfavorable environments. The expression profile at 12 h was more diverse and complex than the other samples and show a greater number of biological processes in the enrichment analysis. These results show that there is a relation between the fermentation period and specific genes being activated or deactivated. This fact likely occurred due to the fermentation kinetic and the conditions of the environment, considering the ethanol content at this point of fermentation process, which probably stimulated several stress response pathways. The expression of the GPD1, GPP1, AYR1 and SUT2 genes, related to lipid metabolism, was possibly a response to the high concentration of ethanol, which disorganizes cell membranes. Several genes related to Glycolysis / Glycogenogenesis are also active, such as TDH3, TDH2 and ENO2. This is an ongoing project aiming the identification of Biotechnological targets to enhance the production of bioproducts by using better the generated biomass per planted area.

Funding:

Link to Video:

,,,,,,

CHANGES INTHE EXPRESSION PROFILE OF TRANSCRIPT ISOFORMS IN MICE HEARTS CAUSED BY TRYPANOSOMA CRUZI INFECTION

Raphael Tavares da Silva, Tiago Bruno Rezende de Castro, Stellamaris Soares, Carlos Renato, Andréa Mara Macedo, Gloria Regina Franco, Nayara Evelin de Toledo

UNIVERSIDADE FEDERAL DE MINAS GERAIS, UFMG - Departamento de Bioquímica e Imunologia, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

Since the description of Chagas disease, caused by the protozoan parasite Trypanosoma cruzi, the reasons for its different clinical manifestations have yet to be completely revealed. Our group has previously shown that different strains of T. cruzi (JG- T.cruzi II and Col1.7G2- T. cruzi I) had a differential tissue distribution in BALB/c mice upon infection. Studies of differential gene expression, seeking to elucidate which host genes could be involved in this phenomenon, evaluated RNA-Seq data of BALB/c hearts infected with either JG, Col1.7G2 or a mixture of both strains and showed that Col1.7G2 is a stronger activator of the immune response, while JG effectively downregulates the oxidative stress response, basal metabolism and protein translation in the host. The mixture-infected group showed both profiles simultaneously. In this study we aimed to quantify differential alternative splicing, as well as identify differentially expressed transcript isoforms from the same transcriptomic data. By comparing Col1.7G2 infected mice with control group, we identified a total of 594 differentially expressed transcripts, including 543 upregulated and 51 downregulated. By comparing JG with control group, a total of 901 transcripts were considered differentially expressed, including 256 upregulated and 645 downregulated. Despite the greater number of protein coding biotypes, many non-coding transcripts from protein coding genes were found. Functional enrichment analysis showed that Col1.7G2 induced a higher inflammatory response while JG exhibit a weaker activation of immune response genes. Furthermore, JG-infected mice showed a reduction in expression of genes responsible for energetic metabolism, mitochondrial oxidative phosphorylation, and protein synthesis, corroborating our previous studies of gene expression analysis. Increase in splicing events were observed, including a rise in the number of skipping exons, retained introns and usage of alternate 5′ and 3′ splice sites. Our future steps include to correlate alternative transcriptome changes with expressed proteins identified by mass spectrometry.

Funding:

Link to Video:

,,

Yclon: Ultrafast clustering of B cell clones from high-throughput immunoglobulin repertoire sequencing data

Liza Figueiredo Felicori Vilela, Alice Ferreira de Souza, JOÃO HENRIQUE DINIZ BRANDÃO GERVÁSIO

UNIVERSIDADE FEDERAL DE MINAS GERAIS, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

