Abstract Book

X-meeting BSB 2013-11-10

Mar Hotel

Recife 3-6 November, 2013

Topics:


X-MEETING BSB 2013
INTERNATIONAL CONFERENCE OF THE RSBC & BRAZILIAN SYMPOSIUM ON BIOINFORMATICS
RECIFE - BRAZIL - 03-06 NOV 2013

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 2

TS-primer: a tool for the identification of PCR-based taxon-specific markers

Gabriela Flavia Rodrigues-Luiz[1], Hugo Oswaldo Valdivia Rodriguez[1], Edward Valencia[1], Robson da Silva Lopes[1], Rodrigo de Almeida Lourdes[1], Thiago de Souza Rodrigues[2], Ricardo Toshio Fujiwara[1], Daniella Castanheira Bartholomeu[1]

[1]Universidade Federal de Minas Gerais, Brazil, [2]Centro Federal de Educação Tecnológica de Minas Gerais, Brazil

Molecular markers represent one of the most powerful tools for the analysis of genomes and have been widely used in diagnostic applications of clinical and environmental studies. One of the most efficient molecular markers is the polymerase chain reaction (PCR) based as they are suitable to high throughput automation and confer high specificity. However, the design of the taxon-specific primers may be difficult and time-consuming, due to the need to identify appropriate DNA regions for the annealing and evaluation of specificity. Microsatellites or Single Sequence Repeat (SSR) are tandem repeated stretches of short nucleotide motifs, usually ranging from 1 to 6 bp, ubiquitously distributed in the genomes of eukaryotic organisms. These regions are more prone to genetic variation and the differences in the length of individual SSR loci can be easily screened by PCR. In fact, this technique has been useful for several studies of strain typing and population genetics. Orthologs are homologous proteins that are related by speciation events and tend to show more functional similarity than other homologs. Orthologous gene sets are used to obtain information about evolutionary conservation and variability of molecular sequences. Leishmania is a genus of flagellate protozoan that cause a broad spectrum of diseases, ranging from self-limiting localized cutaneous lesions to visceral leishmaniasis. Multilocus enzyme electrophoresis (MLEE) has been the gold standard marker for taxonomic studies and strain typing of Leishmania, although this method has several limitations, including the relatively small number of characterized loci and alleles and the requirement of parasite culture. In this work we have developed a pipeline to search and design specie-specific primers targeting the SSR and ortholog sequences identified in genomic data for genotyping using Multiplex. To identify the SSRs we developed the program NT-RepeatFinder, whereas the orthologs groups were obtained from the automated ortholog identification tool, OrthoMCL. To design the primers we have written a script using PERL language and commands of the EMBOSS package, as well as the third-party software e-PCR (Electronic PCR), which was used to identify the best primers. By applying this protocol in the genomes of three species of Leishmania (L. infantum, L. braziliensis and L. major), we generated 6355 sets of primers for L. major genome, 15421 for L. braziliensis and 6324 for L. infantum. Six sets of primers, one set for each species in each approach, were tested by PCR using artificial DNA mixtures of the three Leishmania species as template. These results validated the pipeline since amplification profiles allow discriminating

genomic DNA samples from the three species. This automated bioinformatic pipeline allows the large-scale generation of taxon-specific primers and we are currently developing a web service to provide this application to the scientific community.

Keywords: Bioinformatics; Biomarker; Molecular Diagnostic; Leishmania; Software

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 3

GENETIC DATA SUGGEST THAT ANDEAN QUECHUA AND ARAWAK MATSIGUENGA FROM THE LOWLAND TROPICAL FOREST DIVERGED LESS THAN 5000 YEARS AGO

Mateus Gouveia[1], Marília Scliar[1], Andrea Benazzo[2], Silvia Ghirotto[2], Nelson Fagundes[3], Thiago Leal[4], Wagner Magalhães[4], Latife Pereira[4], Maira Rodrigues[4], Giordano Souza[4], Lilia Cabrera[5], Douglas Berg[6], Robert Gilman[7], Giorgio Bertorelle[8], Eduardo Tarazona[9]

[1]Universidade Federal de Minas Gerais, Brazil, [2]Dipartimento di Biologia ed Evoluzione. Facoltà di Scienze. Università di Ferrara. Italy., Italy, [3]Universidade Federal do Rio Grande do Sul, Brazil, [4]Universidade Federal de Minas Gerais, Brazil, [5]Asociación Benéfica PRISMA, Peru
[6]Washington University School of Medicine, United States, [7]Universidade Peruana Cayetano Heredia, Peru, [8]Dipartimento di Biologia ed Evoluzione. Facoltà di Scienze. Università di Ferrara, Italy, [9]Universidade Federal de Minas Gerais, Brazil

Little is known about the dispersion of first South Amerindians populations and its posterior evolution. We studied a Quechua population from the Peruvian Andes and a Matsiguenga population (Shimaa) from the Peruvian Montaña (a geographic area between the Andes and the Lowland Amazonian). In addition to be settled in very different environments, these populations speak languages from different families (Pan-Andean Quechua and Arawak Matsiguenga) and have different cultures linked to their Andean and Amazonian lifestyles. We resequenced 10 independent genomic regions for a total of ~20 kb per individual in 10 Quechua and 10 Shimaa, and used two model-based bayesian approaches, the Isolation with Migration program and the Approximate Bayesian Computation method, to infer a set of demographic parameters of a model. Our data elucidate a high diversity in Quechuas, and a little diversity in Shimaas, which is a subset of the diversity found in Quechuas (all Shimaas´ SNPs found are shared with the Quechuas). The demographic parameters estimated, suggest that the Quechua and the Shimaa populations diverged less than 5000 years ago from a large ancestral population, most of which founded the Quechua population, while a small fraction founded the Shimaas. The results shows a recent divergence between an Andean population and an Amazonian population, that is, the divergence happened after the peopling of South America and around or after the time known in the Andean history as the start of the development of complex societies. The more plausible explanation for this scenario is that an Andean population has been established in the Peruvian Montaña and over time incorporated the Arawak culture and language of their neighbors, but not their genes. Further studies are needed to show if this is true only for the Shimaas or for other Peruvian Montaña populations.

Keywords: Evolution, ABC, Quechua, Shimaa

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology,Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 6

Application of an entity-feature data mining model to novel drug-drug interaction prediction

[1]Felipe Ferre, [2]Francisco de Assis Acúrcio, [3]Wagner Meira Júnior

[1]Doctoral Program in Bioinformatics. Federal University of Minas Gerais., Brazil, [2]Social Pharmacy Department. Federal University of Minas Gerais., Brazil, [3]Computer Science Department. Federal University of Minas Gerais., Brazil

Background: Bioinformatics and in silico experiments are promising resources for evidenced-based medicine, including drug-drug interaction (DDI) studies. Given the large number of medicines and possibilities for polypharmacy, many drug adverse events such as DDIs are only identified before broad usage by populations. Predictive models can use the knowledge available in large bioinformatics and public health databases to establish novel information about unknown DDIs. Objective: We aim to make a comprehensive evaluation of the entire set of drug combination in pairs for prediction of novel DDI and safe associations using supervised machine learning techniques. Methods: We proposed an integrative binary entity-feature data mining model that determines the most important characteristics of individual drug data using a distance measure for all possible drug-pairs, given a dataset of drugs. We found intrinsic drug patterns in DrugBank combined with ATC/WHO corresponding to a feature profile of known DDIs on drugs.com. We performed supervised machine learning to construct predictive functions using ten fold cross validation. We also used Singular Value Decomposition and feature selection. The relevance of the DDIs predicted was evaluated according to PubMed citations of drug-pairs names. Results: From 1390 drugs and 18,340 known DDIs (training set), the RandomCommittee classifier yielded kappa=0.871, precision=0.959, and 0.985 for the area under the ROC curve. From 947,015 considered unknown DDIs, 12,482 possible DDIs were classified (26.0\% with pubmed citations). Conclusion: This model represents a consistent source of previous evidence for clinical or epidemiological studies.

Keywords: Computational Biology, Artificial Intelligence, Data Mining, Drug Interactions

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 7

Improving the recovery of next-generation sequencing data: an metagenomic approach, as from Anopheles aquasalis libraries, to survey the DNA virosphere and microbiota associated

[1]Luis Eduardo Villegas, [2]Laura Leite, Flavio Araujo, [2]Anna Salim, [2]Guilherme Oliveira, [3]Paulo Pimenta

[1]Laboratory of Medical Entomology , Genomics and Computational Biology Group (CEBio), Brazil, [2]Genomics and Computational Biology Group (CEBio), Brazil, [3]Laboratory of Medical Entomology, Brazil

Human malaria transmission depends upon the completion of the complex Plasmodium cycle within the anopheline vector. Amongst the ~400 species known, only about 30 act as malaria vectors. The vector-parasite interaction is highly complex due to the physical and humoral innate immune responses that challenge the parasite through its journey within the Anopheles. The composition of the vector gut microbiota is one of the major components that determine the outcome of mosquito infections. As observed between the Wolbachia endosymbiont and Brugia microfilariae in Aedes aegypti, the microbial flora in Anopheles gambie stimulate a basal immune activity with known anti-Plasmodium activity. The biological basis of the interaction between american anophelines such as Anopheles aquasalis and Plasmodium parasites are hardly known in depth. This is a direct consequence of the fact that malaria in the New World is a neglected disease by the International aid agencies. In order to maximize the data of the Massive Parallel High Throughput sequencing performed to commence the genome building process of this new model, we used unmapped high quality SOLiD generated reads as a starting point to survey the DNA viral diversity associated with the aquatic stages of the mosquito. We were able to identified (in silico) short viral sequences pertaining to aquatic/marine environments, plant and animal hosts, bacteriophages and insect viruses. To further explore this findings, we submitted longer sequences generated with the Ion Torrent platform to the online metagenomic server MG-RAST. Through this shotgun metagenomics approach we identified marine and bioluminescent bacteria from the Aeromonadaceae, Pseudoalteromonadaceae, Shewanellaceae, Enterobacteriaceae, Vibrionaceae, Victivallaceae families, and the Aeromonas phage phiO18P. In relation to functional profile, we identified bacterial flavohemoglobins, suggesting a beneficial role in maintaining gut redox homeostasis. We believe this viral and microbial sequence survey could be of help for future epidemiological surveillance efforts as well as to explore the impact that bacteria and viruses could have in the Anopheles aquasalis innate immune response and vectorial competence. Finally, in an era where data are generated with higher efficiency than processed, a better connection between the data generated and the biological models is the biggest challenge in nowadays metagenomics. In this study besides the presentation of preliminary results on the virosphera and microbiota associated it was possible to plan a better strategy for future vector related metagenomic projects.

Keywords: metagenomics, gut microbiota, malaria vector

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 8

GASTRIC ADECARCINOMA IN THE PERUVIAN POPULATION: ASSOCIATION AND GENE EXPRESSION STUDIES OF INFLAMMATORY GENES

[1]Thaís Queiroz, [1]Roxana Zamudio, [1]Latife Pereira, [1]Fernanda Freire, [1]Camila Sá, [1]Hanaisa Sant'Anna, [2]Phabiola Herrera, [2]Lilia Cabrera, [3]Carolina Rocha, [3]Felipe Leão, [3]Aristobolo Silva, [4]Robert Gilman, [5]Eduardo Tarazona, [5]Fernanda Kehdy

[1]Instituto de Ciências Biológicas - UFMG, Brazil, [2]Asociación Benéfica PRISMA, Peru, [3]Instituto de Ciências Biológicas - UFMG, Brazil, [4]Department of International Health/ Asociación Benéfica PRISMA, Peru, [5]Instituto de Ciências Biológicas - UFMG, Brazil

Over-expression of IL8, its receptors IL8RA and IL8RB, and PTGS2, resultant of immune inflammatory responses induced by Helicobacter pylori, is associated with major risk of developing gastric adenocarcinoma (GA), which high incidence has been observed in Peruvian population. Besides socioeconomic aspects, host genetic factors also play a role in this risk. The aim of this research is to perform a candidate gene association study to test the hypothesis that host genetic variants in the promoter regions of IL8, IL8RA, IL8RB, and PTGS2 are associated with susceptibility to GA in this population. We sequenced 5'UTR and promoter regions of these genes to examine nucleotide diversity patterns in Native American population. For comparison, 24 Africans and 23 Europeans were also analyzed. Two SNPs associated with cancer in GWAS of MSMB and FGFR2 genes were included in the study. Based on pattern of linkage disequilibrium, we selected 6 tag-SNPs in our samples (261 cases and 344 controls of admixed population of Peru) to be genotyped and their association with gastric cancer was then tested. We found exclusive polymorphisms of our Native American or case-control samples, and a very differentiated haplotype distribution, mainly for IL8RB gene. Nevertheless, association among the candidate SNPs and the gastric adenocarcinoma in Peruvian population was not found. Its high incidence of GA does not seem to be related to susceptibility alleles common in this population, as these results suggest a predominant role for socioeconomic and healthy factors. Although there was no association, we are currently investigating whether there are differences in gene expression between haplotypes of IL8RB, since we verified through analyses in TRANSFAC differentiated transcription factors binding sites caused by polymorphisms. The haplotypes chosen were: reference haplotype (GCG), common among Europeans and Africans; rare haplotype (ATA), found in only one individual; and the most frequent haplotype in Native-American and Admixed population (cases and controls) named NAA (ATG). Sequences of reference and rare haplotypes were cloned in pGL3basic vector and transfected in HEK293A cells to be employed in reporter gene assays. Then, these cells were stimulated with TNFα. Results showed greater level of gene expression for the rare haplotype. The next step is to perform the same analyses using NAA haplotype as well as specific assays to quantify FOXO3, a transcription factor previously found associated to tumorigenesis whose binding site was created in rare and NAA haplotypes. This study is relevant to clarify the functional effects of Native American associated variants, in particular, those that may be associated with complex diseases, such as gastric adenocarcinoma.

Keywords: Gastric Adenocarcinoma, Peruvian Population, IL8RB, Promoter, Haplotype

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 9

Identification of B cell conformational epitopes in the infectome of Toxoplasma gondii as vaccine targets

[1]Arthur Moura, [1]Tiago Mendes, [1]Ricardo Fujiwara, [1]Ricardo Vitor,[1] Daniella Bartholomeu

[1]UFMG, Brazil

Toxoplasma gondii is an intracellular protozoan capable of inducing abortion and morbidity in fetus and cause severe disease in immunocompromised patients. The available drugs do not act in tissue cysts, have severe adverse effects and there are reports on natural drug resistance. Thus, vaccine development against toxoplasmosis has great priority. The objective of this study is the identification of conformational epitopes in the parasite's infectome, in other words, in proteins that interact with host proteins. These epitopes may be targeted as vaccine candidates using antibodies that may block the infectious process by acting on the parasite-host interface. Initially, we predicted the protein-protein interaction network between parasite and human by using three methods: (1) PEIMAP employs local alignment with BLAST to find similarity between the database and query proteins and informs all experimental methods that support the described interaction pair; (2) PSIMAP involves PSIBLAST alignment taking into account all possible conformational domains found in PDB that interact with at least 5 amino acids in a distance less than 5 angstrom; (3) Pfam uses the statistical alignment hidden Markov chain and stores linear domains found in PDB whose near residues has physical and chemical bonds. Furthermore, we refined the potential proteins that participate in interactions using experimental data and subcellular location prediction to select transmembrane, GPI or secreted/excreted proteins. Also we assigned a prediction confidence value based on the frequency with which an interaction is detected and the quality of the proposed method that confirmed its occurrence. The highest score interactions will be experimentally validated including interactions involving hypothetical proteins from the parasite. Each protein pair will be expressed in Escherichia coli and the corresponding polyclonal antibodies will be produced in mice. These antibodies will be used in coimunolocalization and coimunopreciptation assays to validate the predicted interactions. After this validation step, B cell conformational epitope prediction will be performed and those mapped on the interaction interface between the Toxoplasma gondii and the host proteins will be targeted in protection assays in mice.

Keywords: T. gondii, infectome, interspecific interactome, vaccine, imunoinformatics, toxoplasmosis

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Transcriptomics and Proteomics

Abstract #: 10

A strain of Paracoccidioides brasiliensis defective in the thermodimorphic shift: A genomic comparative approach

L. O. Oliveira[1], D. M. Resende[2], L. O. Gonçalves[3], R. C. Cruz[4], J. M. Ribeiro[5], J. C. Ruiz[6], P. S. Cisalpino[1], , , , , , , , , AU

[1]Dpto. Microbiologia, Lab. de Microrganismos, Programa de Pós Graduação em Bioinforática, Instituto de Ciências Biológicas (ICB), Universidade Federal de Minas Gerais (UFMG), and Grupo Informática de Biossistemas - CPqRR - FIOCRUZ Minas, Brazil
[2]Grupo Informática de Biossistemas - CPqRR - FIOCRUZ Minas, Universidade Federal de Ouro Preto, Brazil
[3]Grupo Informática de Biossistemas - CPqRR - FIOCRUZ Minas, Centro Universitário UNA, Brazil
[4]Dpto. Microbiologia, Lab. de Microrganismos and Dpto. Engenharia Mecânica, Lab. de Bioengenharia, Escola de Engenharia, Universidade Federal de Minas Gerais (UFMG), Brazil
[5]National Institutes of Health (NIH), National Institutes of Allergy and Infeccious diseases - Laboratory of Malaria and Vector Research., Brazil
[6]Grupo Informática de Biossistemas - CPqRR - FIOCRUZ Minas, Brazil

Paracoccidioidomycosis (PCM), which is endemic to Latin America, represents a systemic mycosis of medical importance caused by dimorphics fungi of Paracoccidioides complex. Infection is thought to be contracted by inhalation of fungal propagules conidia or mycelial fragments and it is triggered by the dimorphic shift that characterizes this fungus and which consists of its change upon exposure to the body temperature to the yeast form. The morphological change is closely associated with organism virulence but the biological pathways involved are poorly described in the context of system biology. A clinical isolate (B339 - S1 lineage) of Paracoccidioides brasiliensis presenting no dimorphic shift was selected after treatment of the yeast-like form with sulfamethoxazole. This particular strain remains in the yeast-like form at room temperature (250C). To enable the comparative analysis of this strain with a wild type one (B339 - S1 parental lineage exhibiting dimorphism), we are employing a high throughput RNA-seq using Illumina NGS sequencing and developing a RNA sequence workflow analysis that will allow the deep investigation of the transcriptome profiles. This approach has the potential of bringing to light genes that are controlled during the host-pathogen interaction, pathways controlling dimorphism and consequently virulence and pathogenicity. Particularly we will present: a) The clinical isolate (B339 - S1 lineage) selection process employed; and b) comparative genome analysis of the three currently available genomes of Paracoccidioides complex (Pb01, Pb03 and Pb18) in terms of functional annotation and pathways potentially associated with dimorphism and virulence that will be integrated with RNAseq analysis and results in the context of System Biology.

Keywords: Bioinformatics, Comparative Genomics, Dimorphism, Paracoccidioides species complex

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 11

BREAST CANCER PHARMACOGENETIC REGIONS ARE HIGHLY DIFFERENTIATED IN NATIVE AMERICANS

[1]Fernanda Rodrigues-Soares, [1]Rennan Moreira, [2]Mateus Gouveia, [1]Robert Gilman, [1]Roxana Zamudio, [1]Giordano Soares-Souza, [1]Maíra Rodrigues, [1]Wagner Magalhães, [1]Eduardo Tarazona-Santos

[1]UFMG, Brazil, [2]Universidad Cayetano Heredia, Peru

Background: Previous studies by our group showed that some SNPs of the Antiestrogen Pathway (aromatase inhibitors), which is important in breast cancer treatment, might be differentiated in Native American populations. To exhaustively evaluate if there is a haplotype differentiation among Native Americans respect to other worldwide populations, we developed a bioinformatics framework to select cosmopolitan tag-SNPs from HapMap or 1000 Genomes Project for specific genomic regions, that in this case encompasses four genes of the Antiestrogen Pathway (CYP19A1, ESR1, ESR2 e HSD17B1). These SNPs were annotated in detail using the multi-agent annotation system (MASannote) developed by our group. Forty-eight tag-SNPs on these genes were genotyped using a customized kit and the BeadXpress platform (Illumina, US), and analyzed using bioinformatics QC and population genetics analysis procedure developed by our group (see the DIVERGENOMEtool platform available at http:// pggenetica.icb.ufmg.br/divergenome/pagina/dynamicpipeline/tools.php). We analyzed several worldwide populations, including original data from 384 individuals: three Peruvians Native American groups 72 Shimaa, 92 Ashaninkas and 88 Quechuas (the largest linguistic native South American group); four admixed Brazilians populations from Minas Gerais: 19 from Carmésia, 30 from Martinho Campos, 24 from Resplendor and 25 from São João das Missões. We compared these data with public database HapMap from European (CEU) ancestry and African (YRI) ancestry. Ancestry of our South American samples was estimated using a panel of 96 ancestry informative markers genotyped by the BeadXpress platform. Results: Several population genetics analyses consistently show that Native Americans show an extremely differentiated haplotype structure when compared to Europeans, Africans and Brazilian admixed individuals. Conclusions: The highly differentiated haplotype structure of Native Americans for genes involved deserves follow-up, to understand its clinical implication in the therapeutic efficacy of breast cancer therapy with aromatase inhibitors. The study of Native Americans is particularly important because this ethnic group is neglected in genomics and pharmacogenetics initiatives. We are currently evaluating the nucleotide diversity on these genes in detail, performing next-generation targeted sequencing of 16 Native Americans, using Sure Select Agilent technology. Funding: CNPq, CAPES, FAPEMIG.

Keywords: Breast Cancer, Haplotypes, Native American, Human Diversity

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 12

Efficacy of medicinal-plant-derived antibacterial compounds in the treatment of Pharyngeal Diphtheria: An In silco approach for controlling the pathogenesis of Corynebacterium diphtheria

[1]Sandeep Tiwari, [1]Syed shah Hassan, [1]Sintia Silva de Almeida, [1]Vinicius Augusto Carvalho de Abreu, [1]Letícia Castro-Oliveira,[1] Diego CB Mariano, [1]Lukasgoncalves Amorim, [1]Carlosaugusto Diniz, [1]Edgar L aguiar,[1] Siomar C Soares, [1]Vasco Ariston de Carvalho Azevedo

[1]UFMG - Universidade Federal de Minas Gerais, Brazil

Corynebacterium diphtheriae is the causative agent of pharyngeal diphtheria. They possess very important transpeptidases; e.g sortase proteins, which help a majority of the Gram +ve bacteria in decorating the surface with a diverse array of proteins that enable the microbe to effectively interact with its environment. These enzymes help to polymerize and assemble pili proteins to construct multi-subunit hair-like fibers that extend from the cell surface to promote bacterial adhesion and subsequent colonization. Sortases helps in the polymerization of very important virulence factors by deploying them as surface proteins that mediates bacterial adhesion to host tissues, host cell entry, evasion, and suppression of the immune response as well as acquisition of essential nutrients. The presence of different type of sortase proteins in the Corynebacterium diphtheria motivated a search for medicinal plant isolates for the treatment of Pharyngeal diphtheria by considering these transpeptidases as potential drug targets. In this work, the proteomic data of all sortases was used from the genome of C. diphtheria and comparative molding was performed for obtaining the best possible 3D models. All valid structures were used for active site prediction and analysis. The predicted active sites residues of these proteins were compared to their template structure residues using in silico approaches, also a manual search was carried out, where needed. A data-set of active natural antimicrobial compounds derived from different natural resources were retrieved and a library was made and used in this analyses. Molecular properties and prediction of bio-activity of all the compounds were determined. Various drug relevant properties; like mutagenic, tumorigenic, irritant and reproductive effect of the compounds following Lipinski Rule of five, were also checked. The compounds fulfilling the criteria of having these molecular properties and drug-like ability were separated and used in the docking analyses. Compounds demonstrated good docking results with all target proteins were considered to be used as potential therapeutic candidates for the treatment of Pharyngeal diphtheria, owing to experimental validations. The proposed target enzymes of C. diphtheria play a critical role in the pathogenesis caused by a vast number of gram-positive bacteria. Compounds derived from medicinal plants are better enough for their accessibility and low price. This will definitely change the trend in the near future as new inhibitors are discovered and optimized, and the 3D structures of sortase-inhibitor complexes are determined. Structural studies of inhibitors, in complex with sortase enzymes should also further facilitate the application of virtual screening approaches, giving more insight into the inhibitory mechanism of the target proteins.

Keywords: Corynebacterium diphtheria ,Pharyngeal Diphtheria, transpeptidase enzymes, in silico putative targets

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 14

Functional analysis in intronics SNPs

[1]Izinara Rosse Cruz, [2]Juliana Assis Geraldo, [3]Pablo Augusto de Souza Fosenca,[3] Fernanda Caroline dos Santos,[3]Pedro Lamounier de Faria, [3]Raphael Steinberg da Silva, [3]Marlene de Miranda, [4]Guilherme Oliveira, [5]Maria de Fátima Ávila Pires, [6]Maria Gabriela Campolina Diniz Peixoto, [7]Maria [8]Raquel Santos Carvalho

[1]Institute of Biological Sciences – UFMG/Center for Excellence in Bioinformatics – FIOCRUZ-MG, Brazil, [2]Genomics and Computational Biology Group - FIOCRUZ-MG/Center for Excellence in, [4]Bioinformatics – FIOCRUZ-MG, Brazil, [5]Institute of Biological Sciences – UFMG, Brazil, [6]Genomics and Computational Biology Group - FIOCRUZ-MG/Center for Excellence in Bioinformatics – FIOCRUZ-MG, Brazil, [7]EMBRAPA Dairy Cattle – MG, Brazil, [8]Institute of Biological Sciences – UFMG, Brazil

Sequencing of gene regions in the search for polymorphisms that may explain certain phenotypes often leads to the discovery of thousands of SNPs in both exonic and intronic regions. Currently, we know that 80% of non-coding regions have a functional role in regulating gene expression, and therefore, should be taken into account in functional analyses. However, genotyping all polymorphisms, in a larger number of individuals, is a laborious and costly step, which can be minimized with the help of bionformatics. This study was developed in order to ascertain putative functionality of SNPs in the oxytocin gene. This gene was chosen due to its role on milk ejection, as well as in social behavior and parental care. Twenty-four animals of the Bos indicus breed Guzera were sequenced to confirm the SNPs found for this gene. As a result, ten new SNPs were identified and were subjected to in silico analysis in the search for evidence of functionality, such as the degree of evolutionary conservation of the position and/or region where the SNP occurs, putative effects on splicing sites, miRNAs target sites, and other functional domains. G423T, C459T and T658C SNPs altered three alternative splicing sites and the predicted transcripts for these variants showed no recognition domain of the hormone superfamily 5, which can cause changes in gene expression. In addition, G648T, G740A and T658C SNPs are located at target sites of micro-RNA, creating or abolishing miRNA recognition sites. Therefore, these SNPs may influence gene expression. In summary, 63% of intronic SNPs showed evidence of functionality. This approach represents a valuable complement to the emerging genome-wide association studies that generate numerous candidate genes, but no indication of the functional variants. Funding: CAPES, CNPq, FAPEMIG (CBB-1181/0 and TCT 12.093/10, NIH-USA (TW007012), CAPES/CDTS-FIOCRUZ, FIOCRUZ-MG, PRPq/UFMG, EMBRAPA

Keywords: SNPs, functional analysis, oxytocin

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 16

A NEW APPROACH FOR CLASSIFICATION AND PHYLOGENETIC ANALYSIS OF BACTERIA USING SINGULAR VALUE DECOMPOSITION (SVD)

[1]Diego César Batista Mariano, [1]Letícia de Castro Oliveira, [1]Lucas Gonçalves Amorim, [1]Marcos Augusto dos Santos, [1]Siomar de Castro Soares, [1]Carlos Augusto Almeida Diniz, [1]Anderson Miyoshi, [1]Vasco Ariston de Carvalho Azevedo

[1]UFMG, Brazil

Background: In past, evolutionary reconstructions of the tree of life were mainly performed based in identification of the point of divergence between species solely based in shared homologous features. However, this methodology could be very trick due to convergent and divergent evolution. With the advent of molecular techniques, phylogenetic was greatly improved by the use of nucleotide differences in universal reference markers, creating the area of phylogenomics. In the post-genomic era, a second wave of changes brought new approaches to phylogenomics, which now infers the evolutionary divergence by taking advantage of whole-genome data, like: gene content and gene order; orthology; and, DNA string or DNA signature. Phylogenomics inferences based on DNA signature, or genomic signature, take into account the codon usage of the coding sequences, the G+C content and the nucleotide pattern, like di-, tri- and tetra-nucleotides frequencies. The codon usage is mainly affected by the codon/anticodon interaction force and the availability of a given tRNA, where the adoption of AT- or GC-rich codons generates a homogeneous nucleotide pattern through the whole genome, which is different in unrelated organisms. In this work, we analyzed the latent semantic index based on the singular value decomposition (LSI-SVD) of a matrix containing information from the codon usage fraction of the coding sequences (CDS). Results: The resulting data was used as coordinates to plot the genomes in a 3-dimensional chart and a distance matrix was generated from the absolute distances between all genomes. Finally, a phylogenetic tree was created from the distance matrix in order to visualize the evolutionary relationships. The dataset was composed of 60 genomes of Gram-positive and Gram-negative bacteria, and the resulting phylogenetic tree was validated using the already studied evolutionary relationships of the bacteria from the CMNR group (Corynebacteria, Mycobacteria, Nocardia and Rodococcus). Conclusions: The phylogenetic tree generated by this method shows a clear relationship between the bacteria of those genera, in despite of the other organisms; however, a small number of species appear in disagreement. Regarding the high G+C content of the bacteria from the CMNR group, the dataset is under update to consider nucleotide frequencies (G+C, di-, tri- and tetra-nucleotides), which will separate the CMNR group from other bacteria and raise the accuracy of the method. Finally, we intend to develop a public software applying the methodology here used and extend the phylogenetic analysis to other bacterial genomes from NCBI.

Keywords: bioinformatcs, bacterias CMNR, phylogenetic, SVD

Topic: Transcriptomics and Proteomics

Abstract #: 17

Characterization of the SL trans-spliced mRNA sub-population in Schistosoma mansoni by high-throughput RNA-Seq

[1]Mariana Boroni, [2]Michael Sammeth, [2]André Luiz Martins Reis, [3]Marina Moraes Mourão, [4]José Marcos Chaves Ribeiro,[1] Glória Regina Franco

[1]Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, Brazil, [2]Laboratório Nacional de Computação Científica, Petrópolis, Rio de Janeiro, Brazil, Brazil, [3]Grupo de Genômica e Biologia Computacional, Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Minas Gerais, Brazil, Brazil, [4]Section of Vector Biology, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, Rockville, Maryland, United States of America, United States

The spliced-leader (SL) trans-splicing is the process that connects an exon donated from a specialized SL RNA to exons of different pre-mRNA transcripts generating mature mRNAs in a wide range of eukaryotes, from protozoa to chordates. Here we present the characterization of trans-spliced mRNA subpopulations of Schistosoma mansoni by high-throughput RNA-Seq. To determine which genes are subject to this mechanism, we analyzed transcripts that undergo SL trans-splicing from different life cycle stages of S. mansoni, using two distinctive approaches to generate the libraries: i) available RNA-seq reads in the public repository SRA were filtered and reads containing the 36-nucleotide spliced leader (SL) sequence were recovered. The compiled library was named SL Retrieved; ii) cercaria transcripts containing the SL sequence were captured from total RNA and two distinct libraries were constructed and sequenced to produce reads derived mainly from transcripts that initiate with the SL sequence. This cercaria library was named SL Trapping. The obtained reads were aligned to the reference genome and the trans-splicing acceptor sites were identified in the 5' end of annotated and unknown exons where the reads were mapped. We identified 4959 genes undergoing SL trans-splicing that corresponds to 40% of the currently annotated genes at the 5th version of the S. mansoni genome. Most of the events identified occurred in the first exon, as expected, however the SL entrance in exons far away from the first was detected and, in some cases, with a higher frequency than its incorporation in the first exon. Curiously, alternative trans-splicing were also found. Looking for sequence signatures associated with the SL addition sites, the well-known AG dinucleotide was found before the 5' end of the exons that are linked to the SL sequence. To discover whether or not some genes are trans-spliced and if this phenomenon is related to the level of gene expression, we compared the frequency of trans-splicing occurrence with the level of expression of those genes in the parasite. We identified two groups of trans-spliced transcripts: a frequently trans-spliced group, and a rarely trans-spliced group. Most of the frequently trans-spliced transcripts are also highly transcribed. We also found a significant number of genes that generate predominantly non-trans-spliced mRNA molecules but, occasionally, can generate a detectable number of trans-spliced mRNA molecules. When investigating possible correlations between trans-splicing status and gene function, no over-represented functional gene category in the trans-spliced group was observed, suggesting that the SL trans-splicing mechanism could not be associated with a specific gene function. Other interesting finding was the detection of different SL trans-spliced transcripts among distinct parasite developmental stages. All those results indicate that the trans-splicing may be a general and important mechanism of post-transcriptional gene regulation in S. mansoni.

Keywords: RNA-Seq, Trans-splicing

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 18

Computational prediction of co-variation/co-evolution of amino acid residues in the aspartic proteases pepsin simile protein family

[1]Alberto Fernandes de Oliveira Junior, [1]Diego Mariano, [1]Carlos Diniz, [1]Lucas Bleicher, [1]Vasco Azevedo,[2] Floriano Paes Silva Junior

[1]UFMG, Brazil, [2]FIOCRUZ, Brazil

The co-variation/co-evolution of amino acids is defined, from the literature, as the offset of one amino acid at a given position, followed by the change of other amino acids. From this concept, the research on the sites in biological sequences can be used to predict proteins and nucleic acids structures as well as molecular interactions implicated in energetic pathways between functionally important regions, among other applications. The goal of this work is to explore the interdependence of specific amino acid in correlated amino acid sequences of proteins families with pharmaceutical and biotechnology interest like aspartyl proteases pepsin simile (APPS). For the analysis, we used aligned sequences obtained directly from the Pfam database (PFAM code: PF00026). Fragments, i.e., sequences with coverage lower than 85% when compared to a reference sequence, RENI_HUMAN, and sequences with identity lower than 15% to that reference sequence were automatically removed. Finally, all sequences were compared against each other in order to remove redundancy (every time two sequences were found with identity higher than 85%, the smaller one was removed from the alignment). After this pre-processing, the final alignment contained 827 sequences. To perform the analysis, we used the PFstats software developed by Bleicher et al. 2011. The results showed the formation of four communities in APPS. When these amino acids were observed in the 3D structure, the disulfide bonds formation was noticed which can influence the structure / function of the protein. In addition, it was observed that some amino acids are correlated near important regions, such as active site. Phylogenetic analysis was done to enable us to determine groups, from the observation of the scores given to communities. We believe that such correlations may affect the structure / function of such proteins, by amino acids presence or absence.

Keywords: Aspartil Proteases; Coevolution; Alignment Sequence

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 19

Analysis and characterization of co- variation / co- evolution of amino acid residues in the phospholipase A2 protein family

[1]Alberto Fernandes de Oliveira Junior, [1]Edgar Lacerda de Aguiar, [1]Lucas Bleicher, [1]Vasco Ariston de Carvalho Azevedo, [2]Floriano Paes Silva Junior

[1]UFMG, Brazil, [2]Fiocruz, Brazil

To comprehend the mechanisms of evolution of a protein, it is extremely important to understand the co-variation/co-evolution between amino acid residues, and this requires a thorough understanding of many factors that determine the selective forces acting on each amino acid residue. Knowledge of co-evolution sites in biological sequences can be used to predict structures of proteins and nucleic acids as well as, molecular interactions implicated in energetic pathways between functionally important regions, among other applications. The objective of this work was to identify the sites and communities of co-evolving amino acid residues in phospholipases A2 (PLA2) enzymes, a family of proteins with known pharmaceutical and biotechnological interest. For this analysis, we used, aligned sequences obtained directly from the Pfam database (PFAM code: PF00068). Fragments, i.e., sequences with coverage lower than 85% when compared to a reference sequence, PA2GA_HUMAN, and sequences with identity lower than 15% to that reference sequence were automatically removed. Finally, all sequences were compared against each other in order to remove redundancy (every time two sequences were found with identity higher than 85%, the smaller one was removed from the alignment). After this pre-processing, the final alignment contained 373 sequences. To perform the analysis, we used the PFstats software developed by Bleicher et al. 2011. The results showed the formation of five communities in APPS. When these amino acids were observed in the 3D structure, the disulfide bond formation was noticed which can influence the structure / function of the protein. In addition, it was observed that some amino acids are correlated near important regions, such as active site. Phylogenetic analysis was done to enable us to determine groups, from the observation of the scores given to communities. It was observed that the protein not only in the community 2 had a score belonged to serpents. We believe that such correlations may affect the structure / function of such proteins, by amino acids presence or absence.

Keywords: Phospholipase A2 ; Coevolution ; Evolutionary Computation

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 21

MOLECULAR DOCKING AND QUANTUM-CHEMICAL PROPERTIES OF ANTIMALARIAL PEROXIDES DERIVED FROM DISPIRO-1,2,4-TRIOXOLANES

[1]Amanda Ruslana Santana Oliveira, [1]Clauber Henrique Souza da Costa, [1]Maycon José dos Santos Pereira, [1]Dalton Nunes oliveira, [1]Matheus Oliveira Leite de Sá, [1]Ricardo Morais de Miranda, [2]Antonio Florêncio de Figueiredo, [2]João Elias Vidueira Ferreira

[1]Instituto Federal de Educação, Ciência e Tecnologia do Pará, Campus Belém., Brazil, [2]Instituto Federal de Educação, Ciência e Tecnologia do Pará, Campus Tucuruí., Brazil

Malaria is a tropical disease caused by the protozoan parasite Plasmodium falciparum. In Brazil only the Amazon Rain Forest accounts for 97% of the cases. From 2000 through 2011, 99,7% of the cases were notified in this area, which is considered endemic in the country. It is estimated that about one million people dye owing to malaria in the world. From 350-500 million cases are registered each year and Sub-Saharan Africa responds for most of the cases. Today P. falciparum presents resistence to chloroquine and other antimalarial drugs. This work describes a study of dispiro-adamantane-1,2,4-trioxolane with antimalarial activity against k1 clone of P. falciparum multiresistent to cloroquine. A quantum-mechanical treatment with B3LYP/6-31G* was performed to optimize the geometry of the molecules, maps of molecular electrostatic potentials (MEP) were produced (Figure 1) and frontier molecular orbitals, HOMO and LUMO, had their energies computed. The results revealed that in general those compounds with high values for LUMO energy present high biological activity. This result shows the capacity of peroxide ring to attract electron from Fe2+ íon in the process of oxidation of free heme. The maps of MEP showed similar surface next to 1,2,4-trioxolane ring indicating a region of negative electrostatic potentials and suggesting that Fe2+ ion from heme are favored to attack this region of large electronic density. By analyzing the partial atomic charges, the oxygen atom numbered 2 shows the most negative charge, which confirms the assumption that the endoperoxide region is the active site in the process of alkylation of heme, a crucial step in the mechanism related to antimalarial compounds. Molecular docking studies shows that there is a tendency among the least active compounds to present shorter distances and that the more σ(O-O) bond becomes planar the more efficient is the prevention of heme polymerization.

Keywords: Malária, MEP, docking molecular.

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 23

INVESTIGATION OF ANTIMALARIAL ACTIVITY OF DISPIRO-1,2,4,5-TETRAOXANES THROUGH MOLECULAR DOCKING EMPLOYING DOCKTHOR PROGRAM.

[1]Dalton Nunes Oliveira, [1]Amanda Ruslana Santana Oliveira, [1]Clauber Henrique Souza da Costa, [1]Matheus Oliveira Leite de Sá, [1]Fabio Jorge de Nazaré Ferreira, [1]Ricardo Morais de Miranda, [2]Antonio Florêncio de Figueiredo, [2]João Elias Vidueira Ferreira

[1]Instituto Federal de Educação, Ciência e Tecnologia do Pará, Campus Belém., Brazil, [2]Instituto Federal de Educação, Ciência e Tecnologia do Pará, Campus Tucuruí,, Brazil

Malaria is a tropical disease caused by the protozoan parasite of the genus Plasmodium and is transmitted through the bite of infected female Anopheles mosquitoes. There are more than a hundred species of Plasmodium but only four of them are dangerous to human beings: Plasmodium vivax, Plasmodium falciparum, Plasmodium malariae e Plasmodium ovale. Malaria affects millions of people in the world, especially in Africa. According to World Health Organization there are about 250 million cases of malaria in the world and about 880 thousand deaths are registered each year. In Brazil, the Amazon Rain Forest is responsible for 99.7% of the cases, which are attributed mainly to P. vivax. Now considering Para State, the disease affects hundreds of people. The main symptoms of the disease are high fever, headache and aches in many other parts of the body, pale skin, and tiredness. So it is important to study new possibilities to treat patients infected by Plasmodium because nowadays P. falciparum shows resistance to artemisisnin and other antimalarial compounds. Moreover malaria is still a serious problem in Para State. In this work, a theoretical investigation was performed on dispiro-1,2,4,5-tetraoxane derivatives with antimalarial activity against the strain of P. falciparum mult resistant to artemisisnin. A quantum-chemical approach was performed to optimize the geometry of the molecule with B3LYP/3-21G method using Dockthor program. Results made it possible to associate molecular properties to biological activities.

Keywords: Malaria, Dockthor, biological activities.

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 24

THEORETICAL STRUCTURAL STUDIES INVOLVING THE INHIBITORY MECHANISM OF A NOVEL SYNTHETIC QUINOLINONE AGAINST A METALLOPROTEASE FROM BOTHROPS JARARACUSSU VENOM

[1]Fábio Fillipi Matioli,[2]Patrícia T. Baraldi, [3]Angelo José Magro, [4]Silvana Marcussi, [5]Ney Lemke, [6]Leonardo A. Calderon, [7]Rodrigo G. Stábeli, [8]Andreimar M. Soares, [5]Marcos R. M. Fontes, [2]Arlene G. Correa

[1]Departamento de Física e Biofísica, Universidade Estadual Paulista (UNESP), Botucatu-SP, Brazil, Brazil, [2]Departamento de Química, Universidade Federal de São Carlos (UFSCar), São Carlos-SP, Brazil., Brazil, [3]Departamento de Física e Biofísica, Universidade Estadual Paulista (UNESP), Botucatu-SP, Brazil, Brazil, [4]Departamento de Análises Clínicas, Toxicológicas e Bromatológicas, Faculdade de Ciências Farmacêuticas de Ribeirão Preto (FCFRP), Universidade de São Paulo (USP), Ribeirão Preto-SP, Brazil, Brazil, [5]Departamento de Física e Biofísica, Universidade Estadual Paulista (UNESP), Botucatu-SP, Brazil, Brazil, [6]Centro de Estudos de Biomoléculas Aplicadas à Saúde (CEBio), Fundação Oswaldo Cruz (FIOCRUZ Rondônia) e Departamento de Medicina, Universidade Federal de Rondônia (UNIR), Porto Velho-RO, Brazil., Brazil ,[7]Centro de Estudos de Biomoléculas Aplicadas à Saúde (CEBio), Fundação Oswaldo Cruz (FIOCRUZ Rondônia) e Departamento de Medicina, Universidade Federal de Rondônia (UNIR), Porto Velho-RO, Brazil., Brazil, [8]Departamento de Análises Clínicas, Toxicológicas e Bromatológicas, Faculdade de Ciências Farmacêuticas de Ribeirão Preto (FCFRP), Universidade de São Paulo (USP), Ribeirão Preto-SP, Brazil, Brazil

A novel compound, 2-hydroxymethyl-6-methoxy-1,4-dihydro-4-quinolinone, was synthesized from a commercially available aniline in three steps and screened for its antiophidian activity, reporting different degrees of antiproteolytic, anticoagulant, and anti-hemorrhagic properties against snake venoms and isolated toxins. The compound was able to neutralize the hemorrhagic, fibrinogenolytic and caseinolytic activities of class P-I and III snake venom metalloproteases (SVMPs) isolated from Bothrops neuwiedi and Bothrops jararacussu venoms. Clotting and fibrinogenolytic activities induced by isolated thrombin-like enzymes from B. jararacussu and Crotalus durissus terrificus venoms were inhibited after incubation at different ratios. Additionally, myotoxic, edema and phospholipase activities induced by snake venoms or isolated toxins were partially inhibited. The quinolinone also potentiated the ability of the commercial polyvalent antivenom in neutralizing myotoxic effect of the B. jararacussu snake venom. In order to get insights into the structural basis of compound inhibitory mechanism against snake venom toxins, docking and molecular dynamics simulations were performed involving this synthetic quinolinone and a theoretical model of the catalytic domain of BjussuMP-II, a P-III metalloprotease from B. jararacussu venom. The results obtained after these theoretical simulations revealed some insights into the possible inhibitory mechanism of the compound against SVMPs, strengthening the potential of this synthetic molecule as an adjuvant in snakebite serum therapy.

Keywords: 4-quinolinones; chemical synthesis; snake venom; proteases; phospholipases A2; antiserum potentiation

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 25

MOLECULAR DYNAMICS STUDY OF EthR INHIBITORS: AN INVESTIGATION OF N-PHENYLPHENOXYACETAMIDA DERIVATIVES WITH PROMISSOR ANTI-TUBERCUOSIS ACTIVITY.

[1]Maycon José dos Santos Pereira,[1] Amanda Ruslana Santana Oliveira, [1]Clauber Henrique Souza da Costa, [1]Linéia Soares da Silva, [1]Matheus Oliveira Leite de Sá, [1]Ricardo Morais de Miranda, [2]Antonio Florêncio de Figueiredo, [2]João Elias Vidueira Ferreira

[1]Instituto Federal de Educação, Ciência e Tecnologia do Pará, Campus Belém., Brazil, [2]Instituto Federal de Educação, Ciência e Tecnologia do Pará, Campus Tucuruí., Brazil

Tuberculosis (TB) is still one of the main causes of mortality in the world, killing 1.7 million people every year. In 2010 both the incidence and the prevalence of TB were estimated in 8.8 and 12 million cases, respectively, according to The World Health Organization (WHO). Difficulties in detection and cure of the disease are the main drawbacks to control the infection. One third of world population is infected with the latent form of Mycobacterium tuberculosis (TB) and approximately 10% will develop the disease. The EthR transcriptional receptor against TB controls the EthA expression, a bacterial activator of ethanoamide, which is strongly responsible for the low sensibility that M. tuberculosisa present to this antibiotic. Ethionamide is a drug recommended by the WHO to treat MDR-TB. The minimum dose necessary to inhibit the growing of M. tuberculosis is normally sufficient to produce serious side effects, such as gastrointestinal disturbs, hepatitis and psychic diseases like depression, anxiety and psychosis. In this study the interaction between the derivatives of N-phenylphenoxyacetamide inhibitor (O8B) and EthR were analyzed through molecular modeling. Molecular dynamics (MD) simulations describe well the mechanism of interaction O8B-EthR. Results showed that the residue ASN179 present a bond with 2.82Aº length between oxygen atom of the inhibitor and nitrogen atom of the ASN residue (asparagin). This bound display an important role to stabilize the complex. The negative value for interaction energy reveals that the interaction between O8B and EthR is favorable based on thermodynamics. The plot of RSMD (average quadratic shift) shows that the trajectory begins to stabilize around 100ps (picoseconds). Thus it was observed that molecular dynamics simulation were important to refine the O8B-EthR complex under study.

Keywords: Tuberculosis, Molecular Dynamics, biological activity.

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 26

Investigation of the biological activity of statins inhibiting HMGR enzyme against cholesterol by Molecular Docking

[1]Clauber Henrique Souza da Costa, [1]Amanda Ruslana Santana Oliveira, [1]Maycon José dos Santos Pereira, [1]Linéia Soares da Silva, [1]Matheus Oliveira Leite de Sá, [1]Ricardo Morais de Miranda, [2]Antonio Florêncio de Figueiredo, [2]João Elias Vidueira Ferreira

[1]Instituto Federal de Educação, Ciência e Tecnologia do Pará, Campus Belém., Brazil, [2]Instituto Federal de Educação, Ciência e Tecnologia do Pará, Campus Tucuruí., Brazil

Cholesterol is a very important compound for all animals. It is used by cells to synthesize bile acids, digest and absorb lipids and lipossoluble vitamins in the small intestine and synthesize steroid hormones and vitamin E. High LDL cholesterol levels are associated to atherosclerosis, which is responsible for many deaths in the world. Endo e Kuroda (1976) published a study on compounds present in fungi with powerful ability to prevent the production of cholesterol in human liver. This drug is called compactin or mevastatin and act as a powerful inhibitor of HMG-CoA reductase enzyme or estatin (HMGR). Compactin belongs to a class of compounds known as estatins and that now they are used to inhibit HMGR enzyme. These molecules present affinity for HMGR one thousand times greater than for HMG-CoA, showing that estatins are very efficient to treat hypercholesterolemia. The main estatins used in this treatment are compactin, simvastatin, fluvastatin, cerivastatin, atorvastatin and rosuvastatin. In this work a molecular docking study was done with aid of AutoDock 4.2 software. The studied compounds were extracted from the following enzymes: 1HWL (rosuvastatina), 1HWK (atorvastatina), 1HWJ (cerivastatina), found in the RCSB Protein Data Bank (PDB). Molecular docking employed A and B chains of 1HW9 complex. Excellent results were achieved through molecular docking which demonstrate estatins are effective to inhibit HMGR enzymes. Figure 1 shows the comparison for the complex as a result of molecular docking and crystallographic structure from PDB. Molecules showed good interactions with the active site of the enzyme. The theoretical results confirm experimental results from the scientific literature. This study can be taken as a support in the development of new drugs by modifying the non-HMG fragments of estatins.

Keywords: Cholesterol, molecular docking, inhibit HMGR .

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 27

MOLECULAR MODELING STUDY ON TIOFENO-2-IL-1,2,4 –OXIADIAZOLS INHIBITORS WITH ANTI-TUBERCULOSIS ACTIVITY EMPLOYING MOLECULAR DOCKING AND MOLECULAR DYNAMICS

[1]LINEIA SOARES DA SILVA, [1]AMANDA RUSLANA SANTANA OLIVEIRA, [1]CLAUBER HENRIQUE SOUZA DA COSTA, [1]MAYCON JOSÉ DOS SANTOS PEREIRA, [1]MATHEUS OLIVEIRA LEITE DE SÁ, [1]RICARDO MORAIS DE MIRANDA, [1]EDISON ALMEIDA RODRIGUES, [2]ANTONIO FLORENCIO DE FIGUEIREDO,[2] JOAO ELIAS VIDUEIRA FERREIRA

[1]Instituto Federal de Educação, Ciência e Tecnologia do Pará, Campos Belém, Brazil, [2]Instituto Federal de Educação, Ciência e Tecnologia do Pará, Campos Tucuruí, Brazil

Tuberculosis (TB) is certainly one of the oldest diseases that affect humanity. In Brazil it has played a large part in mortality, affecting people from different ages and social classes. According to the World Health Organization (WHO) 8-9 million new cases occur in the world each year and about 3 million people dye owing to TB. In Brazil it is estimated that around 124 thousand cases occur. The country is, along with other 22 countries, responsible for more than 80% of the cases. Even though efforts have been made to inhibit TB, the disease continues to be a big problem in public health systems. Therapies available today include a lot of pro-drugs that must be metabolized by microbacteria in order to produce activity. One of these drugs is ethanoamide, an antibiotic used to treat multi-resistant TB, which is bioactivated by mono-oxygenase flavin adenine (EthA) microbacterian of ethanoamide. This drug is recommended by WHO to treat MDR-TB. So this study intends to investigate through theoretical chemistry how the inhibitory action of ethanoamide takes place and how anti-TB activity can be improved. A theoretical study on structure activity relationship involving ethanoamide was performed to verify the main molecular characteristics relating this compound to anti-TB activity (Figure 1). Then a combination of the approaches involving molecular mechanics, quantum mechanics, molecular docking and molecular dynamics were performed. Initially five compounds had their structures optimized with aid of GAUSSIAN 08 software and B3LYP method using 6-31G** basis set. Molecular docking and molecular dynamics were performed by employing AUTODOCK4e and AMBER software respectively, to study relevant chemical properties and also to elucidate mechanism of action of anti-TB compounds.

Keywords: Tuberculosis,Molecular Modeling,molecular docking

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 28

The role of Dnmt2 in Schistosoma mansoni: Structural clues to an epigenetic mistery

[1]Mainá Bitar,[2] Marcelo Fantappié,[1] Glória Franco

[1]Universidade Federal de Minas Gerais, Brazil, [2]Universidade Federal do Rio de Janeiro, Brazil

DNA methylation at cytosine C5 is an important epigenetic trait of eukaryotes that plays key roles in processes such as: chromosomal inactivation, regulation of gene expression, silencing of transposable and repetitive elements, genetic imprinting and reduction of transcriptional noise. Although there are exceptions to this rule, most invertebrate animals studied so far have predominantly non-methylated genomes. In addition, different levels of methylation have been observed in different organisms, indicating that a genome can be classified as non-methylated, globally methylated or partially methylated (Tweety et al., 1997). In this sense, Fantappié and collaborators have explored the DNA of Schistosoma mansoni in the search for methylations and have finally classified this as a predominantly non-methylated genome. Surprisingly, two isoforms of a DNA methyltransferase (Dnmt2) were found in S. mansoni with the solely difference that isoform 1 (Gi 339776689) presents an 8-amino acid insertion when compared to isoform 2 (Gi 339776691). As it was observed that this parasite does not seem to methylate its genome, the question was raised about possible biological roles for this protein in S. mansoni putatively unrelated or additional to DNA methylation, such as tRNA methylation. To address this question, we have decided to assess the 3D structure of the protein and compare it with other previously defined structures of proteins that are known to methylate either DNA or tRNA molecules. We have employed comparative modeling to generate candidate structures for both S. mansoni Dnmt2 isoforms with Modeller and further analyzed their stereochemical and energetical features, derived respectively with Procheck and ProSA, to identify the best-evaluated candidate structure for each isoform. We have further compared the best-ranked Dnmt2 structures with each other to evaluate the structural impact of the previously mentioned insertion in isoform 1, and with structures from the PDB representing a DNA (3MHT) and a tRNA (2J5B) methyltransferases to identify important residues that could putatively account for a DNA or a tRNA methylase activity of Dnmt2 in S. mansoni. These residues of interest are now being used in docking experiments with Haddock to generate protein-DNA and protein-tRNA complexes that will be further evaluated regarding their structural and energetical profiles. We trust that these in silico experiments will help to reveal the function of Dnmt2 in S. mansoni and will therefore contribute to a better understanding of the role of DNA methylation in this parasite.

Keywords: Comparative modeling, protein structure, methyltransferase

Topic: Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 30

MMFMatrix data structure for huge sequence comparation

[1]José Irahe Kasprzykowski Gonçalves, [1]Felipe Guimarães Torres,[1] Mateus Malaquias Oliveira, [1]Artur Trancoso Lopo de Queiroz

[1]CPqGM/FIOCRUZ-BA, Brazil

Background. In bioinformatics, most sequence comparison exact algorithms use matrix data structure to store information about score and optimal alignment traceback. Those algorithms have a mutual problem to handle huge sequences since data structure size is limited by the address amount that 32 and 64 bits processors could handle and since structural size grows exponentially. Moreover, if only one alignment is performed using the minimum machine resources at the time, on 32bit processors, it is possible to create a 100.000.000 register matrix on memory based structures. It represents only two 10kbp sequences (whole HIV genome sequence, for example). However, if this two sequences are 500bp larger, the system runs out of memory, because it cannot process those many addresses. To improve alignment schedule, we proposed a generic structure implementation to manage all matrix data into memory mapped files to require less computational resources and improve algorithm speed. This implementation was compared with the native memory based structure and the database based generic matrix structure. The memory mapped file (mmf) structure is a file based structure, where the system creates a file already bitmapped. Thus, a single bit can be accessed in any part of file. This structure is accessed by primary Kernel I/O resources. Therefore, it does not consume much time or resource to access an information. Furthermore, this structure type allows multiple simultaneous file access based on the file map structure. Results. The MMF based generic matrix structure was implemented in asp.NET Frameword, C# language with System.IO.MemmoryMappedFiles library and hosted along with Viral Sequence Database Manager code in CodePlex and is public for scientific purposes. The native memory based and a database based structure was implemented for test and comparison. Our comparison showed that the MMF based structure can store a large amount of data almost without memory use. Moreover, the MMFMatrix structure does not require transaction manager (as database one requires) and is faster, using less resources. Conclusions. With the test results, we found that the mmfMatrix structure is a good solution for exact algorithms such as Needleman & Wunch and Smith & Waterman. This technique can reduce in 90% the amount of memory resources used for the alignment and can remove the size limit of the matrix. This implementation was also faster than the native memory implementation, because the native matrix is managed by a third level structure, while the mmf is a primary based structure. We were able to observe that database based structures cannot match these two because it needs a transaction manager, which delays the read and write data process. Supported by: Fapesb, CNPq.

Keywords: MMF, Matrix, Alignment

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 34

In silico identification and characterization of phosphate transporter (Pht1) genes from sugarcane

[1]Arthur Tavares de Oliveira Melo, Isabela Pavanelli de Souza, [1]Ivone de Bem Oliveira,[1] Alexandre Siqueira Guedes Coelho

[1]UFG, Brazil

Sugarcane (Saccharum spp.) is considered the most important crop for global supply of sugar and energy. Brazil is the biggest world producer and leads the market of sugar and ethanol derived from sugarcane. For plant growth and development, phosphorus is the most limiting macronutrient, after nitrogen, mainly due to its reduced rates of absorption along the crop cycle. A gene family related with phosphate uptake in high affinity systems has been identified in Arabidopsis thaliana (AtPht1), Zea mays (ZmPht1) and Oryza sativa (OsPht1). Six of these genes are also present in Sorghum bicolor (SbPht1). Using the high degree of synteny among Poaceae family species, these phosphate transporters genes identified in sorghum were used as a reference for homologous identification of family members in sugarcane (ScPht1). High quality DNA sequences were generated using the HiSeq2000 Illumina platform. A total of 7.14X sequencing coverage were produced. Using bowtie2 and SAMTools, reads from sugarcane Pht1 genes highly similar to those from sorghum were identified. An assembly of these ScPht1 genes, with 13.01 X average coverage was produced. The analysis of the alignments allowed the identification of six genes: ScPht1-1, ScPht1-2, ScPht1-3, ScPht1-4, ScPht1-5 and ScPht1-6, comprising 2239 pb, 1965 pb and 1735 pb, 1844 pb, 1861 pb and 1776 pb, respectively. The orthology between sugarcane Pht1 genes and those from A. thaliana were also investigated. The estimated percentage of nucleotide polymorphic sites in Pth1 sugarcane genes was 0.430. Between sorghum and sugarcane Pht1 genes, the estimated percentage of genetic similarity was 97.13%. ScPht1-1 gene showed high sequence similarity with a cluster SCEQRT1028B07.g from SUCEST EST libraries, annotated as being responsible for high-affinity phosphate absorption. Sixty four putative SNPs were identified along the ScPht1 genes and Ka/Ks ratio should suggest there are any genes under positive selection pressure. The average of GC content in these genes was 56.69% and nearly of 70% of SNPs are G ↔ C substitutions, showing that more than half of these SNPs are transversion substitution. The comparison of genes sequences from sugarcane and sorghum reveled 230 putative SNPs. Of this SNPs characterized between the two species, 144 (62.60%) were characterized as transversion (Tv) substitution and G ↔ C is the main transversion nucleotide substitution type (62%). The phylogenetic tree build using Neighbour-Joining method showed ortology between these all genes in sorghum and sugarcane. These results illustrate how synteny between closely related species can be used in genomics studies, particularly when it comes to grasses. The identification of agronomically important genes is a crucial step to improve the use of genetic engineering techniques to support plant breeding programs.

Keywords: Sugarcane, Pht1 genes, synteny

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 36

GLIOBLATOMA TREATED BY RADIOTHERAPY: GENOMIC ORGANIZATION AND PROTEIN IN THE PROCESS OF INHIBITING CANCER

[1]Eder Simão,[1] Jane Ludwig, [2]Cristhian Bugs, [3]Giovani Librelotto,[4] Evamberto Goes

[1]Centro Universitário Franciscano, Brazil, [2]Universidade Federal do Pampa, Brazil, [3]Universidade Federal de Santa Maria, Brazil, [4]Fundação Universidade Federal do Rio Grande, Brazil

Background: Glioblastoma (GBM) is a type of brain cancer, which can generate metastasis in other regions of this organ. It can also cause brain swelling. The disease is not curable, however the patient's survival may be improved by the use of alternative therapies in conjunction with standard treatment (surgery, chemotherapy, and radiotherapy). Upon initial diagnosis of glioblastoma, standard treatment consists of maximal surgical resection, radiotherapy, and concomitant and adjuvant chemotherapy. A recent study showed that elderly patients with glioblastoma who underwent radiotherapy had improved cancer-specific survival and overall survival compared with those who did not undergo radiotherapy treatment. The responsiveness of GBM to radiotherapy varies. In many instances, radiotherapy can induce a phase of remission, often marked with stability or regression of neurologic deficits as well as diminution in the size of the contrast-enhancing mass. Unfortunately, any period of response is short-lived because the tumor typically recurs within 1 year, resulting in further clinical deterioration of the patient. In this type of brain cancer, as well as other types of tumor, there is a loss of genome maintenance mechanisms (GMM). This is one of the most important aspects of human carcinogenesis. Methods: With the purpose of differentiating and characterizing the functionality of GMM pathways in glioblastoma and in glioblastoma treated with radiotherapy, we decided to analyze 2 datasets obtained from the Gene Expression Omnibus (GEO): 12 samples of glioblastoma, 9 samples of brain normal (GSE35493), 19 samples glioblastoma treated with radiotherapy, and 4 samples of brain normal (GSE7696). The transcriptomics were studied using the tool Viacomplex and related through a graph by using a statistical test in order to identify any alterations in pathways activities. Results: According to the results obtained here, we found a decreased expression activity in the cell cycle pathway associated to glioblastoma treated with radiotherapy. Many alterations in the apoptosis pathway and random alterations in the repair pathways also were observed by us. Conclusions:This suggests that the proteins expressed by the GMM returned to baseline levels. In addition, our findings also suggest that the GMM pathways are affected and can be a useful tool for differentiating between adenoma, cancer, syndromes and other genetic related diseases.

Keywords: Genome maintenance, Cancer and Transcriptomic Analysis

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 38

Homology modeling and molecular dynamics simulation of antiretroviral human APOBEC3H

[1]Kauê Santana da Costa, [1]Nelson Alencar, [1]Alberto Monteiro dos Santos, [1]Anderson Henrique Lima e Lima, [1]Cláudio Nahum Alves, [1]Elcio Souza Leal, [1]Jerônimo Lameira Silva

[1]UFPA, Brazil

Human APOBEC3 proteins are cytidine desaminases that exhibit varying degrees of inhibitory activity against retroviruses, such as HIV and SIV, hepadnavirus HBV, murine leukemia virus (MLV) and retrotransposons. The APOBEC3 are incorporated into budding virions, and in the next cellular infection it causes hypermutation in the viral genome by deamination of the cytosines, thus converting cytosine into uracil (dC to dU). APOBEC3H (A3H) is a monomer enzyme that is highly polymorphic and has multiple haplotypes identified. Its activity is blocked during the HIV infection by interaction with Virion Infectivity Factor, an accessory protein, which is essential for HIV replication. This study aimed to model the three-dimensional structure of A3H using the homology approach in Modeller program. The coding sequence of the A3H (residues 1-200) was obtained from UniProtKB under the accession number: B7TQM8 and a template was selected from the RCSB Protein Data Bank: the C-terminal region of APOBEC3C (PDB ID code: 3VOW, chain A, resolution: 2.15) with identity of 42.31%. The last 12 residues (188–200) at the C-terminal region of the target sequence were removed, since they showed little structure similarity to template sequence. After selecting the model, we performed loop-refinement using Modeller and the model was then submitted on the ModRefiner server to obtain atomic-level energy minimization and a reliable stereochemistry quality. Finally, the zinc ion was added to the structure using the atomic coordinates of homologous structure to form the zinc-finger motif. To investigate the stability of the modeled A3H and the tetrahedral coordination stability of zinc-finger, a molecular dynamics simulation was carried out in the AMBER 12 package and the Cationic Dummy Atom (CaDA) approach and the AMBER FF99SB force field were employed to treat the zinc-finger and the rest of the system, respectively. The stereochemical quality of the proposed model of A3H was evaluated using the PROCHECK tool. The Ramachandran plot showed 91% (152) residues in highly favorable regions (the core), 6.6% (11) in additionally allowed, and 1.2% (2) in generously allowed regions; together, they sum a total 98.8% of residues within allowed regions of the graph and 1.2% (2) in disallowed regions. The results of RMSD shows that the A3H model reaches a plateau at 10 ns of MD simulation and the zinc-finger motif, formed by the residues His54, Glu56, Cys85 and Cys88, maintained the tetrahedral geometry. Therefore, the present model of A3H exhibits an excellent stereochemistry quality, and showed a good ANOLEA and Qmean energy score profile, indicating reliability of the structure prediction. In future, docking studies will be necessary to analyze the interaction between Virion Infectivity Factor of HIV-1 and human A3H structure.

Keywords: APOBEC3H, homology modeling, HIV, molecular dynamics

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 39

Entropy Maximization Applied to Analysis of Bacterial Genomic Sequences

[1]Gesiele Carvalho, [2]Fabrício Lopes,[1] Marie Van Sluys, , , , , , , , , , , , , AU

[1]University of Sao Paulo, Brazil, [2]Federal University of Technology - Campus Cornelio Procopio - Parana, Brazil

Backgroung: In the context of information theory, the Shannon Entropy can be generalized as the measure of information from a particular source or a system, which can be applied to several areas of study and with different purposes. In recent years, various methods based on the entropy principle have been used successfully in different ways. In particular, the analysis of genomic sequences, such as topological entropy of DNA, relative entropy, among others, also received attention in bioinformatics research area. However, these studies just analyze the sequences in general, and now, with the exponential growth in the sequencing of complete bacterial genomes, there is a need to develop a methodology that is able to analyze these data in a direct and easy way, enabling the extraction of relevant information to scientific advance. In this context, the entropy maximization (EM), another method based on Shannon entropy, can be exploited for this purpose. The EM is originally based on principles of statistical mechanics, which aims to provide a probability distribution of a system, in this case in a genome, where the maximum entropy corresponds to the best information (great variation), this technique has been widely exploited in ecological studies. Thus, knowing the capacity EM has to provide important information within a system, our work aims to develop a fast and automated method still not used for DNA analysis, to assess the entropy distribution variation, within bacterial genomes. This method may direct the research to the regions that possibly provide important information about the organism under study

Keywords: Bacterial Genomes, DNA sequence, Entropy

Topic: Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 40

TRANSAMAZON-DB: TRANSCRIPTOME DATABASE OF AMAZONIAN SPECIES

[1]José Deney Alves de Araújo, [1]Andrea Ghelfi

[1]Universidade Federal do Amazonas, Brazil

Amazonia is the largest and most diverse of the tropical forest wilderness areas, and it is not homogeneous in its animal and plant communities. It is a mosaic of distinct endemic areas separated by major rivers, each of them with their own evolutionary relationships and biotic assemblages. Thus, the identification and the high-throughput sequencing of these species, in their natural environment, are of major concern. This is the main goal of the project INCT-ADAPTA, which we are collaborators. In order to deal with the wide amount of date generated by the differential expression analysis of the sequenced Amazonian species we developed TRANSAMAZON-DB. The system was developed in postgreSQL, for the database, an PHP language, for the interface. TRANSAMAZON-DB has three types of users: the administrator; the common users; and the curator, for manual annotations of the database. This database has the identification and annotation of the expressed genes (gene id, gene name, gene symbol) and has also statistical bioinformatics analysis of the differential expression like logFC, p-value, FDR values, graphics of dispersion, histogram and box plot. All these informations were generated by “Tamandua”, a pipeline developed by our research group to analyze transcriptome datasets. This is the first database of transcriptome information of Amazonian species.

Keywords: Database, transcriptome, differential expression, bioinformatics.

Topic: Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 41

A UML Profile for the Semantic Annotation of Web Services for Gene Expression Analysis

[1]Gabriela [2]Guardia, Luís Pires, [1]Cléver de Farias

[1]University of São Paulo, Brazil, [2]University of Twente, Netherlands

Background: In order to acquire new biological insights, biologists frequently need to carry out a number of analysis activities on gene expression data using different software tools and data sources. The vast amount of gene expression data formats and the development of standalone, incompatible software tools for data analysis pose a challenge for data/tool integration. In this context, the development of semantic web services capable of providing the desired analysis functionalities aims at facilitating the integration in the domain. A semantic web service consists of a web service whose description is annotated with semantic information from a domain-specific ontology or conceptual model. Different approaches have been developed to support the semantic annotation of service descriptions, such as OWL-S, SA-REST and SAWSDL. The Semantic Annotations for WSDL (SAWSDL) approach defines a set of mechanisms that enable the assignment of semantic annotations (ontology concepts) to various components of a service description written in the Web Service Description Language (WSDL). These semantic annotations are intended to unambiguously describe different aspects of a service, such as its interface, operations, and input/output message formats through the association of specific ontology concepts. However, the assignment of semantics to documents written in machine-readable XML-based languages such as WSDL may be not a trivial task for non-experts and the use of graphical languages can facilitate this activity. The Unified Modeling Language (UML) is a standard, widely-used graphical language for modeling computer systems. UML provides a well-defined set of modeling elements, which can be extended using a built-in extension mechanism named profile. This work aims at developing a UML profile for the SAWDSL approach in order to provide a set of specific modeling elements that can be used to facilitate and standardize the modeling of SAWSDL annotations assigned to semantic web services developed for gene expression analysis. Results: First, we have studied the WSDL 2.0 language, the SAWSDL approach, and the UML metamodel and profiling mechanism. Second, we have proposed extensions to the UML metamodel in compliance with the SAWSDL approach and defined a profile that implements the proposed extensions. Finally, we have applied the proposed UML profile in the development of a number of models representing the semantic annotation of different web services developed for gene expression analysis. Conclusions: The use of a well-established graphical language such as UML provides a more intuitive manner for semantically annotating service descriptions than only using XML-based notations such as WSDL. Although the profile can be used in any general-purpose UML modeling tool, future research is needed towards the development of a support tool that enables not only the creation and visualization of models but also the automated translation of such models into semantically annotated WSDL service descriptions.

Keywords: SAWSDL; UML profile; semantic web services; gene expression analysis

Topic: Transcriptomics and Proteomics

Abstract #: 42

Searching for circular RNAs in the archaeon Halobacterium salinarum

[1]Livia Zaramela, [1]Felipe ten Caten, [1]Eliane Traldi, Ricardo Vêncio, [1]Tie Koide

[1]Universidade de São Paulo, Brazil

Circular RNAs (circRNAs) have been reported in all domains of life. Despite the initial idea that circRNAs are particularly rare and mainly intermediate products of tRNA and rRNA processing, several studies have shown the involvement of these molecules in different regulatory processes. Recent discoveries brought up an unrecognized regulatory potential of circRNAs. In animals, for example, circRNAs that bind and block microRNAs modulators were described. Furthemore, it was shown that RNA splicing that results in circular RNA isoforms is a general feature of gene expression in eukaryotes. In Archaea, computational approaches have shown that non-coding RNAs are often processed to yield a circular form, and some C/D box snoRNAs involved in rRNA processing are circRNAs. Given this under-explored world, we combined high-throughput sequencing and computational tools to explore systematically the presence of circRNAs in the archaeon H. salinarum. In this study, we selected in several strand-specific RNA-seq data, about 7 millions of reads that do not map in H. salinarum genome by conventional alignment approaches. Then, we searched within the reads for junctions between two sequences separated by 10nt to 2000nt in the reference genome. This approach is supported by the fact that a junction can indicate the connection point between the boundaries of the linear transcript. Looking for this feature, it was possible to map a junction point in 8% of the reads. Preliminary analysis show an enrichment of circRNAs in rRNAs and tRNAs (about 85%), as expected. Additional analysis are being performed to classify all the circRNAs found, which will be then selected for functional characterization.

Keywords: Halobacterium salinarum, circular RNAs, high-throughput sequencing

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 43

Bioinformatics applied to Halobacterium salinarum RNA structurome investigation

[1]Eliane Traldi, [1]Felipe ten-Caten, [2]Lívia Zaramela, [1]Tie Koide, [1]Ricardo Vêncio

[1]USP, Brazil, [2]FMRP - USP, Brazil

Structure implies function. In the universe of biomolecules, this is a highly accepted paradigm. However, when the focus is the transcriptome, the structural configuration is a layer of information that is still poorly used. From the recognition of the different functions performed by the RNAs, it is fundamental that their structure is unveiled so that it is possible to understand how these biomolecules act in a variety of processes that are crucial to the existence and maintenance of biological systems, such as storage, transport and codification of genetic information, in addition to catalysis and regulation activities. A variety of experimental techniques were developed to resolve in detail tridimensional molecular structures as well as to obtain a more simplified elucidation of internal basepairings in bidimensional models. However, in the vast majority of cases, these approaches focus on the structure of only one or a few molecules per experiment, in addition to have technical limitations related to the size of the molecule being investigated. Because of that, development and application of algorithms and computational techniques that allow one to create models and structure predictions in silico was essential. Even with a diverse variety of existent methodologies, the determination of RNA structure still has limitations and occurs in small scale in most of the studies. In order to have a systems approach in which different dimensions of information of the same organism have to be integrated in a dynamic way aiming to obtain a global understanding of living beings, structural studies must be performed in large scale. Structuromics allows one to incorporate the dimension of the RNA Structurome (structure of all RNA) to the emerging field of Systems Biology. In this work, large scale modeling in silico of the RNA structurome is being done, using the extremophyle Halobacterium salinarum as the study organism. ContextFold, an RNA secondary structure prediction tool that uses rich parameterized machine learning models, was selected as the single-sequence method to be used for the structure prediction. This choice was based on the tests of CompaRNA, a server for continuous benchmarking of automated methods for RNA structure prediction, where the performance of several single-sequence and comparative methods are evaluated on two benchmarks, one static generated on structures derived from the RNAstrand database, and one with RNA sequences/structures released weekly by the Protein Data Bank. From this large scale modeling we expect to classify the molecules in structural groups, in addition to investigate a correlation between these groups and information about function and regulation of the transcriptome.

Keywords: RNA structure, structural groups

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 44

Molecular dynamics of two Glutathione S-transferases (AgGSTE2 and AgGSTE5) from the malaria vector <Anopheles gambiae>.

[2]Rafael Trindade Maia, [2]Constância Flávia Junqueira Ayres, [1]Thereza Amélia Soares

[1]Universidade Federal de Pernambuco, Brazil, [2]Centro de Pesquisas Aggeu Magalhães, Brazil

The Glutathione S-transferases (GSTs) are members of a multifunctional superfamily of enzymes of enzymes that catalyzes the conjugation of GSH (the reduced form of glutathione) to xenobiotics, such as to chemical insecticides resistance. The GST epsilon class (GSTE) has eight members (GSTE1, GSTE2, GSTE3, GSTE4, GSTE5, GSTE6, GSTE7 and GSTE8) in Anopheles gambiae, the main malaria vector. Previous studies with four Anopheles species showed that the GSTE2 gene was under strong negative selection while GSTE5 was the only gene in the epsilon cluster on this class that has sites evolving under positive selection. In other study was found an Anopheles gambiae population which was resistant to DDT and had two mutations (I114T, F120L) on GSTE2 enzyme (AgGSTE2) in all individuals. In this study we aimed to compare the molecular dynamics of the Anopheles gambiae GSTE5 (AgGSTE5) and the two isoforms of An. gambiae GSTE2 (AgGSTE2 and AgGSTE2mut). The AgGSTE2 atomic coordinates were obtained from crystallographic structure (PDB ID 2IMI, 2IL3) while a homology model was used for the AgGSTE5. The enzymes were submitted to 50 nanoseconds of molecular dynamic simulations with and without the tripeptide glutathione (ligand GSH,) in a total of six systems. The molecules were centralized in a 9.5 x 8.5 x 9.5 nm3 cubic box and solvated with water molecules (SPC model) and counter-ions were added to neutralize the protein total charge. The systems were maintained under isothermic (300 K) and isobaric conditions (1 Pa) for a relaxing period of 0.2 picoseconds. The GROMOS force field with the G53A6 parameters set was used. The simulations were runned and analyzed on GROMACS software package. Secondary structure (DSSP), Root-mean square deviation (RMSD), root-mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA), principal component analysis (PCA) and clustering analysis were performed for all systems. The results showed that all the six systems displays structural stability. DSSP did not change along simulations indicating that there are no denaturation process. RMSD values showed different profiles for the six systems. RMSF showed that the most flexible regions differed in the proteins. RMSD distribution suggested that the AgGSTE2mut (with the ligand GSH) was the enzyme that displays more conformational heterogeneity. Although the Rg increases in the presence of GSH glutathione for all systems the SASA value (120-135 nm2) did not changed. PCA and cluster analyzes showed that the AgGSTE2mut has more conformational substates compared to its homologous AgGSTE5 and the isoform AgGSTE2. Results allows us to conclude that the three enzymes are promiscuous, and the polymorphisms found on AgGSTE2mut is the reason for notable changes on its dynamics. We also could observe that the presence of glutathione implicates in differences on the protein dynamics. These data are important to provide insights into the functional diversification of GSTs and could help to understand better their role in insecticide resistance.

Keywords: Insecticide resistance, protein motions, enzyme promiscuity

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 45

Holometabolous insects trypsin evolution: phylogenetic proposal and structural characterization

[1]Renata de Oliveira Dias, [2]Allegra Via, [1]Marcelo Mendes Brandão, [2]Anna Tramontano, [1]Marcio de Castro Silva-Filho

[1]University of Sao Paulo – Luiz de Queiroz College of Agriculture, Brazil, [2]Sapienza University in Rome, Italy

Insecta are one of the evolutionary best-adapted groups of organisms in the world and much of their adaptation is due to the ability of exploiting diverse types of energy intake sources. It is believed that this characteristic could be correlated with the set of digestive enzymes expressed by this group of organisms. Trypsins, which belong to a well-studied class of serine peptidases, are mostly responsible for the insect protein digestion by cleaving peptide bonds at the carboxyl side of basic L-amino acids. For the trypsin action, at least three processes are relevant: precursor (trypsinogen) activation, secretion, and specificity of the active trypsin substrate-binding site. In this context, this work aimed to understand how the evolution of holometabous insects shaped the active and inactive trypsins in three orders: Coleoptera, Diptera and Lepidoptera. Three distinct regions were analyzed in terms of evolution: a) the active trypsin, with a focus on the substrate-binding site region and two regions of the inactive trypsin precursor: b) the signal peptide and c) the activation motif. The analysis of the active trypsin sequence encompassed Bayesian phylogenetic propositions, three-dimensional structure prediction, binding site composition characterization and site-by-site selection pressure analysis. The phylogenetic intra-group analysis revealed a high diversification of these genes, including orthologous and putative paralogous in all orders. Moreover, the phylogenetic inter-group analysis and the binding site characterization of these proteins showed an evolutionary profile in the Lepidoptera enzymes contrasting with those of Coleoptera and Diptera. The Lepidoptera trypsin sequences highlighted two distinct groups: Group I - composed of trypsins of insect from distinct families, and Group II – composed exclusively by trypsins belonging to species from the Noctuidae family. The analysis of the predicted three-dimensional structure of the Lepidoptera trypsins allowed us to detect relevant modifications in the substrate-binding region of the Lepidoptera trypsins from Group II, in particular a greater hydrophobicity than trypsins of Group I and other insect orders. These modifications also were predicted to be negatively selected in the site-by-site evolution pressure analysis in Lepidoptera. As for the signal peptide, our phylogenetic analysis showed a high conservation only in close genes. Finally, the analysis of the activation motif highlighted a tendency of Lepidoptera trypsins toward increased auto-activation. This finding is consistent with peculiar characteristics of Lepidoptera feeding habit. Our work showed that the evolution of trypsins in holometabolous insects culminated in a specialized group of enzymes in Lepidoptera thus suggesting that these proteins can be associated with specific adaptive traits observed for this group of insects.

Keywords: Insect; Protease; Trypsin.

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Transcriptomics and Proteomics

Abstract #: 46

Detection of novel non-coding RNAs of the HgcC family in the extremophile Halobacterium salinarum NRC-1 integrating in silico and experimental data

[1]José Vicente Gomes Filho, [1]Lívia Soares Zaramela, [1]Felipe Ten Caten, [2]Valéria Italiani, [1]Tie Koide, , , , , , , , , , , AU

[1]Universidade de São Paulo (USP), Brazil, [2], Brazil

Background: Non-coding RNAs are ubiquitous in all domains of life, being key elements in gene regulation. In Archaea, amongst a plethora of novel non-coding RNAs, there is one interesting family that was first described in AT rich extremophiles, the HgcC family(High GC content RNA), which functions are yet to be described. In H. salinarum NRC-1, one HgcC RNA is cataloged in RFAM database. To improve the studies of this conserved RNA family, we decided to search for RNAs of this family in H.salinarum NRC-1 using bioinformatic tools combined with the analysis of transcriptome data.

Results: Using the core sequence of the already cataloged HgcC RNA, we were able to find 2 intergenic and 4 intragenic transcripts carrying the conserved motif related to HgcC RNAs. Interestingly, all 4 intragenic hits were related to a group of conserved Archaeal/Bacterial Insertion Sequence family, the IS605 family. Another relevant aspect is that the HgcC sequences found in IS605 genes are also the sequences where the transposase TnpA cleaves the DNA to perform transposition. Using Tilling Array data obtained during a growth curve of H. salinarum NRC-1, we found that the regions containing HgcC motifs are differentially expressed and we could map transcription start sites for some of these transcripts, using RNAseq data.

Conclusions: The finding of new HgcC members that are differentially expressed in H. salinarum NRC-1 provide us a wider range of targets for functional analysis. Further experiments, like superexpression and deletion of these RNAs should give us an overview of their impact in the phenotype of H.salinarum NRC-1. Their relationship with IS605 Genes are also going to be investigated, as they might be important players in the transposition regulation.

Keywords: ncRNAs, transposases, extremophile, archaea,

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 49

SIMTar: A tool for prediction of SNPs Interfering on MicroRNA Target sites

[1]Amanda Rusiska Piovezani, [1]Helena Brentani, [1]Ariane Machado-Lima

[1]University of Sao Paulo, Brazil

Background: The identification of SNPs modifying target sites of miRNAs has gained an important prominence in recent years due to the improvements found on the ability of miRNAs as regulatory elements in the genome. In addition, many microRNAs are associated with diseases such as cancer and various psychiatric disorders. The computational resources currently available for this purpose are restricted to the analysis of SNPs in the 3'UTR (UnTranslated Regions) of mRNAs, where the miRNAs typically bind to repress their translation. However, this is a simplification of the problem, once it is already known that miRNAs can also activate or repress gene transcription when bound to gene promoter region, can increase the effectiveness of negative regulation of translation when bound to the gene coding region. In addition, miRNAs can also bind non-coding RNAs. The actual resources are also limited to the identification of SNPs in the miRNAs seed region, and therefore only identify site creations or disruptions. However, SNPs located outside this region can not only create and disrupt target sites but also interfere on the stability of miRNAs binding and therefore on the effectiveness of regulation. Moreover, considering the length of the target site, more than one SNP can occur inside of a binding site and thus, the combination of these SNPs can have an even greater influence on the miRNA binding with its target. In addition, current resources do not display which alleles of SNPs or what combinations of them are causing which effect. Finally, these resources are restricted to the Homo sapiens and Mus musculus. Results: Here we present the computational tool SIMTar (SNPs Interfering in MicroRNA Targets), developed to identify SNPs that alter miRNA target sites and fills the mentioned gaps. Conclusions: SIMTar is the first open source tool using this approach which can be applied to other species than the human.

Keywords: SNPs, microRNAs.

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Transcriptomics and Proteomics

Abstract #: 50

Alternative splicing events are linked to behavioral plasticity in honey bees

[1]Fabricio Baía, [1]Alexandre R. Paschoal, [2]Francis M. F. Nunes, [2]André Y. Kashiwabara

[1]Universidade Tecnológica Federal do Paraná, Brazil, [2]Departamento de Genética e Evolução, Universidade Federal de São Carlos, Brazil

RNA sequencing technologies (RNA-Seq) has challenged our current understanding about the complexity of life. RNA-seq data allow us to analyze a myriad of biological aspect from diverse transcriptomes, such as, levels of expression, novel transcripts, detection of polymorphism or alternative splicing (AS) variants. There are a lot of datasets available in public databases (e.g. NCBI-SRA) which can be used for bioinformatic analysis to explore differentially expressed AS events between two or more conditions from the same species. Honey bees (Apis mellifera) are social insects which display a high phenotypic plasticity from a single genotype. For example, adult workers show distinct behaviors, such as nursing and foraging activities, depending on age, nutrition, and colony needs. These genetic and environmental influences on both behavior and gene regulation, motivated us to investigate to what extent this plasticity is explained by AS events. In this work we developed a bioinformatic protocol to analyze SRA data generated from brains of nurse (SRR071803 to SRR071813) and forager (SRR071814 to SRR071825) bees in order to indentify differentially AS variants. Thus, we used Bowtie 2 to index the Amel_4.5 reference genome which served as an input for TopHat aligments of fastq files. Cufflinks, Cuffmerge and Cuffcompare were used for assemblies, transcript identification and list comparisons. Finally, ASTAlavista program was used to check for potential AS events between differentially expressed genes. In general, around 31% of honey bee genes (4,669) presented spliced variants in adult worker brains. A total of 211 and 208 genes were exclusively expressed in nurses and forager, respectively, emerging as markers of age and behavior. Also, the most frequent AS events found were (i) alternative acceptor; (ii) skip one exon; (iii) retain one intron; (iv) alternative donor. Comparisons indicated Rpb8 (a DNA-directed RNA polymerase) and emb (embargoed) as top key genes. Both are involved in cell-cycle regulation, Notch pathway regulation, and RTK-Ras-ERK pathway regulation. This study represents the first large-scale analysis of AS events regulating behavioral changes in social insects, adding new information toward the comprehension of phenotypic plasticity. Supported by: Universidade Tecnológica Federal do Paraná and Fundação Araucária

Keywords: alternative splicing, honey bee, sequence analysis, RNA-seq

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Transcriptomics and Proteomics,Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 51

Search of Cell-Specific Transcription Factor Binding Sites with DNase Hypersensitivity and Histone Modifications

[1]Eduardo Gusmão, [2]Christoph Dieterich, [1]Ivan Costa

[1]IZKF & Institute for Biomedical Engineering, RWTH University Medical School, Germany, Germany, [2]Max Delbrück Center for Molecular Medicine Berlin, Germany, Germany

Eukaryotic gene expression involves the coordination of a multitude of transcription factors that bind on specific cis-acting DNA elements. The understanding of the complex regulatory networks is crucial for the comprehension of biological processes such as cell differentiation and the onset of diseases. Standard sequence-based computational approaches to find binding sites suffers from a high number of false positive hits, given its incapability to identify active sites. Current research has proven that novel genome-wide assays reflecting chromatin structure, such as DNase I digestion (obtained with DNase-seq) or histone modifications (obtained with ChIP-seq) outperform sequence-based detection of transcription factor binding sites that are active in a particular cell type. Moreover, the discovery of distinct modes of chromatin signatures have strengthened these results and reinforced the usage of epigenetic datasets to such purpose.

Previous methods on detecting cell-specific binding sites from chromatin information can be categorized in two groups: site-centric and segmentation-based. Both use cell-specific experimental data but site-centric methods require sequence information to make factor-specific predictions while segmentation-based methods annotate the genome for likely binding locations for any transcription factor. In this study, multivariate hidden Markov models are used to detect transcription factor binding locations using genome-wide high-resolution DNase I digestion and histone modification data. In this segmentation-based methodology, a complete regulatory map can be generated with a single run using experimental data for a specific cell-type.

We used public data from ENCODE project to predict binding sites in the cell lines H1-hESC, HeLa-S3, HepG2 and K562. Moreover, we performed a comprehensive comparative analysis of competing methods using a well-established gold standard, which was generated by combining sequence motifs with ENCODE's ChIP-seq data for 22 factors. Our method outperformed all other competing methods concerning a Friedman-Nemenyi test applied to the default metric in this scenario: the area under the ROC curve. Furthermore, we observed that our strategy provides a better balance between sensitivity and specificity for detection of active binding sites. For instance, our method presented 83% sensitivity and 96% specificity regarding GABPA binding in H1-hESC, while the segmentation-based method by Boyle et al. achieved 59% sensitivity and 98% specificity. In contrast, site-centric methods were more sensitive but showed low specificity. For instance, Cuellar-Partida et al. achieved 87% sensitivity but only 84% specificity in the same previous scenario, the lowest among all tested methods. Additional experiments demonstrated that our method is independent of cell-type concerning the model's training and that it outperforms all other segmentation-based methods concerning their spatial accuracy, i.e. how close predicted regions are to actual binding sites. Finally, improvement of genome-wide high-resolution data quality as a result of advances in high-throughput sequencing techniques, gives rise to a number of potential applications to the framework presented here.