Next-generation sequencing technologies is revolutionizing our understanding about immunoglobulin (Ig) profile in different immune states. Clonotyping (grouping Ig sequences into B cell clones) is a way of investigating the diversity of repertoires, and how they change upon antigen exposure. Despite its importance, there is no consensus on the best method for clonotyping and the methods developed for that is computationally intractable for large sequencing datasets. This is the case of Change-O, that uses hierarchical clustering for this task. Because of this, we propose here to implement an approach to identify B cell clones from Ig repertoire data, named Yclon, focusing on reducing the runtime and computer memory usage. To do that Ig sequences sharing the same V and J gene segments and the same CDRH3 length were grouped, transformed in n-grams and then into a vector, weighed with tf-idf metrics and compared pairwise using cosine similarity. The resulted square distance matrix was the input for the Hierarchical DBSCAN (HDBSCAN) or for an alternative hierarchical agglomerative method. To test these 2 different methods developed here we used 3 Ig repertoire datasets (dataset 1 contains 82.927 sequences, the 2, 365.370 sequences and the 3, 1.741.413 sequences) and compared the results with the ones obtained by Change-O. For the 15% most abundant clones, regarding dataset 1 our results with HDBSCAN showed that 74% of the sequences within these clones were shared with Change-O, meanwhile using the agglomerative approach we observed 96.7% of shared sequences. This process took 13 seconds (HDBSCAN) and 14 seconds (agglomerative), while Change-O took 912 seconds in a computer with 8Gb RAM, Intel i5 core. For the dataset 2, our result with HDBSCAN shared 92% of the sequences with Change-O results, while with agglomerative 87% of the sequences were shared. For this dataset, Change-O took 4.280 seconds and Yclon made it in only 112 seconds (HDBSCAN) and 94 seconds (agglomerative). Importantly, Yclon were able to process a repertoire with 1.700.000 sequences in 6.300 seconds (HDBSCAN) and 2.422 seconds (agglomerative), but Change-O couldn't process this large dataset. Overall, we find that 2 different clustering approaches developed here, grouped Ig sequences into B cell clones as similar as Change-O did. However, we observed that Yclon was around 70 times faster and even was able to process larger than 1 million sequences Ig repertoire which is a critical part of repertoire studies and enables understanding the structure and affinity maturation of the repertoire.

Funding:

Link to Video:

,,,

Where the pro-proteases PCSKs that may cleave the SARS-Cov-2 spike for improving cell to cell infection are highly differentially expressed in the human body?

Arthur Pereira Fonseca, Glaura da Conceição Franco, José Miguel Ortega, Lissur Azevedo Orsine

UNIVERSIDADE FEDERAL DE MINAS GERAIS, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

Comparative transcriptomics between tissues reveals genes that do not show counts in many but only in some tissues, and these are normally undertaken as Tissue-specific genes. However, there are genes commonly expressed in most or even all tissues, following a distribution that may not include some tissues, in which the expression level would rather be a far outlier. We developed an approach to programmatically detect Tissue-specific genes and genes that are expressed in all tissues but highly differently expressed in some, which are referred as Over-Active. Thus, our approach classifies as "GOAT" genes that is a "Gene Over-Active or Tissue-specific". Covid-19 is an important pandemic disease and all information about SARS-Cov-2 biology is relevant. The virus has gained a small site of cleavage in the spike protein which cleavage may be important to the cell to cell progress of infection. This cleavage might be undertaken by endogenous enzymes of the "substilisin/kexin like proprotein convertase" class, PCSKs. Here we used our approach to determine if PCSKs are GOAT in which tissues or if they are not highly differentially expressed in any of the tissues, using a set of five databases of comparative gene expression: ENCODE, FANTOM5, GTEx, HPA, and IBM, comprising information of, respectively, 13, 56, 53, 32 and 16 tissues.

PCSK1 is GOAT in brain cortex and meninges, hypothalamus, pineal and pituitary gland, therefore showing a clear bias through nervous system. PCSK2 is GOAT in brain and adrenal and thyroid glands. PCSK3 is best known as Furin, and its expression pattern is remarkably not differential, but besides showing different levels amongst tissues, the patter adjusts to a Gaussian distribution, suggesting biological variability with no special role in any tissue. Interestingly, PCSK4 also shows Gaussian expression along tissues, although being GOAT in testis. PCSK5, PCSK7

and PCSK8 (also referred as MBTPS1) profiles are not Gaussian, but also no tissue is outlier. Thus, the group PCSK3, PCSK4, PCSK5, PCSK7 and PCSK8 does not show a tissue of depicted expression besides PCSK4 in testis. PCSK6 is GOAT in liver and spleen, in two out of respectively four and three of the databases. PCSK9 is GOAT in lung and liver, in two and three out of respectively four and five of the databases, and appears as GOAT also in cerebellum, brain and pancreas in single databases.

In conclusion, our study claims attention to PCSK9 in lung, a target tissue, and liver, and neuronal tissues for PCSK1 and PCSK2. There is known inhibitor for PCSK9, the antibody Evolocumabe, but no inhibitor is known for PCSK1 and PCSK2.

,,,,,

Improvement of Angiostrongylus costaricensis genome annotation using RNA-Seq data

Karina Mastropasqua Rebello, Makedonka Mitreva, James McKerrow, Ana Gisele da Costa Neves Ferreira, Fabio Passetti, ESDRAS MATHEUS GOMES DA SILVA

IOC/Fiocruz

Abstract

Angiostrongylus costaricensis is a roundworm species that causes an intestinal inflammatory disease, known as abdominal angiostrongyliasis. The rodents are typically its definitive hosts, where they are usually found in the mesenteric arteries. Humans are accidental hosts, being contaminated by the ingestion of infective third stage larvae present on contaminated water and food. Currently, there is no drug available that acts directly on this parasite, mostly due to the sparce understanding of its molecular characteristics. Thus, aiming to provide a better understanding of its molecular aspects we present here the improved annotation of A. costaricensis protein-coding genes and transcripts using RNA-Seq data. First, the transcripts of both male and female adult worms were sequenced using RNA-Seq Illumina technology, generating short-paired reads. These RNA-Seq reads were aligned to the genome draft (version WBPS15) and used as extrinsic evidence for predicting protein-coding genes and transcripts, using the software BRAKER2. These predicted genes and transcripts were used to increment the WBPS15 genome annotation. The functional annotation of the complete ORFs of the WBPS15 improved was achieved using the software blast2GO from the OmicsBox package. The different gene expression (DGE) analysis between male and female worm genes was performed using the DESeq2 R package. The WBPS15 improved genome annotation comprises 14, 588 genes, 27, 788 mRNAs and 21, 584 complete ORFs. Overall, 72% of complete ORFs sequences were completely annotated by Blast2GO. It was identified 2, 573 genes more expressed in male and 747 genes more expressed in female worms, with adjusted P value (FDR) = 0.01 and Log2 fold change = 2 thresholds. Among the overrepresented terms of male are: non-membrane spanning protein tyrosine kinase activity (GO:0004715), protein kinase activity (GO:0004672) and phosphoprotein phosphatase activity (GO:0004721) and of female (among 7 GO terms over-expressed) are: protein tyrosine phosphatase activity (GO:0004725), mRNA binding (GO:0003729) and protein phosphatase 1 binding (GO:0008157). We believe that this improved genome annotation of protein-coding genes and transcripts contributes to better understand the molecular diversity of A. costaricensis, being this is a key step for the selection of therapeutic proteins.

Funding: CAPES

Link to Video:

,,,,,,,

A novel mitovirus from the sand fly Lutzomyia longipalpis shows sRNA profiles consistent with siRNA pathway activation

Flavia Viana Ferreira, Felipe Ferreira da Silva, Liliane Santana Oliveira Kashiwabara, João Trindade Marques, Aristóteles Góes-Neto, Eric Roberto Guimarães Rocha Aguiar, Arthur Gruber, Paula Luize Camargos Fonseca

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ, UNIVERSIDADE FEDERAL DE MINAS GERAIS, UNIVERSIDADE DE SÃO PAULO, UNIVERSIDADE ESTADUAL DE SANTA CRUZ

Abstract

Hematophagous insects act as major reservoirs of infectious agents due to their intimate contact with a large variety of vertebrate hosts. Lutzomyia longipalpis is the main vector of Leishmania chagasi in the New World, but its role as a host of viruses is poorly understood. In this work, publicly available L. longipalpis RNA libraries were subjected to progressive assembly using viral profile HMMs as seeds. A viral sequence presented a size distribution of small RNAs consistent with the activation of the siRNA pathway. This sequence 2, 697-base corresponds to a monopartite ssRNA(+) genome of a virus called Lul-MV-1. A single ORF encoding an RNA-directed RNA polymerase covers almost the entire genome and uses a typical organellar genetic code with tryptophan being mostly coded by UGA. A phylogenetic analysis positioned Lul-MV1 in a monophyletic clade composed of mitoviruses mostly found in fungal, but also in crustaceans. To determine whether the virus was infecting a fungus from the sand fly microbiota or the phlebotomine itself, we analyzed some molecular characteristics of the genome. Dinucleotide composition and codon usage showed profiles similar to mitochondrial DNA of invertebrate hosts. Also, base preference and size of sRNAs were analogous to those observed in viruses that infect sand flies, suggesting that L. longipalpis is the putative host. Finally, RT-PCR of different insect pools confirmed the presence of Lul-MV-1 in seven out of eight tested samples. Concluding, the strategy used in this work permitted to identify and characterize for the first time of a mitovirus infecting an insect host.