Keywords: transcription factor binding site, DNase-seq, ChIP-seq, hidden Markov model

Topic: Transcriptomics and Proteomics

Abstract #: 52

Statistical analysis applied to GVHD: pathways and genes expression

[1]Franciele Roggia Brondani, [2]Éder Maquel Simão, [3]Evamberto Garcia Góes

[1]Universidade do Rio Grande (FURG), Brazil, [2]Centro Universitario Franciscano (UNIFRA), Brazil, [3]Universidade Federal do Rio Grande (FURG), Brazil

Purpose: Graft-versus-host disease (GVHD) is an immunologically mediated inflammatory reaction, and remains a major cause of morbidity and mortality in patients undergoing allogeneic hematopoietic stem cell transplant1. In addition, the role of various cells, pathways, and others factors associated to GVHD remains an tissue2,3. In this study, a statistical analysis was applied to GVHD in order to better delineate pathways and genes expression related to the development of the disease.
Methods: The National Cancer Institute (NCI) – Nature – Interaction Pathway and Reactome were used as reference databases in order to study Class-I and II major histocompatibility complex (MHC), TNF activation, apoptosis and inflammation pathways related to GVHD. Transcriptome data were obtained using the Affymetrix Human Genome U133 Plus platforms GPL570. Relative activity and relative diversity were used to determine alterations in a specific pathway and, in such cases the fold change of each gene was applied. Analyses were conducted using Via Complex and Bioconductor packages such as limma, biobase, annaffy and hguplus2.db.
Results: We observed alterations in the relative activity and relative diversity regarding to pathways associated to MHC-Class I. MRC1 was a gene located across the fold change, which contained significant alterations. This finding is important because this gene participates in cell recognition, as well as participates in the processes of neutralization of pathogens. The pathway related to the phosphorylation of CD3 and TCR zeta chains also was altered. This pathway is activated by the CD3 and TCR genes, which are related to the T-cells growth. The genes IL2 and CD69 showed high fold change. These findings are important because these genes are associated with cell activation (NK and T-Cells) and toxicity related to GVHD.
Conclusion: Relative activity, diversity and fold change techniques applied to GVHD showed that MHC-Class I is the major pathways of this disease.

Keywords: GVHD, gene, statistical analysis, expression

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics,Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 53

A Quantum Biochemical Willardiines Binding Energy Basis for Partial Agonist Action at Ionotropic Glutamate Receptors

[1]José Xavier Lima, [1]Jonas Nobre, [1]Gabriela Ourique, [1]Umberto Fulco, [1]Eudenilson Albuquerque, [2]Valder Freire

[1]Universidade Federal do Rio Grande do Norte, Brazil, [2]Universidade Federal do Ceará, Brazil

The α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) is an ionotropic glutamate receptor (iGluR) responsible for the rapid excitatory synaptic transmission in the Central Nervous System (CNS). Impairments in AMPA function are correlated with the development of many brain disorders, such as stroke, schizophrenia and autism. The use of 5-substituted willardiines has been a powerful tool to the understanding of partial agonist activation and desensitization mechanisms. In this work, we employed quantum biochemistry computation, based on the Density Functional Theory (DFT) approach, to unveil the protonation state of willardiines complexed with the GluR2. Results demonstrate that the total binding energies of willardiines in the GluR2 ligand-pocket is correlated with their agonist action. However, the matched correlation was obtained only after considering the uncharged uracil ring form of willardiines through an extended analysis of 9.5Å willardines-GluR2 binding pocket radius, enough to stabilize the total energy. It indicates that the main contributions to the total willardiines-GluR2 binding energy are due to the residues in the following order Glu705 > Arg485 > Tyr450. Furthermore, Met708, which is positioned close to the 5-substituent, shows attractive interaction with the willardiines HW and FW, and repulsive interaction with the willardiines BrW and IW. The present data contribute to the understanding of the willardiines binding mechanisms and provide a quantum biochemical energy-based model of partial agonist binding in GluR2.

Keywords: Binding energies; DFT calculations within the MFCC approach; Willardiines; Ionotropic Glutamate Receptor

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 55

VERMONT: a tool to analyze mutations

[1]Alexandre Fassio, [1]Sabrina Silveira, [1]Valdete Gonçalves-Almeida, [1]Yussif Barcelos, [1]Elisa Lima, [1]Flávia Aburjaile, [1]Laerte Rodrigues, [1]Wagner Meira Jr., [1]Raquel Melo-Minardi

[1]Universidade Federal de Minas Gerais, Brazil

Some DNA mutations (i.e., substitutions, insertions and deletions), which occur naturally due to evolutionary pressure, are known to affect protein function. A significant open problem in science and in Bioinformatics consists of collecting, integrating and processing a huge amount of information, and ultimately presenting it in a simple and visual manner in order to facilitate both comprehension by scientists and the solution of intriguing problems. How may specific modifications in protein amino acid properties help to identify potentially significant mutations? Depending on where these mutations occur, a protein can lose its function or become inactive. Some mutations can cause destabilization and significant structural modifications. The goal of this work is to develop a visualization tool capable of identifying mutations and pointing out possible consequences for protein function. We propose VERMONT (ViewER MutatiON Tool), an interactive tool to visualize mutations and their possible consequences for protein structure and function. We modeled the problem as spotting residue conservations together with the conservation of physicochemical and topological properties. The proposed interactive visualization provides a macro view of the multiple structure-based sequence alignment as well as several other structural parameters. The tool allows users to view, at a glance, a multivariate set of residue parameters, expanding or compressing panels and zooming out to see full-length of sequences, or zooming in to focus on parts of them. Users may also filter residues individually or in groups of similar properties by highlighting or attenuating them, which helps in spotting patterns and exceptions. Using the proposed visualization tool, we were able to predict mutations we believe have a significant probability of causing damage to protein function and some these seem to be more severe and have a high likelihood of causing function loss.

Keywords: mutations, protein function damage, visualization tool, information visualization

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 56

Three-dimensional solution of C-terminal extension of cysteine proteinase B of Leishmania (Leishmania) amazonensis

[1]Deborah Santos, [1]Ernesto Caffarena, [1]Carlos Alves

[1]Fiocruz, Brazil

Background: Leishmania (Leishmania) amazonensis is an important etiological agent of cutaneous leishmaniasis in humans in South America, including Brazil. This parasite has adaptive mechanisms that can be guided by their proteinases, which features the cysteine proteinase (CP) as the most extensively studied CPB among CPs described in Leishmania spp. This work is focused on the COOH -terminal region of CPB (cyspep). It is already known that its derived peptides present modulatory properties on the immune system of mice, although this mechanism has not been adequately described yet. This present study aims at proposing a structural model for the COOH -terminal region of the CPB of L. (L.) amazonensis targeting its structure-function relationship. The study was conducted by in silico to develop a three-dimensional model using comparative modeling, threading and ab initio methods, and the structural stability was analyzed by molecular dynamics. Results: Through secondary structure prediction and disulfide bridges information, four structures were selected, one derived from a template by applying comparative modeling (Comp), one from threading (T1) and two ab initio models (Ab5, Ab7); all of them were subjected to a 200 ns molecular dynamics simulation. At the end of this process, the threading model secondary structure was close to the one predicted by the servers, however the ab initio models go apart themselves from it. The homology model lost the alpha helix structures, maintaining only minor beta structures, similar to what was formerly predicted. Regarding the disulfide bonds, models Ab5 and Ab7 showed better outcomes when compared to the others, remaining approximately 39 % and 46 % of the time, respectively, at a distance between the pair of cysteine residues 73-85, predicted by the DiANNA server. The predicted models obtained by the Scratch server, Comp and Ab5 remained approximately 35 % of the time at a distance less than 5.5 Å between the pair of cysteine 31-52 and the T1 model within 50 % of the time between the pair 21-44. Conclusions: Out of the different techniques for building three-dimensional structures, the ab initio models were more faithful to the secondary structure prediction. None of the analyzed models stabilized over simulation time indicating the need for higher run-times. The models here evaluated can be considered as putative structures, in a way that it is necessary to use experimental data to provide more reliability to the models.

Keywords: Leishmania cysteine proteinase; COOH-terminal extension; protein structure; molecular dynamics

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 58

SUFFIX ARRAYS AND GENOME ASSEMBLY

[1]Felipe A. Louza, [2]Steve Hoffmann, [2]Peter F. Stadler, [3]Guilherme P. Telles, [1]Cristina D. A. Ciferri

[1]University of São Paulo, Brazil, [2]University of Leipzig, Germany, [3]University of Campinas, Brazil

Suffix arrays play an important role in several string processing problems. In bioinformatics, suffix arrays have been used in different tasks, from similarity searches to genome assembly. Recent advances in sequencing technologies are reducing the time and the costs to generate massive amounts of sequenced data. In general, the volume of such data exceeds the capacity of the internal memory. Then, it becomes necessary to develop methods to treat large data volumes considering algorithms and data structures in external memory. Many strategies for genome assembly have been devised. However, most of them build up on the idea to traverse read-overlap or de Bruijin graphs. Recently, Simpson and Durbin (2010) used the FM-index to construct string graphs, and showed that it can efficiently derive the genome assembly. However, maintaining the data structures used by the FM-index in internal memory requires a large memory space and can impair their usage. This scenario is typical with NGS data amounts. In this work we present an extension to the eGSA algorithm to handle large collections of small strings (e.g. reads). And we propose a new algorithm to construct string graphs using suffix arrays in external memory. The proposed algorithm is being implemented and validated through comparative tests with related works. As future work, we intend to extend our construction to identify repetitive regions.

Keywords: suffix array, genome assembly, string graph

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 59

Protein-Protein Interaction (PPI) network prediction based on structural information of proteins coded in genomes of Leishmania species

[1]Antonio Rezende

[1]Centro de Pesquisas Aggeu Magalhães - FIOCRUZ, Brazil

One of the main goals of modern biology is make possible a high prediction power for perturbation results that have not been assessed yet. One of the approaches to reach this aim is the cellular biological network study, for instance, protein-protein interaction networks (PPI networks). These studies provide information concerned which proteins of a genome interact each other, and how they do that.Therefore, the main goal of this work is to increment and to reinforce the PPIs modeled for Leishmania braziliensis, Leishmania major and Leishmania infantum (Rezende et al 2012) using structural information, and in the end to performing a more effective search for new targets for developing drugs and vaccines against these microorganisms. Up to now, the Phyre2 software is being evaluated to model protein structures of Trypanosomatids, thus 1000 L. major proteins randomly selected were modeled. In parallel, the tool ZDOCK for molecular docking is also being assessed. For that 158 binary complexes were downloaded from PDB. These complexes had their chains separated in different files, and these ones were used as input for ZDOCK. The results were compared to original PDB complexes applying the MMalign, which align complex structures in a pair-wise way. ROC curves are been applied to assess the performance of this docking tool. The Phyre2 generated 3084 models, and among them 93% had a confidence value higher than 0.9, which is considered a confident prediction, however these high confident model had a average of coverage of 54.4%, and 39% of the confident models had a coverage higher than 70%. For the evaluation of ZDOCK, the alignment score between PDB complex and predicted complex was considered against two measures recovered from dock analysis, they are ZDOCK score and amplitude. Each alignment was manually evaluated, and alignments with score higher than 0.91 indicated a correct prediction by ZDOCK. From the two measures extracted from docking analysis, the amplitude presented a better performance when analyzed using ROC curve, thus its AUC value presented a value equal to 0.615. Therefore, the results point to a feasible utilization of protein modeling and molecular docking in a large scale to predict PPI networks. The next step will be the evaluation of I-TASSER modeling and the RosettaDock tools. In the end, when the new data for Leishmania PPI networks are added to the previously ones, the new PPIs will be evaluated for their number of interaction, modularity feature and topological index for all proteins mapped in the networks.

Keywords: Leishmania, protein interaction, protein structure

Topic: Transcriptomics and Proteomics

Abstract #: 60

RNA-sequencing analysis of Trichophyton rubrum transcriptome in response to sublethal doses of acriflavine

[1]Gabriela F Persinoti, [1]Nalu TA Peres, [1]Tiago R Jacob, [2]Antonio Rossi, [3]Ricardo ZN Vêncio, [1]Nilce Martinez-Rossi

[1]Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Brazil, [2]Department of Biochemistry and Immunology, Ribeirão Preto Medical School, University of São Paulo, Brazil, [3]Department of Computer Science and Mathematics (DCM/FFCLRP), University of São Paulo, Brazil

Background
The dermatophyte Trichophyton rubrum is an anthropophilic filamentous fungus that infects keratinized tissues and is the most common etiologic agent isolated in human dermatophytoses. The clinical treatment of these infections is challenging because it is prolonged, costly, and there is a limited number of cellular targets. Moreover, most antifungal drugs commercially available acts on ergosterol. One drug that apparently does not act on ergosterol and hast antifungal activity is acriflavine. To understand the mode of action of cytotoxic drugs against fungi, we evaluated the time-dependent effects of acriflavine on T. rubrum transcriptome using high-throughput RNA-sequencing (RNA-seq) technology.
Results
RNA-seq analysis generated approximately 200 million short reads that were mapped to the Broad Institute's Dermatophyte Comparative Database before differential gene expression analysis was performed. By employing a stringent cut-off threshold of -1.5 and 1.5-log2-fold changes in gene expression, a subset of 490 unique genes were found to be modulated in T. rubrum in response to acriflavine. Among the selected genes, 69 genes were modulated at all exposure time points. Functional categorization indicated the putative involvement of these genes in various cellular processes such as oxidation-reduction reaction, transmembrane transport, and metal ion binding. Interestingly, genes putatively involved in the pathogenicity of dermatophytoses were down-regulated suggesting that this drug interferes with the virulence of T. rubrum. Moreover, we identified 159 novel putative transcripts in intergenic regions and two transcripts in intron regions of the T. rubrum genome, indicating the presence of novel transcripts expressed in response to acriflavine.
Conclusion
The results provide insights into the molecular events underlying the stress responses of T. rubrum to acriflavine, revealing that this drug interfered with important factors involved in the establishment and maintenance of fungal infection in the host. In addition, the identification of novel transcripts will further enable the improvement of gene annotations and open read frame prediction of T. rubrum and other dermatophyte genomes.

Keywords: Trichophyton rubrum, RNA-seq, Acriflavine, Virulence factors

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 61

ANCESTRAL ORIGIN AND VIRULENDE MARKERS OF Helicobacter pylori STRAINS AND HOST GENETIC STRUCTURE AS PREDICTORS OF GASTRIC CANCER

[1]Dulciene M M Queiroz, [1]Charles Anacleto, [1]Cynthia G Trant, [2]Andrea M Pinto, [3]Rafael S Calixto, [1]Kadima N Teixeira, [4]Fernanda S Kehdy, [4]Eduardo T Santos, [5]Roney S Coimbra, [1]Gifone A Rocha, [1]Andreia M Rocha, [6]Aldo A M Lima

[1]Laboratório de pesquisa em Bacteriologia - Departamento de Propedêutica Complementar, Faculdade de Medicina – Universidade Federal de Minas Gerais, Brazil, [2]Laboratório de pesquisa em Bacteriologia - Departamento de Propedêutica Complementar, Faculdade de Medicina – Universidade Federal de Minas Gerais; Departamento de Microbiologia,[3]Instituto de Ciências Biológicas – Universidade Federal de Minas Gerais, [4]Departamento de Biologia Geral, Instituto de Ciências Biológicas – Universidade Federal de Minas Gerais, Brazil, [5]Informática de Biossistemas, Centro de Pesquisas René Rachou – FIOCRUZ, Brazi, [6]Unidade de Pesquisas Clínicas Instituto de Biomedicina, Faculdade de Medicina - Universidade Federal do Ceará, Brazil

Background: We aimed to investigate the phylogeographic origin of H. pylori strains from gastric cancer patients as well as the genetic structure of the patients to determine whether they are predictors of gastric cancer in an admixed population from South-East Brazil. Phylogeographic origin was evaluated in 103 H. pylori strains from patients with gastric cancer (n=27), duodenal ulcer (n=28) and gastritis, control group (n=48), by sequencing of both strands of 397 to 690 bp per gene of the atpA, efp, mutY, trpC, ureI, and yph house-keeping genes. The sequences were aligned (MUSCLE program) and deposited in the multi-locus sequence typing-MLST database. Neighbor joining tree of H. pylori strains (1,201 classified in ancestral haplogroups and our 103 strains) was created by software's Phylip and MEGA 5.1 using the Kimura model with 10,000 bootstraps. To determine the ethnicity of each patient, 106 validated SNPs were evaluated by Sequenon iPLEX Plataform. We estimated individuals' ancestry using parental groups: African, European and Brazilian Amerindians employing the software Structure 2.3.3. Data were analyzed by Fisher, $\chi 2$, Student and correlation tests. Results: H. pylori strains were classified as hpAfrica1 (73-70.9%) and hpEurope (30-29.1%). hpAfrica1 strains were observed in 88.9% (gastric cancer), 85.7% (duodenal ulcer) and 47.9% (gastritis) patients ($p<0.001$). However, when the patients were stratified by bacterium virulence markers, the association disappeared ($p>0.25$). s1i1m1 vacA associated with gastric cancer and s1m1/m2 with duodenal ulcer ($p<0.001$). The percentage of each ancestry was similar in patients with gastritis and gastric cancer ($p>0.46$). European ancestry correlated with corpus gastritis ($r=0.2$, $p=0.05$) and intestinal metaplasia ($r=0.2$, $p=0.04$) in the s1 vacA gastritis group. European ancestry also correlated with European origin of H. pylori strains ($r=0.5$, $p=0.01$). Conclusion: H. pylori virulence markers, more than H. pylori ancestry "per se" and genetic structure of the population, are the most important predictors of gastric cancer in the studied admixed population. Supported by: CAPES, INCT, CNPq, FAPEMIG. Acknowledgments: PDTIS-FIOCRUZ – Platform RPT04B – Bioinformatics.

Keywords: Helicobacter pylori , vacA, Gastric cancer, Duodenal ulcer, Host genetic structure

Topic: Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 62

Scientific text mining aimig at the identification of bioactive compounds with therapeutical potential against Chagas disease, Malaria and Dengue

[1]Milene Pereira Guimarães de Jezuz, [1]Ernesto Raúl Caffarena, [1]Oswaldo Gonçalves Cruz

[1]Programa de Computação Científica, Presidência, Fundação Oswaldo Cruz, Brazil

Keywords: Text Mining; Neglected diseases; scientific workflow

Topic: Transcriptomics and Proteomics

Abstract #: 63

HYBRID DE NOVO ASSEMBLY STRATEGY FOR NO-MODEL PLANT TRANSCRIPTOME

[1]Adriano Viegas, [1]Edith Moreira, [1]Sheila Gordo, [1]Rommel Ramos, [1]Artur Silva, [1]Wilson Silva Júnior, [1]Daniel Pinheiro, [1]Horacio Schneider, [1]Iracilda Sampaio, [1]Sylvain Darnet

[1]Universidade Federal do Pará, Brazil

Background: The next-generation sequencing platforms (NGS) are powerful, efficient and low cost tools to obtain a complete trasncriptome. These vantages allow new sequencing strategies, specially in agrigenomics, with no-model plants with few or without genetic data. However, the complexity of de novo transcriptome assembly, assembly without genome reference, is still one limitation of this approach. In this context, the optimization of de novo transcriptome assembly is very studied and many bioinformatics tools are available for this purpose. This work is about the de novo assembly optimization of different datasets of short reads, from SOLiD and Illumina NGS platforms, from a no-model plant, the black pepper. For this, two approaches were used: one is classical, where the contigs were assembled for each datasets and in a further step assembled in supercontigs; the second one is described as hybrid assembly and in this option, all read datasets are pooled and the assembly is performed integrating all data in colorspace, the SOLiD NGS format. Results: The two bioinformatics tools used for de novo hybrid approach were CLC workbench and the kmer-additive pipeline based on velvet, oases and STM-. The short reads dataset is formed by two SOLiD 50pb runs, totaling 3.6 Gb and 3.4 Gb, and one paired-end 100bp Illumina run, totaling 9,8 Gb. The classical approach was based on separate reads assembly and use of IAssembler for supercontigs achievement, the result was 97,420 contigs, with N50 of 1196 bp and 14,198 unigenes were identified when compared with Arabidopsis genome. The hybrid de novo approach using CLC workbench done 100,373 contigs with N50 of 584 and 15,199 unigenes. The result of the same approach with Velvet kmer-additive, Oases and STM tools in colorspace, is 857,350 contigs with N50 of 402 and 20,615 unigenes. The gain of CLC hybrid approach is 3% for contigs number and 7% for unigenes, comparing to classical approach, and the gain with hybrid velvet approach is 880% and 45% respectively. The N50 is lower in hybrid approach than classical, which should be correlated with higher fraction of small contigs present in hybrid contig dataset. Conclusions: These results are showing that in hybrid methods, where reads from different NGS platforms are assembled together, should exist a "synergism effect" between the different reads and this approach should be more efficient to extract biological information and identified sequences present in a transcriptome of no-model plant.

Keywords: Next-generation sequencing platforms, trasncriptome, hybrid de novo assembly

Topic: Transcriptomics and Proteomics

Abstract #: 65

Metabolomics data normalization with EigenMS

[1]Yuliya Karpievitch, [2]Sonja Nikolic, [3]Lindsay Edwards, [1]Richard Wilson, [2]James Sharman

[1]University of Tasmania, Australia, [2]Menzies Research Institute, Australia, [3]King's College London, United Kingdom

Liquid chromatography mass spectrometry (LC-MS) is has become the analytical platform of choice for cell or bodily fluid metabolome studies. LC-MS metabolomics data are generally very noisy due to the effects of a number of systematic biases. These include batch effects, day-to-day variations in instrument performance and signal intensity loss due to column and inlet clogging. Here we analyze a serum metabolomics dataset and physiological measurements were collected for healthy subjects and persons with Type 2 diabetes. We propose the use of singular value decomposition-based normalization method (EigenMS) for metabolomics data. The method works in several stages. First, it preserves the treatment group differences in the metabolomics data by estimating treatment effects with an ANOVA model. Singular value decomposition of the residuals matrix then determines the bias trends in the data. The number of bias trends is estimated via a permutation test and the effects of the bias trends are eliminated. This is the first application of EigenMS to metabolomics data. We show that a singular value decomposition-based normalization method removes bias of unknown complexity from LC-MS metabolomics data allowing for improved differential analysis. Moreover, normalized samples better correlate with other normalized samples and the physiological data collected for the same biological samples. We thus advocate the use of singular value decomposition-based normalization for metabolomics data.

Keywords: Metabolomics, normalization, mass spectrometry

Topic: Transcriptomics and Proteomics,Signaling and Metabolic Networking; Ontologies; Systems Biology,Human Pathophysiology; Animal Models of Disease

Abstract #: 66

MOLECULAR MECHANISM DRIVING RETROPERITONEAL ADIPOCYTE HYPERTROPHY AND HYPERPLASIA IN RESPONSE TO HIGH-SUGAR DIET

[1]Karina Queiroz, [2]Roney Coimbra, [1]Amanda Rios, [3]Nivea Paiva, [3]Claudia Carneiro, [1]Elísio Evangelista, [1]Renata Sá

[1]Laboratório de Bioquímica e Biologia Molecular, Núcleo de Pesquisas em Ciências Biológicas, Universidade Federal de Ouro Preto, Brazil, [2]Informática de Biossistemas, Centro de Pesquisas René Rachou – FIOCRUZ, Brazil, [3]Laboratório de Imunopatologia, Núcleo de Pesquisas em Ciências Biológicas, Universidade Federal de Ouro Preto, Brazil, Brazil

Keywords: adipogenesis, high sugar diet, PCR array, gene expression

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 67

Metarhizium anisopliae genome sequencing and a comparative secretome analysis

[1]Rafael Lucas Guedes, [1]Oberdan Cunha, [2]Claudia Thompson, [1]Rangel Souza, [1]Luiz Almeida, [2]Augusto Schrank, [1]Ana Tereza Vasconcelos

[1]LNCC, Brazil, [2]UFRGS, Brazil

Background: A large number of fungi species are known to cause disease in several hosts, including mammals, arthropods, and plants. In order to evade from hosts immunity and degrade extracellular components for nutrition, these organisms are required to secrete a plethora of enzymes and toxins. Here we report the sequencing, assemby and genome annotation of an biocontrol agent Metarhizium anisopliae E6 (MAA) along with a comprative secretome analysis with fifteen other fungi species spanning different life-styles, such as entomo, plant, human and mycoparasites and saprophytes. Results: MAA was sequenced using 454-Roche technology and assembled with Newbler Assembler. Gene prediction was carried out with the automated pipeline SABIA, using the combination of AUGUSTUS, Glimmer and GeneMark, yelding 10.814 predicted coding regions, a quantity similar to the observed in taxonomically close organims. A secretome refined prediction was applied to all studied fungi by the combination of several bioinformatic tools (SignalP, TargetP, TMHMM, PredGPI, WoLFPSort, ProtComp and PROSITE) and selected proteins assigned to KEGG Orthology (KO) or PFAM families for functional inference. MAA refined secretome represented 3.8% of complete proteome and similar proportions were found for all species, ranging from 3.1% (Metarhizium acridum and Cordyceps militaris) to 4.8% (Magnaporthe oryzae). The amount of secreted sequences associated to KO groups varied from 33.1% (M. oryzae) to 62.6% (Aspergillus niger), indicating a considerable portion with yet unknown function for all fungi. Interestingly, considering exopeptidases (cleavage of a single amino acid from the amino or carboxylic ends) and endopeptidases (cleavage of a non-terminal amino acid), it is a common characteristic from all secretomes a prevalence of endopeptidases, suggesting that faster mechanisms for inactivating substrate proteins prevails. A homology search using OrthoMCL was performed and a considerable amount of MAA homologs, ranging respectively from 7.9% to 29.6% for Fusarium oxisporum and Trichoderma virens, showed no detectable signal peptide, which can be explained by a real difference between N-terminal portions or alternative start codon predictions. Additionally. it has been shown for Saccharomyces cerevisiae that longer (more complex) duplicated proteins are more likely to be retained due to a higher probability of generating a new biological function. Considering mean proteins size for all sixteen studied fungi species, we also found that duplicated genes are larger than single copy genes for both secreted and not secreted sequences and as more copies, longer the protein. Also, independently of copy quantity, secreted proteins are generally smaller than the average for the rest of the proteome (All pvalues < 0.000 on one-tailed Student t test). Finally, phylogenomics analysis reveals no significant differences between life-styles proteome and secretome phylogenies. Conclusions: An economically important fungi genome is presented with a comparative secretome analysis.

Keywords: Metarhizium anisopliae, secretome

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 68

COMPARATIVE ANALYSIS OF HOMOLOGOUS GENOMIC REGIONS FROM A BRAZILIAN SUGARCANE CULTIVAR (RB867515) AND R570

[1]Isabela Souza, [1]Ivone de Bem, [1]Arthur Melo, [1]Alexandre Coelho

[1]Universidade Federal de Goiás, Brazil

Sugarcane is widely recognized as one of the most important energy crops in the world. Ethanol derived from sugarcane has shown an increasing demand in the Brazilian trade for its utilization as a non-fossil renewable fuel. The international trade of sugar, on the other hand, is supplied by only four countries, led by Brazil. Despite its importance, the development of sugarcane genomic studies has been facing hard problems due to the genetic complexity involved. Besides the high level of polyploidy and common occurrence of aneuploidy, it is known that a significant portion of the sugarcane genome is represented by highly repetitive elements. Despite the amount of challenge, an international effort to sequence the entire sugarcane genome is underway. A commonly raised question is if a reference genome sequence from a specific sugarcane genotype would really be useful for different research groups, since the divergence among different genetic backgrounds seems to be non negligible. In this work we try to contribute to answering this question using high quality Illumina paired end sequences to investigate the level of genomic divergence between a Brazilian genotype (RB867515), the most widely cultivated genotype in Brazil, and an Australian genotype (R570), commonly used in many genetic and genomic studies. NGS reads from RB867515 were aligned against sequences of already published assemblies of eight overlapping R570 BACs using Bowtie2. The positive RB867515 reads were used in a de novo assembly using MaSuRCA. The 16 BAC consensus sequences, eight from R570 and eight from RB867515, were then aligned using Mauve and Mega. The eight assemblies generated comprised a total of 142, 126, 138, 142, 157, 81, 85 and 81 kb, representing 82%, 95%, 92%, 91%, 95%, 97%, 94% and 93% of the R570 BACs, respectively. The percentage of mismatches among homologous sequences from R570 and RB867515 was, for each pair: 0.27%, 0.92%, 1.13%, 1.32%, 1.05%, 0.65%, 0.89% and 1.01%. Despite the high level of conservation of homologous sequences from R570 and RB867515, since no structural rearrangements were observed within these groups, we detected a remarkable divergence among different groups of homology within the same homeology group. The results herein suggest that the homologous genomic regions from R570 and RB867515 are highly similar and that a much higher divergence exists among different groups of homology within the same homeology group than previously anticipated.

Keywords: genomics, sugarcane

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 69

A GEPHI PLUGIN FOR SBGN MODELS

[1]Daniel Capeletti, [2]Marcelo Cezar Pinto, [1]Jose Claudio Biazus

[1]Universidade Comunitária da Região de Chapecó (Unochapecó), Brazil, [2]Universidade Federal Fronteira Sul (UFFS), Brazil

In Systems Biology, which is the study of interactions between biological system components, there is a need to create and share graphical models of these system in a way that it can be useful to other researchers. For this purpose there is the SBGN (Systems Biology Graphical Notation) which aims to standardize a graphical representation of biological system and its components, and with SBGN-ML, that is a XML representation of the system, it is possible to be interpreted by a computer. According to this context, the visualization and exploration of networks and complex system tool, Gephi, brings us a platform that allows the representation of a system , manipulation of structures and the network data analysis as well. This is a free tool and its use is widespread in the area of bioinformatics. Our study is aimed at using Gephi interface via a plugin, to read a SBGN-ML document . Once the document was read and validated in the right way, we would use the tools provided by Gephi to generate a graphical model of the system, within the standards of SBGN. The plugin that we developed in this work is able to read SBGN-ML representations considering all its components (e.g. pertubing agent, state variables, delay etc.) and validate if the document is written correctly, in other words, if it is according to SBGN standards and respecting the rules implemented by XML. The next step would be then, generate its graphical model, however, to generate such model it is necessary to change the source code of Gephi. This change is not possible through a plugin, making it difficult to create a complete editor for SBGN based on Gephi. With the available tools that already exist, it is possible to create graphics that represent the SBGN, like Biographer, however, it is not possible to read a SBGN-ML to create its graphical representation. Adapt Gephi to allow graphical editing of SBGN is possible by changing its source code, but we must be very careful because we can compromise the other features of the tool. This work also provides the knowledge of how to analyze and validate SBGN-ML documents, this way, it is possible to choose a tool that already represents SBGN models and apply such knowledge to generate a complete editing SBGN model tool.

Keywords: SBGN, Gephi, SBGNML, Systems Biology

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 70

EVOLUTIONARY AND FUNCTIONAL GENOMICS OF PSEUDOGENES IN TRYPANOSOMATIDS

[1]Márcio Albuquerque Silva, [2]Fernando Alvarez-Valín, [1]Ana Carolina Ramos Guimarães

[1]Fiocruz, Brazil, [2]Sección Biomatemática, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

Abstract: Trypanosomatids are protozoan parasites belonging to the order Kinetoplastida, an ancient group in the phylogenetic tree of eukaryotes. The sequencing of the genomes of some pathogenic trypanosomatids belonging to the genera Leishmania and Trypanosoma, has undoubtedly contributed to the understanding of the biology of these organisms as well as of relevant aspects of the evolution of their genomes. The current availability of several complete genomes of trypanosomatids, at various degrees of divergence, allows for robust comparative and evolutionary analyses, offering new opportunities to better understand important biological processes, or even reveal other, yet unknown, biological processes in these organisms. In this work, we aim to perform a comparative and evolutionary analysis of the processes and mechanisms related to the evolution of pseudogenes in Trypanosoma and Leishmania species. We will try to identify genes or groups of genes which are prone to defunctionalization, and therefore to pseudogenization, as well as genes or groups of genes which have not been affected by these processes, and therefore are indispensable in these organisms. Such study will allow us to acquire a qualitative and quantitative overview of the processes of acquisition and loss of genes over time in these parasites. Another important aspect to be explored in this work is the possible acquisition of new functions by pseudogenes. The traditional view is that once function is lost, the fate of these sequences is progressive degradation. However, there are considerable evidences suggesting that pseudogenes may acquire new functions, particularly when related to gene expression regulation of other members of the same multigenice family. Thus, using deep sequencing data (Illumina), we will determine the expression patterns of pseudogenes in two trypanosomatids, T. cruzi and T. vivax, in order to get clues about their possible new functions. Hence, in this work, we aim to contribute to a better understanding of the dynamics of genomes of trypanosomatids by (i) obtaining a qualitative and quantitative overview of the processes of acquisition and loss of genes in these organisms over time, tracing the history of each particular gene family among the lineages studied, as well as (ii) by analyzing the evolutionary dynamics of pseudogenes, i.e., the processes of pseudogenization and neofuncionalization in these parasites.

Keywords: PSEUDOGENES,TRYPANOSOMATIDS,EVOLUTIONARY AND FUNCTIONAL GENOMICS

Topic: Transcriptomics and Proteomics

Abstract #: 71

Transcriptome analysis of interaction between rice and Magnaporthe oryzae by RNA-seq

[1]Rosangela Bevitori, [1]Marilia Vilela, [1]Marcelo Narciso

[1]Embrapa Arroz e Feijão, Brazil

One of the obstacles to expansion of cultivated area of rice that is associated with high productivity and yield stability is the susceptibility of cultivars available on the market by fungus Magnaporthe oryzae (Hebert). The causal agent of rice Brusone in rice. Few studies have been concerned with investigating genes by transcriptome analysis in interaction between M.oryzae X rice. One reason is that the molecular basis of this interaction are not yet fully established due to mutation of the pathogen. The aim of this study was to profile expression by transcriptome analysis of regions of the genome of the fungus that may be involved in the interaction between pathogen and host.

Keywords: Rice, M.oryzae fungus, RNAseq

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 72

Characterization of Endogenous Retroviruses in Primates

[1]Andrei Rozanski, [1]Fábio C. P. Navarro, [1]Pedro A. F. Galante

[1]Bioniformatics Group - MOCHSL, Brazil

Primate genomes are largely populated by Transposable Elements (TE). Among these TE, Endogenous RetroViruses (ERVs) and other LTR elements account for ~8% of the human genome sequence. Whether, in one hand these viral sequences may be associated to several diseases such as cancer, schizophrenia and immune disorders, on the other hand they have also played critical role in the evolution of several organisms, such as cooperating to the development of mammalian placenta. Therefore, a large-scale and comprehensive approach focusing on evolutionary and genomic aspects of ERVs may contribute to a better understanding of their roles in the human and other genomes. In this study, we have performed a set of systematic analyses of ERVs in six primate genomes (Human - Hs, chimpanzee - Pt, gorilla - Gg, orangutan - Pa, rhesus – Rh, and marmoset - Cj). First, we identified all ERVs in each organism. Next, we selected those fixed ERVs and, then, we explored better some of their features in the human and chimpanzee genomes. For human and chimpanzee, we found ~10k ERVs, being ~400 (on average) still intact, and potentially actives. We also found an underfixation of ERVs into coding gene region and some ERVs integrated next to promoter regions of coding genes. We also identified a set of conserved ERVs among the primates and a set of 14 human specific elements. As may be expected due to their recent activity, only ERV-K accounts for ~43% of these human specific events. One of these human specific events were identified as polymorphic. In summary: we believed that using this plethora of data and these systematic large-scale analyses, we have gained insights regarding the ERVs' roles and contributions into the human and other primate genomes evolution. Supported by FAPESP.

Keywords: Endogenous Retroviruses, Primates, ERVK, Genomics, Transposable Elements

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 73

Human miRNA-gene network reveals a Small-World topology pattern

[1]Ana Beatriz Soares Brasil Sampaio Costa, [1]Aline Ferreira da Silva, [1]Bianca Minetto Napoleão, [2]Francis M. F. Nunes, [2]Fabrício Martins Lopes, [1]Alexandre Rossi Paschoal

[1]Federal University of Technology - Paraná – UTFPR – Cornélio Procópio, Brazil, Brazil, [2]Departamento de Genética e Evolução, Universidade Federal de São Carlos – UFSCar – São Carlos, Brazil, Brazil

MicroRNAs (miRNAs) are a class of small non-protein-coding RNA (ncRNA) that negatively regulate the mRNA and protein levels of living cells. Despite miRNAs are the most studied ncRNAs, little is known about their network topologies based on experimental or manually curated data. Here we hypothesized the existence of at least one topology for miRNA-gene interactions in humans. Toward this end, first we recovered data from Cancer miRNA Regulatory Network (http://cmrn.systemsbiology.net). By using R and PERL scripts we applied complex network theory in order to extract complex network measures from data (e.g. Betweenness centrality, degree distribution, among others measures) and compare the resulting network (biological one) with other known complex networks models (e.g. Small-World, Scale-free, Random). These measures are calculated to distinguish structured network. After that, we linked these miRNA-gene results to other available information found in miRNA databases (miRBase, KEGG, miR2Disease, miRCancer, Pharmaco-miR) to improve the functional annotation. A web site system was built to help users with different scientific purposes. Based on the mathematical modeling, our results revealed that the human miRNA-gene network can be better approximated by the Small-World topology. The most connected miRNAs were hsa-miR-124, hsa-miR-192, hsa-miR-34a, hsa-miR-21 and hsa-let-7b, while CEP55, TP53INP1, CDK6, HMGA2 and SHCBP1 were the hub genes. Together, these key genetic elements are strong candidates to maintain general cellular functions and when dysregulated they are involved in the development of cancers. Supported by: Fundação Araucária, Projects 23.466 - 339/2012 (from "Programa Universal – Pesquisa Básica e Aplicada – Chamada Projetos 05/2011") and PIBIC-UTFPR.

Keywords: Systems biology, Topology, Network, Bioinformatics, miRNA, Non-coding

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 74

DNA Symphony: A new method to represent genomic sequences

[1]Rosario Medina Rodriguez, [2]Harieth Bernedo Cordova, [3]Jesus Mena-Chalco

[1]University of Sao Paulo, Brazil, [2]San Pablo Catholic University, Peru, [3]Federal University of ABC, Brazil

Representation of DNA sequences is a simple and powerful tool to identify and analyze features on complex genomic sequences. For instance, complete genomic sequences from biological species can be graphically represented by a ``genomic signature''. This representation provides information about the oligonucleotide frequencies considering different size of K-mers. Moreover, genomic sequences can also be represented by an audio signal, which is obtained by translating each oligonucleotide or protein into a predefined range of audio frequencies. This type of conversion will help to reveal some auditory patterns which could be useful for future (artistic) research. Although audio representation strategies provide an interesting result, they only use part of the genomic sequence. To date no method exists which contemplates the complete genome sequence. This work proposes a new method for audio representation of genomes by composing a polyphonic signal using a set of complete genomic sequences. The main goal of our work is to combine/integrate music and biology to generate musical compositions from genomic datasets preserving the original genome's structure and organization. Our approach, first extracts the genomic signature of each sequence. Then, in order to obtain the audio signal, we transform two-dimensional genomic signatures into an one-dimensional sequence by normalizing each value into an audible spectrum. Finally each signal, depending on the number of pre-defined sequences, is played on a different channel to generate a polyphonic track. In this way, our approach focuses more on DNA sequence organization which could be beneficial for analysis and pattern detection. Experimental results showed that this approach can be applied to a wide variety of complex biological datasets. Therefore, we can generate polyphonic musical compositions to be presented to a wider audience (interdisciplinary groups) not only composed by scientists. To demonstrate the feasibility of this approach, given a set of genomic sequences two audio signals will be compared. One obtained from a multiple alignment using BLAST and another one extracted from a consensus genomic signature. The stronger the correlation between them, the better the representation of the set of sequences are. As future directions, we plan to optimize the algorithm to transform the genomic signatures, and extend the method to perform pattern analysis. Collected experimental data are available at www.vision.ime.usp.br/~rmedinar/DNASymphony.

Keywords: chaos game representation of frequencies, genomic signature, dna audio representation, dna symphony.

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 76

NEW SINGLE NUCLEOTIDE POLYMORPHISMS IN GUZERÁ BREED REVEALED BY WHOLE-GENOME RE-SEQUENCING

[1]Izinara Rosse Cruz, [2]Juliana Assis Geraldo, [2]Francislon Silva de Oliveira, [2]Laura Rabelo Leite, [3]Flávio Araújo, [4]Adhemar Zerlotini, [5]Beatriz C. Lopes, [6]Wagner A. Arbex, [6]Marco Antônio Machado, [6]Maria Gabriela Campolina Diniz Peixoto, [6]Rui da Silva Verneque, [6]Marta Fonseca Martins, [6]Marcos Vinícius G.B. Silva, [7]Roney Santos Coimbra, [1]Maria Raquel Santos Carvalho, [8]Guilherme Oliveira

[1]Universidade Federal de Minas Gerais-UFMG, Brazil, [2]Universidade Federal de Minas Gerais-UFMG/Genomics and Computational Biology Group - FIOCRUZ-MG/Center for Excellence in Bioinformatics – FIOCRUZ-MG, Brazil, [3]Genomics and Computational Biology Group - FIOCRUZ-MG, Brazil, [4]EMBRAPA Informática Agropecuária - SP, Brazil, [5]EPAMIG – MG, Brazil, [6]EMBRAPA Dairy Cattle – MG, Brazil, [7]Biosystems Informatics, Research Center René Rachou - FIOCRUZ, Brazil, [8]Genomics and Computational Biology Group - FIOCRUZ-MG/Center for Excellence in Bioinformatics – FIOCRUZ-MG, Brazil

Background: Currently, livestock accounts for 9% of Brazil's GDP and the traditional breeding has provided a gain in production of 1% per year. With the proposal to increase the efficiency of breeding programs, genotyping chips containing thousands of SNPs as putative markers for traits of interest were developed. These chips contain variations observed in various commercial breeds, but do not include the Guzerá that contributes significantly to milk production and is well adapted to local climate conditions. In this context, the objective of our work was to sequence and assemble the Guzerá genome in order to identify race-specific variations that might be used in breeding programs. Results:Our group performed the genome assembly of Guzerá breed that was performed using mate-paired libraries, with 1-2 and 3-4 kb inserts, by SOLiD v3 e v4 second-generation sequencing. The sequences were mapped to the publicly available reference genome of Bos taurus (UMD 3.1) using the LifeScope software. The average depth of coverage achieved from mapping was 26X (ranging from 14.5 to 52.3) for each chromosome and 87% of the reference genome was covered. A list of putative SNPs was generated from the mapped reads, using the diBayes SNP Detection module included with LifeScope. These SNPs were filtered according to the following criteria: (1) SNPs with an overall quality less than 20 were removed; (2) minimum rate of 20% for the new allele; (3) variants with too low or too high read depths were removed: the minimum of 4 reads of depth and the maximum as the mean read depth + 3 times the standard deviation; (4) variants within 5 bp of each other were removed. 4.040.476 SNPs (54%) remained after the filter and they were compared with dbSNP138. It was observed that 51.5% of them were new. Conclusions:A significant number of distinct SNPs were observed between the Guzerá and Hereford breeds. These differences may be associated with the production traits and adaptability that characterize the zebu breeds and may be useful in breeding programs. Supported by: CAPES, CNPq (307975/2010-0, 309312/2012-4), FAPEMIG (CBB-1181/0, TCT 12.093/10, REDE-56/11), NIH-USA (TW007012), CAPES/CDTS-FIOCRUZ. Acknowledgments: PDTIS-FIOCRUZ - Platforms RPT04B (Bioinformatics) and RPT01F (NGS).

Keywords: SNPs; animal breeding; next-generation sequencing

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 77

Identifying variations in nucleic acid sequences using in silico single strand melting curve

[1]Raffael Oliveira, [1]Ricardo Almeida, [1]Márcia Dantas, [1]Daniel Lanza, [1]João Paulo Lima

[1]UFRN, Brazil

Background: Since the creation of the PCR technique, many variations of this technique have been successfully implemented in research, and in the clinical laboratory always have significant advances that help to give better results, with more confidence. Currently it is possible to perform large-scale PCR with minimum cost and with high accuracy in the results. An interesting approach that has been exploited to differentiate DNA sequences is the melting curve analysis. In this analysis, the profile generated by the release of an intercalating fluorophore during the denaturation of double-stranded DNA (dsDNA), reproduces, at low resolution, the nucleotide sequence of a double strand. To date, there are no methods for identifying nucleic acids based on information obtained by the denaturation of a single strand of DNA (ssDNA). Most studies that involve single strand nucleic acid secondary structure are focused on solving particular RNA functions, characterize noncoding RNA species or new classes of siRNAs. To validate the hypothesis that the denaturation of ssDNA can generate more information about the nucleotide variations, than denaturation of a dsDNA, softwares that reproduce the profile of denaturation were used to denaturate nucleic acid sequences of members of the viral family Totiviridae, which include virus with medical and zootechnical importance, that infect shrimps and protozoa. Results: Initially a phylogenetic tree was made using the aminoacid sequences of RdRp of Totiviridae family members. Within this tree, one group that represents Trichomonasvirus was selected, and their respective nucleic acid sequences were submitted to melting curve analysis using two open-acess softwares: RNAheat and MELTSIM, which consider the analyte being an ssDNA and dsDNA, respectively. Two different regions of the genomes were used for melting curve analysis. One of these regions corresponds to a 200pb region that shows a conserved RNA secondary structure between all sequences of Trichomonasvirus. The other region does not have conserved secondary structure of RNA and encodes the viral capsid protein. After grouping the curves, clearly is noticed a difference between members when we see the ones generated by RNAheat; when analyzed the ones provided by MELTSIM, the difference is not so clear, meaning that we cannot spot the difference between strains of viruses with this software, only using RNAheat. Conclusion: The results show that we can easily identify point differences in nucleic acid sequences comparing in silico melting curves and that ssDNA provides a more detailed profile than dsDNA. After confirming these results using in vitro assays, this new approach will be applied to develop new strategies to detect nucleic acid polymorphisms by PCR.

Keywords: melting curve, trichomonasvirus, totiviridae, secondary structure, rnaheat, meltsim

Topic: Transcriptomics and Proteomics,Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 78

GENE CO-EXPRESSION AND BIG DATA APPROACH TO EVALUATE THE MOLECULAR CHARACTERIZATION OF PE-2 YEAST ADAPTED TO HIGH ETHANOL CONDITIONS IN INDUSTRIAL FERMENTERS

[1]Marcelo Brandão, [1]Lucas Lopes, [1]Marcio Silva-Filho

[1]Escola Superior de Agricultura Luiz de Queiroz, USP, Piracicaba, SP, Brazil

Background: Biology is becoming increasingly data-intensive as high-throughput genomic and transcriptomic assays become more accessible and more feasible for in-lab daily analyses. Data-centric approaches that compute on massive amounts of data (often called "Big Data") to discover patterns and to make relevant predictions is gaining adoption. The bioinformatics scientists has seen on the Big Data a new way to analyze and understand the complex patterns involved on the organisms system Biology. In the current world renewable energy scenario, bioethanol is emerging as one of the better options in the short and medium term alternatives to non-renewable resources such as fossil fuels. The Brazilian bioethanol production mainly depends on sugar cane fermentation, and the yeasts responsible for this job are often recycled throughout the entire sugar cane harvesting season and fermentation period. So, the availability of yeast that can handle multiple cycles of fermentation, high ethanol concentration and elevated heat is important to the biofuel industry. Results: Here, we use microarrays to transcriptionally profile the industrial yeast line PE-2, whilst being subjected to high levels of ethanol in industrial fermentation. Gene expression levels were compared at 0/6h, 0/12h, and 0/18h along with 6/12h, 6/18h, and 12/18h. The results show a large array of genes exhibit significantly altered gene expression patterns during fermentation. We also measured important biochemical metrics for the fermentation such as ethanol, glycogen and trehalose percentage and the weigth of glycerol, glucose, frutose and sacarose. Using the weighted correlation network analysis (WGCNA) we were able to describe the correlation patterns among genes across fermentation periods. The proposed clusters were annotated according to its components ontologies, biochemical pathways and the biochemical metrics in each sample period. With all this information we were able to generate a gene correlation network which clearly demonstrates 3 major hubs, been one of them inwardly related to the glucose metabolic pathway. Checking the microarray expression patterns we identify the expression of genes involved in this metabolic pathway undergo significant changes that differ from other yeast strains during fermentation and could contribute to PE-2's adaption to high ethanol levels during industrial fermentation. Conclusions This approach has the long term aim to engineer new yeast strains and improved operational procedures for increased ethanol yields.

Keywords: WGCNA, Big data, Saccharomyces cerevisiae, ethanol stress, industrial fermentation

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 79

MOLECULAR ASPECTS OF DOCKING INTEGRINS INHIBITORS

[1]Heitor Modenesi Fraga, [1]Jorge Hernandez Fernandez

[1]Universidade Estadual do Norte Fluminense Darcy Ribeiro, Brazil

Background: Integrins are important molecules in cellular adhesion and migration and important information of structural descriptors in interaction surface is mandatory for development of new therapeutic agents. Integrin interaction surface is centered by Mg2+ bivalent cations, described as the central MIDAS (Metal Ion-Dependent Adhesion Site) flanked by ADMIDAS (Adjacent MIDAS) and LIMBS (Ligand-Induced Metal Binding Site). Integrins alpha6beta1 and alphaVbeta3 were described as laminin and colagen receptors respectively, and therefor different in ligand recognition surface. In this study structural models of these integrins were used in docking experiments on AutoDock 4.2. Structure protonation, working surface and ligand parametrization were defined in ADT 1.5.6. As ligands, small peptide based structures of desintegrin interaction loop from SVMPs (Snake Venom Metalloproteinase) were used. Results were analyzed by ΔG, inhibition constant (Ki) and interaction shape, considering as positive results only the structures interacting with integrin MIDAS site. Results: Obtained results showed a better interaction of used peptide inhibitors with alpha6beta1 ranging on nM values for Ki, and showing μM values when docked with alphaVbeta3 integrin. A9a inhibitor showed better interaction pattern, contacting both alpha6 and beta1 integrin chains. Hydrogen bonds and electrostatic interactions, mostly with MIDAS and ADMIDAS sites were observed in obtained A9a-alpha6beta1 interaction, making this inhibitor the better structure for further development of specific and selective alpha6beta1 integrin antagonist. Conclusion: Our docking experiments showed that small peptide structures based in desintegrin-like interaction loops are good lead structures for specific and selective integrin ligands. The development of these structural studies may result in new drugs targeting inflammation and metastasis pathologies.
Suported by: CNPq, FAPERJ.

Keywords: Docking, integrins, inhibitors

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 80

Theoretical Study of Monoterpenes by Dynamic Molecular and Free Energy in the Enzyme 3-hydroxy-3-methyl-glutaryl-CoA Reductase

[1]carlyle lima, [2]Silvana de Oliveira Silva, [2]Nelson alberto nascimento de alencar

[1]ufpa, Brazil, [2]Universidade Federal do Pará, Brazil

The enzyme 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) reductase is distributed in organs and organic metabolic processes, it's main function is the production of cholesterol by the organism. Now days the investigation of new drug is in focus to the nature, mainly the monoterpenes, they present a high affinity with this enzyme. The research was based on the action of four monoterpenes described in the literature for having a strong action against the enzyme target of the study: D-Limonene, Perillic Alcohol, Perillic Aldeid and Perillic Acid. Here, we using molecular dynamics (MD) and free energy to investigate the interactions most important between HMG-CoA reductase (1DQA, PDB code) and four monoterpenes. The MD was performed by a time of 3 ns with hybrid QM/MM method, where Co-enzyme and inhibitor were treated with the semiempirical AM1 Hamiltonian, while the rest of the system was described using the combination of the OPLS-AA and TIP3P. Because the active site exhibit preferential binding positions at both the ends and the monoterpenes present favorable binding regions at one end only, in two different positions: S (bond on the end of active site) and C (region to link to co-enzyme), trying to understand what were the amino acids more favorable to the bond. The best inhibitors were detached and subjected to calculations of MD after simulation showed greater stability of monoterpenes in position S with the Glu559 O residue deposited near Co-enzyme, performing favorable reception of electrons by the tested oil with an average distance 2.5 Å distance and Lys735 that is favorable electronic exchange, with a value of the distance of 1.97 Å, this stability is achieved similarly to the inhibitor complexed in crystallography of the enzyme. The free energy values are consistent with those of molecular dynamics to the way in which the most promising complex, proves more stable suggesting that monoterpene containing the aldeid functional group, holds the most promise as an anti-HMG. The results indicate a good behavior of these structures active site.

Keywords: HMG-CoA reductase, Molecular Dinamyc and Free Energy

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 81

APPROXIMATION ALGORITHMS AND HEURISTICS FOR THE PROBLEM OF GENOME REARRANGEMENT BY REVERSALS

[1]Eduardo Moreira Fernandes, [1]Elói Araújo

[1]Universidade Federal do Mato Grosso do Sul, Brazil

We are studying the genome rearrangement problem using reversal operations. We are interested in its theoretical aspects and finding good heuristics to solve it. This work includes comparison between a known heuristic and each one of three new algorithms proposed by us.

Keywords: genome rearrangement, reversal, algorithms, heuristics

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology,Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 82

A new method for microRNAs compound functional similarity based on ontologies

[1]Mariana Yuri Sasazaki, [1]Joaquim Cezar Felipe

[1]University of São Paulo, Brazil

Background: MicroRNAs play critical roles in many important biolog-ical processes. There are several microRNA-related database systems, besides the Ontology for MicroRNAs Targets (OMIT). These data help investigations about microRNA functions in gene regulation. In this work we will create a framework and implement a method to calculate microRNAs functional similarity, based on all information annotated by OMIT, such as: if a microRNA is oncogenic or tumor suppressor, the organism it belongs, the experiment in which it was found, its asso-ciations with diseases, target genes, proteins and pathological events. In addition to being applied on OMIT, our method is supported by other ontologies, such as MeSH and Gene Ontology (GO).
Methodology: Consider that we want to estimate the functional simi-larity of two microRNAs M1 and M2. First, we use an algorithm to search for all annotations about microRNAs M1 and M2 at OMIT. Then, Wang method will be applied to each information category about both M1 and M2, resulting in specific similarities based on their respective ontologies. For example, if there are annotations about the target gene category for both M1 and M2, the specific similarity for this category will be calculated based on GO. However, we must treat redundancy cases verifying if there is no direct association between target gene and disease for each microRNA. After calculating similarities for the specific categories, we create a vector in which each element represents the specific similarity for a category. Finally, we apply a weighted distance function on M1 and M2 vectors to get the compound functional similarity value. The validation of the proposed method will be con-ducted by using a set of similarities predefined by specialists, based on the annotation of each microRNA pair that is being compared.
Results: After a bibliographic survey, we selected some microRNA repositories (miRBase, HMDD, TarBase) and pre-processed their data so that they could be integrated into a single relational database based on OMIT structure. From Genetic Association Database (GAD) we could verify the existence of redundancies. GO and MeSH ontologies were obtained from their respective websites. We also used a species taxonomy from miRBase website to create an Organism Ontology. All of these ontologies were integrated to our database. A web framework was created to search, include, update and delete microRNAs annotations on our database. Next step is the implementation of Wang method and distance functions to calculate the compound functional similarity between two microRNAs on our framework.
Conclusion: Since there is no method using all microRNA information at once to calculate microRNA similarity, this work aims helping bio-medicine researchers to better understand the important roles of mi-croRNAs, their functions and associations to human diseases.

Keywords: MicroRNA, Functional Similarity, Gene Ontology, MeSH

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 83

METHOD, SOFTWARE AND DATABASE FOR MOLECULAR SEROTYPING OF Streptococcus pneumoniae: A TOOL SET TO ASSESS THE EFFECTIVENESS OF VACCINATION PROGRAM IN BRAZIL

[1]Dhian Renato Almeida Camargo, [2]Fabiano Sviatopolk Mirsky Pais, [1]Barbara Rezende Pereira da Mata, [3]Michelle Lara Samuel, [4]Ângela Cristina Volpini, [4]Guilherme Correa de Oliveira, [3]Marluce Aparecida Assunção Oliveira,[1] Roney S Coimbra

[1]Biosystems Informatics, Research Center René Rachou - FIOCRUZ, Brazil, [2]Center for Excellence in Bioinformatics, Research Center René Rachou - FIOCRUZ, Brazil, [3]Service for Bacterial and Fungal diseases - FUNED-MG, Brazil, [4]Genomics and Computational Biology Group, Research Center René Rachou - FIOCRUZ, Brazil

Background: Ninety-one Streptococcus pneumoniae serotypes have been described, but the conjugated anti-pneumococcal vaccine distributed by the Brazilian public health system covers the ten most prevalent in this country. Pneumococcal serotype shifting after massive immunization is a major concern. Monitoring this phenomenon requires efficient serotyping schemes. Classical pneumococcal serotyping is expensive, centralized in a few reference centers, and error prone due to cross-reactivity between similar capsular antigenic polysaccharides (CPS). We present herein a new tool set for molecular serotyping of S. pneumoniae that includes: 1) a molecular biology method which produces serotype-specific fingerprints; 2) a database to store the reference fingerprints: and 3) a software to predict the serotype of clinical samples by comparing their fingerprints with those in the database. The molecular method is based on the restriction fragment length polymorphisms of the PCR-amplified cps loci (cps-RFLP) which encodes the enzymes responsible for CPS synthesis. Results: We predicted the cps-RFLP patterns of 90 serotypes with previously sequenced cps loci by in silico digesting these genomic regions with the all known endonucleases, including XhoII, and HinfI. The latter had already been proposed for molecular serotyping of a subset of pneumococcal serotypes. Using our previously published web service Molecular Serotyping Tool (MST–www.cebio.org/mst), the restriction patterns were aligned and their distance calculated as the sum of the penalties for the edit operations that transform one pattern into the other. Taking into account the median of the distances between all pairs of cps-RFLP patterns and the number of indistinguishable pairs for each enzyme, XhoII (mean= 36.06 ± 17.93; indistinguishable pairs= 7) was more discriminant than HinfI (mean=10.56 ± 6.65; indistinguishable pairs= 34). Means were significantly different ($p < 10\text{-}15$) by the Wilcoxon rank-sum test. The rare cps-RFLP patterns pairs that MST could not distinguish when using XhoII correspond to some of the serotypes with reported cross-reactivity to antisera used in classical serotyping (9L/9N, 12B/12F, 15B/15C, 18B/18C, 22A/22F, 32A/32F, 33A/33F). Published oligonucleotide sequences complementary to dexB and the aliA, conserved genes flanking the cps loci, were used to amplify these regions of 46 epidemiologically unrelated clinical isolates previously serotyped. They represent the 31 serotypes prevalent in Minas Gerais, Brazil. A unique amplified fragment was observed for each serotype, with sizes ranging from 17 to 25 kbp. Clearly identifiable and reproducible cps-RFLP patterns comprising four to 17 bands were obtained after XhoII digestion of the amplicons and fragments separation by electrophoresis in agarose gels. Experimental and predicted cps-RFLP patterns were loaded to the database linked to MST. Conclusions: Experimental and predicted cps-RFLP patterns are strongly correlated. This tool set will be made freely available allowing real time epidemiological surveillance of serotype shifting, a pre-requisite to continuous improvement of pneumococci vaccines. □

Keywords: Streptococcus pneumoniae, Molecular serotyping, cps-RFLP, Web service

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 84

GENOME-WIDE DYNAMIC ANALYSIS: PROMISES AND PITFALLS OF ENZYME KINETICS

[1]Luismar Porto, Julia Castro

[1]Federal University of Santa Catarina, Brazil

Background: Kinetic modeling and parameter determination involving metabolism and cell physiology in general is one the most important bottlenecks of genome-wide dynamic network analysis. It has been postulated a decade ago that the cost of having a complete Escherichia coli quantitative kinetic model would be around ten times the cost of the Human Genome Project, an unsayable amount of money. Even with constantly decreasing analytical costs in terms of quantitative biochemical and biological assays, the day of mechanistic laboratory kinetic data for full scale, genome-wide analysis is yet to come in an unforeseeable future. On the other hand, kinetic equations (and parameters) are essential for the successful predictions of, e.g., metabolic control analysis (MCA), a metabolic engineering tool capable, in principle, of point out key enzymatic control points in metabolic reconstruction network models. Therefore, alternative approaches to mechanistic models and direct experimental parameter determinations are necessary and may prove to be invaluable to study cell dynamics and patho/physiological states, besides been a powerful tool to help develop synthetic biology machines. Results: Gene expression levels have been used as a measurement of total enzyme concentrations applied to the Michaelis-Menten kinetic equation in order to estimate the enzyme-substrate interaction, the Michaelis constant, proposed one hundred years ago. We took Michaelis-Menten kinetics as a reference model and analyzed other alternatives, and how they would improve our current capability of estimating kinetic and thermodynamics constants from metabolic engineering tools and current high-throughput gene expression technologies. Conclusions: State of the art kinetic parameter evaluation is central to modern cell physiology studies. Our approach, although restricted, may be useful in describing metabolic dynamic correlations and should strengthen the ability of biochemical kinetics to elucidate cell function for different physiological cell states.

Keywords: network dynamic analysis, metabolic engineering, enzyme kinetics

Topic:

Abstract #: 85

DRAFT RECONSTRUCTION OF THE CORE METABOLIC NETWORK IN Gluconacetobacter hansenii ATCC 23769

[1]SAMARA SILVA DE SOUZA, [1]LUISMAR MARQUES PORTO

[1]Federal University of Santa Catarina - UFSC, Brazil

Background In the last decade, reconstruction and applications of genome-scale metabolic models have truly influenced the field of systems biology, based on a combination of genome sequence information and detailed biochemical information, providing a platform on which high throughput computational analysis of metabolic network can be accomplished. The Gram-negative bacterium Gluconacetobacter has been extensively used for cellulose synthesis. Recently Gluconacetobacter hansenii ATCC 23769 genome sequence has become available (GenBank number: CM000920 and taxonomy ID: 714995), allowing studies of this organism and its production of bacterial cellulose, which has been used in the medical field as wound dressing, artificial skin material and other applications. For a global understanding of the capabilities of G. hansenii it is essential the construction of a metabolic model that allows the integration of typical experimental data along with genomic and high-throughput post-genome data. The constraint-based reconstruction and analysis (COBRA) approach was used to analyze and build in silico metabolic network reconstructions. Physiological and chemical constraints, such as uptake rates and reversibility of reactions, represent the constraints of the model. Flux balance analysis (FBA) was used to determine the steady-state flux distribution of a constraint-based network by maximizing an objective function.

Keywords: Gluconacetobacter hansenii ATCC 23769, metabolic network, genome-scale, flux balance analysis, constraint-based

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 86

BIOINFORMATIC TOOL FOR GENETIC MARKERS SELECTION FOR ANCESTRY INFERENCE

[1]Sérgio Paiva, [1]Marcus Silva, [1]Valdir Balbino, [2]Antonio Rezende, [1]Ronald Moura, [1]Sérgio Crovella, [1]Lucas Brandão, [1]Antonio Coelho

[1]UFPE, Brazil, [2]Oswaldo Cruz Foundation - Research Center Aggeu Magalhães, Brazil

Admixture mapping is a powerful gene mapping approach for an admixed population formed from ancestral populations with different genetic background. The ancestry profile of an individual is known to impacts on outcomes of association studies that are designed to identify risk for common diseases in human populations. The use of Ancestry Informative Markers (AIMs) is a more accurate method for determining genetic ancestry for the purposes of population stratification. The main aim of this study is to develop an efficient web-based tool specifically designed to retrieve AIMs from the 1000 Genomes project. The tool includes an automated and simple scripting at the click of a button functionality that enables researchers to choose the most appropriate markers using various ancestry informative measures: the Absolute Allele Frequency Differences (delta, δ), Ancestry Informative Content (f value), and Informativeness for Assignment Measure (In). The allele frequencies of the markers in the populations, which are stored in VCF format files available on the 1000 Genome Project, were treated and separated using the VCF python library in order to calculate the measures of ancestry. A total of 8.417.669 Single Nucleotide Polymorphisms (SNPs) and their frequencies in European (EUR), Amerindian (AMR), Asian (ASN) and African (AFR) “super populations” were stored in a relational database using the MySQL as Data Base Manager System (DBMS). For data accession, calculation of the ancestry measures and user interface, we used the PHP language. Access to data can be made by apache server, and it is available for academic community (www.bioinfo.ufpe.br/snps/index.php). Our results can aid in decisions about the type, quantity, and specific choice of markers to be used in studies of ancestry. In summary, we conclude that this tool has been designed for helping users with limited bioinformatics skills to take advantage of publicly available genomic data to extract and identify AIMs among ancestral populations by using one or more measures of marker informativeness criteria.

Keywords: Ancestry Informative Markers

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 87

NMR-BASED METABOLOMIC ANALYSIS OF THE SECRETOME OF HEAD AND NECK CARCINOMA CELLS TREATED WITH 5-FLUOROURACIL

[1]Larissa Menezes dos Reis, [2]Bianca Rodrigues da Cunha, [2]Giovana Mussi Polachini, [2]Ana Carolina Buzzo Stefanini, [2]Eloiza Helena Tajara da Silva

[1]UNESP, Brazil, [2]FAMERP, Brazil

The 5-Fluorouracil (5-FU) is an antimetabolite, analogue of uracil, widely used in the treatment of various cancers, including head and neck. The cytotoxicity of this drug is mainly attributed to participation of its metabolites to blocking synthesis of protein and, although much is known about this mechanism of action, there are few data about their effects on metabolism and secretome of the neoplastic cell. In the present study, we analyzed the secretome of head and neck carcinoma cells using metabolomic approaches to study metabolic changes in treated cells with 5-FU. Futhermore, we choose a gene involved in energetic metabolism for quantification using Real-Time PCR. The Hep-2 cell line (originally described as derived from a human larynx carcinoma) was incubated on MEM culture media with 1.5 mM 5-fluorouracil (5-FU) for 48 hours. Thereafter, the culture media was aspirated, replaced by fresh serum-free culture media and incubated for 24 hours. The conditioned media from treated and control cells (FUCM and CCM, respectively) were collected, filtered through 0.2nm filters and thus analyzed by nuclear magnetic resonance (NMR) spectroscopy. The results revealed statistically significant increase levels of glucose and glutamine as well as decrease of lactate and alanine in the secretome of cells treated with 5-FU compared to their controls. Real-Time PCR reveals decreased expression of PKM2, uppermost isoform of pyruvate kinase gene in proliferating cells, with a role in regulation of energetic metabolism. In part, these findings may be caused by a shift from glycolysis e glutaminolysis (cancer cells) to oxidative phosphorylation (normal cells) with loss of tumor features, opposite to the effect of Warburg. In conclusion, our results provide evidence that 5-FU can modulates the metabolism of cell line.

Keywords: Head and Neck Cancer, Secretome, Metabolome

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 89

An algorithm to find "anonymous loci" in fully and partially sequenced genomes: bench validation and primer decay studies

[1]Igor Costa, [2]Bryan Jennings, [1]Francisco Prosdocimi

[1]Instituto de Bioquímica Médica, UFRJ, Brazil, [2]Museu Nacional, UFRJ, Brazil

Background
The ideal features of DNA sequence-based for coalescent analyses in phylogeographic and phylogenetic studies include the following: (i) neutral evolution; (ii) independent assortment in meiosis, (iii) single-copy in the genome; (iv) conservation among conspecific individuals or closely-related species. Although Anonymous Loci (AL) markers fill all these requisites, their development – and hence their widespread adoption by researchers – has been slow owing to the technically challenging and time consuming methods required to find them.

Methods
To address this need, we developed a new algorithm to find AL in fully or partially sequenced genomes. The results were validated in chicken model and a primer decay study for AL amplification has been performed in primates. Anonymous Loci Finder (ALF) algorithm receives as input a whole and/or partial genome data in FASTA format, an optional INFO file containing the position of the genes predicted in the respective genome and another set of parameters. ALF searches for anonymous genomic regions located far from genes to avoid gene linkage (default distance: 10Kb), tests whether the region selected is single-copy and run internal validations. By default the algorithm prints 50 anonymous regions with 2Kb in size that have proven to be single-copy and unlinked to genes or other ALs. However, hundreds or thousands of such loci can be quickly found using this methodology.

Results
We then used our algorithm to predict 500 ALs for the chicken and human genomes. Our program also created forward and reverse primers to amplify a 550bp region inside each of the ALs found. We selected 11 of those primers from the chicken genome to test the amplification in vitro. All the 11 primer pairs amplified successfully a chicken sample. Sanger sequencing of the amplicons was used to verify that the sequences obtained matched the reference genome.
In order to test AL conservation among clades, primers created for the human ALs were amplified using in silico PCR (isPCR) tool against the genome of 9 other primate species (chimpanzee, gorilla, orangutan, gibbon, macaque, marmoset, tarsier, bushbaby, mouse lemur) with nearly complete genome available in ENSEMBL database. We observed the amplification of 337, 296, 179, 149, 80, 32, 3, 3, 0 loci, respectively, which fitted well against an exponential decay curve over time (R^2 = 0.9983). Using the Anonimous Locus amplification ratio as a measure of pairwise distance, we managed to recover a well-supported primate phylogenetic tree.

Conclusion:
Our results show that predicted ALs can be helpful to study species with more than 20 MY divergence from a common ancestor. The publication of ALF will help to elucidate evolutionary patterns of divergence amongst biological species.

Keywords: Locus Anonimous; Genetic Markers; Phylogeny; Bioinformatics;

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 90

USING VIACOMPLEX TO CHARACTERIZE CELL PATHWAYS ALTERATIONS IN ALZHEIMER'S DISEASE

[1]Bruno Vendrusculo, [2]Eder Simão, [1]José Carlos Merino Mombach, [3]Cristhian Augusto Bugs

[1]Universidade Federal de Santa Maria, Brazil, [2]Centro Universitário Franciscano, Brazil, [3]Universidade Federal do Pampa, Brazil

Background: Alzheimer's disease (AD) is defined by the cooperation of abnormal aggregates composed of phosphorylated Tau protein and of abnormal cellular changes including neurite degeneration (loss of neurons) and loss of cognitive functions. Some of these anomalies are strongly associated with disease progression. However, many of the mechanisms linked to cognitive decline remains unclear. Methods: With the purpose of characterizing the functionality of genome maintenance pathways and the Cdk5 and Tau pathways in Alzheimer's disease (AD), we analyzed the dataset GSE1297 obtained from the Gene Expression Omnibus (GEO):9 samples of control tissue, 7 samples of incipient AD, 9 samples of moderate AD and 7 samples of severe AD. The transcriptomic data were analyzed using the tool ViaComplex and a statistical test that identifies significant alterations in pathways activities. Results and Conclusions:The analysis identified several altered pathways. Based on preliminary results, we observed the constant expression of DNA repair and altered programmed cell death according to the evolution of the AD. We also observed the persistence of changes in cell cycle and Tau pathways during the progression of the neurodegeneration.

Keywords: Alzheimer, Statistical Analysis, Genome Maintenance Mechanisms

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 91

In silico reconstruction of metabolic network of Paracoccidioides lutzii

[1]Waldeyr Silva, [2]Maria Emília Walter, [3]Maria Sueli Felipe

[1]IFG, Brazil, [2]unB, Brazil, [3]UCB, Brazil

Background Paracoccidioides lutzii is a dimorphic fungus that causes paracoccidioidomycosis (PCM), a systemic mycosis in Latin America. Genome Project Pb, developed at the MidWest Region of Brazil, identified about 6,000 genes, part of them annotated as enzymes. Using these enzymes, the first in silico reconstruction of the primary metabolic network of P. lutzii was proposed in 2012. In this work, we propose a pipeline to predict gene clusters involved in the secondary metabolism of P. lutzii. Methods The P. lutzii genome was obtained from BROAD Institute. Perl scripts were developed to manipulate these original data, which were taken as input for two programs (SMURF and antiSMASH) that identify gene clusters. The enzymes of these predicted clusters had their ECs identified (IUBMB), which were used to locate several metabolic pathways, potentially occurring in fungi in MetaCyc using Pathway Tools. Besides in these identified pathways, some non identified enzymes could be annotated using BRENDA data base. Results Previously, 275 pathways, and 1413 annotated enzymes were identified. Our work identified 20 pathways: Heme biosynthesis from uroporphyrinogen-III I, Aerobactin biosynthesis, Selenate reduction, Reductive TCA cycle I, Retinol biosynthesis, Mevalonate pathway I, Isobutanol biosybthesis, Serine racemization, Phenol degradation I (aerobic), Ethylene biosynthesis III (microbes), Trehalose degradation II (trehalase), Trehalose degradation VI (periplasmic), Methanol oxidation to formaldehyde IV, 3 pathways for tRNA charging, Phytol degradation, Pyruvate fermentation to butanol II, Ubiquinol-6 biosynthesis (eukariotic), Galactose degradation I (Leloir pathway). Besides we reannotated 9 enzymes: phenol 2-monooxygenase, N6-hydroxylysine O-acetyltransferase, 2-oxoglutarate synthase, phosphoenolpyruvate carboxylase, pyruvate, water dikinase, pyruvate synthase, branched-chain-2-oxoacid decarboxylase, D-Alanine-poly(phosphoribitol) ligase, trans-2-enoyl-CoA reductase (NADPH). Conclusions The study of secondary metabolism in fungi requires a different approach to the primary metabolism. The main reason is that secondary metabolites are expressed in gene clusters. The identification and prediction techniques for primary metabolic pathways do not apply to secondary metabolism pathways. In this context, a pipeline for the identification of specific secondary metabolic pathways has been proposed and achieved satisfactory results.

Keywords: Metabolic network, fungi, reconstruction, pipeline

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 92

Gene selection and classification for cancer microarray data with multiple ordering criteria

[1]Juliana Silva Bernardes, [2]Carlos Eduardo Pedreira

[1]COPPE-PEE, Brazil, [2]COPPE-PEE e Faculdade de Medicina, Brazil

Gene expression has been proved to be a valuable resource for classification of complex diseases such as cancer. However, microarray data pose great challenges to accurate prediction for two reasons: there is a large amount of inherent noise and variability in samples and difficulties also arise from high dimensionality as compared to a relatively small sample size. Under such circumstances, feature selection methods become an essential element to improve model performance and avoid over-fitting. Basically, feature selection methods are divided into three categories: filter, wrapper, and embedded methods. Filter approaches, due to its simplicity and computational efficiency, is a widely used dimensionality reduction technique, especially for large dataset where other methods are computationally too expensive. We propose a new filter strategy for gene selection based on multiple ordering criteria. First, we used four objective functions (mutual information, receiver operating characteristic scores, bhattacharyya distances and wilcoxon test) to rank genes and then we found the top ranked ones by employing the Pareto front algorithm. Finally, the top-ranked genes were used to cancer diagnostic prediction. We tested our method on four cancer prognosis datasets, and achieved either comparable or superior performance compared to other filter methods and to the SVM-RFE (Recursive feature elimination algorithm that uses SVM weight vector). Our studies suggest that multi-objective approaches are suitable for feature selection, since they combine the consistency and otherness of different evaluation criteria. By adopting more than one evaluation criteria sequentially one can improve the efficiency of feature selection.

Keywords: microarray, gene selection, filter method, multi-objective optimization

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 93

Using metadynamics simulations to predict the binding affinity between major histocompatibility complex class I and peptides derived from C-terminal extension of cysteine proteinase B of Leishmania (Leishmania) amazonensis

[1]Artur Brandt, [1]Paulo Ricardo Batista, [1]Oswaldo Cruz, [2]Carlos Roberto Alves, [2]Ernesto Cafarena

[1]Programa de Computação Científica, Presidência/Fiocruz, Brazil, [2]Laboratorio de Biologia Molecular e Doenças Endêmicas, IOC/Fiocruz, Brazil

Background: One of the essential steps during the infection of the Leishmania is the capacity of peptides derived from its cysteine proteinase type B (CPB) C–terminal region to bind to MHC Class I cleft. It has been proposed that during the intracellular life stage of the parasite interactions occur between of some fragments of C-terminal region of CPB and the immune system of the vertebrate host, specifically the major histocompatibility complex class I (MHC) proteins. Molecular dynamics (MD) can be employed to study these interactions, but simulations on a molecular scale of MHC/peptide complexes are not trivial. Generically, MD simulations would be meaningful only if the run is long enough to visit all the energetically relevant configurations, but it can require an unpractical amount of computer time. Metadynamics belongs to this class of methods that can accelerate the sampling of configuration space in which the sampling is enhanced by the introduction of a historical bias-dependent potential. Results: We applied metadynamics simulations to predict the binding free energy for four crystallographic MHC Class I/peptide complexes. Using this approach, we were able to mimic the dynamics of peptides exiting the MHC cleft and we were able to reconstruct the free energy surfaces for these complexes. This protocol was able to distinguish clearly between bound and unbound states of MHC/peptide complexes, predicting the binding free energy for each case. Additionally, we applied the same methodology to six synthetic MHC/peptides complexes, and we compared the results with values experimentally obtained from SPR based essays. With metadynamics simulations, we also showed that peptides stably bind to H-2 cleft, indicating the spontaneity of complex formation. Conclusion: This protocol confirms the predictability power of metadynamics to calculate the binding affinity between MHC class I H-2/peptide complexes. In this case, the details of simulations can be used to go further and investigate the dynamics of peptide. This methodology could be used as a new tool for enhancements in drug design.

Keywords: Molecular Dynamics, metadynamics, MHC, Leishmania

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 95

THE HISTORY OF HUMAN ORGANS EVOLUTION TOLD BY THEIR PROMINENT GENES

[1]Katia P Lopes, [1]J Miguel Ortega

[1]Lab. Biodados, Dept. Bioquímica e Imunologia, ICB, UFMG, Brazil

Background. The occurrence of genes can be restricted to some clades. Some human genes are shared with prokaryotes; others are present only in Eumetazoa, while some are restricted to placentary animals. We can determine the occurrence of homologues of human genes by inspecting a database UEKO, which comprises the Kegg Ortology database enriched by us with UniProt clusters. By using a local copy of Unigene database, we can infer the expression level of human genes in different organs. Thus we set up to determine the time of appearance of the genes that constitute the organs gene expression (prominent versus all), aiming to tell the human organs evolution. Results. The typical profile of insensitivity of gene expression is logarithmic. We examined genes that show at least 1, 10, 50 or 100 ESTs per 100K. This corresponds to the following number of genes, respectively: Brain (5460, 1095, 115, 32), Lung (5401, 1099, 77, 28), Ovary (4799, 1173, 94, 26), Pancreas (4852, 989, 100, 39), Placenta (4976, 1154, 115, 38), Prostate (5176, 1135, 86, 30), Skin (4649, 1295, 134, 42) and Testis (5261, 1183, 74, 14). Accessing the group or orthologues and determining the Last Common Ancestor of it, we determined the fraction shared with Cellular organisms, Eukaryota, Placentary, up to Homo sapiens. Our data show that the area underneath these curves of appearance of genes differs: Placenta (26, 26, 23, 21), Brain (27, 27, 27, 27), Prostate (27, 27, 27, 27), Pancreas (26, 27, 26, 24), Ovary (27, 27, 27, 28), Testis (27, 28, 27, 27), Lung (26, 27, 26, 27) and Skin (27, 27, 27, 27). Remarkably, some placentary genes are originated as late as by the époque of Simiiformes. One interesting detail is that the percentage of the tissue prominent genes (> 100 ESTs/100K) that is shared with prokaryotes is: Placenta (13%), Pancreas (26%), Prostate (30%), Skin (31%), Brain (34%), Lung (36%), Ovary (38%), Testis (43%), showing that different portion of these organs comprise very ancestral genes. Conclusions. The history of organs was told by the origin of the tissue most prominent genes, showing that the set of genes that rather characterize lung, brain and placenta, for instance, reached the set repertoire in different moments of human evolution. Moreover, a distinct portion of the prominent genes in human organs are shared with prokaryotes. When all expressed genes are used, we noticed that the analysis lack the tissue specific signature, approaching the distribution of appearance of the entire repertoire of genes. Thus, the prominent tissue expressed genes appear with a distinct kinetics along human evolution. Supported by: Capes, FAPEMIG, CNPq.

Keywords: human organs, ESTs, gene expression

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 96

LOCAL ASSEMBLY OF SUGARCANE EXPANSINS GENES USING SORGHUM AS A REFERENCE

[1]Ivone de Bem Oliveira, [1]Isabela Pavanelli Souza, [1]Arthur Tavares de Oliveira Melo, [1]Ludmila Ferreira Bandeira, [1]Alexandre Siqueira Guedes Coelho

[1]Setor de Melhoramento de Plantas, Escola de Agronomia, Universidade Federal de Goiás, Brazil

Keywords: Homology, Phylogeny, Genomics

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 99

COMPARATIVE ANALYSIS OF DIFFERENT PROTEIN STRUCTURE REFINEMENT PROGRAMS

[1]Tales Feitosa, [1]Marcus Batista

[1]Federal University of Segipe, Brazil

COMPARATIVE ANALYSIS OF DIFFERENT PROTEIN STRUCTURE REFINEMENT PROGRAMS TAL Feitosa1, MVA Batista1 1Laboratory of Genetics and Conservation of Natural Resources, Department of Biology, Federal University of Sergipe In silico comparative modeling is one of the most important tools to elucidate the working mechanisms of a protein. During the process, the modeled structure is submitted to several procedures to refine its three dimensional conformation so that it agrees with its natural state in vivo. It is known that, the natural state of a protein accommodates their atoms in a state of low energy. So, several softwares are available for such refining tasks based on energy minimization. With the advances of technology and the increased interest in this field by the scientific community, these softwares are increasingly available through the web in the form of servers or stand alone programs. These computer programs submit the modeled protein to a series of tests carried out by a number of logarithmic calculations to improve its three dimensional conformation. However, the accuracy of these algorithms could vary from software to software. Therefore, the present study focused on comparing the accuracy of the most used free structure model refining tools, the online servers ModRefiner and 3Drefine. These servers were used to refine two structure models created by different homology modeling programs with different validation tools. ModRefiner was able to improve the first modeled structure with 89.4% of residues in most favorite regions in the Ramachandran Plot to 93.8%, and the second model from 94.2% to 95.1%. 3Drefine improved the same first model to 94.7% and the same second model to 95.1%. When both refined structures were aligned in STRAP, small differences could be detected among the entire structural segment. Another method to validate a model is assessing of the Ramachandran Plot using Z-Score. ModRefiner presented a Z-Score of 1.303 and the 3Drefiner presented a Z-Score of 1.395. Despite the 3Drefine presented a slightly better result than ModRefiner, both refinement servers presented satisfactory results. The process of protein modeling can have different outcomes depending on the diverse variety of methods and situations these methods are applied to. So it might be prudent to use a more diverse range of softwares when refining a protein under the same circumstances so that there are more suggested results to choose from when testing the outcomes on the bench. Another important task is to report these results to the groups that developed such softwares so they could improve the algorithms. Supported by: COPES/POSGRAP/UFS.

Keywords: protein refining programs, Modrefiner, 3drefine

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 100

STRUCTURAL COMPARISONS BETWEEN E6 ONCOPROTEINS FROM BOVINE PAPILLOMAVIRUS-5 AND BONIVE PAPILLOMAVIRUS-1

[1]Tales Feitosa, [1]Marcus Batista

[1]Federal University of Segipe, Brazil

STRUCTURAL COMPARISONS BETWEEN E6 ONCOPROTEINS FROM BOVINE PAPILLOMAVIRUS-5 AND BONIVE PAPILLOMAVIRUS-1 TAL Feitosa1, MVA Batista1 1 Laboratory of Genetics and Conservation of Natural Resources, Department of Biology, Federal University of Sergipe Homology modeling is one of most efficient tools for predicting unknown protein structures. The understanding of the structure of viral proteins is an important strategy to study the molecular interactions between the parasites and their host. Bovine papillomavirus (BPV) are circular dsDNA viruses that infect keratinocyte stem cells of cattle from all around the world. At the present moment, thirteen BPVs have been characterized, and three of them are known to induce carcinogenesis. This oncogenic characteristic is due to the activity of a set of three proteins (E5, E6 and E7). Bovine Papillomavirus-5 (BPV-5) presents the E6 homologous gene but has not been linked to abnormal cell proliferation of the host cell. On the other hand, Bovine papillomavirus-1 (BPV-1) stands among the carcinogenic types described in literature. Therefore, the purpose of this study was to make use of structural bioinformatics tools in order to compare the modeled structure of E6 from BPV-5 with the structure of the oncogenic E6 protein from BPV-1, so we could increase the knowledge about the transforming properties of these proteins. When we compare the E6 from BPV-1 and BPV-5, we found an E-value of 2e-20 and they share 38% of identity when aligned with BLAST tool. The E6 model from BPV-5 was created using a homology modeling approach by satisfying special restrains using the MODELLER program. The model presented 95% of residues in most favored regions in Ramachandran Plot and its Z-score was 0.571. BPV-5 E6 has 134 amino acid residues and model showed two zinc finger regions situated within amino acids 15-51 and 88-124, and BPV-1 has 137 amino acids residues and also has 2 zinc finger regions situated within amino acids 17-53 and 90-127. Both structures have five CxxC conserved catalytic sites that are responsible for binding to LxxLL motif from adhesion proteins in the host cell. The presence of these conserved catalytic residues in E6 protein family has been suggested to be a characteristic of all mammalian papillomaviruses. The data shown by E6 model from BPV-5 agrees with that possibility. In general, BPV-5 E6 presented some structural conserved regions with BPV-1, associated with the functional motifs. However, structural variable regions could also be identified. Despite the structural variability, functional domains seem to be conserved. The understanding of the structural properties of oncoproteins from BPV could increase the knowledge on the function of these proteins, which could help in the discovery of novel possibilities of treatment. Supported by: COPES/POSGRAP/UFS.

Keywords: Bovine papillomavirus, homology modelling, structural comparison

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 102

WHEN THE TOPOLOGICAL DOMAINS FIRST APPEARED IN HUMAN ANCESTORS?

[1]Diego Trindade de Souza, [1]Sergio Russo Matioli, [2]José Miguel Ortega

[1]USP, Brazil, [2]UFMG, Brazil

The most cited database of topological domains is CATH. It comprises onlyproteins with PDB entries. We developed approaches to assign CATH domains to human proteins,determined the groups of orthologous to which they belong, and estimated its first occurrence in the ancestral lineage that lead to humans (e.g. Metazoa, Gnathostomata, Euteleostomi, Tetrapoda, Eutheria…).CATH database contains 1313 distinct topological domain entries. To map their occurrence into all human proteins (73972 UniProt entries which are "Complete"), alignments were collected using 80% coverage and identity thresholds. This procedure resulted in 11547 human proteins with CATH Topological Domains. Then we made use of the UniRef Enriched Kegg Orthology (UEKO) database (GR Fernandes and collaborators, not published) and sent the taxonomy identifiers to a LCA WebService (H Velloso and collaborators, not published), both database and WebService produced by our group. Thus, the origin of the proteins bearing Topological Domains was determined. Respectively, we found that 1801, 232, 488, 2 and 2 of them originated before the origin of the clades Euteleostomi, Tetrapoda, Eutheria, Catarrini and Hominoidea, belonging to the following number of KOs: 373, 74, 55, 2, 2. Moreover, grouping the occurrence by CATH domain, in order to detect the most ancient occurrence, it was detected that the proteins originated in Catarrini and Hominoidea contain domains also present in human proteins shared with bacteria, thus very ancient domains. However, we found 20, 1 and 1 domains which had first appeared in, respectively, Euteleostomi, Tetrapoda and Eutheria, very recent clades in human evolution. The only one which appears in Eutheria is the 2.40.15 (Proto-oncogene – Oncogene Product P14tcl1). This is the first initative, in our knowledge, to map well known Topological Domains to human proteins and determine the first occurrence of them along the ancestors lineage, aiming to discover when the Topological Domains first appeared. Although most of Topological Domains appear in proteins shared with bacteria and/or archaea (52.6%), some of them are recent. The depicted scenario opens the opportunity of driving subsequent efforts to understand domain achieving along evolution

Keywords: origin of new genes, genomic evolution

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 103

GENOME ASSEMBLY OF FOUR NOVEL VIRUSES USING 454 PIROSEQUENCING DATA, THROUGH A DE NOVO ALGORITHN

[1]Davi Inada, [1]João Vianez, [1]Alex Lima, [1]Valéria Carvalho, [1]Layanna Oliveira, [1]Clayton Lima, [1]Márcio Nunes

[1]Evandro Chagas Institute, Brazil

Introduction: The viruses belonging the orthobunyavirus genus are known for harm the human health. The genetic reassortment event has been detected in many of these viruses in their in nature state, leading to a large phenotypic diversity. Methodology: In this study, was performed the nucleotide sequencing of four novel viruses of the Orthobunyavirus genus (Guama virus, Moju virus, Bimiti virus and Ananindeua virus), the development of an effective genome assembly strategy for this data set and the genomic annotation. The genomes were sequenced by the 454 GS FLX system. For the de novo assembly two programs were used, Newbler 2.6 and 7.0 Celera. The gap closing was performed by the program CLC Genomics Workbench, for the curation and annotation process, Geneious R6, Tablet 1.13.07.31 and BLASTx were used. Results and Discussion: In the assembly of the four viruses' genomes using Newbler 2.6, only 2,5% of the generated contigs presented as a viral contig in a total of 2841 contigs. While the Celera 7.0, 18,5% of 214 contigs. Only Celera 7.0 program was not successful to recovery the sequences from small segment of the Bimiti virus, Moju virus and Guama virus. It was evident that the great prevalence in the contaminant data in the sequencing process affected directly the genomic coverage, the assembly by Newbler 2.6 presented an coverage average of x50,35 and the assembly by Celera 7.0, an coverage average of x11.98. Conclusion: In addition to the pipeline described in this study, the genomic data generated in this study may assist the future studies, particularly the study of evolution of these viruses, which will enable us to gain a greater understanding of their evolutionary mechanisms.