Funding:

Link to Video:

,,,,,,,,

Autophagy related genes influence global survival in Ewing's sarcoma

Mauricio Gomes, Ricardo M. Ferreira, Caroline Brunetto de Farias, André Tesainer Brunetto, Mariane da Cunha Jaeger, Rafael Roesler, Marialva Sinigaglia, Rita Maria Cunha de Almeida, Matheus Dalmolin

UFRGS, UFRGS, ICI - Instituto do Câncer Infantil

Abstract

Ewing's sarcoma (ES) is a highly aggressive tumor, affecting the bones or soft-tissues, being the second most common pediatric bone neoplasia. It is characterized by chromosomal fusion involving the EWS gene and transcription factors of the ETS family, usually FLI1. Although treatment for localized disease is proven to be effective, the long-term survival of patients with metastatic or recurrent ES is still very low. Thus, a detection of prognostic markers of ES outcome at the time of diagnosis, could be oriented towards a more effective and ES outcome at the time of diagnosis could be oriented towards a more effective and individualized treatment protocol according to the tumor's aggressiveness profile. The analyzes were performed using public data available from the Gene Expression Omnibus (GEO). Gene expression data from ES biopsies, collected at the time of diagnosis and containing metadata about the outcome of patients: COG (GSE63155), EuroEwing (GSE63156) and Italians (GSE17679) were analyzed. Each cohort was classified into two groups: (i) SOB (survivors), overall survival greater than five years and (ii) NSOB (non-survivors), overall survival less than five years. The comparison between the two outcomes was performed using The Transcriptogramer V.1 software, other analyzes were performed in the R environment. There was a significant difference in the gene expression profile between the groups (SOB and NSOB) of the COG and Italian cohorts. The group of differentially expressed genes from each of the two cohorts was processed through several steps, available in a list of genes. From this list, 43 genes have their expression levels associated with overall survival (Kaplan Meyer with $p < 0.05$) validated in the three independent cohorts. The protein-protein interaction network (string-db version 11) between 43 genes was generated, and 11 genes were added to connect all nodes in the network. Gene Ontology (GO) terms related to autophagy, such as autophagy and macroautophagy had the largest number of representatives (17 and 16) within the network and the lowest adjusted p-values (<0.0001). Autophagy was also present among Kegg and Reactome pathways. These findings indicate that autophagy is a very relevant process within our network and can be a potential marker of ES outcome at the time of diagnosis.

Funding: Intituto do Câncer Infantil Instituto do Câncer Infantil e Pronon - SIPAR: 25000.202.751/2016-65

Link to Video:

9 — Systems Biology and Networks

,

Dynamics of the feedback loops required for the phenotypic stabilization in the epithelial-mesenchymal transition

José Carlos Merino Mombach, Daner Acunha Silveira

UNIVERSIDADE FEDERAL DE SANTA MARIA

Abstract

The epithelial-mesenchymal transition (EMT) is a complex mechanism in which cells undergo a transition from epithelial to mesenchymal phenotypes (there is also an intermediary hybrid state) in response to microenvironmental alterations and aberrant stimuli triggered by molecules such as TGF-β. Recent studies in breast cancer progression reported new feedback loops and new participant molecules such as microRNAs 340 and 1199. In this work, we propose a logical model of EMT contemplating the influence of these new published molecules on the regulatory core of EMT. The model results were compared with theoretical and experimental data for the human breast epithelial cell line MCF10A presenting excellent agreement. We propose that the miRNAs 340 and 1199 should be considered phenotypic stability factors of the hybrid state based on the positive feedback loops they form with ZEB1. Our results highlight new mechanisms related to the EMT dynamics in response to TGF-β stimulus in epithelial breast cells and might help the design of therapeutic strategies for breast cancer.

Funding:

Link to Video:

,

Study on Tsallis Entropy applied on Network Inference

Fabrício Martins Lopes, Cassio Henrique dos Santos Amador

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ, Universidade Tecnológica Federal do Paraná (UTFPR)

Abstract

The inference of gene networks is still an open field for research. One of the different methods to evaluate the connection in a network is based on entropy. Previous works showed, through a series of experiments, that non-extensive entropy, or Tsallis entropy (TS), a type of theory generalization, give better results then Boltzmann entropy, if you use conditional entropy with the q parameter around 2.45. This parameter, when it is equal to 1, makes TS returns the same entropy as the Boltzmann one. In this work we prove analytically that systems with binary discretization (for example, when a gene is either ON or OFF) shows a greater ratio of signal to noise when evaluated with non-extensive entropy with q around 2.45. Also, we present a method to use non-extensive entropy to study systems with discretization values higher then 2, with the analysis of E. Coli gold standard network as an example.