Keywords: Orthobunyavírus, bioinformatic, genome assembly.

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 104

FGAP, a new gap-filling program for finishing draft genome sequences with superior performance

[1]Vitor Cedran Piro, [2]Helisson Faoro, [2]Vinicius Weiss, [2]Maria Berenice Reynaud Steffens, [2]Fabio de Oliveira Pedrosa, [2]Emanuel Maltempi Souza, [1]Roberto Tadeu Raittz

[1]Laboratory of Bioinformatics, Federal University of Paraná, Brazil, [2]Department of Biochemistry and Molecular Biology, Federal University of Paraná, Brazil

Finishing is the most time consuming and labor intensive step in genome sequencing. Several in silico methods have been proposed aiming to solve finishing problems such as error correction, contig ordering, gap filling, assembly validation and refining. Gap filling or gap closing involves the identification of sequences to fill in gaps between adjacent contigs or between scaffolds. The presence of such gaps may be due to the absence of the respective reads in the database, which requires new sequencing data, or to inherent inability of the assembler to deal with repeated regions and low coverage, which can be improved by gap filling approaches. We developed a new tool named FGAP to make use of assemblies obtained with different assemblers or from different sequencing platforms to close gaps of a draft genome sequence. This tool was compared with the programs GapCloser, GapFiller and IMAGE, developed for the same purpose. FGAP searches for sequences overlapping contig ends of a proposed scaffold draft genome. GapCloser, GapFiller and IMAGE have similar approaches: they identify paired reads that map at the contig ends of a given gap to extend them by performing local assemblies. All methods were tested and evaluated with an assembly of E. coli str. K-12 substr. MG1655 sequenced using paired-end Illumina HiSeq 2000 and single-end 454 GS FLX reads. Both sets of reads were assembled together using SOAPdenovo2, generating a draft genome sequence with 123 gaps in 41 scaffolds plus 32 non-ordered contigs, totaling 4.63Mb. For FGAP, two datasets were independently generated: an assembly of Illumina reads used as single-end and a separated assembly of 454 reads. For all other programs the paired-end Illumina information was used for gap extension. The validation was made comparing 100bp upstream and 100bp downstream from the new insert against the reference E. coli K12 genome using nucleotide BLAST. Gaps were considered correctly closed when the coverage of the aligned sequence was 100% and identity higher than 95%. The results obtained were (gaps closed/gaps validated): FGAP (106/99), GapCloser (101/92), GapFiller (98/73), IMAGE (101/79). The results show that FGAP has a superior performance with more gaps closed and a higher percentage of validated gap closures. In addition, FGAP does not depend on paired-end or mate-pair information, is more flexible, is faster, provides an intuitive output for easy validation and can be used directly as a web application. Finally, when Illumina paired-end information is available, the combination of FGAP and GapCloser, which together allowed closing 109 out of 123 (~89%) gaps, showed to be a good choice for closing gaps since their results were partially complementary. The tool is freely available at www.bioinfo.ufpr.br/fgap/

Keywords: Genome finishing, gap filling, genome assembly

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 105

Protein sequence clustering to identify functional analogous and verification of false-positives specific enzymes from Trypanosomatids.

[1]Larissa Catharina Costa, [1]Nicolas Carels, [1]Wim Degrava, [1]Ana Carolina Ramos Guimarães

[1]Fundação Oswaldo Cruz, FIOCRUZ-RJ, Brazil

Enzymes are proteins that catalyze chemical reactions. Analogous enzymes are interesting entities for the development of studies on gene and pathway evolutions. They have independent evolutionary origin, and also represent potential targets for the development of new drugs. This work used the AnEnπ tool to identify analogy instances by grouping primary sequences of proteins from their enzymatic function. For clustering sequences, two databases were used: the KEGG (public releases from June 2009 and June 2011) and SwissProt (August 2008 and October 2012 releases), where both databases contained a data set representing all the enzymatic activities described. The results obtained with data sequence clustering from different versions databases corroborate the expected result since the number of sequences, organisms and enzyme activities was greater in more recent versions of both databases, indicating that during updates there might have been additions to them. At the same time, deletion cases of sequences and decreased enzyme activities during these updates may have occured. Therefore, the clustering of sequences is indispensable to maintain the data updated, because it will be possible to proceed the search for analogous enzymes from these data. In order to detect possible cases of false-positives in identifying specific enzymes obtained from the clustering performed by AnEnπ, an analysis of predicted specific enzymes from parasites (Leishmania, Trypanosoma brucei and T. cruzi) was performed and compared with the human genome using BLAST tool. From the data generated by BLAST, protein sequences that had similarities to the database of the Human Genome were subjected to a comparison with the database of Ensembl proteins in order to identify regions of similarities between them. The occurrence of a hit between the parasite protein sequences and protein sequences of the human genome, does not mean that the enzymatic activity is present in the human genome, becoming the sequence specific to the parasite homologous to human. To determine the existence of this function, a coverage analysis of the alignment is being performed between the sequences of the parasite that obtained a hit with human protein sequences, in order to check whether they are truly specific of the parasite or if they are homologous to sequences of human.

Keywords: Clustering; Analogous Enzymes, Specific Enzymes, Protein Sequences, Trypanosomatids,

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 106

Comparative analysis of secreted proteins in helminths reveals a dynamic history

[1]Yesid Cuesta-Astroz, [1]Francislon Silva, [1]Laila Alves Nahum, [1]Guilherme Oliveira

[1]Fiocruz-Minas (CEBio), Brazil

Comparative analysis of helminths is important to understand genomic biodiversity and evolution of parasites and hosts under different selective pressures in diverse habitats. Parasite secreted proteins are able to modify host environment and modulate its immune system. Our main goal is to understand the diversity and evolution of secretomes across different helminths. We aim at understanding how secretome diversity is shaped according to different niches and lifestyles and thereby identify specific features that allow parasite survival in different environments. In this project, we initially performed an in silico secretome prediction from the predicted proteome data of four species including free-living and parasitic Nematoda (C. elegans and B. malayi) and Platyhelminthes (S. mediterranea and S. mansoni). We identified secreted proteins associated with invasion, infection, signaling, adhesion, and immunoregulation such as protease inhibitors, cytokines, among others. The main Pfam clans found in the predicted secretomes included the Classical C2H2 and C2HC Zinc Fingers (CL0361), EGF (CL0001), immunoglobulin (CL0011), and protease inhibitor toxin (CL0454) superfamilies. Metabolic information retrieved from KEGG analysis using Blast2GO revealed a great deal of molecular diversity of the helminth secretomes in respect to molecular function and biological process. Altogether, shared and unique features were identified among the helminth secretomes. As we continue this work, we will contribute to the understanding of the evolution of parasitism including host-parasite interactions.

Keywords: secretome, molecular evolution, biodiversity, bioinformatics, phylogenomics.

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 107

COMPARATIVE MODELING AND MOLECULAR DYNAMICS OF THE ENVELOPE PROTEIN CEPA BE AN423429 VIRUS ENCEPHALITIS OF SAINT LOUIS

[1]Fabiano Reis da Silva, [1]Jedson Ferreira Cardoso, [1]Davi Toshio Inada, [2]Sandro Patroca Silva, [1]João Augusto Pereira da Rocha, [2]Edivaldo Costa Sousa Junior, [2]João Lídio da Silva Gonçalves Vianez Júnior,[2] Marcio Roberto Teixeira Nunes

[1]UFPA, Brazil, [2]IEC, Brazil

Background The Saint Louis Encephalitis Virus (SLEV) belongs to the Flaviviridae family, genus Flavivirus. It is widely distributed in the Americas, causing outbreaks in some regions. After human contamination, the infection can evolve and cause disease with symptoms such as fever, headache, dizziness, malaise and nausea and, in some cases, can lead an individual to death. This virus has been isolated in the state of Pará since 1960. Some species of wild birds and mosquitoes, mainly of the genus Culex are hosts of this virus. The SLEV genome consists of single-stranded RNA of positive polarity of approximately 11 kb. This genome produces a polyprotein which is subsequently cleaved, thereby forming three structural proteins. One of those proteins is the envelope protein (E), which shows capacity for fusion with host cell membranes, playing an important role in the pathogenicity of the virus. Results In the present work, we carried out comparative modeling and molecular dynamics simulations of the envelope protein (E) from SLEV strain BE AN423429, which was isolated at Belém-PA and sequenced at SAARB/IEC (Genbank acession number GU808548). The structure of the model was evaluated and validated using a Ramachandran Plot, Anolea, Qmeam and RMSD value. 83.1% of the residues were located in the core regions of an Ramachandran plot. Qmean gave an energy value of 0.573 and the majority of the residues had negative energy values in the Anolea plot. RMSD between the model and template was 0,05 Å. After 10000 steps of energy minimization (steepest descent), the system was subjected to 30 ns of molecular dynamics with the software package GROMACS 4.6.1. We found that at pH 6.0, the pH that the protein adopts a trimeric form, the ectodomain of the protein has a high flexibility region, specially near amino acids 102 and 51. The electrostatic potential of the structure at the end of the molecular dynamics simulation remains identical to one found in the crystal. Conclusions We could determine a theoretical model of the envelope protein from strain BE AN423429 and evaluate its behaviour through molecular dynamics simulations. The model opens possibilities to future studies, including rational drug design, that we intend to conduct in the near future.

Keywords: Virus, Modeling, Molecular Dynamics

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics,Transcriptomics and Proteomics

Abstract #: 108

Web-tool for Ab initio Protein Structure Prediction using Multi-Objective Evolutionary Algorithms

[1]Alexandre Defelicibus, [2]Rodrigo Antonio Faccioli, [3]Leandro Oliveira Bortot, [2]Alexandre Claudio Botazzo Delbem

[1]Programa de Pós-Graduação Interunidades Bioengenharia - EESC-FMRP-IQSC - USP, Brazil, [2]Institute of Mathematical Science and Computation - USP, Brazil, [3]Laboratory of Biological Physics, Faculty of Pharmaceutical Sciences - USP, Brazil

Keywords: Galaxy, PSP, evolutionary algorithms

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 110

A NOVEL ENTROPY BASED TOOL TO SELECT PHYLOGENETIC INFORMATIVE MOLECULAR MARKERS

[1]Marcus Batista, [2]Sergio Paiva Junior, [2]Valdir Balbino

[1]Federal University of Sergipe, Brazil, [2]Federal University of Pernambuco, Brazil

The quality of multiple sequence alignments plays an important role in the accuracy of phylogenetic inference. It has been shown that removing ambiguously aligned regions and/or highly variable characters, can improve the overall performance of many phylogenetic reconstruction methods. When assessing the phylogenetic relationships between groups of organisms, a good way to decrease the computational time and minimize the bias introduced by genomic regions of uncertain homology is to detect and remove those regions from the multiple sequence alignment in order to find only the most phylogenetic informative regions of the genome. In this context, the objective of this study was to implement a novel entropy based computational tool that selects phylogenetic informative molecular markers. Therefore, we present new software that is designed to select regions in a multiple sequence alignment that are suited for phylogenetic inference. For each position in the alignment, the entropy was calculated using the Shannon entropy. All sites with low information complexity, defined as those that exhibited entropy values under a given threshold are considered as suited for phylogenetic analysis. In this way, a new alignment is created, which contains only those sites with low entropy values. Parameters to decrease degeneracy, use sliding windows, and penalize gaps are available. An important available strategy is to use a histogram to help the threshold definition. Simulations with several datasets from various organisms like viruses and arthropods have showed that the trimmed alignments produced accurate phylogenetic trees with less computational cost. Entropy based methods have been shown to be more accurate than other trimming approaches. Although another study presented software the uses entropy to trim alignments, our tool does not use weights and recoding methods. Our idea is to provide a simple and comprehensive algorithm that is based on the statistical determination of the threshold. So, we were able to develop an efficient method that can be used to select phylogenetic informative molecular markers. Even though some studies have shown the effectiveness of other methods which eliminate regions that disrupt the dataset phylogenetic signal, it is important to develop and apply new methods and approaches that increase this efficiency to make the estimation of large phylogenies more accessible. Supported by: CNPq and COPES/POSGRAP/UFS.

Keywords: Entropy; Phylogenetic analysis; Molecular marker selection tool

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 111

AN ENTROPY BASED CONSENSUS DEGENERATE PRIMER DESIGN TOOL APPLIED TO THE AMPLIFICATION OF PHYLOGENETIC INFORMATIVE REGIONS

[1]Marcus Batista, [2]Sergio Paiva Junior, [2]Valdir Balbino

[1]Federal University of Sergipe, Brazil, [2]Federal University of Pernambuco, Brazil

Designing degenerate PCR primers for templates of unknown nucleotide sequence may be a very difficult task. Proper primer design is essential for projects where PCR amplification and/or DNA sequencing play an important role, and a number of algorithms and design recommendations have been proposed. However, it is difficult to implement degenerate primers for highly divergent DNA templates. In addition, in molecular phylogenetics and population genetics studies, it is necessary that one can select appropriate markers and design primers in order to amplify those markers. So, it is important to develop novel approaches and tools that could design efficient degenerate primers that could be used in evolutionary studies. Therefore, the aim of this study was to develop an entropy based consensus degenerate primer design tool applied to the amplification of phylogenetic informative regions. First, it is necessary identify the regions in the genome of interest. To do this, for each position in the alignment, the entropy was calculated using the Shannon entropy. All sites with low information complexity, defined as those that exhibited entropy values under a given threshold are considered as suited for phylogenetic analysis. Once those regions are selected, the algorithm uses entropy to determine the best regions to establish degenerate primers. The locations of forward and reverse primers are found around the selected region, sorted by their entropy values. The algorithm was used to design primers in order to detect Bovine Papillomaviruses (BPV) DNA. Simulations showed that the primers were capable of detecting all BPV types. In order to confirm their efficacy, the primers were tested experimentally and they successfully detected BPV DNA. Therefore, we have developed an entropy based computational tool that efficiently design degenerate primers that can be used to amplify phylogenetic informative regions, which could help evolutionary studies of several organisms. Supported by: CNPq and COPES/POSGRAP/UFS.

Keywords: Entropy; Phylogenetic analysis; Primer design tool

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 112

Theoretical Study of half-reactions in the Class 1A Dihydroorotate Dehydrogenase from Trypanosoma cruzi

[1]Natália de Farias Silva, [1]Jeronimo Lameira, [1]Claudio Nahum Alves, [2]Sergio Martí

[1]Universidade Federal do Pará, Brazil, [2]Universidad Jaume I, Spain

Chagas' disease is considered a health problem affecting millions of people in Latin America. However, cases have been detected in North America, Canada, Europe and some Western Pacific countries. This disease is caused by the parasite Trypanosoma cruzi that can be transmitted to humans mainly by the insect vector Triatona infestans, also called barbeiro. Recently DHODH class 1A from Trypanosoma cruzi (TcDHODA) was shown to be essential for the survival and growth of T. cruzi and proposed as a drug target against Chagas' disease. Based on amino acid sequence similarities, DHODHs can be divided into two families: Family 1 and 2. This DHODH catalyzes the oxidation of S-dihydroorotate (DHO) to Orotate (ORO), in the fourth step and only redox reaction in the de novo pyrimidine biosynthesis pathway. The proposed catalytic cycle of DHODH consists of two half-reactions. In the first half-reaction DHO is oxidized to ORO, with reduction of the Flavin Mononucleotide (FMN) cofactor. In the second half-reaction Fumarate (FUM) is reduced to Succinate (SUC). The DHO oxidation half-reactions may occur via a concerted or stepwise mechanism. Herein, the catalytic mechanism of TcDHODA involving DHO oxidation (first half-reaction) and FUM reduction (second half-reaction) was studied using hybrid quantum mechanical/molecular mechanical (QM/MM) Molecular Dynamics (MD) simulations and Bidimensional PMF (2D PMF) calculations. The free energy profile at the AM1/MM level shows that DHO oxidation takes place by means of a stepwise mechanism for TcDHODA, which a proton is abstracted from C5 DHO to Cys130 and a hydride is transferred from C6 DHO to N FMN. In FUM reduction, a proton is abstracted from Cys130 to C2 FUM and a hydride is transferred from N FMN to C3 FUM, where bond breakage and formation occur simultaneously and the free energy profile using the AM1/MM levels shows that the second half-reaction in TcDHODA occurs via a concerted mechanism. Our results demonstrate the great importance of detailing the half-reactions for TcDHODA and help our understanding of the particular role of active site residues in the DHODHs families. These results are expected to provide useful information for the rational design of new inhibitors of T. cruzi.

Keywords: TcDHODH, half-reactions, Molecular Dynamics, 2D-PMF

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics,Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 113

AUTOMODEL: INTERACTIVE SERVICE FOR PROTEIN MODELING

João Luiz de Almeida Filho[1], Jorge Hernandez Fernandez[1]

[1]Universidade Estadual do Norte Fluminense, Brazil

AUTOMODEL: INTERACTIVE SERVICE FOR PROTEIN MODELING JL Almeida Filho1, JH Fernandez1 1 Universidade Estadual do Norte Fluminense Nowadays, prediction of protein three-dimensional structure is made by nuclear magnetic resonance (NMR) and X-ray crystallography. However, the use of both methods is expensive and not always a good result is obtained. Moreover, they cannot be performed on a large scale which leads to a search for a computational method of predicting three dimensional structures. The molecular modeling, in turn, is a feature on extensive development in bioinformatics, it is increasingly being used in many more areas, for example in pharmacology, live sciences and biomaterials. However, modeling softwares are complex and have a user interface harder to use, hindering the daily practice of structural biologists and biochemists, inexperienced users and students. Thus, in order to facilitate the experience of researchers with this type of modeling, we are developing the Automodel, which is a semi-automatic online service that allows the user to make modeling with greater control of parameters using an intuitive interface use. The AutoModel engine is the software Modeller, a modeling tool known in academic environment. As a differential of AutoModel compared to other online modeling systems, the Automodel has a greater control of the modeling and has a module of models analysis, based on refinement of loop regions. Automodel is a service multi-platform that is being developed and your client will run in Microsoft Windows, Linux and Mac OS. Supported by: CNPQ, FAPERJ, CAPES, INCTEM

Keywords: Protein Modeling, Modeller, Web Service

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 114

In silico and in vitro characterization of the enzymes gluconate kinase and uracil phosphoribosyltransferase from Trypanosoma cruzi.

Vanessa de Vasconcelos Sinatti Castilho[1], Marcelo Alves Ferrreira[1], Ana Carolina Ramos Guimarães[1], Wim Maurits S. Degrave[1]

[1]Instituto Oswaldo Cruz (IOC)/ FIOCRUZ, Brazil

Chagas' disease is a public health problem in Latin America with a strong impact in Brazil. According to estimates by the Pan American Health Organization (PAHO), around 8 million people in the Americas are infected with Trypanosoma cruzi at a rate of 12,000 deaths per year. Despite this situation, the drugs used for treating this disease are active only in the acute phase and have low efficiency and many side effects. Therefore, researches for new therapeutic targets are an alternative for the development of new drugs. In this context, we used AnEnPi, a tool designed to identify, annotate and compare analogous enzymes combining different bioinformatics algorithms. Analogous enzymes are the result of independent evolutionary events. Their structures should be significantly different, and opens the possibility to design a ligand that may inhibit the parasite's enzyme and has no or low effect on the human's enzyme. Through this approach, two proteins were identified in T. cruzi, which have the following enzymatic activities: uracil phosphoribosyltransferase (UPRT) and gluconate kinase (GK). UPRT is a pyrimidine salvage enzyme that converts uracil and 5-phosphoribosyl-α-1-pyrophosphate (PRPP) into uridine monophosphate (UMP) and pyrophosphate (PPi). GK is an enzyme from the pentose phosphate pathway that catalyses the phosphorylation of D-gluconate into 6-phospho-D-gluconate. Each of these enzyme activities was identified by AnEnPi, as well as by TriTrypDB and KEGG as having two copies of putative genes in the T. cruzi genome. T. cruzi protein sequence analyses of T. cruzi using bioinformatic tools, such as PFAM and ProDom confirmed these activities by identifying their structural motifs and functional domains. PROSITE was used to identify the sequences of potential motifs for the active site, binding sites and post-translational modifications. Subsequently, the molecular characterization of the enzymes GK and UPRT of T. cruzi was performed by cloning and heterologous expression in E. coli, strain BL21 (DE3), using the vector pBAD-TOPO and pET28a. The polyclonal antisera raised against the purified recombinant proteins was effective in the recognition of their native forms present in the epimastigote form of T. cruzi. The immunofluorescence assay revealed that GK and UPRT enzymes were possibly located in the cytoplasm and inside cytoplasmic vesicles along the parasite, which corroborated in part with the predictions of subcellular localization made in silico with SignalP and PSORT. Moreover, the secondary structure of both enzymes was predicted through PSIPRED, and the tertiary structure was obtained by homology modeling with MODELLER. This work aims to contribute to a better understanding of the parasite's metabolism in order to help the search for new therapeutic targets against Chagas' disease.

Keywords: gluconate kinase; uracil phosphoribosyltransferase; Trypanosoma cruzi

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 115

CELL CYCLE CHARACTERIZATION BY TRANSCRIPTOGRAM

Isabela Correa Berger[1], Rita Maria Cunha de Almeida[1]

[1]Universidade Federal do Rio Grande do Sul, Brazil

The metabolic state of a cell may greatly vary at different stages of cell cycle. In this work we characterized the metabolic state of the cells during the cell cycle with the use of an gene expression analysis technique, the transcriptogram. Measurements of expression by microarrays identify which genes are being transcribed at the instant that the expression measure is performed and are available in public databases such as ArrayExpress. However, microarray measurements are very noisy, and the data processing performed by transcriptogram can significantly improve the signal to noise ratio. The measurements are performed on samples of synchronized cells at different times of the cycle. The results detect significant changes in the transcription of gene groups of interest, such as metabolic pathways or Gene Ontology terms. When applied to Saccharomyces cerevisiae cell cycle has made it possible to characterize the various phases of the cell cycle in sufficient detail to discern phases from each transcriptogram. We reproduce this application to Homo sapiens, which facilitated the identification of areas of greatest change and identifying transcriptional regions with altered biological processes. The transcriptogram demonstrated to be a useful tool for analysis of data from microarrays transcription phase of the human cell cycle.

Keywords: cell cycle, metabolic pathways, gene ontology terms

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 116

Structural modeling of enzymes from the metagenome of the snail Achatina fulica potentially involved in cellulose degradation

Raysla Alves Pires[1], Manuela Leal da Silva[2]

[1]Universidade Federal do Rio de Janeiro - UFRJ, Brazil, [2]Instituto Nacional de Metrologia, Qualidade e Tecnologia - INMETRO, Brazil

Brazil is a world reference in the production of ethanol; however ethanol produced from biomass (cellulosic) is not economically viable yet. Due to its advantages like as reducing the use of petroleum,the minimization of environmental pollution, and coexistence with the culture of food without competition, research in this sector has grown considerably, thus representing a promising solution for the problems current like use of fossil fuels. Thus, it is important to know how the enzymatic hydrolysis happens at the molecular level because we can use enzymes for this technology in the industrial level. This study analyzed sequences from Achatina fulica's metagenomic and searched for enzymes potentially involved in the degradation of cellulose. After, these enzymes were analyzed and their structural models were build using the technique of comparative modeling. Results: Among the enzymes obtained from the metagenome, we chose to analyze the endo–1,4–D–glucanase from Stenotrophomonas sp. This protein had its 3D model built based the structure of the endoglucanase of Escherichia coli K–12 (PDBCod 3QXQ). The choice by the best model from the 100 models candidates constructed by MODELLER was made using two criteria: RMSD between mold and model (generated by PyMOL) and Ramachandran graph (generated by PROCHECK). Conclusions: There was obtained a three dimensional model of excellent quality for the enzyme endo–1,4–D–glucanase from Stenotrophomonas sp., with RMSD 0.167 Å between template and model and Ramachandran plot with 95.1% of the amino acids in favorable regions (Red) 4.6% in allowed regions (Yellow) 0.3% in generously allowed regions (Beige) and 0.0% in the forbidden regions (White). With the better model generated in hands, the next step will be to study of molecular dynamics to elucidate and understand the potential catalytic degradation of cellulose. In addition, other enzymes from the metagenome of the snail Achatina fulica are potentially involved in cellulose degradation and should be analyzed.

Keywords: Structural modeling, Cellulose degradation.

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 117

DEVELOPMENT OF HIGH-THROUGHPUT ASSAYS WITH APLICATION IN FUNCTIONAL METAGENOMICS

Rachel Mazzei Moura de Andrade Lins[1], Elisa Cavalcante Pereira[1], Floriano Paes Silva junior[1], Alberto Martín Rivera Dávila[1]

[1]Fundação Oswaldo Cruz, Brazil

Background: Traditional methods applied to laboratory cultured microorganisms are usually time consuming and limited to specific taxa. Therefore, technologies accessing uncultured microorganisms provide a more efficient strategy to access the enormous diversity that is lost when only the conventional methods of isolation are used in the search for bioactive molecules, like enzymes. As a result, independent methods are increasingly being used to search for new enzymes of uncultured organisms, generating knowledge on microbial ecology, currently designated as metagenomics. The present study consists of a high-throughput metagenomic approach to identify target genes associated with the production of glycoside hydrolases. For this purpose, in silico and functional strategies were applied using biological material isolated from the seawater of Praia dos Anjos, in Arraial do Cabo - RJ, Brazil. Results: Metagenomic DNA (mgDNA) was sequenced using the GS FLX + 454 generating 1.2 million sequences with an average of 600 bp and a total of 682 Mb of raw data. Regions containing conserved domains of the target genes were identified in silico. Simultaneously, another screening strategy was developed to improve the ability to detect sequences of interest through functional assays. Thus, a preliminary study was conducted using chitinolytic enzymes as a model to test a strategy to identify sequences based on HMM profiles (Hidden Markov Model). The search against public metagenomic databases CAMERA (Community Cyberinfrastructure for Advanced Microbial Ecology Research and Analysis) and IMG/M (Integrated Microbial Genomes) recovered a total of 708, 104 and 256 metagenomic sequences of chitinases families GH-18, 19 and 20, respectively. The analysis of these sequences using RPS-BLAST, InterPro and IMG/M, revealed the presence of conserved domains in 75%, 97% and 98% of the sequences of the GH-18, GH-19 and GH-20, respectively. The occurrence of chitinases signatures was also detected in 82% (GH-18), 89% (GH-19) and 99% (GH-20) of the metagenomic sequences, reinforcing the hypothesis that this strategy was efficient and well succeeded in the search for the sequences of interest. Hence, this approach was also applied in the search for other enzymes of biotechnological interest as xylanases and cellulases. Results obtained for xylanase indicated the prevalence of phylum Bacteroidetes, which are organic-matter decomposing organisms. Cellulase enzyme analysis demonstrated fungal cellulases predominance, which was expected since fungi are the largest producers of cellulase. Conclusions: The in silico strategy used for chitinase identification in public databases was, therefore, effective in the search for the other enzymes mentioned so far. Nevertheless, the hits obtained on our data obtained by pyrosequencing were few, suggesting a discrete representation of organisms producing these enzymes, and thus indicating that the use of a functional approach might provide a better performance to identify the target enzymes in our genomic libraries.

Keywords: Metagenomics, HMM profile, in silico assays, functional assays, bioinformatics

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 118

Comparative analysis of atomic interactions in protein-protein interfaces

Pedro Magalhães Martins[1], Wellisson Rodrigo Santos Gonçalves[1], Valdete Maria Gonçalves de Almeida[1], Marcelo Matos Santoro[1], Carlos Henrique da Silveira[2], Raquel Cardoso de Melo Minardi[1]

[1]Universidade Federal de Minas Gerais, Brazil, [2]Universidade Federal da Paraíba, Brazil

The interfaces of protein-protein complexes contain important information about molecular recognition. Atomic interactions that occur in these region are likely to have specific characteristics favoring the formation of agglomeration of atoms, such as hydrophobic patches. Discovering conserved patterns is a major step towards understanding and predicting how substrates and inhibitors recognize each other. Most of the works about such interactions are conducting using coarse-grained information (residue level), however, protein interactions occur on fine-grained or atomic level. The main goal of this work is to quantitatively demonstrate which types of atomic interactions are more common among protein-protein complexes, and what are the residues most commonly found in each atomic interaction, along with its polarity.

Keywords: compare, interfaces, protein-protein, atomic interactions

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 119

Prediction of genetic interactions in Escherichia coli

Esther Camilo[1], Marcio Acencio[1], Ney Lemke[1]

[1]UNESP, Brazil

A genetic interaction is defined by the emergence of a surprising phenotype when two genes are disrupted together. This interaction is called positive, if after the double deletion the growth is improved. Otherwise, if the interaction is negative, after the double deletion the organism get sick or even die. The knowledge of such gene pairs have applications in drug targets discovery and the understanding of disease mechanisms. However, the experimental determination of such genes is an expensive experimental work even for Escherichia coli which has about 16 million of gene pairs arrangement. Moreover, less then 50% of experiments performed to this end yield reliable data. To accelerate research, in silico approaches have been employed. Here, we have tested two different computational techniques: Flux Balance Analysis (FBA), and Machine Learning (ML). In FBA approach, three main steps were performed: the reconstruction of the metabolic network (obtained from literature), the deletion of reactions related to the genes of interest and finally the optimization of the metabolic flux aiming to maximize the cell growth. In the network approach, we firstly built the gene integrated network consisting of protein-protein physical interactions, transcriptional regulational interactions and metabolic interactions, then we computed two topological measures - degree and betweenness - for each gene and finally, we trained our machine learning predictor algorithm with experimental data of positive and negative genetic interactions. While the first approach yielded a very low correlation between the FBA prediction and the experimental data, through the second approach we could observe that degree joint centrality from the regulatory network, that is the gene 1 degree plus the gene 2 degree and, the distance between these genes can be used to discern positive from negative genetic interactions. The constructed decision tree model is able to recover 66% of positive interactions with a precision of 63% and recover 60% of negative interactions with a precision of 63%. Moreover, the decision tree model shows that pairs which degree sum is greater than 129 are strong candidates to be negative interactions. The low precision obtained by the FBA can be explained by the lack of information about transcriptional regulation on the model. To make this integration consistently, we should have experimental data about genetic interaction that was at the same time in the metabolic network. At the present, from the about 16 million E. coli gene pairs arrangement, only 246 are in this situation. So, conversely the common sense, more experiments for model organisms need to be performed to more accurate computational prediction.

Keywords: genetic interaction machine learning

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 120

MOLECULAR DOCKING STUDY OF NEOLIGNANS WITH POTENTIAL ACTIVITY AGAINST DIHYDROOROTATE DEHYDROGENASE

João Augusto Pereira da Rocha[1], Jedson Ferreira Cardoso, Fabiano Reis Silva[1], João Lidio da Silva Gonçalves Vianez Junior[2], Marcio Roberto Teixeira Nunes[2], Luiz Guilherme Machado de Macedo[1]

[1]UFPA, Brazil, [2]IEC, Brazil

Background Leishmaniasis is among the six diseases of the World Health Organization (WHO) and World Bank´s special program for research and training in tropical diseases. According to the WHO, this disease currently threatens 350 million people in 88 countries and causes around 60 thousand deaths each year. The leishmanial parasite relies on de novo pyrimidine biosynthesis. The proteins involved in this pathway can be used as potential targets in the planning and development of new drugs against Leishmaniasis. The nucleotides play an essential role in the functioning of cellular metabolism and they are also the building blocks of deoxyribonucleic acid (DNA) and ribonucleic (RNA). The enzyme Dihydroorotate Dehydrogenase from leishmania major (LmDHODH) catalyzes the oxidation of dihydroorotate (DHO) to produce orotate, a key step in synthesis of UMP (Uridine monophosphate), which is the precursor of pyrimidines. In this study, we performed virtual docking simulations of 20 neolignans with activity against the leishmanial parasite using Molegro Virtual Docker 5.5. The ligands were docked after removal of the crystalographic ligand in the crystalographic structure of LmDHODH in the holo conformation (PDB code 3TQ0). Results The docking energies were calculated using the Moldock score function and ranged from -88 to -130 kcal/mol. The RMSD between the docked ligands and the crystalographic inhibitor ranged from 0.39 to 1.7 Å. The docking simulations showed that the neolignans were stabilized trough hydrogen bonds and hydrophobic interactions with the residues LYS44, ASN68, MET70, GLY71, LEU72, LEU129, SER130, CYS131, ASN195 and SER196 of LmDHODH. Conclusions Our results indicate that the neolignan structure are able to interact with LmDHODH. Moreover, the interactions occur with residues previously described as able to interact with other known inhibitors of the enzyme. In future studies, we will conduct molecular dynamics simulations and free energy calculations in order to confirm our preliminary results.

Keywords: Molecular Docking, Leishmania, DHODH

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 121

Methylation analysis of Infinium HumanMethylation450 beadchip

Henrique Cursino Vieira[1], Ana Cristina Victorino Krepischi[2], Erica Sara Souza de Araújo[2], Tatiane Rodrigues[3], Adriano Polpo de Campos[4], Carlos Alberto de Bragança Pereira[1], Daniel Ciampi de Andrade[5], Ricardo Galhardoni, Helena Brentani[5]

[1]Institute of Mathematics and Statistics, Universidade de São Paulo, Brazil, [2]International Research Center, AC Camargo Cancer Center, Brazil, [3]Institute of Biosciences, Universidade de São Paulo, Brazil, [4]Department of Statistics, Universidade Federal São Carlos, Brazil, [5]Department of Neurology, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Brazil

DNA methylation is a major epigenetic modifications studied. The Illumina Infinium platform HumanMethylation450 offers the lowest cost to interrogate the methylation status of ~ 480.000 CpG dinucleotides distributed throughout the genome. To study the methylation pattern, the data are extracted by software GenomeStudio and can be analyzed in software or exported for statistical analysis usually performed in one of many pipelines described in the literature. The problem of differential methylation in the analysis is to identify the end of CpGs with statistical significance, but isolated, is not reliably the biological phenomenon of DNA methylation. In our work, we propose a differential methylation analysis that considers an interdependent relationship between CpGs. Our pipeline uses normalization and correction proposed by Nizar Touleimat and is divided into two main parts: use of the method IMA and method SEeD. We modified the IMA pipeline for importing standardized data and determination of differentially methylated regions between the two groups using Student's t test and correction by FDR. The purpose of the SEeD is based on the construction of CpGs cores adjacent in the genome, based on selection of CpGs with confidence values of the analysis MedOr from 80% to 99%, and the methylation pattern of difference between groups concordant (gain or loss) for CpGs constituents of each seed. Thus, pellets were generated high confidence in the identification of differentially methylated regions. We use data obtained DNA methylation for two projects: (p1) lymphocyte DNA of 24 fibromyalgia patients and 23 controls, (p2) 16 samples of invasive ductal breast carcinomas, 10 positive and 6 negative axillary lymph nodes, and control group 7 breast tissue samples control. The analysis of p1 resulted in 1,028 differentially methylated CpGs distributed over 307 seeds, whereas in only 94 CpGs IMA analysis isolates were detected. In the analysis of differential methylation of p2, we obtained 22,934 CpGs differentially methylated between tumors and control tissue, over 5,966 seeds, while the IMA 19,492 identified as differentially methylated CpGs. Currently, the microarray for DNA methylation analysis platforms are low cost and wide coverage. The challenge lies in the analysis of the vast amount of generated data for identification results with biological significance. We consider the hypothesis already demonstrated in several studies of the interdependence of CpG sites near the pattern of methylation and built so a pipeline that adds this information to the end result. We identified clusters of CpGs that are differentially methylated regions with high confidence. The pipeline has the capacity to expand the number of sites identified as differentially methylated, especially in the case of p1, in which the differences are necessarily much smaller because it is methylation of lymphocytes and not cancer. The biological evaluation of these findings is still needed.

Keywords: methylation, pipeline, analysis confidence

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 122

Transcription rate of ribosomal DNA in Escherechia coli using a stochastic model sequence-dependent

Rafael Takahiro Nakajima[1], Pedro Rafael Costa[1], Ney Lemke[1]

[1]UNESP-Botucatu, Brazil

Background: Ribosomal genes are essential in cellular metabolism, as a consequence of their importance, they are exquisitely controlled and highly transcribed. Multiple rounds of transcription initiation are the main mechanism to guarantee a fast enough ribosomal RNA synthesis supplying ribosomes for the cell's particular growth condition. Condon et al. estimated the transcription rate of these genes studying the elongation of the ribosomal RNA operon of E. coli. In their experiments, they measured the incorporation of [3H] adenine into the final tryptophan - tRNA (tRNATrp) after addition of rifampicin by hybridization to filters containing an excess of tryptophan probe. The time evolution of the tRNATrp activity shows a plateau achieved when the last RNA polymerase finish the elongation and indicate the required time for it. Based on this result, they determined the RNAP elongation rate: 90 bp/s. We made 1200 simulation of the elongation of the ribosome RNA operon using our stochastic sequence-dependent model for multiple rounds of transcription to ensure greater accuracy of it. We estimated the NTP concentration, the average stalling force of RNAP and the number of active RNAPs during fast exponential growth in E. coli. In case of collision between RNA polymerases, the trailing RNAP "pushes" the leader, which lead to a cooperative behavior that attenuates the backtracking. Results: The simulation reported the time evolution of RNA accumulation obtained of the experimental behavior for the main findings. Besides when we increased the concentration of the nucleoside triphosphates the average time of transcription decreased more than two times, consequently increased the transcription rate. Conclusions: The results show the influence of [NTP] in transcription: it increases the velocity of the processes and corroborate Schneider et al. results however the asymptotic value is smaller than the experimental estimates. Since our model is based on in vitro parameters, our results indicate that other factors are required to explain the highly transcribed rate observed in vivo. Good candidates are transcription factors (FIS) and sequences upstream of the promoter (UP element) that experimentally have increased transcription rate.

Keywords: rrn operons, RNAP activity, multiple rounds of transcription, molecular motor, pause sites.

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 123

A COARSE-GRAINED SIMULATED ANNEALING APPROACH FOR RNA SECONDARY STRUCTURE PREDICTION

Pedro Costa[1], Ney Lemke[1]

[1]IBB - Unesp, Brazil

Background: Accurate tools for predicting the RNA structures are an important issue for molecular biology. Ribozymes, riboswitches and functional noncoding RNAs are commonly found nowadays and their role as an important part of cell machinery is undeniable. In contrast to proteins, the pairing rules of the RNA results in a hierarchical folding pattern, in which the secondary structure gives the most important contribution to the free-energy of the molecule. The state-of-art secondary structure prediction programs ignore kinetically trapped structures and because of their computational implementation, they have incomplete thermodynamic parameters and do not take into consideration pseudoknots. In this work we propose a simulated annealing approach for RNA secondary structure prediction. The method estimates the free energy of the RNA complex arrangements using nearest neighbor method. Our approach is based on the parameters of Turner group (2004) from the Nearest Neighbor Database, NNDB. The structural motifs considered are hairpin loops, internal loops, bulge loops, multi-branch loops and exterior loops. Our approach determines for a given structure the set of all the possible topologies that results of adding or removing a single stem from the current conformation. It assigns for each one a transition rate based on the difference between their free-energy and the free-energy of the current structure using the Boltzmann weight. Then, it performs a Monte Carlo simulation to set one of them as the new conformation of the RNA molecule, and repeat these steps until the stopping criterion is achieved. We tested our approach on some example structures from the NNDB and we performed simulations using a sequence of a classical cloverleaf structure for a transfer RNA proposed by the Vienna webserver. The stopping criterion for these simulations was set as seven iterations. The 480 simulations take less than 1 min to run on a 2.6 MHz 12-core desktop. Results: Our program accurately recovered the expected free-energy values for the NNDB examples and also recovered the cloverleaf structure from the Vienna sequence as the minimum free-energy structure, but with a slightly different value (4%). The program also returned a cluster of suboptimal cloverleaf structures, and other cluster with structures with very similar values of free-energy. Conclusions: The results confirm the consistency of our approach, since we can obtain accurate results with a reasonable computational cost. We intend to make the program available on line as soon as possible and add pseudoknots to the implementation.

Keywords: RNA Secondary Structure

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics,Human Pathophysiology; Animal Models of Disease

Abstract #: 124

STUDY OF THIAZOLYLHYDRAZONE MOLECULES WITH ANTITRYPANOSOMAL ACTIVITY USING MOLECULAR DOCKING METHOD

Juliane Nascimento[1], Jaqueline Bianca Duarte[1], Karine Silva[1], Renato Costa[1], Fábio Molfetta[1]

[1]Universidade Federal do Pará, Brazil

Chagas disease is a parasitic infection caused by the protozoan hemoflagellate Trypanosoma cruzi (T. cruzi). This disease affects a large number of people and it is estimated that seven to eight million people are infected in the world, the majority in Latin America, in which the disease is endemic. In addition, the disease is considered neglected by the World Health Organization (WHO). Currently, there are only two available drugs, benznidazole and nifurtimox in the treatment of the disease. These compounds have side effects and are not very active in the chronic phase of the disease. Among a number of drug targets being investigated are cruzain, the major cysteine protease active in the parasite. Cruzain is a cathepsin-L-like protease of the papain family thought to be important for intracellular replication and differentiation of the T. cruzi parasite. In this study, 17 compounds were selected from literature because these molecules present IC50 values. These compounds were subjected to molecular docking calculations through the Molegro Virtual Docker (MVD) 5.0 program. The protein structure used in the docking study was obtained from the Protein Data Bank (1ME3 code). The docking results demonstrated that more well scored molecules showed interactions with Gln19, Cys25 and Gly66 amino acid residues, which are required in enzyme catalytic activity. From this, based on the interactions of hydrogen with amino acid residues in the active site, conformation of molecules, values of energy affinity and RMSD (Root Mean Square Deviation) between docked and crystallographic conformation, we choose three molecules to subject to subsequent molecular dynamics simulations. Thus, the development of this project could lead to powerful drugs against Chagas disease.

Keywords: Docking method, Cruzain, Thiazolylhydrazone molecules

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 125

Computational identification of genomic correlates of anticancer therapeutic response

Lindsay Stetson[1], Yanwen Chen[1], Jill Barnholtz-Sloan[1]

[1]Case Western Reserve University, United States

Background A challenge in precision medicine is the transformation of genomic data into knowledge that enables stratification of patients into treatment groups based on predicted clinical response. An effective preclinical model system that enables prediction of the anticancer drug response phenotype could speed the emergence of personalized medicine.

Results Three large-scale pharmacogenomic studies have screened anticancer compounds against greater than 1000 distinct human cancer cell lines. We have combined these datasets to generate and validate omic predictors of the drug response phenotype. A penalized linear regression model and two non-linear machine learning techniques, random forest and support vector machine, were used to generate predictive signatures. The precision and robustness of each drug response signature was assessed using cross-validation across the three independent datasets. We have developed robust and clinically relevant prediction signatures for eleven out of the fifteen tested drugs tested (17-AAG, AZD0530, AZD6244, Erlotinib, Lapatinib, Nultin-3, Paclitaxel, PD0325901, PD0332991, PF02341066, and PLX4720). The random forest machine learning approach outperformed both elastic net regression and support vector machines.

Conclusions The resulting classification of genetic predictors of drug response could be used to stratify patients into treatment groups based on their individual tumor biology, to inform the experimental investigation of our clinical collaborators, and to provide a powerful platform from which existing anticancer therapies could be improved and new therapeutic paths identified.