Funding:

Link to Video:

,,,

Network-based identification of subtype-specific candidate genes and associated drugs for new therapies in colorectal cancer

Nicole de Miranda Scherer, Luís Felipe Ribeiro Pinto, Mariana Boroni, Cristóvão Antunes de Lanna

INCA - Instituto Nacional de Câncer, Instituto Nacional de Câncer

Abstract

Colorectal cancer (CRC) is the fourth most incident carcinoma worldwide, being the second in Brazil. Incidence is related to hereditary factors, eating habits, overweight and obesity, and physical inactivity. The variety of etiologic factors results in highly heterogeneous tumors with distinct prognosis and response to treatment. Different classification strategies have been proposed to characterize tumors more efficiently. The Colorectal Cancer Subtyping Consortium (CRCSC) recently identified four consensus molecular subtypes (CMS1-4) from primary CRC transcriptomic data. Identification of disease-related genes with high potential for drug interactions may assist in discovering new targets and more effective therapeutic strategies. This enables repositioning of previously approved drugs to treat other diseases and may reduce the time required to approve new treatments. However, few works have assertively proposed new treatments based on this classification system. For this reason, the aim of this work is to identify candidate genes and associated drugs for the development of new therapies for different molecular subtypes of colorectal cancer from large-scale genomic and transcriptomic data. Gene expression data from 623 patients generated by The Cancer Genome Atlas (TCGA) were used, totaling 623 samples from primary tumor tissue and 51 from tumor-adjacent tissue. Tumor samples were classified into 4 groups using the CMSClassifier package, with posterior subdivision of CMS4 samples into epithelial and stromal. Unique differentially expressed genes (DEGs) in each CMS subtype were identified with DESeq2 and InteractiVenn. Co-expression modules were constructed using weighted gene correlation network analysis (WGCNA), correlated with subtypes and normal samples, and used in the construction of protein-protein interaction networks using the STRING base. Interactions with low confidence were filtered out and subgraphs were identified within each module based on modularity using the igraph package. Candidate target genes were selected based on degree, closeness betweenness, and pagerank centralities. Each of these centralities was measured, converted to z-score, and combined with log2 fold change. For each measure, the top 20 genes with the highest score were selected for drug-gene interaction identification in the DGIdb database. Drug-gene interactions were validated using tumor-derived cell line sensitivity data from Genomics of Drug Sensitivity in Cancer (GDSC), classified into CMS subtypes from expression data available from the Gene Expression Omnibus (GEO) database using CMScaller, a cell line-specific classifier. Subtype-specific drug-gene interactions as well as interaction overlaps were evaluated in order to propose candidate drug combinations for further testing. These results demonstrate the potential for the evaluation and implementation of new therapeutic strategies in CRC and the possibility of implementing these analyses in other tumor types.

Funding: Capes, Ministério da Saúde

Link to Video:

,,,,,,

Multi-omics-based identification of SARS-CoV-2 infection biology and candidate drugs against COVID-19

Debmalya Barh, Marianna E. Ivanova, Vasco A de C Azevedo, Aristóteles Góes-Neto, M. Michael Gromiha, Preetam Ghosh, sandeep tiwari

INAPG- França, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

SARS-CoV-2 has ushered a global pandemic with no effective drug being available at present. Although several FDA-approved drugs are currently under clinical trials for drug repositioning, there is an on-going global effort for new drug identification. Here, using multiomics (interactome, proteome, transcriptome, and bibliome) experimental data and subsequent integrated analysis, we present the biological events associated with SARS-CoV-2 infection and identified several candidate drugs against this viral disease. We found that: (i) Interactome-based infection pathways differ from the other three omics-based profiles. (ii) viral process, mRNA splicing, cytokine and interferon signaling, and ubiquitin mediated proteolysis are important pathways in SARS-CoV-2 infection. (iii) SARS-CoV-2 infection also shares pathways with Influenza A, Epstein-Barr virus, HTLV-I, Measles, and Hepatitis virus. (iv) Further, bacterial, parasitic, and protozoan infection pathways such as Tuberculosis, Malaria, and Leishmaniasis are also shared by this virus. (v) A total of 50 candidate drugs including the prophylaxis agents and pathway specific inhibitors are identified against SARS-CoV-2 infection. (vi) Anticancer antibiotics, steroids, Estrogen, analgesics, antipsychotic drugs, anticholesteremics, antihemophilic factors, and immunosuppressants are the key drug categories. (vii) Ozone, Nitric oxide, and photosensitizer drugs are also identified as possible therapeutic candidates. (viii) Curcumin, Retinoic acids, Vitamin D, Arsenic, Copper, and Zinc may be the candidate prophylaxis agents. Nearly 80% of our identified agents are suggested to have anti-COVID-19 effects or under clinical trials. Our identified drugs, that are not yet tested, need validation with caution while an appropriate drug combination from these candidate drugs along with a SARS-CoV-2 specific antiviral agent is needed for effective COVID-19 treatment.

Funding:

Link to Video:

,,,

Comparative analysis of genetic networks of Anticarsia gemmatalis Hübner, 1818 (Lepidoptera: Erebidae) obtained from transcriptomes of strains resistant and susceptible to the protein Cry of Bacillus thuringiensis

Laurival Antônio Vilas Boas, Fabrício Martins Lopes, Rogério Fernandes de Souza, Freddy Eddinson Ninaja Zegarra

Universidade Tecnológica Federal do Paraná (UTFPR)

Abstract

The larvae of the pest Anticarsia gemmatalis Hübner, 1818 (Lepidoptera: Erebidae) are capable of generating serious economic losses in the soybean industry and, although efficient methods are known to combat them, such as chemicals, less aggressive techniques are chosen to human health. Biological technology that uses organisms or by-products of biocontrollers, such as Cry toxins from Bacillus thuringiensis, is capable of acting efficiently and specifically on the insect pest without generating serious environmental impacts. However, cases of resistance of A. gemmatalis to Cry toxin have been reported. Therefore, the objective of this project is to perform a comparative analysis of genetic networks of transcripts of A. gemmatalis obtained from a resistant colony and another sensitive to the Cry protein of B. thuringiensis subsp. Kurstaki Lineage HD-73. The project is based on the data set of the work of Forim et al., 2017, which reports information on transcriptomes of A. gemmatalis resistant and susceptible to this toxin. Each treatment and control will be carried out in triplicate for resistant and susceptible, resulting in 12 experimental units. Our consensus software are adopted to analyze differentially expressed genes. The produced data set (active and inhibited genes) are performed adopting the DimReduction software to infer the genetic networks. Subsequently, the metabolic pathways corresponding to the target genes are investigated using the computational tools of the KEGG database. As a result, some relevant genes in the process of resistance and susceptibility to Cry protein are identified, as well as the metabolic pathways in which they participate.

Funding:

Link to Video:

,,

txGeneNetwork: transcript-level functional analysis using network biology

Lucio Rezende Queiroz, Gloria Regina Franco, Izabela Mamede Costa Andrade da Conceição

UFMG - Departamento de Bioquímica e Imunologia, UFMG, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Abstract

The differential expression analysis outcomes are highly variable due to the vast amount of differentially expressed genes from a single transcriptomics experiment. It is a challenging task to correlating these genes with a concise biological function or molecular pathway. Multiple gene set enrichment (GSE) methods were developed to address that problem. However, most GSE analyses result in hundreds of enriched gene sets, with considerable overlap of each gene set's genes. Thus, defining biologically relevant pathways that correlate with a given study design is a challenging task. Another common issue in most gene expression analyses is that gene-level only quantification is performed. A gene can produce multiple distinct coding and non-coding transcripts through alternative splicing, a process present in all multicellular eukaryotes. Gene-level summarized quantification sometimes leads to incorrect conclusions about a gene's contribution in a biological process, ignoring individual isoform contribution to the analyzed condition. Therefore, integrating transcript-level expression could provide invaluable insights into the significant contribution of a given gene. Current methods and workflows based on biological networks used to visualize gene expression data and integrate with GSE do not explore the relationships between genes and their isoforms, especially non-coding transcripts. An additional consideration is that most network biology workflows consist of several steps in different tools, accessed in variable interfaces, that can be hard to reproduce. We present a comprehensive, freely available, and open-source R/Bioconductor workflow to access and interpret transcripts' functional relevance using network biology. We correlate the differentially expressed transcripts and genes from transcript-level expression data, also performing GSE analysis based on transcripts effect size. Different methods are offered for functional enrichment analysis. We then use a graph-based network approach to interconnect the relations between pathways, related genes, and the transcript isoforms which arise from those genes. We then integrate isoform subtypes, the direction of their expression, properties, and topology of the network with other network science metrics to rank biological processes and transcript relevance. The final network forms an intuitive and functional visualization for GSE analysis. Its topology could bring insights into transcript function, isoform switch, and integration between gene sets, helping unravel transcriptomic signatures for coding and non-coding transcripts.