Keywords:

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics,Human Pathophysiology; Animal Models of Disease

Abstract #: 126

METHOD OF VIRTUAL DOCKING AND MOLECULAR DYNAMICS QM/MM APPLIED TO SEARCH AMAZON MOLECULES WITH POTENTIAL ACTIVITY AGAINST THE ENZYME DIHYDROORATATE DEHYDROGENASE FROM LEISHMANIA MAJOR

Jaqueline Bianca Duarte[1], João Augusto Rocha[1], Karine Silva[1], Fábio Molfetta[1]

[1]Universidade Federal do Pará, Brazil

Leishmaniasis is a disease caused by protozoa of the genus Leishmania which affects about 14 million people in the world, causing approximately 2 million new cases per year. Recent studies suggest that the enzyme Dihydroorotate Dehydrogenase from Leishmania major (LmDHODH) is essential for the survival of protozoan parasites of Leishmaniasis, once it catalyses the fourth step of the synthesis Uridine monophosphate, precursor of all nucleotides in the parasite, which makes this enzyme a potential target in design of new drugs. Recent studies of combinatorial chemistry have compared natural products to synthetic compounds, whereas natural products have some characteristics that offer advantages in drug design, so it is favorable to use computational techniques to screening of natural compounds in the search for LmDHODH as potential inhibitors. In this study, about 800 structures were cataloged in a Molecular Database of Amazon Region (BMA), created by the Laboratory of Molecular Modeling (LMM), were subjected to molecular docking in LmDHODH enzyme by eHits program. Through ranking proposed by the program, which based on the affinity energies, were selected twenty best ligands, and these were subjected to an analysis of hydrogen and hydrophobic interactions. Posteriorly, were selected two structures (ligands 324 and 493) to be subjected to Molecular Dynamics (MD) through hybrid method QM/MM by computational library fDynamo. From this, we carried out 2000 and 1500 ps MD simulation for the structures 493 and 324, respectively. Both structures showed RMSD (Root Mean Square Deviation) stabilized at the end of the simulation time. Finally, the interaction energy by residues shows hydrophilic interactions with the residues ASN68, SER130, ASN195 ASN128, LEU72 and CYS131, reported in the literature like residues that contribute to the stabilization of inhibitory structures on the site of LmDHODH. Besides, interactions with residues ALA20, ARG51, LYS137 and GLU277 not cited in the literature yet, but can be exploited in the search to new compounds against Leishmaniasis. Thus, we can conclude that virtual docking was able to effectively allocate the BMA's structures at the site of the enzyme with good energy affinity, besides the structures subjected to MD had been stable kept interactions with important residues in LmDHODH.

Keywords: Virtual Docking, Molecular Dynamics, LmDHODH

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 128

ON THE PURSUIT OF miR137 TARGET SEQUENCE ACQUISITION IN RECENTLY ORIGINATED GENES

VERÔNICA RODRIGUES DE MELO COSTA[1], HENRIQUE DE ASSIS LOPES RIBEIRO[1], SALEH TAMIN[2], DAT VO[2], MEI QIAO[2], KÁTIA DE PAIVA LOPES[1], LUIZ OTAVIO PENALVA[2], JOSÉ MIGUEL ORTEGA[1], , , , , , , , , AU

[1]UNIVERSIDADE DE MINAS GERAIS, Brazil, [2]The University of Texas Health Science Center at San Antonio, United States

Background. The recent progress on total RNA sequencing (RNAseq), methods of differential gene expression analysis (either statistical or making use of the poli(A) immune precipitation for focusing on unique mRNA fragments) and pattern recognition algorithms that allow for miRNA target sequences, provide a scenario for selection of genes regulated by miRNA. One possible approach is to express a mimetic miRNA and therefore detect genes with modified expression, which may contain the target sequence, or comprise subsequently regulated genes. Results. MicroRNA miR137 affected the expression of 647 genes when a mimetic miRNA was expressed in glial cells. In 312 of these genes a miR137 target sequence has been predicted by at least one software (miRanda, TargetScan or PicTar). We used the UniProt accession for each transcript to run the software SeedServer and grouped orthologous sequences. Retrieving the taxonomic information we mapped their occurrence and determined the Lowest Common Ancestor (LCA) clade. We also studied the appearance of the miRNA molecular machinery and found that human sequences are related to proteins originated by the origin of Bilateria. Remarkably, several human genes regulated by miR137 are more recent (30%). We were surprised by the fact that the distribution of their origin along evolution was very similar to the complete set of human genes, suggesting that their choice was very broad. Moreover, genes originated after the origin of miRNA processing machinery might have either gained the target by mutations, or by recombination with ancient 3' ends. To investigate these two points we checked if most of the 647 genes would be expressed in a tissue specific manner, thus comprising of the focus of a broad choice. Indeed, 235 of them (36%) showed tissue specific expression (>100 ESTs/100K in Unigene database). Remarkably, 34 are prominent in just one of 95 tissues. To investigate the origin of target sequences in recent genes we sought for evidence of ancient ancestry of the C-term in recent genes, accompanied by recent ancestry of their complementary N-term. We made use of PSI-BLAST searches with Annothetic tool, which reports the taxonomic distance between query and subject. Annothetic returned, for Euteleostomi originated proteins, alignments of C-term to more ancient proteins, while the complementary N-term returned hits to Euteleostomi ones, supporting the model of 3' end recombination, incorporating ancient C-term from a gene already regulated by miR137. Conclusions. We pursued the miR137 target sequence acquisition in recently originated genes and our data support two points: (i) the mechanism might comprise several statistically differentially expressed genes however it is focused on tissue specific prominent ones; (ii) miR137 target sequence can be gained from more ancient genes through exon shuffling. Supported by: Capes, FAPEMIG, CNPq.

Keywords: RNAi, target genes, evolution, orthologues, bioinformatics, gene expression

Topic: Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 129

IN SILICO ANALYSIS OF BREAST CANCER METHYLATED GENES: PROMISING TARGETS TO TRIPLE NEGATIVE THERAPEUTIC

Nara Araújo[1], Roberta Godone[1], Carlos Castelletti[1], José Lima Filho[1], Danyelly Martins[1]

[1]LIKA/UFPE, Brazil

Background: Breast cancer is characterized by a heterogeneous clinical disease in which therapeutic options and treatment strategies depend on the biological properties of the disease subset. The management of the breast cancer treatment is based upon three major disease subtypes: (i) ER-positive that are treated with anti-estrogen therapy, (ii) HER2-positive that are treated with HER-2 target therapy and (iii) triple negative cancers (ER-/PR-/HER2-). Patients with triple negative subtype exhibit a poor prognosis, reflecting the relatively more aggressive behavior of this form of breast cancer. Furthermore, chemotherapy is the unique alternative of treatment due to lack of targeted therapies. Thus, many efforts are being made in the discovery of novel targets to therapeutic for triple negative breast cancers. Multiple genetic and epigenetic alterations are involved in the development of cancer. Among the epigenetic alterations, DNA methylation in the promoter region has been shown to play a fundamental role in breast tumor progression, as silenced genes have been identified that fall into each of the six 'acquired capabilities of cancer'. So, the identification of genes that are hypermethylated and consequently become down-regulated is of clinical significance once these genes can serve as potential targets for therapeutic interventions. This study aims to identify methylated genes among triple negative breast cancer through bioinformatics resources. Materials and Methods: Text and data mining for methylated genes related to breast cancer were performed using, respectively, GoPubMed® and PubMeth tools. For text mining, it was used as inclusion criteria only journals with impact factor greater than or equal to 5. The achieved genes were analyzed individually in respect to their signaling pathways using KEGG as tool. Results: The screening resulted in 159 breast cancer methylated genes, but only 25 of the silenced genes methylation-dependent were found to be involved in important signaling pathways related to cancer, such as ErbB, Estrogen and PI3K-Akt. In this list of 25 genes, 3 were found to be related to triple negative breast cancer: ERBB4, ESR1 and BRCA1, which are involved in the signaling pathways above, respectively. Conclusions: Currently, treatment to triple negative breast cancers is a challenge due to the lack of targeted therapies. This study, through the identification of these methylated genes in triple negative cancers, may aid to the discovery of a therapeutic in a DNA methylation manner once epigenetic alteration are easier to be modified than genetic alteration.

Keywords: breast cancer; DNA methylation; epigenetics; triple negative

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 130

Evolutionary and functional impact of a primate specific retrocopy into the Transferrin locus

Fábio C. P. Navarro[1], Ana Paula S. Urlass[2], Val C. Smith[3], Anamaria A. Camargo[4], Ross T. MacGrillian[3], Pedro A. F. Galante[2]

[1]Centro de Oncologia Molecular, Hospital Sírio-Libanês, São Paulo, Brazil; Dep. de Bioquímica, Universidade de São Paulo, São Paulo, Brazil., Brazil, [2]Centro de Oncologia Molecular, Hospital Sírio-Libanês, São Paulo, Brazil, Brazil, [3]The University of British Columbia – Vancouver, Canada, Canada, [4]Cancer Research, São Paulo, Brazil, Brazil

Transferrin (TF) is an enzyme responsible for transport and distribution iron among body organs. Circulating iron in the human body is mostly bound to TF enzymes and mutations on TF this gene are frequently associated with diseases, such as, anemia and atransferrinemia. Given its fundamental role on organism metabolism, TF is highly conserved in the eukaryotic kingdom, being present in insects, fish and mammals genomes, for example. TF enzyme is composed by two major transferrin domains, which are maintained by seven disulfide bonds on each lobule, which individually binds to a Fe3+ atom. Here, we investigate a ACSL3 retrocopy located into the TF gene of primates. Interestingly, we found that this retrocopy insertion happened 35-40 million years ago, during the primate evolution. By analyzing multiple alignments and conservation on primate genomes, we found putative selective pressures shaping the transferrin locus in order to create new TF isoforms also containing the ACSL3 retrocopy sequence. By using Second Generation Sequencing and publicly available data, we detected at least three described transferrin transcripts. Curiously, due to two isoforms present premature stop codons, we checked their expressions using cytoplasm's RNAs, and, as expected, only an isoform was found. Nowadays, we are characterizing this undegraded TF isoform (tTF), which is composed by only one transferrin domain (N-lobe) and a new 3'UTR sequence, one of the first described natural occurring monolobular transferrin present on mammals. We found that tTF is also binding to iron, therefore, must be very similar to the N-lobe of TF enzyme. Further investigations should be addressed to measure the frequency of tTF on natural tissues, as well as its functions on primates. Supported by CAPES

Keywords: Transferrin, Evolution, Retrocopy, Primate

Topic: Transcriptomics and Proteomics

Abstract #: 131

Differentially regulated microRNAs and their associations with chemoresistance in colorectal cancer

Gustavo França[1], Luiza Andrade[2], Camila Ramos[2], John Mariadason[3], Raphael Parmigiani[2], Anamaria Camargo[4], Pedro Galante[2]

[1]Universidade de São Paulo - USP; Molecular Oncology Center - IEP - Hospital Sírio Libanês, Brazil, [2]Molecular Oncology Center - IEP - Hospital Sírio Libanês, Brazil, [3]Ludwig Institute for Cancer Research – Melbourne Branch, Australia, [4]Ludwig Institute for Cancer Research – São Paulo Branch;

Colorectal cancer (CRC) is one of the three major causes of death by cancer worldwide. The diagnosis is mainly performed by histopathological analyses and the treatment is carried out based on the observed tumor stage. In advanced cases (stages III and IV), surgical excision combined with adjuvant chemotherapy is the recommended procedure. Nevertheless, due to genetic heterogeneity of the disease, the clinical response of patients undergoing adjuvant therapy is extremely variable, making it problematic to define the proper treatment. To investigate possible mechanisms behind the drug response variability, we screened for differentially expressed microRNAs (miRNAs) by deep-sequencing in a panel of 11 CRC cell lines characterized as sensitive and resistant to 5-Fluorouracil (5-FU) and Oxaliplatin, two of the most widely used chemoterapeutic compounds in clinical routine. We have identified 6 and 8 miRNAs differentially expressed between resistant and sensitive cells to 5-FU and Oxaliplatin, respectively. Among the most confident predicted targets for these miRNAs, genes involved in regulation of transcription, apoptosis and pathways in cancer were significantly overrepresented. Previous works have shown that miR-342 is downregulated in CRC cell lines and tissues. Consistently, we have found that miR-342 is suppressed in cells resistant to Oxaliplatin, which was further confirmed by qRT-PCR. Moreover, we noticed that miR-342 is located within the third intron of EVL, a gene involved in homologous recombinational repair of double-strand breaks (DSBs), suggesting that it may help to repair the damage caused by the chemoterapeutic agent. We could also detect the expression correlation between miR-342 and its host gene EVL by qRT-PCR. We propose that miR-342 itself, or, in conjunction with EVL, is a key player that directly affects drug response by promoting apoptosis. To confirm our hypothesis, experimental validation is ongoing.

Keywords: microrna, colorectal cancer, deep-sequencing, gene expression

Topic: Transcriptomics and Proteomics,Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 132

In silico analysis of the raf kinase inhibitory protein (RKIP) expression impact in breast cancer patients

Raul Torrieri[1], João Paulo Lima[1], Rui Manuel Reis[1]

[1]Molecular Oncology Research Center, Barretos Cancer Hospital, Brazil

Background: Breast cancer is a major cause of death in women. It is a highly complex tissue composed of neoplastic and stromal cells. The raf kinase inhibitory protein (RKIP) has been associated with tumor progression and metastasis in several human neoplasms, being currently categorized as a metastasis suppressor gene. We aimed to perform an extensive analysis of the clinical impact of RKIP expression, using the large datasets available for the scientific community at the compendium of cancer transcriptome profiles, OncomineTM (Compendia Bioscience, Ann Arbor, MI). We studied the expression levels of RKIP in 33 different breast cancer expression microarray experiments, totalising 6980 samples. We selected 5 variables (age, tumor grade, PAM50 classification, hormonal receptor status and triple negative status) with known clinical importance for Cox regressions and meta-analysis. Variables were dichotomized and RKIP expression was categorized in quartiles. The 1st quartile was taken as reference. With the development of appropriated R scripts, we identified distinct groups with significantly differential expression. Results: Meta-analysis of HR from multivariate Cox regressions (for each study) indicates that RKIP low expression (1st quartile of expression measure) is consistently associated with poor prognosis. The resultant Hazard Ratio for RKIP high expression is 0.74, 95% CI [0.64, 0.85], with p-value < 0.0001 and 12% of heterogeneity. We also observed a significant difference in risk of metastasis recurrence between patients with low RKIP expression (1st qaurtile) and high expression (4th quatile), with a log-rank p-value of 0.01. Finally, we observed that luminal patients presented higher RKIP expression when compared with non-luminal ones. Conclusions: The developed R scripts are able to efficiently explore Oncomine datasets in order to identify distinct groups with differential expression, perform Cox regression and meta-analysis to look for association of gene expression with prognostic implications. With the developed strategy, we were able to conclude, based on analysis of a large number of samples, that RKIP down regulation is consistently associated with breast cancer poor prognosis, with a significant impact on risk of metastasis and death occurrence.

Keywords: RKIP, breast cancer, in silico expression analysis, Oncomine, R scripts, metastasis supressor

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 135

Human Genes As Candidates to Subtyping of Complex Phenotypes in the Combined Etiology DOISm (Diabetes, Obesity, Inflammation and Metabolic Syndrome)

Emanuel Diego Penha[1], Ana Paula Bezerra[1], Christina Pacheco[1], Samara Silva-Santiago[2], Kaio Farias[1], Mônica Silva[3], Ana Carolina Pacheco[3], Fabiana Araújo[4], Diana Magalhães de Oliveira[1]

[1]Universidade Estadual do Ceará - UECE, Brazil, [2]Faculdade Mauricio de Nassau, Brazil, [3]Universidade Federal do Piauí - UFPI, Brazil, [4]Dataprev – Empresa de Tecnologia e Informações da Previdência Social, Brazil

Background. Nutrigenomics examines complex interactions between genetic/epigenetic and nutritional variables. Since it reflects gene–diet interactions, nutrigenomics data tend to be large, complex and noisy, requiring ever-increasing bioinformatics expertise to deal with their high dimensional and non-linear datasets. Nutrient modulation of genomic expression alters diet response, nutrient requirement and impacts on non-communicable chronic diseases (NCDs) such as obesity (a chronic low-grade inflammatory condition), type 2 diabetes mellitus (T2DM) and metabolic syndrome (Smet). The term DOISm (Diabetes, Obesity, Inflammation, metabolic Syndrome) specifies these disease phenotypes in a combined etiology. Through creation of an exhaustive, curated panel of human genes implicated in DOISm (Dataset 1), we have catalogued the full confluence of genes concomitantly involved in these phenotypes by identifying their underlying genetic variation in order to help in subtyping complex phenotypes, which depends on multiple factors, thus necessitating multiscale approaches such as gene module association studies (GMAS), a complementing method to GWAS in understanding complex diseases. GMAS focus on how genes work together in groups rather than singly. Gene networks can be constructed using gene co-expression data and the expression of gene modules (ie, the eigengene expression pattern) predicts phenotype, since they correlate with traits and, by associating eigengenes with specific traits, the gene module represented by the eigengene can be broken down into its constituent submodules, which can be analyzed in detail to discover which and how novel genes/pathways regulate phenotypes. Results. Through mining several GWAS reports and pertinent databases, we collected data about DOISm related genes. Conditional logistic regression estimated odds ratios (ORs) for conceiving the combined etiology. Out of 1441 human genes related to DOISm (Dataset 1), we have found 769 to be directly related to T2DM and 1057 directly related to obesity, whereas 288 were directly related to inflammation and only 139 directly related to Smet (Supplementary Tables S1-4). Restricting our data to the gene confluence of DOI (diabetes, obesity and inflammation), we found only 78 human genes concomitantly related to the three complex phenotypes (Table 1), but not directly associated to Smet, while 217 genes (Table 2) are confluent for any given two comorbidities. The gene network structure was, then, built to comply with gene modules for DOISm confluence submodules (Figure 1). In another analysis, we used the Cox proportional hazards regression model to estimate hazard ratios (HRs) for future DOISm in groups undergoing one or two comorbidities. Conclusions. The gene network structure is known to be relatively invariant—that is, individual genes do not jump to different modules in response to a stressor (or any specific phenotype). Rather, the gene content of modules remains similar, but the expression level of the genes in the module (represented by the eigengene) is apt to change due to nutrient interactions. In DOISm case, it may turn out that changes in eigengene pattern, rather than the module dataset, will be a better predictor of complex phenotypes, due to the functional relevance of DOISm confluent genes as good candidates for eigengene differential analyses (Figure 1). Supported by: CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico and FUNCAP – Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico.

Keywords: Complex Phenotypes; DOISm (Diabetes, Obesity, Inflammation and Metabolic Syndrome); gene module association studies (GMAS); Eigengenes

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics,Transcriptomics and Proteomics

Abstract #: 136

Annotation and Modelling a Glycoside Hydrolase enzyme from Achatina fullica's gastric Metagenome

Fernando Limoeiro Lara de Oliveira[1], Rosemberg de Oliveira Soares[1], Diego Enry Gomes[1], Manuela Leal da Silva[1]

[1]INMETRO, Brazil

Production of ethanol from Lignocellulosic Biomass is recurrently target of studies due to a deman d for enewable energy sources. Bioethanol is one of the main targets of brazilians researches. The bagasse, a residue from ethanol production is largely available, and make it possible to re-use it as a energy source is a very attractive research objective. The handicap on extracting energy from Biomass is given by the difficulty on the extraction of fermentables from cellulose, and researches on lignocellulolytic enzymes such as cellulases and xylanases may help eke out this demand. The metagenome of gastric juice from african giant snail (Achatina fulica) was made and several enzymes were identified. Among them, a enzyme from Glycoside Hydrolase family 4 (GH4), which was our research target for it's involvement on the Lignocellulosic Biomass degradation.

Keywords: Bioinformatics , proteomics , hydrolases , comparative modelling, bioethanol

Topic: Transcriptomics and Proteomics,Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 138

Prediction of ncRNAs using Machine Learning Approach

Marcos Fonseca[1], Felipe ten Caten[1], Tie Koide[1], Ricardo Vêncio[1]

[1]University of São Paulo, Brazil

In order to deal with the cell dynamicity, many functional RNAs play important roles in gene expression regulation, including transcription regulation, mRNA stability and translation. The functional characterization of non-coding RNAs (ncRNAs) has increasingly become fundamental in the comprehension of gene regulation mechanisms. Some ncRNAs, for example, interact with proteins in order to perform their biological processing. Therefore, an efficient identification of these functional RNAs in genomic sequences is highly significant. High-throughput sequence technologies provide a great source of transcripts expression information and also it can potentially be data source for computational approaches that considering multiple data types, seek for a general model able to predict a new set of candidate ncRNAs. In this work, we consider the method proposed in Lu et al., 2011 aiming to integrate available expression data, RNA secondary structure information, sequence properties and conservation in a Machine Learning (ML) model. This strategy was applied in the archaeal model organism Halobacterium salinarum to assign new putative ncRNA molecules. Available genome annotations, which include coding DNA sequence (CDS), untranslated region (UTR) and already known ncRNAs, were used to build the ML model from a decision tree technique. The remaining regions were used in the prediction considering several sub-regions with fixed size, corresponding to a ncRNA mean size. According to results, starting from a cross-validation evaluation, the ML approach was able to correctly classify nearly all ncRNAs in the training set. Moreover, the model was rectified since the ML approach errors in fact were mislabeling instances. Using the revised model, an independent test set derived from UCSC Genome Browser website (http://genome.ucsc.edu/) was applied and 45% of these elements were predicted as ncRNAs. It is worth mentioning that the derived ncRNAs were obtained from a prediction approach. Thus, none of them were experimentally validated. Merging multiple results arising from different predictions sources, an accurate set of ncRNAs was obtained as the primary contribution of this work. These data provide significant grounds for experimental validation.

Keywords: Machine Learning, Data Integration, ncRNA Predictions

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 139

Determination of structural influences of single nucleotide polymorphisms on TLR1/TLR2 heterodimers and the susceptibility of patients with leprosy through comparative modeling and molecular dynamics

João Herminio Martins da Silva[1], Milton Ozório Moraes[1], Carolinne De Sales Marques[1], Disraeli Cavalcante Araújo Vasconcelos[1], Ernesto Caffarena[1]

[1]Fiocruz, Brazil

Keywords: Leprosy, Molecular Dynamics, Toll Like Receptors, SNP

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 140

A new architecture for BioAgents annotation tool

Daniel da Silva Souza[1], João Paulo Ataíde Martins[1], Maria Emília Machado Telles Walter[2], Célia Ghedini Ralha[2], Roberto Coiti Togawa[3], Natalia Florencio Martins[3], Tainá Raiol Alencar[2]

[1]IESB, Brazil, [2]University of Brasília, Brazil, [3]Embrapa/Cenargen - Genetic Resources and Biotechnology, Brazil

Identifying functions and biological characteristics is a central task in genome sequencing projects. In bioinformatics workflows, this task is developed in the annotation phase, basically using sequence comparison algorithms together with the biologists' knowledge to infer functions of the sequences of a studied organism, from similar sequences that already had their functions determined. This work aims to propose a new architecture to BioAgents, a tool to annotate biological sequences using production rules to infer the biologists' reasoning simulated in a multiagent environment. Experiments were done with real data from the Paracoccidioides brasiliensis project to evaluate the new architecture's performance.

Keywords: annotation, biologists' reasoning, multiagent environment, new architecture

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 142

Mutation Landscape in the Surfaceome of 23 Colorectal Cancer Cell Lines

Elisa Donnard[1], Paula Asprino[1], Fernanda Koyama[1], Fabiana Bettoni[1], Sandro de Souza[2], John Mariadason[3], Raphael Parmigianni[1], Pedro Galante[1], Anamaria Camargo[3]

[1]Molecular Oncology Center HSL, Brazil; [2]Universidade Federal do Rio Grande do Norte, Brazil, [3]Ludwig Institute for Cancer Research, Australia

Background: Colorectal cancer cells contain a large number of non-synonymous somatic mutations ranging from hundreds (microsatellite stable) to thousands (microsatellite unstable). These non-synonymous mutations can offer new therapeutic targets for known drugs and can also generate immunogenic epitopes, which can be used as vaccines for tumor control. The subset of genes coding for cell surface proteins (surfaceome) is a highly interesting target for the analysis of non-synonymous mutations and has not been properly explored to this date. Results: Using 23 human colorectal cancer cell lines, we captured and sequenced the portion of the exome corresponding to 3,657 genes coding for transmembrane proteins located in the cell surface. For each cell line we called SNVs (single nucleotide variations) and used a recurrence analysis together with the dbSNP database to remove common polymorphisms. The remaining non-synonymous SNVs were submitted to several algorithms to assess their functional impact such as Mutation Assessor, SIFT and POLYPHEN. This analysis allows classification of the SNVs as deleterious or possible activating mutations. Genes with activating mutations were searched against known drug target databases (e.g., dGene and DrugBank) to identify possible therapeutic targets (druggable SNVs). The protein fragment containing each non-synonymous SNV was used as input to RANKPEP and NetMHC to identify new immunogenic epitopes present in these cancer cell lines. Surfaceome SNVs may also alter phosphorylation sites, using a known kinase-substrate database (PhosphoNetworks.org) we evaluated the mutated genes present in these and if the mutation occurred in the phosphorylation site described. Lastly we analyzed the SNV position relative to the cell membrane, seeing as juxtamembrane mutations have been implicated in tumor aggressiveness. Conclusions: Our results provide an interesting insight of the mutation landscape in these different cell lines, including disruption of phosphorylation sites and clusters of juxtamembrane mutations. We identified a number of new possible therapeutic targets for known drugs and a set of immunogenic epitopes that can be explored as vaccines. Supported by: FAPESP

Keywords: Colorectal, SNV, Surfaceome

Topic: Transcriptomics and Proteomics

Abstract #: 143

Differential expression analysis of the transcriptome Piper nigrum L. infected by Fusarium solani f. sp. piperis

Edith Moreira[1], Sheila Gordo[1], Daniel Pinheiro[2], Simone Rodrigues[3], Jersey Maués[1], Iracilda Sampaio[1], Sylvain Darnet[1]

[1]UFPA, Brazil, [2]Universidade de São Paulo, Ribeirão Preto, SP, Brazil, Brazil, [3]EMBRAPA, Brazil

Black pepper is one spice consumed worldwide which have great economic potential. Brazil is the most producer of this culture, and the state of Pará, the largest domestic producer. The specie is vulnerable to attack by the pathogen Fusarium solani f. sp. Piperis which causes a disease known as root rot. For years the control of this disease has been investigated, but so far little is known about the plant response to pathogen attack. Understanding the molecular mechanisms involved in this response may represent a direction for the current breeding programs of this species. In this paper we use the RNA-seq technique for analyzing differentially expressed genes after infection of P. nigrum by Fusarium solani. Were produced more than 67 million short reads in SOLiD V-3 sequencing platform. De novo assembly was done using the program Velvet and the package OASES. For optimization of the assembly, we use the method "additive multiple-k" and STM-. Reads filtered of Piper nigrum L. infected were attached to the reads preprocessed of experiment of Piper nigrum L. published by Gordo et al., 2012 and used as input for assembly. The contigs obtained were filtered with seqclean and sequences shorter than 100bp were excluded. To cluster fragmented contigs and get Unigene sequences was used iAssembler. We obtained 16941 Unigenes (2.8 Mbp) used as reference in expression analysis. The raw reads of P. nigrum Infected (I Pn) and P. nigrum without infection (Pni) were aligned against the reference obtained using the program Bowtie. The result of the alignment was used as input for the package Cufflinks. The package Degseq was used to analyze differentially expressed genes. To make the automatic functional annotation of differentially expressed transcripts in the categories of Biological Process, Molecular Function and Cellular Component we use Blast2GO. The analysis of differentially expressed transcripts showed several genes of interest produced in response to infection, including genes in response to stress, genes involved in oxidative processes, signal transduction, transcription factors and resistance genes, opening new applicability and perspectives in programs breeding for Piper nigrum.

Keywords: Piper nigrum L., Fusarium solani, SOLiD, RNA-seq, Transcriptome, differential expression

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Transcriptomics and Proteomics,Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 144

Genotype-Environment Interactions Reveal Causal Pathways That Mediate Genetic Effects on Phenotype

Julien Gagneur[1], Oliver Stegle[2], Chenchen Zhu[3], Petra Jakob[3], Manu Tekkedil[3], Raeka Aiyar[3], Ann-Kathrin Schuon[3], Dana Pe'er[4], Lars Steinmetz[3]

[1]LMU, Munich, Germany, [2]EBI, Cambridge, United Kingdom, [3]EMBL, Heidelberg, Germany, [4]Columbia University, NY, United States

Unraveling the molecular processes that lead from genotype to phenotype is crucial for the understanding and effective treatment of genetic diseases. Knowledge of the causative genetic defect most often does not enable treatment; therefore, causal intermediates between genotype and phenotype constitute valuable candidates for molecular intervention points that can be therapeutically targeted. Mapping genetic determinants of gene expression levels (also known as expression quantitative trait loci or eQTL studies) is frequently used for this purpose, yet distinguishing causation from correlation remains a significant challenge. Here, we address this challenge using extensive, multi-environment gene expression and fitness profiling of hundreds of genetically diverse yeast strains, in order to identify truly causal intermediate genes that condition fitness in a given environment. Using functional genomics assays, we show that the predictive power of eQTL studies for inferring causal intermediate genes is poor unless performed across multiple environments. Surprisingly, although the effects of genotype on fitness depended strongly on environment, causal intermediates could be most reliably predicted from genetic effects on expression present in all environments. Our results indicate a mechanism explaining this apparent paradox, whereby immediate molecular consequences of genetic variation are shared across environments, and environment-dependent phenotypic effects result from downstream integration of environmental signals. We developed a statistical model to predict causal intermediates that leverages this insight, yielding over 400 transcripts, for the majority of which we experimentally validated their role in conditioning fitness. Our findings have implications for the design and analysis of clinical omics studies aimed at discovering personalized targets for molecular intervention, suggesting that inferring causation in a single cellular context can benefit from molecular profiling in multiple contexts.

Keywords: QTL, eQTL, causal inference, systems genetics

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 147

Genome Assembly of a seasonal oscine, the Sabiá-laranjeira (Turdus rufiventris)

Pablo Gomes de Sá[1], Rommel Ramos[1], Francisco Prosdocimi[2], Nicholas Lima[2], Luiz Gonzaga[3], Marcus Teixeira Teixeira[4], Carla Fontana[5], Maria Sueli Felipe[4], Ana Vasconcelos[3], Claudio Mello[6], Artur Silva[1], Maria Paula Schneider[1]

[1]UFPA, Brazil, [2]UFRJ, Brazil, [3]LNCC, Brazil, [4]UnB, Brazil, [5]PUCRS, Brazil, [6]Oregon Health & Science University, United States

Among the known bird species, oscines are one of the few groups that produce complex vocalizations due to vocal learning. This cognitive feature is associated with a high degree of encephalization and contributes to speciation and the generation of biodiversity. A more in-depth knowledge of the genetics related to vocal learning will be essential to better understand the biological basis of this trait. The sequencing of the Sabiá-laranjeira (Turdus rufiventris) genome is of particular interest for better understanding the mechanisms and variants associated with vocal learning, As a representative oscine, it evolved vocal learning and associated brain circuits. Furthermore, it is a seasonal singer, thus providing an important biological contrast to the zebra finch (Taeniopygia guttata), a non-seasonal oscine whose genome had been previously sequenced. The genome of a male Turdus rufiventris was sequenced by SOLiD 5500xl platform with one paired-end library and five mate-paired libraries with insert sizes of 2, 3, 5, 10 and 20 kb. The 3' region of the reads was trimmed due to the low Phred quality, and the reads with average quality below Phred 20 were removed. After this filtering step, a sequencing coverage of about 42x was achieved. The assembly was performed with SOAPdenovo2, using the computational structure at the National Laboratory for Scientific Computing (LNCC), with k-mer value of 31 for all genomic libraries. The amount of contigs contain ~766 mb, representing 64% of the estimated genome size of the zebra finch (~1195 mb). As a next step, more sequencing will be produced and incorporated in the draft genome to improve the quality of the assembly.

Keywords: Genome, De novo Assembly, Turdus rufiventris

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 148

Stingray@Galaxy: a pipeline for NGS data functional annotation in high performance computational environment

Rodrigo Jardim[1], Diogo A. Tschoeke[1], Alberto MR Dávila[1]

[1]Fiocruz, Brazil

Background: Large amounts of data are generated by high-throughput technologies as Next Generation Sequencing. Many types of users, mainly groups centered on functional annotation and re-annotation of genomes, deal with this kind of data. Usually these activities are structured in tasks: data preparation, data preprocessing, assembly, data analysis and functional annotation/re-annotation. Given the large volume of data generated, such tasks require huge computational processing power. In addition, at each task it is necessary to use several bioinformatics software that sometimes need to be installed or even compiled, consuming researches time that would prefer to be centered on data analysis. In order to facilitate the use of bioinformatics tools, the Galaxy project has been developed by the community as a web-based platform for running and displaying data from different data sources. By developing the Stingray@Galaxy system, we have incorporated into Galaxy, tools for genome assembly as MIRA, VELVET and ABySS, as well as the Stingray functional annotation pipeline (briefly: similarity analysis by Blast and RPSBlast as well as OrthoMCL-based annotation), providing useful tools for genome assembly, viewing, analysis and functional annotation of NGS data. Results: Several genome assembler and analysis tools available in Stingray were incorporated in our local Galaxy system as the necessary steps for functional annotation/re-annotation of genomes. Some workflows were created in Stingray@Galaxy to facilitate the processing and analysis of genomicsdata. A computing environment (Cloud and Cluster computing) for high performance was made available for the execution of the bioinformatics software available at Stingray@Galaxy. Conclusions: The Galaxy tool provided the possibility for the researchers to concentrate on their main task: data analysis. Stingray@Galaxy showed to be a complete tool for processing, visualization and functional annotation/re-annotation data of partial or complete genomes in a high performance computing environment.

Keywords: ngs, functional annotation, stingray, galaxy

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 149

Identification of small polymorphisms within human exons: a preliminary analysis

Gabriel Wajnberg[1], Carlos Gil Ferreira[2], Fabio Passetti[1]

[1]Bioinformatics Unit, Clinical Research Coordination, Instituto Nacional de Câncer (INCA)., Brazil
[2]Clinical Research Coordination, Instituto Nacional de Câncer(INCA), Brazil

Background: Alterations in the DNA sequence can lead to mutations such as single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs). INDELs can occur by changing one to thousands of nucleotides in length. If the aforementioned alterations are located within the coding region of the genes, INDELs can produce transcripts that can modify splicing sites, the encoded amino acid or cause frame shift. The 1000 genomes project sequenced 1,092 human genomes in different populations which permited to search for genomic alterations (such as SNPs and INDELs) not only related to the reference sequence genome for the human species. Previous publications have described INDELs up to 100 nucleotides in length in DNA samples from 24 humans using microarrays and only 24% out of 2,000,000 known INDELs described in dbSNP have been identified by the 1000 genomes project. Our group conducted a preliminary prospection using data from the 1000 genomes project and the dbSNP. For many years, only large structural variants have gained attention like, for example, the amplification and loss of large chromosomal regions (starting from 10kb). Our aim was to identify small polymorphisms up to 100 nucleotides in length in different populations and associate this type of polymorphism with cancer. Results: We analyzed 38,219,238 polymorphisms (36,820,992 SNPs and 1,398,246 INDELs) and 47,252 small deletions with a size between 7 and 99 bp (5.63% of total deletions) from the 1,092 genomes analyzed. The Luhya population from Kenya (LWK) has the largest quantity of total polymorphisms (4,589,023 per genome). Furthermore, it has the largest quantity of small deletions with a size between 7 and 99 bp per genome, from a total of 97 genomes analyzed (13,486 per genome). Most of these deletions (10,016) have the possibility to change the coding frame. However, the Han Chinese population from Beijing (CHB) has the smallest quantity of total polymorphisms (3,728,220 per genome), from a total of 97 analyzed genomes. In contrast, the Iberian population from Spain (IBS) has the smallest quantity of this type of small deletions per genome, from a total of 14 genomes analyzed (11,301 per genome). 8,378 of them can cause a frame shift if it occurs within an exon. Simultaneously, we identified 51 small deletions (less than 100 bp in length) that occur in exons using the dbSNP. We identified three small deletions with a size non-multiple of three that occur in the central regions of the BRCA1 gene and in the gene MUC4, genes linked to cancer. Conclusion: Our preliminary analysis detected a group of small polymorphisms from the 1000 genome project and dbSNP. These small deletions in exonic regions may alter functional sites of proteins leading to loss of function and can be associated with cancer.

Keywords: small, polymorphisms, exons, snps, indels, cancer, genome

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 150

Study of the calcium binding site and its interaction in cysteine protease using QM / MM

Ronaldo Correia da Silva[1], Adonis de Melo Lima[1], Jéssica Tuany Perreira Almeida[2], Maryene Wilde Alves Melo[3], Vanderson de Souza Sampaio[4]

[1]Universidade Federal do Pará, Brazil, [2]Faculdade Integrada Brasil Amazônia, Brazil, [3]Faculdade Integrada Brasil Amazônia, Brazil, [4]Fundação de Medicina Tropical Doutor Heitor Vieira Dourado, Brazil

Introduction. Calpain is a cysteine protease activated by intracellular calcium (Ca2+) with known involvement in diseases such as cancer, heart attack and stroke. Hyper activation of the enzyme is able to initiate a series of destructive cycles that can cause irreversible damage to the cells. An increase in the concentration of mitochondrial Ca2+ under different pathophysiological conditions is capable of initiating a series of destructive cycles which can cause irreversible damage to cells. Methods. Homology modeling can be used to build the terciary structure of a protein based on the primary structure. The QM/MM approach and classical MD simulations have become the method of choice for modeling reactions and interactions in biomolecular systems. In addition we have investigated the calpain interactions with calcium ions using hybrid QM/MM and molecular dynamics simulations. Results. The 3D structure selected as a template for constructing the model of calpain-10 was μ-calpain from PDB access code 2G8E. It has 36.73% identity and 50% similarity with the target sequence. The model was validated considering the stereochemical quality in the MolProbity server. During the simulations, the ion affinity was compared with the residues of the template and the target. All selected residues belong to known structures, such as calcium binding loops. The results show that the template has a larger number of residues with an acidic character, keeping distances of up to 2.67 Å during the simulations. On the target, the residues that showed the highest affinity for the ion were Asp77, Glu286, Glu308 and Glu310. These results were expected, as evident differences were observed in the physico-chemical properties of the residues, suggesting that mutations occurring in acidic residues could contribute to the distinct functions of these molecules in different organisms. Conclusions. The human calpain-10 enzyme model obtained through homology modeling suggests that the active site of this enzyme is conserved and the main interactions are similar to those observed in μ-calpain. The amino acid mutations observed in the calcium binding site located in the loop in Domain I are in accordance with the results of other authors that have proposed that calpain-10 is probably less dependent on calcium than the typical calpains. We hope these results may be useful for the design of new inhibitors of calpain-10.

Keywords: cisteine protease, homology, QM/MM

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 151

CMRegNet: A database of transcriptional regulatory network

Vinicius de Abreu[1], Sintia Almeida[1], Diego Mariano[1], Leiticia Castro[1], Lucas Amorim[1], Carlos Diniz[1], Siomar Soeres[1], Syed Hassan[1], Jan Baumbach[2], Vasco Azevedo[1]

[1]Universidade Federal de Minas Gerais, Brazil, [2]University of Southern Denmark, Denmark

Microorganisms use a number of mechanisms to handle the changing environmental conditions for maintaining their functional homeostasis and to overcome the stressful situations. Such mechanisms use molecular strategies coordinated for transcriptional regulation networks – TRN, to manage these unfavorable conditions. The complexity of the regulatory networks results from the basic interaction among RNA polymerase, sigma factors, promoter regions, transcription factors - TF, transcription factors binding sites - TFBS and a set of target genes regulated. A more profound study of these transcriptional regulation networks and its main characters may lead to a greater understanding of organisms according their cellular behavior. Additionally, the increasing number of the complete genomes of prokaryotes, enables the scientific community to apply extensively the computational comparative genomic approaches to predict the transcriptional regulatory networks and elements like TFBS, in bacteria. The main characteristics of this study involve analyses and descriptions of previously known regulators in model organisms, to organisms not known yet and ab initio prediction of new regulators. However, even for model organisms, we are still far from having the first complete regulatory network, by because, we could not simulate the various environments (as the cells live) in the laboratory or because of the noise from the existing techniques, the fact is that model organisms as E. coli that has the largest number of publications with experimentally validated data when compared to any other organism, we found only about a third of its TRN defined (2). In some cases such as Mycobacterium leprae that has the longest duplication timeoubling of all known bacteria and also does not grow in artificial culture medium, we are far from having a defined TRN. Thinking about it, the aim of this work is to solve some of the challenges found in the process of reconstruction and transfer of the transcriptional regulatory network among multiple species, using two bacterial species as model organisms belonging to the CMNR group. These are Corynebacterium glutamicum ATCC 1303 and Mycobacterium tuberculosis H37Rv, where posteriorly, the network data will be transferred to other related organisms from the same genus or species, using the approach of two models with different lifestyles, with the clear aim of improving the quality of the predictions. For this purpose, an ontological integrated database in online platform called CMRegNet, was developed which provides a friendly interface for accessing the contents of the database, enabling various types of queries, support reconstruction, analysis and visualization of regulatory networks in different hierarchical levels.

Keywords: CMRegNet, transcriptional regulatory network, TFBS, Mycobacterium, Corynebacterium

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 152

Theoretical Study of Monoterpenes by Dynamic Molecular and Free Energy in the Enzyme 3-hydroxy-3-methyl-glutaryl-CoA Reductase.

Carlyle Lima[1], Nelson Alencar[1], Isaque Medeiros[1], Barbara Silva[1], Davi Brasil[1], Claudio Nahum[1]

[1]Universidade Federal do Pará, Brazil

Theoretical Study of Monoterpenes by Dynamic Molecular and Free Energy in the Enzyme 3-hydroxy-3-methyl-glutaryl-CoA Reductase. CR Lima[1], NAN de Alencar[1], I.G. Medeiros[1], B. C. P Silva[1], DSB Brasil[1], CN Alves[1] [1] Universidade Federal do Pará The enzyme 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) reductase is distributed in organs and organic metabolic processes, it's main function is the production of cholesterol by the organism. Now days the investigation of new drug is in focus to the nature, mainly the monoterpenes, they present a high affinity with this enzyme. The research was based on the action of four monoterpenes described in the literature for having a strong action against the enzyme target of the study: D-Limonene, Perillic Alcohol, Perillic Aldeid and Perillic Acid. Here, we using molecular dynamics (MD) and free energy to investigate the interactions most important between HMG-CoA reductase (1DQA, PDB code) and four monoterpenes. The MD was performed by a time of 3 ns with hybrid QM/MM method, where Co-enzyme and inhibitor were treated with the semiempirical AM1 Hamiltonian, while the rest of the system was described using the combination of the OPLS-AA and TIP3P. Because the active site exhibit preferential binding positions at both the ends and the monoterpenes present favorable binding regions at one end only, in two different positions: S (bond on the end of active site) and C (region to link to co-enzyme), trying to understand what were the amino acids more favorable to the bond. The best inhibitors were detached and subjected to calculations of MD after simulation showed greater stability of monoterpenes in position S with the Glu559 O residue deposited near Co-enzyme, performing favorable reception of electrons by the tested oil with an average distance 2.5 Å distance and Lys735 that is favorable electronic exchange, with a value of the distance of 1.97 Å, this stability is achieved similarly to the inhibitor complexed in crystallography of the enzyme. The free energy values are consistent with those of molecular dynamics to the way in which the most promising complex, proves more stable suggesting that monoterpene containing the aldeid functional group, holds the most promise as an anti-HMG. The results indicate a good behavior of these structures active site.