Funding:

Link to Video:

,,,,,,

Phylogenetic and phenotypic relationships of the Triatoma sordida subcomplex (Hemiptera: Reduviidae: Triatominae)

Jader de Oliveira, Heloisa Pinotti, Lucas Abrantes da Silva, Kaio Cesar Chaboli Alevi, Cleber Galvão Ferreira, João Aristeu da Rosa, Tiago Belintani

UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO, IB-UNICAMP

Link to Video

,,,,,,,,

The nuclear and mitochondrial genomes of Frieseomelitta varia – a highly eusocial stingless bee (Meliponini) with a permanently sterile worker caste

Flávia C. de Paula Freitas, Anete P. Lourenço, Daniel Guariz Pinheiro, Francis M. F. Nunes, Marcia M. G. Bitondi, Zila L. P. Simo es, Angel Roberto Barchuk, Klaus Hartfelder, Alexandre R Paschoal

UTFPR- Universidade Tecnológica Federal do Paraná

Link to Video

,,,,,

Poor clinical outcome in metastatic melanoma is associated with a microRNA-modulated immunosuppressive tumor microenvironment

Natasha Jorge, MARCO ANTONIO MARQUES PRETTI, Martín Hernán Bonamino, Patricia Abrão Possik, Mariana Boroni, Jéssica Gonçalves Vieira da Cruz

IOC/Fiocruz, INCA - Instituto Nacional de Câncer

Link to Video

,

FERRAMENTA BASEADA EM CUCKOO FILTER PARA REMOÇÃO DE REDUNDÂNCIA EM DADOS DE SEQUENCIADORES DE SEGUNDA GERAÇÃO (NGS - NEXT GENERATION SEQUENCING)

Adonney Allan de Oliveira Veras, Antonio Sérgio Cruz Gaia

UNIVERSIDADE FEDERAL DO PARÁ

Link to Video

Using structural signatures to propose mutations in β-glucosidase enzymes used in the production of biofuels

Diego Mariano

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Link to Video

Epidemiological and genomic evaluation of chikungunya virus circulating in Rio de Janeiro

Joilson Xavier dos Santos Junior

Ufmg- Universidade Federal de Minas Gerais

Link to Video

”’

BIOINFORMÁTICA NA INVESTIGAÇÃO DE RNAS NÃO-CODIFICANTES E ELEMENTOS TRANSPONÍVEIS EM PLANTAS

Douglas Silva Domingues, Alan Péricles Rodrigues Lorenzetti, Daniel Longhi, Alexandre R Paschoal

UNIVERSIDADE DE SÃO PAULO, UTFPR- Universidade Tecnológica Federal do Paraná

Link to Video

,,,,,,

Evolutionary analysis based on the Pasteurella multocida genome from Veterinary isolates

Marcus Vinicius Canário Viana, Dennis Carhuaricra, Siomar de Castro Soares, Lenin Maturrano, Vasco A de C Azevedo, Flavia Figueira Aburjaile, Raquel Enma Hurtado Castillo

INAPG- França, UNIVERSIDADE FEDERAL DE MINAS GERAIS, IOC/Fiocruz, UNIVERSIDADE FEDERAL DE MINAS GERAIS

Link to Video

”’

PRIORITIZING PROMISING COMPOUNDS IN VIRTUAL SCREENING CAMPAIGNS

Michael Keiser, Rafaela Ferreira, Raquel Cardoso de Melo Minardi, Alexandre Victor Fassio

ucsf, UNIVERSIDADE FEDERAL DE MINAS GERAIS, UNIVERSIDADE DE SÃO PAULO

Link to Video

Index of Authors