Keywords: HMG-COA, Dockin

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 153

THE APPLICATION OF AN ENTROPY BASED APPROACH TO ASSESS THE PHYLOGENETIC RELATIONSHIP OF ISOLATES FROM GROUP A HUMAN ROTAVIRUS

Debora Barreto[1], Marcus Batista[1]

[1]Federal University of Sergipe, Brazil

The Group A Human Rotavirus has been the main cause of gastroentenrite in children. It is responsible for 611,000 deaths per year worldwide. The viral particle has no envelope and it presents an icosahedral symmetry. The viral genome consists of 11 genomic segments of double-stranded RNA that encode 12 proteins, six are structural (VP1, VP2, VP3, VP4, VP6, and VP7), and the other six are nonstructural (NSP1, NSP2, NSP3, NSP4, NSP5, and NSP6). In 2006, a vaccine called Rotarix was introduced in the vaccination card and it decreased the mortality of children with diarrhea. However, previously low frequent genotypes are becoming more frequent, after the vaccine implementation. So, the understanding of these viruses evolution is crucial to evaluate the effectiveness of established vaccines, and/or to assist in the formulation of new treatment methods. Nevertheless, the complete evolutionary scenario of the Rotavirus diversification is not elucidated. In this context, new computational methods should be developed to assist the phylogenetic analyzes. Therefore, the aim of this study was to make use of an entropy based method to select phylogenetic informative regions of the genome of several isolates from Group A Human Rotavirus in order to infer their evolutionary relationships. The genetic variability of all genes was assessed and it was found that VP4, VP7, and NSP1 are the most variable genes. VP4 had 670 conserved and 1808 variable sites; VP7 presented 129 conserved and 852 variable sites; and NSP1 had 1331 conserved and 538 variable sites. A qualitative analysis was carried out and it was observed that VP1 has a conserved domain that is important to form an enzyme complex; VP2 has two domains that present variation and this region contains leucine, which interacts with a target in the host DNA; VP3 has a conserved domain and a variable motif that are important for RNA replication; VP4, VP6 and VP7 have high genetic variability. NSP1 has a variable motif, which also appears to function directly linked to zinc; NSP2, NSP3 and NSP4 presented motifs that were conserved among the isolates. In the phylogenetic analysis of VP4 gene, we observed that the isolates cluster themselves accordingly to P genotypes, P [4] was the most frequent, and new genotypes have emerged over the years. We compare the analysis of VP4 with the entropy based method, and it was observed that there were no major differences in topology and bootstrapping values. However, the entropy based method could achieve this phylogenetic tree in a very fast way. So, novel methods and approaches should be developed in order to make large phylogenetic inferences available. Supported by: COPES/POSGRAP/UFS.

Keywords: Rotavirus; Entropy; Phylogenetic analysis

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 154

Molecular Docking Study on Enzyme Tyrosinase Inhibitors with Seven: Prospects for rational design of drugs

Carlyle Lima[1], Nelson Alencar[1], Sarah Furtado[1], Barbara Silva[1], Davi Brasil[1], Claudio Alves[1]

[1]Universidade Federal do Pará, Brazil

Tyrosinase is an enzyme widely distributed among living organisms, and is involved in the production of melanin in addition to performing other multiple catalytic functions. In mammals, the enzyme responsible for the conversion of tyrosinase 3,4 dihydroxyphenylalanine (DOPA) and then DOPA to quinone. Metalloprotein is, the binding site is quite typical in all living organisms, consisting of two coppers driven by three histidine amino acids each. Tyrosinase can also induce neurotoxicity apoptotic stress pathways, which relate to serious skin diseases such as albinism cancer, melanoma and skin darkening. Most inhibitors current is derived from phenol or structures with chelating properties. In this work we propose to investigate the best fit conformation based on the binding affinity of the enzyme-ligand complex. In the present study, we have selected seven inhibitors of literature: Hesperitin (HS), thiobarbituric acid (AB), terephthalic acid (TA), isophthalate (AI), arabinose (AR), phthalic acid (PA), rutin (RT) and more two as control for comparison: kojic acid (AK) and Tropolone (TP), in order to discover which conformation of the enzyme-inhibitor complex is has a higher affinity. To perform this investigation we used the program Molegro (MVD), which has function score based on genetic algorithm Larmakiano. The crystallography of this enzyme is deposited in the Protein Data Banck: 2Y9X, derived from Agaricus bisporus in deoxygenated form. Generated a total of 20 structures RMSD deviation of less than 1Å, to select only the best structure of each of inhibitors based on scores. Thus it was found that all were docados inhibitors in the active site cavity and 2.7 Å average distance of the ions which make up the copper binding site. All structures analyzed maintained interaction with the majority MET280, maintaining the hydrogen bond at a distance of 2.5 Å away waste deposited near the entrance of the active site, well with ASN 260, which remained mostly hydrogen bonds at a relatively greater distance about 2.8 Å away. Only inhibitors HS, AT, RT, showed values of scores above the control, since in its present structure portions of hydroxyls By the presented results, we can infer that the program can generate MVD suitable geometry for enzyme-inhibitor complexes with higher accuracy, since it represents a more forceful affinities that even with tyrosinase.

Keywords: Doking, Tyrosinase, Drugs

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 155

Draft genome isolated from Corynebacterium ulcerans FRC58 secretion of bronchitis in a patient Troyes (France).

Andréia Silva[1], Rommel Ramos[1], Rafael Baraúna[1], Diego Graças[1], Adriana Carneiro[1], Allan Veras[1], Pablo Sá[1], Maxime Thouvenin[2], Vasco Azevedo[3], Edgar Badel[4], Nicole Guiso[4], Renato Costa[1], Artur Silva[1]

[1]UFPA, Brazil, [2]Centre Hospitalier de Troyes, France, [3]UFMG, Brazil, [4]Intituto Pasteur, France

Background: Corynebacterium ulcerans is a catalase-positive toxigenic bacteria and nitrate-negative, pertaining to the group of CMNR. This species can cause various clinical manifestations in both humans (association with milk consumption barely boiled or pasteurized or direct contact with infected animals) and in domestic or wild animals (cattle, cats, dogs and other). It is an emerging pathogen that has been isolated from infectious processes in various countries such as Brazil, Japan, Germany and England. The signs and symptoms observed very often resembles the framework of classical diphtheria. The C. ulcerans contains genes encoding phospholipase D (PLD) and diphtheria toxin (DT), which are the major virulence factors also produced by Corynebacterium diphtheriae and Corynebacterium psedutoberculosis respectively

Keywords:

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 156

The comparative genomics of Protozoa: identification and categorization of core, group and species-specific genes

Diogo A. Tschoeke[1], Rodrigo jardim[1], Sergio Serra da Cruz[2], Maria Luiza Machado Campos[3], Marta Mattoso[4], Alberto M.R. Dávila[1]

[1]Fiocruz, Brazil, [2]PPGMMC/UFRRJ – Federal Rural University of Rio de Janeiro, Brazil, [3]PPGI/UFRJ – Federal University of Rio de Janeiro, Brazil, [4]COPPE/UFRJ – Federal University of Rio de Janeiro, Brazil

Background: Protozoa are defined as single celled eukaryotic organisms showing an extremely diversity and variety. Approximately 200,000 species are described and nearly 10,000 are parasitic. The pathogenic species cause diseases such as malaria, sleeping sickness, Chagas disease, leishmaniasis, amoebiasis and giardiasis. Comparative studies among Protozoa are important because they may identify similarities and differences in these species. Identification of orthologs is central to functional characterization of genomes because orthologs typically occupy the same functional niche in different organisms. Thus, we can infer what genes are shared and the ones that are specific to each organism increasing our understanding on the biology of species. Results: 204,624 non-redundant proteins from Plasmodium, Entamoeba, Trypanosoma, Leishmania, Giardia, Theileria, Toxoplasma, Trichomonas and Cryptosporidium, totalizing 22 species, were submitted to OrthoMCL resulting in 26,101 homologs groups. Among them, 21,119 groups are orthologs. Among them, 348 orthologs are shared by all 22 species, representing the Protozoa core proteome (PCP). When those 348 orthologs are categorized using KOG/NCBI, 72 orthologs (20.7%) belong to the functional category “J”, followed by “O” (61 orthologous groups, 15.53%), and the “R” category, with 45 groups (12.9%). Similar analyses, using the prokaryotes ortholog groups (COG/NCBI) were performed, showing the “J” category as the most abundant with 103 groups (29.6%), followed by the "R" and “L” category. Most of these orthologs are related to the maintenance of the cell and information processing. Among the 348 PCP, 79% (275) showed similarity with COG. We observed what PCP are closer to Archaea or Eubacteria and the “J” COG category was closer to Archaea than Eubacteria, and these genes are involved in the information process such as translation and DNA maintenance. While the COG categories “G” and “T” showed to be closer to Eubacteria regarding proteins with operational function, like metabolic enzymes. The Kinetoplastid Core Proteome (KCP) inferred has 5,000 orthologous groups and 3,396 (67.92%) are Kinetoplastid Specific Proteins (KSP). Functional categorization of KSP, using KOG/NCBI, demonstrated that “R” category is the most abundant, besides 46.29% (1,592/3,396) of these orthologs are annotated as “hypothetical”. Apicomplexa Core Proteome (ACP) has 986 orthologous groups and 224 (27.82%) are Apicomplexa Specific Proteins (ASP). The most abundant functional category in ASP was “R”, whereas 40.63% (92/224) were classified as hypothetical proteins. Entamoeba Core Proteome (ECP) has 5,915 orthologous groups and 4,441 (75.08%) groups, are Entamoeba Specific Proteins (ESP). The functional category “R” was the most abundant and 2,905 (65.41%) ESP were annotated as hypothetical. Conclusions: In PCP most of the orthologs are related to cell maintenance, among them a set are closer to Eubacteria and other to Archaea. Among the specific orthologs groups, most of the orthologs have hypothetical functions, however functions like ABC-transporter, RNA-helicase and Rab-GTPase could be also found.

Keywords: protozoa, comparative genomics

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 158

On the particularities of Protozoa: inferring family expansions and orphans proteins

Diogo Tschoeke[1], Rodrigo Jardim[1], Alberto M.R. Dávila[1]

[1]Fiocruz, Brazil

Background: Protozoa is the common name given to unicellular Eukaryotes and comprises about 200.000 species, showing an extremely diversity and variety. Most species are free-living and nearly 10.000 are parasitic. The pathogenic species cause diseases such as: amoebiasis, Chagas disease, giardiasis, malaria, leishmaniasis and sleeping sickness. Comparative studies among Protozoa can help to identify similarities and differences at the genomic level. Identification of paralogs is important to characterization of genomes because paralogs undergo a functional diversification by duplication, via the processes of neofunctionalization and subfunctionalization; Results: 204,624 non-redundant proteins from Plasmodium, Entamoeba, Trypanosoma, Leishmania, Giardia, Theileria, Toxoplasma, Trichomonas and Cryptosporidium, totaling 22 species, were submitted to OrthoMCL resulting in 26,101 homologs groups. Among them 4,982 are inparalogs and 7,679 co-orthologs (groups that contain recent paralogs). The Protozoa species that presented the highest number of inparalogs groups, was Trichomonas vaginalis, with 2,933 groups. The functional categorization shows that the most abundant KOG category was the "T" category. T. vaginalis also showed 948 co-orthologs totalizing 3881 paralogs. However Trypanosoma cruzi showed the highest duplication number, totalizing 5777 paralogs, 4963 co-orthologs and 814 inparalogs groups, and the most abundant KOG category in inparalogs was "T". When we look only for the larger expansions, that is, those families with at least 80 members we found the following: B. bovis had a variant erythrocyte surface antigen-1. G. lamblia shows Kinase, NEK and Variant-specific surface protein. P. chabaudi: Pc-fam-2 protein. P. falciparum: rifin and var. P. knowlesi: SICA-like antigen. P. vivax: variable surface protein Vir24. P. yoelii: hypothetical protein. T. brucei: expression site associated gene (ESAG) 4 protein and variant surface glycoprotein. T. cruzi: trans-sialidase, mucin TcMUCII, dispersed gene family protein 1 and retrotransposon hot spot protein. Finally T. vaginalis genome showed: ankyrin repeat protein, hypothetical protein TVAG_289600, TVAG_580790 and TVAG_235700, some of them with more than 1000 copies. We also found a number of orphans proteins (those proteins that showed no similarity within the cut-off values to be clustered by OrthoMCL). The species that showed the highest number of orphan proteins were: Plasmodium chabaudi : 6961, followed by Trichomonas vaginalis that, despite having almost 60,000 proteins, showed 4309 orphan proteins. It is important to notice that approximately 99% of orphan proteins in 22 species are described/annotated as hypothetical proteins or unknown function; Conclusions: T. vaginalis and T. cruzi showed higher expansions, considering the number of families and its size. Signal transduction mechanisms was the most abundant category in inparalogs, and proteins related to membrane have usually the largest expansions. Almost 50% of Plasmodium chabaudi proteins were classified as orphans.

Keywords: comparative genomics, protozoa, paralogs

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Transcriptomics and Proteomics

Abstract #: 159

B4D - Blast2GO for Dummies

Adonney Veras[1], Pablo Gomes de Sá[1], Rafael Baraúna[1], Adriana Carneiro[1], Diego Graças[1], Diogo Almeida[1], Vasco Azevedo[2], Artur Silva[1], Rommel Ramos[1]

[1]UFPA, Brazil, [2]UFMG, Brazil

The next-generation sequencer (NGS) increase the amount of genomic data with lower cost. Beyond the advantages about the structural genomics, other studies have been developed due the NGS like the funcional genomics (transcriptomics and proteomics), which allow us understand the function and the behavior of an organism under some condition, an important resource to develop biological and medical products based on these. However, scripts and/or computational tools are required to handle and analyze these data due to the number of the biological database which can be integrated to improve the results, like can be observed for Corynebacterium pseudotuberculosis, a pathogenic bacteria which causes high economic losses around the world and have been used to structural and functional genomics. About the functional analysis which can be applied over transcriptomics experiments we can cite the gene ontology through the Blast2GO, that is important by classify the genes regarding the biological process, cellular component and molecular function. But there are some limitations to perform the access and handle the results. Therefore, we developed the B4D (Blast2GO for Dummies) which allows insert the blast2go results in the database to cluster them and produced reports based on filters defined by the user. The tool can identify the products shared by two or more groups of sequences (core) as well the exclusives (accessories) through the product assigned by Blast2GO, a valuable resource when transcriptomics studies based on RNA-Seq denovo approaches are conducted. For C. pseudotuberculosis the RNA-Seq data of the strains 1002 and 258 under 4 different conditions (50oC, 2M, pH and control) were submitted to assembly by denovo methods and the transcripts loaded in the Blast2GO: after blast, mapping and annotation steps the results were exported and used as input for B4D which returned 47 exclusive products for 258 and 139 for 1002. These information can help understand the main pathogenic factors associated with each bacteria strain further on that are shared by them.

Keywords: Transcriptomics, Corynebacterium pseudotuberculosis, Blast2GO, De novo

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Transcriptomics and Proteomics

Abstract #: 160

THERE IS CODON USAGE BIAS AMONGST E. coli GENES

Lucas Ferreira[1], José Miguel Ortega[1]

[1]UFMG, Brazil

Background. Codon usage bias is a largely studied phenomenon, and several reasons for its existence have been raised, including the control of the velocity of translation, since frequent codons seem to reflect a large concentration of available tRNA capable of binding to them and therefore fast translation rate. We have populated a local database of E. coli operons integrating data of Sigma Factors which control them, and the origin of the genes therein, e.g. if a gene is shared with archaea, the lowest common ancestor (LCA) is named cellular organisms, but some genes are shared only within the order Enterobacteriales. Therefore, after grouping genes by Sigma Factor and LCA we set up to investigate the codon usage bias, suspecting that some class of genes could make use of alternative tRNA molecules. Results. A remarkable difference appeared when comparing Sigma 24 and Sigma 19 controlled genes with the global codon usage pattern or with the other Sigma Factors controlled genes, including the housekeeping Sigma 70. Sigma 24 and Sigma 19 are implicated with heat-shock signals. Focusing specifically in one amino acid, Arginine, in Sigma 70 controlled genes there is a strong preference for CGU (43%) and CGC (41%), but in genes controlled by Sigma 24 the usage for AGA and AGG raises from 1% and 2%, respectively, to 6% and 9%, and, for CGA and CGG, raises from 4% and 7% to 17% and 15%, respectively. This result is compatible by a tRNA switch from GCG to the UCU anticodon. Selecting sub-groups of the Sigma 24 and Sigma 19, some showed more bias. Therefore we classified the Sigma 24 regulated genes which are present in UEKO database (286 genes) according to their LCA (we did not analyzed this for Sigma 19 since only 5 genes are in UEKO). Remarkably, the codon usage bias is observed in genes shared with archaea (LCA: Cellular Organisms). Genes controlled by Sigma 24 that are originated in bacteria or other more recent E. coli clades (e;.g.: restricted to the phylum proteobacteria, the class gammaproteobacteria, etc.) did not show a bias. This suggests that the E. coli heat-shock genes comprise genes shared with archaea, which are adapted for a different codon usage bias and putatively use different tRNAs. Conclusions. Our work approached a hitherto not investigated question and has shown that there is a codon usage bias amongst E. coli genes. Remarkably, Sigma Factors which control heat-shock (Sigma 24 and Sigma 19) prefer to use for carrying Arginine the tRNA with UCU anticodon, opposite to GCG, preferred by Sigma 70 and other Sigma Factors controlled genes. We also verified that the strongest bias was concentrated on genes shared with archaea. Thus, there is codon usage bias amongst genes in, E. coli and it is implicated with gene expression regulation and with the époque of origin of these genes.

Keywords: E. coli; sigma factors; Sigma24; Sigma70; Codon Usage;

Topic: Signaling and Metabolic Networking; Ontologies; Systems Biology,Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 161

EVOLVING THE VIRTUAL MUSEUM OF ATTINI ANTS TOWARDS TO FUZZY ONTOLOGY FOR SEMANTIC WEB

Diego Negretto[1], Erik Antonio[2], Maurício Bacci[1], Milene Ferro[1]

[1]Universidade Estadual Paulista "Júlio de Mesquita Filho", Brazil, [2]Universidade Federal de São Carlos, Brazil

The leaf-cutter ants in the Attini tribe are agricultural pests due to the habit of cutter fresh vegetation to nourishment of the nests. The construction of leaf-cutter species database have play a special role to store, retrieve and disseminate information related to collections, taxonomy and morphological characterization through high-definition images. In this context, was developed the Virtual Museum of Attini Ants (http://evol.rc.unesp.br/formiga/) to provide support for search and retrieval of such information. However, the information retrieved from this Virtual Museum could not be used to express automatically castes of the Attini ants due to the use of the trivial queries. Therefore, the aim of this paper is to presents improvements in this legacy database to make it more feasible to use more complex and non trivial queries. This improvement was performed in three steps. In the first one, we applied a reengineering approach of the MySQL legacy database. This reengineering explored elements of the abstraction, comprehension and relationship between the entities. In addition, we produced the Entity-Relationship Model (ER-model) describing the preliminary meta-taxonomy used in that legacy database. In the second one, we developed semantic ontology for Attini ants using Methontology framework, which resulted in a set of artifacts in the XML/OWL2 (Extensible Markup Language/Ontology Web Language). Furthermore, we developed a core of the functionalities for recovery semantic information written in SPARQL/RDF (Resource Description Framework). Finally, we applied fuzzy logic to minimize the imprecision related to "crisps" ontology's previously defined. As a result of the fuzzy logic, two fuzzy gates (triangular and right shoulder) were made and depicted in Fuzzy Ontology OWL2 to classify the Attini ants into castes of workers: gardeners, generalists, foragers or soldiers. The classification was based on the average of the encephalic size of ants in a particular caste. In summary, the reengineering, use of ontology mapping and fuzzy ontology allowed to group and characterize semantically the workers in castes of the Virtual Museum of Attini ants.

Keywords: Bioinformatics, Database, Ants, Fuzzy Ontology, Artificial Intelligence

Topic: Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 163

Can essential proteins be predicted by their physicochemical features?

Maurício Lopes Casagrande[1], Marcio Luis Acencio[1], Ney Lemke[1]

[1]Botucatu Biosciences Institute - UNESP - Univ Estadual Paulista, Brazil

Background: The identification of proteins essential for survival is important for the understanding of the minimal requirements for cellular life and for drug design. As experimental studies with the purpose of building a catalog of essential proteins are time-consuming and laborious, a computational approach which could predict protein essentiality with high accuracy would be invaluable. Most approaches for protein essentiality prediction are based on homology mapping to experimentally verified essential proteins in model organisms or based on sequence features of the encoding gene. In this work we sought to investigate whether physicochemical features of proteins can predict their essentiality. Results: For this purpose, we first collected all known essential and non-essential proteins of the baker's yeast Saccharomyces cerevisiae, then we calculated the frequency of amino acids according to their physicochemical features (hydrophobicity, polarity, polarizability, charge, normalized van der Waals volume and solvent accessibility) as defined by the PROFEAT database (http://jing.cz3.nus.edu.sg/cgi-bin/prof/prof.cgi) and finally used the Wakaito Environment for Knowledge Analysis machine learning environment to present the training data–proteins with their physicochemical features–to the decision tree algorithm J48. The decision tree model generated can recover 68% of essential proteins with a precision of 62%. Moreover, the decision tree model suggests that, in general, a yeast protein to be essential should contain more than 27% and 12% of, respectively, polar and negatively charged amino acids. Conclusions: Despite the apparently low prediction performance, these preliminary results are promising since to the best of our knowledge this is the first time that physicochemical properties were used to predict essential proteins.

Keywords: Essential proteins, physicochemical features, machine learning

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 164

MOLECULAR CHARACTERIZATION AND COMPARATIVE GENOMICS OF THE Neisseria meningitidis C ISOLATES ASSOCIATED WITH MENINGITIS OUTBREAKS IN THE ESTATE OF MINAS GERAIS / BRAZIL IN 2011

Dhian Camargo[1], Laura Leite[2], Flávio Araújo[3], Bárbara da Mata[4], Michele Samuel[5], Marluce Oliveira[5], Guilherme Oliveira[6], Roney Coimbra[4]

[1]Biosystems Informatics, Research Center René Rachou - FIOCRUZ; Service for Bacterial and Fungal diseases – Ezequiel Dias Foundation -MG, Brazil; [2]Center for Excellence in Bioinformatics, Research Center René Rachou - FIOCRUZ, Brazil; [3]Genomics and Computational Biology Group, Research Center René Rachou - FIOCRUZ, Brazil; [4]Biosystems Informatics, Research Center René Rachou - FIOCRUZ, Brazil; [5]Service for Bacterial and Fungal diseases – Ezequiel Dias Foundation -MG, Brazil, [6]Genomics and Computational Biology Group, Research Center René Rachou - FIOCRUZ; Center for Excellence in Bioinformatics, Research Center René Rachou - FIOCRUZ, Brazil

Background: Neisseria meningitidis is a leading cause of mortality due to infectious diseases worldwide. This study aimed at characterizing the meningococci C isolates associated with a meningitis outbreak occurred in Ouro Branco, estate of Minas Gerais, in October 2011. The outbreak affected 18 people and caused several deaths. Four isolates epidemiologically related to this outbreak and 34 not related, but isolated in Minas Gerais in 2011 were included in this study. Thirty-eight isolates were grouped into three genotypes identified by RAPD, two of them were epidemiological related with the outbreak. Eight isolates representing the genotypic, geographic and temporal diversity among the collection had their whole genome sequenced with Life Technologies Ion Torrent PGM™ (one chip 316 per isolate). Reference guided draft assemblies of the genomes were accomplished using the softwares TMAP, MIRA3 and CLC Main Workbench. The coding regions were predicted with FgenesB and Prodigal, and the genes were annotated using BLAST2GO. Comparative analysis of the eight genomes sequenced in this study and reference Neisseria with published complete genomes (14 N. meningitidis, two N. gonorrhoeae and one N. lactamica) was performed with Artemis and MBDG (http://mbgd.genome.ad.jp/). Results: The eight genomes of N. meningitidis C from Minas Gerais were highly similar, with 2,05 to 2,07 Mb genome size, 1718 to 1758 coding sequences (CDS) and 51,9% to 52.1% GC content. These eight isolates belong to the same Multilocus Sequence Types (MLST) type ST 3780, ST-103 complex. They belong to serotype 23, but to serosubtype families 14 and 22. Comparative genome analysis showed up to 4X more overlapping CDSs or suspicious starting codons due to sequence "frameshifts" in the eight genomes sequenced herein when compared to the reference genomes, precluding further SNPs analysis and phylogenetic reconstruction based on the core proteome. The ortholog analysis performed with MBGD unrevealed the core and accessory proteomes of the 25 isolates. Hierarchical clustering (Manhatan distance and UPGMA) of the 25 isolates taking as input the matrix of presence/absence of 226 high confidence orthologous groups comprising the predicted accessory proteome of Neisseria showed that the eight isolates from Minas Gerais are indistinguishable. This finding is also supported by phylogenetic reconstruction based on polymorphisms in the seven housekeeping genes used for MLST. Conclusions: The Ouro Branco outbreak was caused by a clone ST 3780 from ST-103 clonal complex, a known invasive lineage which has been reported in Brazil since 2000. Whole genome sequencing of N. meningitidis using the Ion Torrent PGM™ (chip 316) resulted in a sub-optimal assembly of the genomes and, thus, did not add value to the epidemiological approach of the present work. Supported by: FIOCRUZ-Minas; FUNED-MG; FAPEMIG; CNPq; Life Technologies. Acknowledgments: PDTIS-FIOCRUZ - Platform RPT04B, Bioinformatics MG.

Keywords: Neisseria meningitidis, Genotyping, Comparative Genomics, RAPD, Next Generation Sequencing

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 165

3D-Structure Modeling of C-type Lectins from the Mangrove Oyster

Milena Marcela Domingues Pereira Schettini[1], Juliana Alves Americo[1], Mauro de Freitas Rebelo[1]

[1]Universidade Federal do Rio de Janeiro (UFRJ), Brazil

C-type lectins are carbohydrate-binding proteins. Among their several functions, they act recognizing pathogens and triggering host innate defense mechanisms. Thus, C-lectins can be used in Biotechnology as a “molecular shield”, since they neutralize pathogens by binding to sugar residues that microorganisms use to access host cells. These proteins also have pro or anticoagulant effects, depending on what platelets receptors they interact with. The main purpose of this project is to identify C-type lectins from the Brazilian oyster, Crassostrea rhizophorae, with biotechnological potential by in silico approaches. In order to obtain its 3D-structures, comparative modeling was used after the complete transcript sequence determination. Here we present eleven novel C-type lectin constructed models. Three of them (CrClec-3, -4 and -10) showed the EPN motif classically found in lectins which recognize mannose-type ligands. CrClec-5 revealed the QPD motif specific for galactose-type ligands. Other ligand motifs were found, as RPT (from CrClec-1), DPN (CrClec-2), YPG (CrClec-6), QPN (CrClec-7), EPD (Crclec-8), QPG (CrClec-9), IYH (CrClec-11), highlighting the increased repertoire present in bivalve lectins compared to mammals. Furthermore, CrClec-11 shares 27% (e-value 5e-8) identity with the crystal structure of the snake venon Agkisacucetin, which inhibits platelet adhesion and aggregation. To evaluate if CrClec-11 can potentially display this anticoagulant effect, docking and dynamic simulations with platelets proteins involved in the coagulation process will be conduct.

Keywords: bivalve invertebrate carbohydrate innate defense coagulation

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 166

MULTIATTRIBUTE MODEL FOR GWAS USING SUPPORT VECTOR REGRESSION WITH PEARSON UNIVERSAL KERNEL

Fabrízzio Oliveira[1], Fernanda Almeida[2], Marcos Vinicius Silva[2], Rui Verneque[2], Wagner Arbex[2], Carlos Cristiano Borges[1]

[1]Universidade Federal de Juiz de Fora, Brazil; [2]Empresa Brasileira de Pesquisa Agropecuária, Brazil

This work proposes a new methodology to simultaneously select the most relevant SNPs markers for the characterization of any measurable phenotype described by a continuous variable using support vector regression (SVR) with Pearson Universal kernel (Puk). The advantage of using a phenotype quantified by a continuous variable rather than a binary variable is to capture different levels of such characteristics from different genotypes. The proposed methodology is multiattribute towards considering several markers simultaneously to explain the phenotype. Currently , most GWAS studies quantifies the average impact of each marker on the phenotype through linear regressions between a marker and phenotype (monoattribute), in order to indicate the most significant markers in relation to phenotypic trait in question . However, such methods assume that the effects of each marker on the phenotype are only additives, disregarding the possible occurrence of complex interactions such as dominance and epistasis between markers. The Puk has the property of mimicking the behavior of linear kernels, RBF and polynomial, since their parameters are chosen appropriately. Thus, change is the question of what is the kernel to be used and what their parameters for choosing the best kernel parameters Puk to maximize inductive learning of SVR.

Keywords: Support vector regression. Pearson Universal kernel. Genetic markers. GWAS.

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 167

Theoretical Study of The Encapsulation of Nifedipine Derivates.

Antônio Nascimento[1], Sarah Furtado[1], Bárbara Silva[1], Frederico Silva[2], Carlyle Lima[1], Davi Brasil[1], Nahum Alves[1]

[1]Universidade Federal do Pará, Brazil, [2]Instituto de Ensino Superior da Amazônia, Brazil

Theoretical Study of The Encapsulation of Nifedipine Derivates. A. E. S.N Junior[1], S.V Furtado[1], B. C. P Silva[1], F.G.S.S Filho[2], CR Lima[1], DSB Brasil[1], CN Alves[1]. [1] Universidade Federal do Pará. [2] Instituto de Ensino Superior da Amazônia. The β-cyclodextrin (β-CD) is a cyclic oligosaccharide originating from the enzymatic degradation of starch, which has hydrophilic cavities in the structure giving it a high efficiency in the formation of inclusion complexes. Therefore, there is feasibility of the use of β-CD upon encapsulation of drugs in order to enhance the effect of drugs with low solubility or are quickly absorbed by the organism, such as Nifedipine, Felodipine and Isradipine. This paper is to propose a form of encapsulation in-silico drug cited from β-CD, among conformations that were suggested in Sousa et al., 2008. The orientations of structures of complexes were arranged in a similar manner for each of the inhibitors, allowing comparison of the performance of the tunnel in each conformation. To this end, we used the Gaussian 09 program with DFT method with operator B3LYP and the 6-31G, for parameterization of the structures and extracting loads. The Amber 13 program was also used in the calculations of energy minimization to elucidate which compound and mode of interaction with the more stable β-CD. From the results obtained it can be concluded that the most probable conformation of compounds derived from Nifedipine and β-CD is that with a 2:1 ratio, wherein the β-CD interacts with the aliphatic chains of the inhibitor. However, the less likely it is that conformation where the interaction between the cyclic groups in the molecule with β-CD ratio of 1:1 to 2:1 in Isradipine and Felodipine. In general, the inhibitor shows that the encapsulation structure is less favorable Felodipine. Therefore, it is observed that the Felodipine doesn't interact satisfactorily with the β-CD due to chlorine atoms in its structure. Already Isradipine, can interact with β-CD in a manner similar to Nifedina due to nitrogen compounds that have both. Supported by: Capes, CNPq Supramolecular Self-Assembly Of Cyclodextrin And Higher Water Soluble Guest: Thermodynamics And Topological Studies. F.B. de Sousa,A. M. L. Denadai, I. S. Lula, C. S. Nascimento Jr., N. S. F. Neto, A.C. Lima, W.B. De Almeida and R.D. Sinisterra. J. Amer. Chem. Soc. 130 (2008)

Keywords: nifedipine, beta-cyclodextrin, encapsulation, in-silico

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 168

COMPARATIVE ANALYSIS AND EXPLORATION OF THE TRICHOMONAS VAGINALIS AND GIARDIA LAMBLIA PROTEOMES

Lennon B. Almeida[1], Diogo A. Tschoeke[1], Alberto M.R. Dávila[1]

[1]Fiocruz, Brazil

Background: Since the 90s, there have been international efforts to obtain complete genome sequences, which led to genomic sequencing ana analyses of a large number of organisms, including protozoa. Those are unicellular Eukaryotes, comprising approximately 200.000 species, showing an extreme diversity and variety. Besides, most species are free-living and only a few are pathogenic. Comparative studies among Protozoa can help to identify similarities and differences at the genomic level, identifying and studying their homologous genes (orthologs and paralogs) can help to figure out which genes are shared between these species, and which ones are specific to each organism. Trichomonas vaginalis is a flagellated protist, member of the lineage Parabasilia, microaerophilic eukaryotes that lack mitochondria and peroxisomes. It causes trichomoniasis, a sexually transmitted infection , with approximately 170 million cases annually worldwide. Giardia lamblia is a unicellular organism that infects the jejunum of humans and a variety of other mammals. The giardiasis develops when cysts are ingested, is common among people with poor fecal-oral hygiene. The main modes of transmission include contaminated water supply or sexual activity. Results: 59872 proteins from T. vaginalis and 6525 from G.lamblia, totalizing 66397 proteins in these proteomes, were submitted to the OrthoMCL software to infer the relationships between these proteins. OrthoMCL inferred 5860 homologs groups, 4852 paralogs and 1008 orthologs. Inside these groups proteins with functions related to metabolism were found, as well as structure and information processing (ribosomal protein). As example of paralogs in G. lamblia we found 40S ribosomal protein and dynein heavy chain. Among T. vaginalis paralogs we can found actin and beta-tubulin that showed to be so divergent to their homologs in G. lamblia. As orthologs examples we found 40S ribosomal protein S5, L27, adenylate kinase family protein, and structural proteins such as dynein intermediate chain. For G. lamblia 344 paralogous groups were found, while in T. vaginalis 4508 paralogs groups were found. Moreover, 3630 putative orphans proteins were found in G. lamblia, from which, 3127 (86.14% 3127/3630) have hypothetical function. In T. vaginalis 10622 putative orphans proteins were found and 89.5% of these proteins (9507/10622) are annotated as hypothetical. Homologous proteins are being compared with data from Gene Ontology to obtain results on the functionality of proteins. Conclusion: There are many proteins shared (1008 orthologs) by these parasites, most of them being essential for the survival of the organisms. T. vaginalis showed a high number of paralogs proteins in relation to G. lamblia, and as expected most orphans proteins from both organisms displayhypothetical functions in their description.

Keywords: protozoa, comparative genomics

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 169

Towards a distributed/parallel comparative genomics workflow: Elastic-OrthoSearch

Nelson Kotowski[1], Rodrigo Jardim[1], Rafael Cuadrat[1], Diogo Tschoeke[1], Alberto M. R. Dávila[1]

[1]Computational and Systems Biology Laboratory, IOC/Fiocruz, Brazil

Background. Homology inference helps on identifying similarities, as well as differences among organisms, providing a better insight on how closely related one might be to another. In this sense, comparative genomics pipelines are widely adopted tools, designed using different bioinformatics applications and algorithms. OrthoSearch, a scientific batch workflow, targets at homology among Protozoa species through protein-profile comparison and reciprocal best hits. The increasing amount of data manipulated by it requires a considerable computational power and time to generate its results. In order to address these issues, we propose Elastic-OrthoSearch, a distributed/parallel-ready homology inference workflow, which maintains OrthoSearch's original premise while providing new features.

Keywords: comparative genomics, distributed computing, homology, protozoa, parallel computing, pipeline

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 170

Complete genome of Bordetella pertussis from Brazil: a phylogenomic study

Diego Cambuy[1], Michel Abanto[1], Fernanda Freitas[1], Ana Vicente[1]

[1]Fundação Oswaldo Cruz, Brazil

Complete genome of Bordetella pertussis from Brazil: a phylogenomic study. Cambuy, D.D. , Abanto, M.M., Freitas, F. , Vicente, A.C.P. Pertussis, more commonly referred as whooping cough is mainly caused by Bordetella pertussis. Despite over 50 years of vaccination, pertussis still remains an important cause of morbidity even in high-vaccinated countries. Previous analysis by MVST and MLST revealed that the current pertussis outbreaks in Brazil are driven by the predominant linage found worldwide. The genome B. pertussis size is around 4.1 Mb with a GC content of 67%. In this study, a Brazilian strain, BpBr2013.1, was sequenced by the GS 454 Junior sequencer. A total of 198000 reads were generated with an average length of 450 bp (20X coverage), the mean quality of the reads was 35.23. The de novo assembly performed by Roche assembler software Newbler generated a total of 275 contigs. Orthologs search with complete and draft genomes of B. pertussis (strains Tohama I ,B1834, B1831, B1917, B1920, B0558, and B1193) available at GenBank, the draft BpBr2013.1 genome, Bordetella parapertussis, and Bordetella bronchiseptica was performed using the InParanoid program. Since InParanoid uses coding sequences for its analysis, all genomes were (re)annoted by RAST, in an attempt to avoid annotation bias. The core genome consisted of 2651 genes, and their concatenated sequences were used for a phylogenomic reconstruction. A Maximum-likelihood tree was generated by RaxML (version 7.7.8), with the best evolution model (GTR+ GAMA+I) chosen by Model Generator (version 0.85). The results confirmed that pre-vaccination strains are distinct from post-vaccination strains, and some strains from the latter group carry a novel allele for the toxin pertussis promoter (ptxP3). The BpBr2013.1 strain, which carries the ptxP3 allele, also belongs to that clade along with the post-vaccination strains. This is the first study conducted outside The Netherlands that uses a whole-genome sequencing approach to determine the phylogenetic relationships between pre- and post- vaccination B. pertussis. We reassure that despite little variation in its genome, B. pertussis is still able to persist in vaccinated environments probably due to evolution/selection under the vaccine pressure. This information is basic for decision making on vaccine preventable diseases programs.

Keywords: Genome Sequencing, Bordetella, Phylogenomic

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 171

Developing a pipeline for rapid detection of aerobic anoxygenic photoheterotrophic bacteria (AAP) in oceanic metagenomes

Rafael Cuadrat[1], Hans-Peter Grossart[2], Alberto Dávila[1]

[1]FIOCRUZ - IOC, Brazil, [2]Leibniz-Institute of Freshwater Ecology and Inland Fisheries – IGB – Neuglobsow, Germany

Background: Aerobic anoxygenic photoheterotrophic bacteria (AAP) are considered of great interest because of their ability to produce various pigments and to harvest light energy. Thus they are of great importance for carbon and energy cycling in various natural environments. As photoheterotrophic bacteria, AAP needs oxygen to grow and they synthesize the pigment bacteriochlorophyll A (Bchl A), which is used in bacterial photosynthesis (without oxygen production). Several studies have been conducted with the purpose of estimating the abundance of AAP in oceans, using different approaches, such as detection of Bchl A by fluorescence, marker genes amplification by PCR and metagenomics. Thereby, a great variability in abundance of AAP has been found (between 0.1% and 10% of total bacteria).

Keywords: metagenome, AAP, pufM, pufL

Topic: Transcriptomics and Proteomics

Abstract #: 172

Comparative analysis of the differential expression data (RNAseq) of Corynebacterium pseudotuberculosis applying different bioinformatics tools

Mariana Santana[1], Anne Pinto[1], Vinicius Abreu[1], Pablo Sá[2], Rommel Ramos[2], Ulisses Pereira[1], Edson Folador[1], Siomar Soares[1], Silvanira Barbosa[2], Artur Silva[2], Vasco Azevedo[1]

[1]Universidade Federal de Minas Gerais, Brazil, [2]Universidade Federal do Pará, Brazil

Corynebacterium pseudotuberculosis is a Gram positive pathogenic bacteria, that causes lymphadenitis in small ruminants, which brings great economic losses to global agribusiness. In this context, Pinto et al (2012) characterized the transcriptional differential profile of C. pseudotuberculosis 1002 (Cp1002) in a condition that simulates the acidity of the host macrophage. Thus, was possible to discover which molecules are involved in the infection process that allows the bacteria to adapt in this environment. But authors in literature highlight the need to compare different softwares to verify the accuracy of the data. For the differential expression experiment, the bacteria grew to DO600nm = 0.2 in BHI medium under optimum conditions or in BHI supplemented with HCl, (pH 5.0). RNA was extracted, converted to cDNA, amplified and then sequenced in SOLiDTM 3 plus. For the analysis, the data were submitted to the Bioscope software and then to DEGseq. This result was compared to those obtained by the softwares Rockhopper and CLC using the software MySQL. A table with different and equal results was generated and 838 and 1252 genes are present in each, respectively. From those tables, some genes were selected due to their importance for the bacteria. The msrB gene, whose function is the reduction of methionine-R-sulfoxide, showed that the acid condition was twenty times more expressed than the control, in the DEGseq, two in CLC and twenty in Rockhopper. Other studies demonstrated that pathogens without a functional msr gene present a reduced ability to adhere, survive in the host and resist to oxidative stress. The pld gene, which encodes the phospholipase D, showed to be five times more expressed than the control in DEGseq and CLC, but showed to be no significant in Rockhopper. This protein is involved in the spread of bacteria in the host. Another important gene is the katA, which encodes a catalase responsible for reducing the concentration of hydrogen peroxide in the cell, presented a similar result in all the softwares, around four times more expressed in the acid condition. In general, the results of the softwares DEGseq and CLC showed to be more alike than when compared to Rockhopper. The first two had a 41,53% similarity, while comparing the last one with the first (19,56%) and second (40,38%), respectively. This work demonstrates the need to compare the differential expression data in different softwares and thus reach a more accurate result.

Keywords: Corynebacterium pseudotuberculosis; Differential expression; transcritomics, softwares

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 173

Bioinformatics tools in the screening of transcription factors involved in Fetal Hemoglobin levels regulation

Gisele Cristine de Souza Carrocini[1], Larissa Paola Rodrigues Venancio[1], Claudia Regina Bonini-Domingos[1]

[1]Ibilce-Unesp, São Paulo State University, São José do Rio Preto, Brazil

Background: In adults, fetal hemoglobin (Hb F) expression comprises up to 1% of the total, suffering influence of transcription factors in its regulation. We aimed to track out regulatory elements of γ-globin genes by comparative analysis of sequences (Phylogenetic footprinting), in order to know the genetic determinants that modulate Hb F levels. We compared sequences of Homo sapiens, Pan troglodytes, Macaca mulatta and Cebus apella, deposited in the National Center for Biotechnology Information. Alignment and analysis were based on conservation programs BLAT and mLAGAN, visualized by VISTA Browser. For evaluation of the conservation was used as parameters 70% identity to at least 100 bp non-coding regions. To identify transcription factors was used rVISTA program. Results: Among the 529 motifs of transcription factors indicated, 12 show connection with the regulation of γ-globin genes: AML-1 (RUNX-1) (21q22), CDC5 (6p21), c-MYB (6q22-23), CP2 (12q13), DLX7 (BP1) (17q21), EKLF (KLF1) (19p13), FKLF (KLF11) (2p25), GATA-1 (Xp11.23), GATA-2 (3q21), MAFK (7p22), NF-E2 (12q13) e NF-E4 (7q22). Among the 12 transcription factors highlighted in our analysis, the majority behaves as transcriptional activator of γ-globin gene, acting as positive regulators of Hb F. In contrast, c-MYB and KLF1 act as negative regulators of Hb F levels. The γ-globin gene transcriptional repression exerted by KLF1 is mediated by activation of the Bcl11a gene, which encodes the transcription factor BCL11A, when interacting with FOG1, NURD and SOX6, forms a protein complex that inhibits Hb F expression. The performance of Hb F positive regulators can occur in different ways: a) directly, by transcriptional activation of γ-globin genes; b) by transcription factor (NF-E4) interaction with the LCR, promoting the maintenance of chain coding range; c) the formation of protein complexes (CDC5 and NF-E4); d) and the involvement of transcription factors (GATA-1 and NF-E4) in the delay of switching from Hb F to Hb A expression. Conclusions: Knowledge about the regulatory elements of Hb F can contribute in different therapeutic strategies discovery that may improve on clinical pathways when considering patients with hemoglobinopathies.

Keywords: Fetal hemoglobin, transcription factors, gamma-globin gene regulation

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 175

DRAFT GENOME SEQUENCE OF PASTEURELLA MULTOCIDA STRAIN 11246

Mauricio Egídio Cantão[1], João Xavier de Oliveira Filho[2], Jalusa Deon Kich[1], Cátia Silene Klein[1], Ricardo Zanella[1], Marcos Antônio Zanella Mores[1], Raquel Rebelatto[1], Nelson Mores[1]

[1]EMBRAPA - Suínos e Aves, Brazil, [2]Universidade Federal do Rio Grande do Sul (UFRGS), Brazil

Background, Pasteurella multocida Strain 11246 serotype A is the most common bacterial agent isolated from lesions caused by pneumonia in pigs. Although this pathogen is considered a secondary opportunistic agent in enzootic pneumonia caused by Mycoplasma hyopneumoniae, there are evidences showing its involvement as a primary agent. However, little is known about its pathogenisis. In this context, to detect potential virulence associated genes, we sequenced the genome of P. multocida isolated from pneumonia lesions, in a group of specific pathogen-free animals (SPF) experimentaly exposed to this phatogen; Results, the paired-end sequences were produced by Illumina MiSeq platform (2x250 bp). Low quality reads and adapters were removed using SeqyClean V 1.2.3, in addition to that, all sequences with phred quality score < 25 and length < 180 bases were also removed out. After quality control, 1,485,870 reads were assembled using the MaSuRCA Assembler V. 2.0.3.1. De novo assembly produced 10 Scaffolds with 2,244,069 bp in length with GC content of 40,4%. N50 of final scaffold reached 530 kb, with 638 kb being the largest scaffold. The genome contain 2,014 predicted coding regions, 5 ribosomal 16S RNA, and 47 predicted tRNAs. The sequence identified the presence of kmt gene in our samples, this gene is a specie-specific for Pasteurella multocida. The presence of two additional genes hyaD and hyaC with 99% of identity with P. multocida A:1 strain X-73 further classify this bacteria as sorotype A. Twelve virulence-associated genes of P. multocida were identified in the sequenced genome: A) outer membrane and porin proteins(oma87, psl, ompH); B) a type 4 fimbriae (ptfA); C) a filamentous hemagglutinin (pfhA); D) neuraminidases (nanB, nanH); E) iron acquisition related factors (exbBD-tonB, hgbA, hgbB), and F) superoxid dismutases (sodA, sodC). In total, out of the 2,014 predicted genes, 4 did not match when compared with genes at NCBI-NR with e-value 1e-05. It can be an indicative that those genes are excluvive of our genome; Conclusions, we have identified 4 unique genes in the studied genome. In addition to that, 12 genes were associated with virulence of P. multocida. Further investigationg are being conducted by our group to use those genes as a genetic marker for the pathogenicity of this bacterium. With the availability of the P. multocida sequence we will be able to conduct comparative, epidemiological and evolutionary studies.

Keywords: Pig; Pasteurella multocida; genome

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 176

EVIDENCES OF DISTING WAVE OF GENE ORIGINS IN RECENT EVOLUTIONARY EVENTS

Ricardo Vialle[1], José Ortega[1]

[1]Universidade Federal de Minas Gerais, Brazil

The ability of clustering homologous genes allows depicting their occurrences on distinct clades and the inference of when they have appeared. Compilation of events of gene origin by our group suggest a quantitative pattern of appearance in waves, since some periods of human evolution account for concentrated events. Here we address the question qualitatively, asking whether or not the origin of distinct EC categories vary along evolution, and if there is a bias for secondary structure content in recent proteins as compared to more ancient ones. Our data support a distinct pattern of gene origin in recent genes. We have analyzed the appearance of enzymes with distinct EC numbers along evolution by means of determining the distribution of homologous genes along taxonomic groups, using the UEKO database. Remarkably, the scenario is distinct in respect to the first EC digit, since some classes are more prone to generate modern genes. After the Euteleostomi époque, there appeared, respectively for EC classes starting with 1 (oxidoreductases), 2 (transferases), 3 (hydrolases), 4 (lyases), 5 (isomerases) and 6 (ligases): 5, 9, 22, 0, 0 and 2 genes. The more modern Lyase (EC 4.4.1.20) appears in Euteleostomi and the second more modern Lyase (EC 4.1.2.30), in Chordata, while three Isomerases (EC 5.3.99.4, 5.3.99.5 and 5.1.3.19) appear in Euteleostomi after the appearance of one (EC 5.1.3.17) in Metazoa. Furthermore, we found that the percentage of 2D structure varies along evolution. We analyzed five époques: Cellular Organisms, Eukaryota, Opistokonta-Chordata, Craniata-Euteleostomi and the more modern clades up to Homo sapiens. For beta-strand the occurrence varied around 20-25%. For these époques, alpha-helix often shown a double distribution: (40%), (35%, 45%), (10%, 40%), (15%, 40%), (5% and a large concentration of sequences with little helix, and a pick with 40%). Conversely, the segment without structure compensates for the decrease in alpha-helix content, varying from around 42.5% up to 50%. Thus, the remarkable change along evolution in 2D structure are a small raise of the not structured fraction and a remarkable presence of proteins with low alpha-helix content; while some proteins still maintain around 40% of helical structure, many show less than 25%, what is rare in ancient proteins, shared with prokaryotes. In our knowledge this is the first initiative to depict the scenario of appearance of distinct catalytic enzymes. All EC human classes are shared with cellular organisms; however, just some classes originated genes along the recent periods of human evolution, after the origin of bones by the Euteleostomi époque and the conquest of Earth environment. Moreover, the 2D structure extracted from PDB files show a bias for reduction in alpha-helix in this period, thus supporting a distinct wave of gene origins in recent evolutionary events.

Keywords: Secondary structure, EC Number, Evolution

Topic: Transcriptomics and Proteomics,Signaling and Metabolic Networking; Ontologies; Systems Biology

Abstract #: 178

Characterization of secondary metabolism in copaíba (Copaifera multijuga)

Emily Bandeira[1], Taina Raiol[1], Maria Emilia Walter[1], Spartaco Astolfi Filho[2], Marcelo Brígido[1]

[1]University of Brasilia, Brazil, [2]Federal University of Amazonas, Brazil

Background: The resin-oil of Copaíba, produced by the copaíba trees as a defense mechanism against its predators, have been used in popular traditional medicine over 500 years. Among the many applications of the oil, dozens of them were described as medicinal properties. In some of the cases, these properties were even scientifically proven, such as anti-microbial, anti-inflammatory and antineoplastic activities. Through the Copaiba's transcriptome, the metabolic pathways could be determined, particularly those involved in the terpenoid production, the main oil components. The cDNA library, constructed from the Copaíba leaves, were sequenced using the 454 technology. A pipeline was developed gathering several bioinformatics tools for all analysis stages (filtering, assembly and annotation). Using the automatic annotation of all protein coding genes, the metabolic pathways present in this plant could be mapped. Results: 454 sequencing produced 638,576 reads, which were grouped into 504,273 contigs and 39,907 singlets. In order to characterize the metabolic pathways that compose the Copaíba's secondary metabolism, metabolic maps were built using the identified ECs (Enzymes Comission numbers). A large number of sesquiterpenes, the main components of the Copaíba oil, could be identified. Among the other components, we found metabolics maps of diterpenoids, flavone and flavonol, flavonoids, glycolysis, isoflavonoids, monoterpenoids, phenylpropanoids, terpenoids, ubiquinone, zeatin, anthocyanin and carotenoids. Conclusions: Our preliminary analysis have already shown the presence of many of the essential resin-oil components. Additional enzyme sequence databases will be included in our analysis to improve the characterization of secondary metabolic pathways.

Keywords: secondary metabolism, Copaifera multijuga, copaíba

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Transcriptomics and Proteomics

Abstract #: 180

De novo Transcriptome of Two Species of Anastrepha Fruit Flies (Diptera: Tephritidae) and Analysis of Gene Expression and SNP Calling

Victor Rezende[1], Carlos Congrains[1], André Lima[1], Emeline Campanini[1], Aline Nakamura[1], Janaína Oliveira[1], Samira Chahad-Ehlers[1], Iderval Sobrinho Junior[1], Reinaldo de Brito[1]

[1]Federal University of São Carlos - UFSCar, Brazil

Background: Several species of the Anastrepha fruit flies are of great economic importance for causing damage to a variety of fleshy fruits. Some of the most important species in the genus belong to the closely related and recently diverged fraterculus group. Some species in this group have similar morphological attributes that make it difficult to identify them, making it important to identify new genetic/molecular markers that can differentiate the species. In this work, we compare levels of gene expression and identify single nucleotide polymorphisms (SNPs) aiming to isolate genes and SNPs that show significant differences between two Anastrepha species (A. obliqua and A. fraterculus). To do so, we built multiple Next Generation Sequencing (NGS) libraries from head tissues of these species, at different reproductive stages for both sexes and assembled de novo all reads by species for mapping and subsequent SNPs discovery as well as for an analysis of differential gene expression. Results: cDNA libraries were generated from total RNA extracted from fly heads pooled according to species, gender and reproductive stages, totaling 10 libraries per species. Over 168 million reads of 100 paired-end sequences were generated by runs on an Illumina HiSeq 2000. The Illumina reads were trimmed and filtered by quality and about 16% of all reads were discarded. The remaining 140 million reads were assembled using the Trinity short read assembler, which resulted in 154,787 contigs with N50 of 2012 and an average length of 1,027 bp. Over 45,000 contigs had more than 1,000 bp and almost 23,000 had more than 2,000 bp, with the longest having 25,704 bp. The use of SAMtools and Bowtie in the alignments of the reads of each species mapped against total assembly identified 190,487 intraspecific SNPs that were polymorphic only in A. fraterculus, 96,568 in A. obliqua and 19,499 in both. We found 8,092 contigs with at least one SNP present in both species, and 413 of these had fixed interspecific differences between species. Gene expression analysis showed 1,689 transcripts with significantly difference between the species, within 1,096 overexpressed in A. obliqua and 593 in A. fraterculus. 174 transcripts show differential gene expression, as well as different SNPs for each species, some of the most important annotated contigs are serine proteases (such as Ser6), cuticular proteins (Cpr47Ef) and receptors of activity (CG2789 and CG5428), as well as some uncharacterized protein. Conclusions: These results generate a set of candidate genes that are potentially important to help us understand the evolution and differentiation of A. obliqua and A. fraterculus which might help us identify genetic markers that would be relevant to study the evolution and identification of species in the fraterculus group.

Keywords: Transcriptome, RNA-Seq, De novo assembly, Next generation sequencing, Gene expression, SNP calling, Tephritidae, Anastrepha, Fraterculus group, Head tissue.

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 181

Phylogenetic bioindicators for prediction the enterotypes of the human intestinal microbiome

Henrique Veras[1], Gabriel Fernandes[1]

[1]Catholic University of Brasília - UCB, Brazil

Phylogenetic bioindicators for prediction the enterotypes of the human intestinal microbiome HCT VERAS1, GR FERNANDES1 1 Universidade Católica de Brasília – UCB Background: Humans live in constant association with microorganisms. The amount of microorganisms present in the human body exceeds our own cell number. The use of high-throughput DNA sequencing technologies and independent culture molecular approaches have been enlarging the understanding concerning the communities of microorganisms and the association of these with the host. The human gastrointestinal tract contains one of the most complex bacterial communities. It was proposed recently that the intestinal microbiome could be categorized in three enterotypes (ET). In our study, we used data metagenomics quantitative to identify phylogenetic patterns in the intestinal microbiome to develop prediction models for the enterotypes. To reach this aim, statistical tests were applied to the data regarding abundance of bacteria in level taxonomic corresponding to genus. Results: We identified genus of bacteria with significantly different abundance and important correlations between them. The combination of the ratio among genus was used as bioindicator parameter of the respective enterotype. Through the logistic regression test we identified that the prediction model for ET1 was influenced significantly by the ratio between Bacteroides/(Prevotella + Ruminococcus). In the model for prediction of ET2, it was the ratio between Prevotella/Bacteroides that influenced with significance the indicators of this enterotype. And for the model of ET3, we identified the ratio between (Akkermansia + Alistipes)/(Bacteroides + Prevotella) as significant parameter this enterotype. These models were assessed against two groups of independent data (85 samples of data Illumina and 154 samples of data rRNA 16S) and associated with the value of cut-off 5%, 20% and 95% respectively. Besides the value of cut-off for each models, the crossed validation allowed the association of the model with the measure of prediction positive value (PPV) for ET1, specificity value (SPV) for ET2 and prediction negative value (PNV) for ET3. Conclusions: The prediction models provide a more feature for identifying enterotypes of the intestinal microbiome. Models to categorize enterotypes 1 and 2 present the most significant chances. We proposed the experimental validation of these models for the qPCR technique. These two microbial biotypes have great potential for use as mediators of intestinal processes related to health and disease and influence eating habits. With this established methodology, it would be possible to do the diagnosis of the enterotypes individually. Supported by: Universidade Católica de Brasília – UCB Fundação de Apoio à Pesquisa do Distrito Federal – FAP/DF

Keywords: Metagenomics; Bioinformatics; Intestinal Microbiome; Enterotypes; Bioindicators

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 182

EXPLORING THE LIPASE DIVERSITY IN THE AQUATIC METAGENOME OF ARRAIAL DO CABO, RIO DE JANEIRO

Mayla A Costa[1], Rafael RC Cuadrat[1], Rodrigo Jardim[1], Alberto MR Dávila[1]

[1]Fiocruz, Brazil

Background: The planet Earth is formed for ¾ of water and only ¼ of soil. The water has variety of ecosystems, housing considerable biodiversity of plants, animals and microorganisms. The marine microorganisms are the most abundant life forms in the ocean. The microbial world includes a diversity of organism, and are constantly discovered new microorganisms. Studies about biodiversity of various environments demonstrate that only 0,1% to 1,0% of the bacteria presents in the environments are cultivated using traditional farming methods. The metagenomics approach is being widely used to access the physiology and genetics of microorganism still no cultivable and therefore a viable alternative for the discovery of new enzymes and metabolic pathways. From the studies of metagenomics were identified more than 80 new lipases. Lipases are enzymes belonging to the family of hydrolases, acting in the aqueous-organic interface demonstrating considerable levels of activity and stability in aqueous environments and non-aqueous. They act on lipids, catalyzing chemical reactions and are produced by living organisms including animals, plants, fungi and bacteria. Among the functions performed by this enzyme we can cite natural, industrial and medical functions. Recently, DNA pyrosequencing technology together with the improvement of bioinformatics tools has allowed the DNA sequencing of communities of non-cultivable bacteria. Results: We have found a total of 1,178 lipase sequences distributed among several families: I (178/15, 11%), II (4/0, 34%), III (2/0, 17%), IV (28/2 , 38%), V (497/42, 19%), VI (184/15, 62%) and VII (184/15, 62%) and VIII (101 / 8.57%)). Conserved domains were inferred by RPSBlast, then found 132 sequences of putative lipases, as follows: I (5/4%), II (2/1%), III (0/0%), IV (16/12%), V (24/17%), VI (46/33%) and VII (46/33%) and VIII (0/0%). Conclusions: These results show that our HMM-based methologywas able to recover of lipases sequences from metagenomic data of Arraial do Cabo. However, more specific HMM profiles will probably need to be built in order to to detect more sequences form the different families (I, III, V, VI and VII). On the other hand, HMM profiles for families II and IV demonstrated good sensitivity. For the VIII family, lipase sequences were not found, but our preliminary results suggest that the sequences retrieved with HMM profiles built for VIII family are related to β-lactamase. According to the literature lipases of this family are related to β-lactamases.

Keywords: lipase, enzyme, hmm

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 183

Phylogeography of the sardine Opisthonema oglinum (Clupeidae, Clupeiformes) from Atlantic neotropical through mtDNA sequences

Cleonilde Queiroz[1], Jersey Maués[2], Iracilda Sampaio[1], Horacio Schneider[1]

[1]IECOS - UFPA, Brazil, [2]UFPA, Brazil

The genetic variability of sardine flag (O. oglinum)was estimated populations collected along the Brazilian coast. This study was conducted using DNA sequences of the control region of mitochondrial DNA. The results showed that the species is in genetic equilibrium and showed the existence of a single gene pool because it was not observed significant genetic structure, with high gene flow among the seven populations. Tests of neutrality (Tajima's D and Fu's FS) were negative. The D values were not significant, while FS were significant, indicating population expansion. Considering these results it can be inferred that this sardine flag of the Brazilian Atlantic coast is a single genetically homogeneous stock, widely distributed geographically.

Keywords: Opisthonema oglinum, population genetics, control region, mtDNA

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Transcriptomics and Proteomics

Abstract #: 184

Assembly and annotation of a thermophilic bacterium of Bacillus cereus group

Joao Victor Oliveira[1], Helena Silva[1], Taina Raiol[1], Nalvo Almeida[2], Marcelo Brigido[1], Marlene De-Souza[1], Lidia Moraes[1], Maria Emilia Walter[1], Peter Stadler[3]

[1]University of Brasilia, Brazil, [2]Federal University of Mato Grosso do Sul, Brazil, [3]University of Leipzig, Germany

Background: Bacillus cereus is a gram-positive, spore-forming rod and an opportunistic pathogen, involved in poisoning food and systemic and local infections, specially in immunologically compromised individuals. Only very few strains of B. cereus group are able to grow at temperatures higher than 48oC. B. cereus strain FT9 was isolated in a hot spring located at the Midwest Region of Brazil. It is a potential new thermophilic strain that might present heat resistance genes of biotechnological interest. Objectives: The objective of this work is to assemble and annotate the Bacillus cereus FT9's genome, making some preliminary analyses concerning its circular chromosome, which determine its main genomic features. Methods: The scaffold assembly was performed using a pipeline composed of softwares PRINSEQ, MIRA, Segemehl, BLAST and IGV, while in annotation step, GLIMMER, BLAST and the PATRIC database were employed. Results: From 561,784 reads, 376,531 (67.02%) were used by MIRA using the mapping technique, and 165,948 (29.54%) were used by MIRA de novo method. The final consensus has 5,223,665 bp, corresponding to the circular chromosome. Other information are: the chromosomal GC content is 35.54%, the total number of ORFs is 5,743, the average ORF size is 697 bp, the total number of conserved ORFs with protein function assigned is 5,083 (88.51%) and with unknown protein function is 108 (1.9%), and the total number of non conserved ORFs is 552 (9.6%). Using RNammer, 11 rRNAs operons were identified in the chromosome, which is consistent with other species from Bacillus cereus group. In addition, the phylogenetic relationship among species of the Bacillus cereus group were investigated, and strongly suggested that the B. cereus FT9 strain could be classified as a Bacillus cereus strictu senso. Conclusion: In this work, we presented a probable new strain of B. cereus group, identified as strain FT9, having characterized its complete circular chromosome sequence. We could identified about 88% of conserved ORFs with protein function assigned. Further analysis will be performed through homology analysis in order to further determine the taxonomic classification of strain FT9.

Keywords: Bacillus cereus strain FT9, termophilic bacteria of Bacillus cereus group, genome assembly and annotation

Topic: Transcriptomics and Proteomics

Abstract #: 185

Transcriptome analysis of root Piper nigrum L. (Piperacea) using RNA-Seq

Sheila Maysa Gordo[1], Edith Cibelle Moreira[1], Sylvain Darnet[1], Iracilda Sampaio[1], Horácio Schneider[1], Jersey Maués[1], Daniel Pinheiro[2]

[1]UFPA, Brazil, [2]USP - Ribeirão, Brazil

The extract of piperine is the most known principles of active black pepper and is used in research as anti-cancer drug and anti-inflammatory agent. The species also is second most exported agricultural product in Pará, being a culture adapted to the climate and soil conditions of the region. We sequenced the transcriptome of this specie in order to discover and annotate transcripts of interest for agriculture, ecology and medicine. We have used the short reads obtained with the RNA-Seq platform SOLID V3, to perform the assembly of root transcriptome. The short reads were filtered in accordance with the quality Phed value. The reconstruct transcripts from short RNA-Seq reads was performed with De novo assembly aproach in the Velvet and Oases software using Multiple K and STM- Method using the reference proteoma of Aristolochia fimbriata for scaffolding contigs that map onto the protein. The identification of microsatellites was performed with MISA script. We generated 22363 transcripts and 10338 unigenes with ortoMCL. Furthermore, were identified 168 microsatellites. Blast2Go and orthoMCL methods were used to annotate. Two root proteomes identified were about 615 proteins. The 4472 predicted proteins showed about 52% homology with the Arabidopsis proteome. The sequencing of non-model plants by using NGS Platform produced the first genes database of Piper nigrum. These data significantly increased the information of molecular genetics for plants in particular the group Magnoliids and family Piperales. These data can be analyzed with higher accuracy and produce essential information for research in the field of genetic improvement medicine and ecology of this species.

Keywords: transcriptome, root, piper, ngs, black pepper

Topic: Human Pathophysiology; Animal Models of Disease

Abstract #: 186

High mutation rate in proteins networked with HIV1 suggests a mechanism for resistance acquirement in macaque lineage

Tetsu Sakamoto[1], José Miguel Ortega[1]

[1]Universidade Federal de Minas Gerais, Brazil

Background HIV1 is known to have a limited host range. Besides humans, HIV1 can infect chimpanzees by artificial inoculation. Recently, genetic engineering of virus and/or host cell has allowed HIV1 to infect other mammal cells like T-cells from cat and rabbit, indicating their potential in being model organisms for HIV1 studies. In a similar way, there were attempts to generate infectious HIV1 in rhesus macaque, but these seem to be more difficult to achieve. Those facts suggest that cat and rabbit have almost all machinery necessary for the virus propagation. In contrary, rhesus macaque could lack host factors or a region of the protein essential for virus proteins interaction, resulting in low infectivity of HIV1 in this organism. In this work, we aimed to search for host proteins associated to HIV1 that might have conferred HIV1 resistance in rhesus macaques. To achieve this, we performed similarity analysis of all human proteins associated to HIV1 infection with their orthologues in cat, rabbit and rhesus macaque. Results Human proteins associated to HIV1 were retrieved from “HIV-1 Human Protein Interaction Database”. By aligning each of those protein sequences with their respective orthologues in cat, rabbit and rhesus macaque, we identified and quantified the amino acid sites that were conserved in human, cat and rabbit orthologues but mutated in rhesus macaque. Among 2494 proteins analyzed, we selected roughly 60 proteins with significantly high number of mutation sites specific in rhesus macaque. Most of selected proteins are described as auxiliary proteins during the infectious process of HIV like virus entry (integrin-alpha X, CXCR1) and transactivation of HIV1 LTR promoter (BAF200, CDK11, granulin, NELF, NFAT, PP2A, HIC, SNFL2, SP3, transglutaminase). There were also proteins associated with apoptosis (BID, DFFB, glutamate receptor, cytochrome c oxidase subunits, tubulin chain) and with neuronal dysfunction caused by HIV (adenylate cyclase, glutamate receptor, RyR, LRP, tyrosine hidroxylase). Moreover, sequence alignment of those proteins with orthologues of other non-human primates showed that some of mutated sites encountered in rhesus macaque orthologues are encountered in other primates, suggesting that those sites could confer resistance against HIV1 infection to other primates. Conclusions Data are indicative of mutations selected in macaque lineages that are associated with HIV1 resistance, since amongst man, cat and rabbit proteins those regions are conserved. Moreover, these data support the hypothesis that high mutation rate on those proteins in macaque lineage resulted in adaptive resistance to HIV1. Supported by: FAPEMIG

Keywords: HIV resistance, animal models, primates, rhesus macaque, mutation

Topic: Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 187

Arachis Transcriptome Survey using Drupal

Ana Paula Zotta Mota[1], Ana Cristina Miranda Brasileiro[2], Patricia Messenberg Guimarães[2], Orzenil Bonfim da Silva-Junior[2], Roberto Coiti Togawa[2]

[1]FAP-DF, Brazil, [2]EMBRAPA, Brazil

South America is the center of origin of the cultivated peanut and of more than 80 species of Arachis wild relatives. These wild species, differently to the cultivated species (Arachis hypogea L.), are diploids, and some of them harbor resistances to nematodes and fungi. The difference in ploidy and the genome size has hindered the characterization and introgression of wild alleles intro cultivated species. Our group, (in association with groups in North America, India and Europe) has used RNAseq technology and other NGS, to produce transcriptome profiles from resistant A. stenosperma challenged with the gall nematode M. arenaria, and A. duranensis under hydric stress. Analysis of differential expression between challenged and control plants showed a number of candidate genes of tolerance/resistance to these constraints. Results from this analysis were organized on a CMS using Drupal (http://drupal.org), with a Tripal module (http://www.gmod.org/wiki/Tripal). Tripal is an important GMOD tool which allows the establishment of online genomic database. Using Drupal/Tripal, the user can retrieve information from the created database such as clustering statistics, fasta sequences for each analysis, results for the KEGG orthology, InterPro and Blast search. The user also can access some other tools like blast search a viroblast implementation (http://indra.mullins.microbiol.washington.edu/viroblast/viroblast.php) and e-PCR both using the same local database.The use of Drupal/Tripal is a good strategy to organize and distribute information integrating the existing data in one unique address. The use of tools like blast and e-PCR using local databases improves the quality of the results. The platform provides an important resource facilitating further studies for the scientific community working with Arachis. The platform can be accessed at the following address: http://lbi.cenargen.embrapa.br/arachis.

Keywords: Arachis, Drupal

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 188

Studies on the structure and conformation of the transmembrane domain of the P2X7 receptor

Rafael Ferreria Soares[1], Ernesto Raul Caffarena[1], Pedro Celso Nogueira Teixeira[1]

[1]Oswaldo Cruz Foundation, Brazil

The ionotropic P2X7 receptor (P2X7R) is expressed in the central nervous system, immune system and epithelial cells. When activated by its substrate, ATP opens a cation-selective channel where ions K+, Na+ and Ca+ + can pass through the cell membrane, leading to the activation of intracellular second messengers. Exposure to high concentrations of its agonist (> 100μM) allows the passing of high molecular weight dyes through the cell membrane. However, there is no data in the literature about its 3D structure, limiting the understanding of the molecular mechanisms of pore opening. This study aims to determine the structure of the transmembrane region of the human P2X7R structures using the 3D structure of P2X4r zebrafish (zfP2X4b) in its closed and opened states (PDBs codes: 4DW0 and 4DW1, respectively), determined by X-ray crystallography. The model was obtained by the homology modeling software MODELLER 9.v11 followed by molecular dynamics studies by the insertion of protein structures in a lipid bilayer molecules of dipalmitoylphosphatidylcholine (DPPC) using Gromacs V.4.6.3. These in silico studies were conducted by a divide-and-conquer approach, which intended to study the structures of transmembrane segments TM1 and TM2 inserted in hydrophobic environment. We analyzed two molecular dynamics (MD) simulations of 100ns each. The first MD1 composed by DPPC bilayer solvated in water (SPC model). The second MD2 containing the transmembrane portion of zfP2X4b inserted in the bilayer.

Keywords: Purinergic receptor, Molecular modeling, Molecular dynamics

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 189

COMPARATIVE MODELING OF E1Ca-ATP-Mg STATE OF SERCA2a FROM HUMAN CARDIAC MUSCLE

Vinicius Carius de Souza[1], Carlos Roberto Ladeira[1], Vinicius Schmitz Nunes[1], Rodrigo Weber dos Santos[1], Carlos Cristiano Hasenclever[1], Priscila Vanessa Zabala Capriles Goliatt[1]

[1]University of Juiz de Fora, Brazil

The SERCA2a (Sarcoplasmic/Endoplasmic Reticulum Calcium ATPase 2) is an ATPase protein which transports calcium from cytosol to the sarcoplasmic reticulum (SR) of cardiac muscle, through ATP hydrolysis. It promotes the control of cytosolic calcium levels allowing a potential difference which plays a key role in the contraction/relaxation of cardiac muscle. This enzyme is composed by a transmembrane domain (TM), the binding site of Ca+2, and a cytoplasmic (ECD), the active site domain. The TM is formed by 10 helices and the ECD is subdivided in: (i) domain-A, that presents a highly conserved sequence known as TGES; (ii) domain-N, the nucleotide bind site; (iii) domain-P, where is located the phosphorylation residue (D351). SERCA2a assumes four distinct states: (E1) when occurs the bind of two Ca+2 into TM domain, one ATP in domain-N and one Mg+2 to D351; (E1-P) we verify modifications in domains A and P, allowing the nucleotide hydrolysis; (E2-P) when occurs the liberation of Ca+2 into the lumen of SR and replacement of this by ions H+, also occurs dephosphorylation of residue D351; (E2) ions H+, phosphate and Mg+2 are released into the cytoplasm. In this work, we present a comparative three-dimensional (3D) model of human SERCA2a (at the E1-Ca+2-ATP-Mg+2 state), considering their membrane environment. The 3D model of SERCA2a was constructed (via Modeller9v10) using as template the structure of PDB1T5S from Rabbit, considering the ligands (phosphomethylphosphonic acid-adenylate ester – ACP, Ca+2 and Mg+2). The local alignment between these two sequences presents 84% of identity and 92% of similarity. The protein model was added into the POPC membrane model, through the creation of a pore in the membrane. After that, two solvation box with ions were added to simulate the E1-Ca+2-ATP-Mg+2 state (during cardiac muscle contraction): (i) cytoplasmic environment, with 10μM CaCl2; (ii) lumen environment, with 1mM CaCl2. Due to periodic boundary conditions, a second membrane was added at the top of the system to promote the restriction of the ions in the z-axis of solvation boxes. The total system (693,533 atoms), was submitted to a pre-molecular dynamics simulation, using the CHARMM force field and the NPT ensemble at 1atm. The minimization step was subdivided into: (i) restrained minimization of ions and protein (0.5ns); (ii) all-atom minimization (0.5ns). After minimizations, we were able to observe the energy stabilization of the system, although this has not yet reached 1atm pressure. The next step is heating the system from 292K (crystallographic temperature) to 310K (usual human heart temperature), and then, we going to submit the system to the equilibration procedure followed by the production step of molecular dynamics simulation. We expect to observe structural movements, mainly in ECD domain, that could help in the analysis of the E1-P state of SERCA2a.

Keywords: SERCA, Sarcoplasmic Reticulum, cardiac muscle, Comparative modeling and Molecular dynamics

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 190

MicroRNA-190-5p: expression profile and cellular function in Schistosoma mansoni

Victor Fernandes de Oliveira[1], Matheus de Souza Gomes[2], Fabiano Carlos Pinto de Abreu[1], Roberta Verciano Pereira[1], Liana K. Jannotti-Passos[3], Charles Spillane[4], William Castro Borges[1], Renata Guerra-Sá[1]

[1]Universidade Federal de Ouro Preto, Brazil, [2]Universidade Federal de Uberlândia, Brazil, [3]Centro de Pesquisas René Rachou, Brazil, [4]National University of Ireland, Ireland

Mature microRNAs (miRNAs) are small, non-coding regulatory RNAs that can repress post-transcriptionally mRNA levels of target genes. Previous results of our group using computational approach identified a set of 35 Schistosoma mansoni miRNAs evolutionarily conserved. In the present study, we report the characterization of miR-190-5p in S. mansoni. For this, we firstly used bioinformatic approaches to identify miR-190-5p in S. mansoni and to confirm the hypothesis that like other species the biogenesis of this miRNA is also related to talin gene expression. Our analysis indicates that sma-mir-190 is located within putative talin gene structure between exons 7 and 8 in S. mansoni. The S. mansoni mir-190 precursor displayed 90% sequence identity with S. japonicum, and 100% identity with the mature miRNA sequence of S. japonicum. We also observed that miR-190-5p structure has been conserved across a large phylogenetic distance, from the choanoflagellate, Monosiga brevicollis to humans. We used miRanda and data from Schisto DB to target prediction and observed a set of 76 putative genes target. Quantitative RT-PCR was used to measure expression of both miR-190-5p and talin gene in cercariae, adult worms and in vitro cultivated schistosomula. Our data showed high expression of talin gene in cercariae and schistosomula when compared to adult worms and a similar profile of miR-190-5p expression in all stages analysed suggesting that this miRNA can be an essential to parasite biology. Taken together, these results suggest a novel insight into the association between miR-190-5p and their possible targets in the S. mansoni biology.

Keywords: miR-190-5p, Schistosoma mansoni, predict targets, differential expression.

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 191

MOLECULAR DYNAMICS SIMULATION OF THE ACRB EFFLUX PUMP PROTEIN

Nubia S. Prates[1], Lande V. Silva Jr. [1], Pedro E. A. Silva[1], Karina S. Machado[1]

[1]Universidade Federal do Rio Grande, Brazil

Background Penicilin can be considered as the first antibiotic developed for the treatment of infections in the human organism. Since then, humanity has benefited from this discovery. However micro-organisms have been developing resistance to antibiotics, which may turn treatments with conventional drugs inefficient, what is an issue of global public health. All the stages considered, the development of new drugs takes approximately 10 years and spend enormous financial resourses. One of the methods used to reduce such time and budget is defined as rational drug design. In these in-silico simulations, the interactions between a protein or target receptor and inhibitors potential candidates are analyzed. This is a process defined as molecular docking. In order to turn these experiments closer to in-vitro tests, there are innumerous attempts to incorporate the flexibility of the target proteins, for instance using different protein conformations generated by molecular dynamics simulation (MD). Among the many important targets, currently under investigation, we are studying the AcrB efflux pump protein. This protein is found in bacterial cells and is an active transporter of compounds which are captured at the periplasmic space or even at the cytosol. Besides, the efflux pump mechanism is related to the acquisition of bacterial resistance to innumerous compounds, such as dyes, heavy metals, antibiotics, among others; thus making this study important to aid in the research of new drugs to treat infections and other diseases, as Tuberculosis. Results In order to understand the flexibility and regions of higher conformational change of the AcrB protein we performed a MD simulation of this protein using the GROMACS package. We obtained the initial structure from Protein Data Bank (PDB ID: 1IWG). In the first stage of the AcrB MD simulation we generated a 10 ns trajectory. To analyze and validate this initial trajectory we studied the RMSD and the radius of gyration. In an upcoming study, we are going to use these new protein conformations to introduce the AcrB flexibility when looking for new drug candidates for this receptor. However, MD generates a lot of data that need to be interpreted and somehow reduced. Thus, we are going to apply clustering algorithms for effective use of the AcrB flexibility in molecular docking experiments. Conclusion Analysis of the data indicates that the simulated system is unstable, showing a temporal evolution of the structure. Subsequently, other features will be evaluated for the simulation conducted and we are going to extend the simulation until to at least 50 ns. Finally, we will reduce the conformational space of this receptor for effectively using it in molecular docking experiments and searching for new drugs.

Keywords: molecular dynamics simulation, AcrB Efflux Pump

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 192

Development of integrated environment for analysis and annotation of DNA sequences

Liliane Santana Oliveira[1], Alan Mitchell Durham[1], Arthur Gruber[1]

[1]USP, Brazil

Recent advances in the sequencing technology has produced a vast increase in the amount of sequenced data. However the discovery of DNA composition is just the first step of the study of an organism's genome. Once the nucleotide sequence is known, we need to identify the various elements present in the genome. Although there are several tools described in the literature to help researcher in the annotation process, there is a lack of a complete tool with which the user can select, process, analyze, annotate and store sequences the same environment. Artemis, for example, is a tool for visualization and annotation of sequences. It allows users to process sequences, but this processing is done either by web tools where results can't be automatically added to the annotation, or by developing plug ins that are very difficult to code. Apollo is another example of an tool for annotation support. It provides visualization and, but automatic processing is restricted to Blast and PSI-Blast. The Gaggle genome browser allows only the annotation of protein genes, is not connected to a database, and annotations cannot be added automatically. Finally, GenDB is a tool to create workflows to process and to annotate sequences. However the processing follows fixed steps. None of these systems are integrated with a system of controlled vocabulary, so annotation is registered using free text, which makes the posterior data mining a potential challenge. In this project we report the development of PATO (Platform Annotation Tool), a graphic platform for genome annotation. PATO allows the user perform the various tasks of annotation of genome sequences in a single environment. The user select sequences based either on information like source organism, project, or results of precious processing. Each dataset can be submitted to either a pipeline developed dynamically or to one previously stored in the database. Datasets can also be inspected visually and annotated using GO, SO and a locally developed ontology. Visualization separates the results of various programs that are used in the final annotation from those that were ignored in the annotation process. This last feature allows PATO to support the process of curating either manually or automatically annotated sequences. The platform is integrated to the EGene system, which is responsible for processing of sequences and integrates more than 50 annotation programs. The platform is also integrated with a database that allows the management of a large number of sequences.

Keywords: Annotation, selection, curation

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Transcriptomics and Proteomics

Abstract #: 194

Discovery of new microRNA-like molecules in high throughput sequencing data

Liliane Santana Oliveira[1], Alexandre Rossi Paschoal[2], Natália Torres[3], Flávia Maziero Andreghetto[3], Maria Fátima Guarizo Klingbeil[4], Monica Beatriz Mathor[4]

[1]USP, Brazil, [2]UTFPR, Brazil, [3]Hospital Albert Einstein, Brazil, [4]IPEN, Brazil

The implication of post-transcriptional regulation by microRNAs in molecular mechanisms underlying human diseases is well documented. It is known that high throughput sequencing technologies allow for the discovery of novel molecules, due to their inherent sensibility and accuracy. In this work we propose the analysis of sequenced reads based on structural alignment against known microRNA families and ab initio prediction using HHMMIR and RNAfold. We considered the size of putative precursor sequences and the location of the sequenced read within the precursor for data interpretation.

Results

Small RNA libraries for two cell types – a cancer cell and normal oral keratinocytes – were sequenced in a SOLiD equipment. Reads were matched against annotated databases using Small RNA Analysis Pipeline Tool v5. tRNA, rRNA, DNA repeats, and molecules matching miRBase were filtered out. For the identification of novel miRNA-like molecules, all reads having more than 10 copies in our libraries were tested. Reads were mapped to the genome and, at each genomic locus, a longer sequence, extended 100 nt upstream and 100 nt downstream from the read, was extracted for secondary structure analysis. Of the 448 sequences tested, 13 presented some evidence in at least one of the approaches used. Since we sequenced small RNAs, we postulated that any microRNA sequenced in the process was a mature miRNA, and therefore had to be part of a stem in the precursor microRNA secondary structure. Five of the selected sequences matched these criteria. Genomic mapping of these candidates showed that two of them corresponded to adjacent genomic locations. The original reads of these two molecules matched to both sides of the same predicted stem in a microRNA family, constituting, thus, a strong candidate for a new functional molecule. Additionally both candidates were more expressed in the cancer cell line when compared to keratinocytes, indicating a possible role in cancer.

Conclusions

Broadly used bioinformatics approaches were combined in this work for the discovery of microRNA-like molecules in high throughput sequencing data. Within over 400 sequences, 13 presented evidences and only two were selected as strong candidates. This study provides insights on difficulties underlying microRNA discovery in biological samples and highlights the need for careful interpretation of results obtained from tools used for microRNA prediction.

Keywords: microRNA, cell, sequencing

Topic: Transcriptomics and Proteomics

Abstract #: 195

Automated detection of transcript start site (TSS) in dRNA-seq data

Felipe ten Caten[1], Ricardo Zorzetto Nicoliello Vêncio[1], Tie Koide[1]

[1]University of São Paulo, Brazil

Beyond provide essential information about gene response against environmental and genetic modifications, RNA-seq data can be a powerful tool to help the identification of new genetic elements like coding regions, UTRs and ncRNAs. In order to better characterize, in a global way, the elements that compound transcriptomes it is crucial develop methodologies able to process transcriptional information, detect the variations on signal expression and associate this variations to genomic elements with functionality. One of the main point in characterization of genomic elements is the determination of transcript start site (TSS). However, as transcriptome is a plethora of different kinds of RNAs in different stages of development, the measurement of all RNAs levels lead to a high variable signal of expression. Thus, process and segment this signal, in order to obtain the initial position of transcripts, is a hard task. To reduce transcriptome complexity and allow the detection of TSS, Sharma et. al., 2011 proposed a methodology based on selection and sequencing of primary RNAs (differential RNA-seq). Comparison between the sequencing results of libraries enriched for primary RNAs and libraries of total RNAs allowed the manual detection of TSSs in different growth conditions of human pathogen Helicobacter pylori. In order to provide a fast way to identify TSSs in dRNA-seq data and achieve a global identification of new elements in transcriptomes, we implement an algorithm able to perform a signal segmentation to identify peaks associated to transcript start sites. The initial step is the production of a signal from the coverage of reads start position. The first nucleotide of a transcript able to be sequenced will generate a great number of reads starting in the same genomic position, therefore analyze a signal that considers just these positions, instead the total read coverage, reduce the signal variability and make it easier the TSSs detection. The next step is the segmentation of this data using an algorithm which investigate the signal and find the positions with the greatest number of read beginnings, when compared with neighboring positions. Applying this approach at dRNA-seq data of one grow condition of H. pylori we are able to detect with precision the positions enriched for beginning of reads. When compared with signal of total RNA, this positions enable the identification of regions with significant characteristics. An example is the regions with multiple peaks for reads initiation inside a single transcriptional unit, like operons, which can indicate a presence of different TSS positions inside a cluster of genes. The future purpose is prepare dRNA-seq libraries and apply the automated TSS detection in extremophile Halobacterium salinarum in order to identify regulatory RNAs that compound the transcriptional landscape of this important organism for systems biology.

Keywords: transcription start site, RNA-seq, dRNA-seq, ncRNAs

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 196

COMPARATIVE MODELING AND MOLECULAR DYNAMIC OF THE ISOFORM 1 OF THE ATP-DIPHOSPHOHYDROLASE FROM Schistosoma mansoni

Vinicius Schmitz Pereira Nnunes[1], Priscila Faria-Pinto[1], Carlos Cristiano Hasenclever Borges[1], Priscila Vanessa Zabala Capriles[1]

[1]Federal University of Juiz de Fora, Brazil

Schistosoma mansoni is the main responsible for the human Schistosomiasis in South America, Caribbean and Africa. The Schistosomiasis Research Agenda advises about the importance to detect proteins that could be considered as new putative drug targets, mainly due to reports of parasite resistance to the most used pharmacotherapy, the praziquantel. The possible drug targets to be considered are the ATP-diphosphohydrolases (or apyrases). They are enzymes that hydrolyze both the ADP and ATP in the presence of calcium or magnesium, and they possibly participate in the evasion of the host immune system by parasite. S. mansoni has two isoforms of apyrase, Sm1 and Sm2, being the Sm1 the second most expressed protein on the parasite's tegument. Its location, expression rate and function, associated with the role of ATP and ADP in the processes of immune cell activation, reinforcing the importance of the investigation of Sm1 as drug target candidate for Schistosomiasis treatment. We present a three-dimensional (3D) model of Sm1, considering their membrane environment. We selected the templates PDB3CJA and PDB3ZX3 (from Rattus norvegicus) according to identity and coverage, the presence of ligands (phosphoaminophosphonic acid-adenylate ester-ANP, calcium) and conserved waters. We used predictors to confirm the presence of two transmembrane helices (TM1 in N-terminal and TM2 in C-terminal) and one extracellular (ECD) domain. The ECD is subdivided into ECD-I (five helix around five anti-parallel strands), and ECD-II (eight helix around six anti-parallel strands), and the active site is located in the cleft between them, with the catalytic residue E201 and ANP. Due to the absence of templates to model the TM regions, we used informations about secondary structure and Ca-Ca contacts as restrictions in the comparative modeling step. We submitted to a pre-molecular dynamics simulation (p-MD) the system (total of 315,024 atoms) composed by the Sm1 model inserted into POPC membrane (x x y=120 x 120Å), solvated with TIP3P water model, and with following salts: NaCl (150mM), KCl (5mM), MgCl2 (2mM) and CaCl2 (5mM). The steps of p-MD using CHARMM forcefield were: (i) protein restrained minimization (0.5ns); (ii) all-atom minimization (0.5ns); (iii) heating from 277K (crystallographic temperature) to 310K (human host temperature) (1ns); (iii) all-atom equilibration (1ns). The results of the p-MD show that the quality of the Sm1 model increases along the time but the membrane is not yet equilibrated. In the next step, we will extend the p-MD until membrane equilibration, and then, we will perform the production phase of molecular dynamics simulation to study protein movements. We expect analyze the preferable conformational states and cavities' behavior of Sm1, and apply this knowledge in the protein-ligand docking studies of Sm1 and a set of ligands that have been tested experimentally (in vitro) as inhibitor's candidates.

Keywords: Schistosomiasis, ATP-diphosphohydrolases, Comparative Modeling, Molecular Dynamics

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 197

Investigation of the origins of stem cell differentiation

Newton Gontijo Sampaio[1], Diego Trindade de Souza[2], José Miguel Ortega[1]

[1]Universidade Federal de Minas Gerais, Brazil, [2]Universidade de São Paulo, Brazil

INVESTIGATION OF THE ORIGINS OF STEM CELL DIFFERENTIATION NS Gontijo1, DT Souza1,2, JM Ortega1 1 Lab. Biodados, Dept. Bioquímica e Imunologia, ICB, UFMG 2 Lab. de Bioinformática. Dept. de Genética e Biologia Evolutiva, IB, USP Background. Gene Ontology terms have been attributed to proteins, some of them after manual curation. By using genes associated to a general phenomenon, one can use the reliable annotated proteins to create clusters of homologous genes and, by using their taxonomic distribution, to map along evolution the clade (e.g. Eumetazoa, Primates, etc.) of appearance. Thus, one can understand the evolution of the Stem Cell Differentiation processes. Results. Stem Cell Differentiation is the Gene Ontology term GO:0048863. It presents ten children terms that have manually curated entries (varying from three up to 46): GO:0010002 cardioblast differentiation; GO:0060218 hematopoietic stem cell differentiation; GO:0061017 hepatoblast differentiation; GO:0048762 mesenchymal cell differentiation; GO:2000737 negative regulation of stem cell differentiation; GO:0014016 neuroblast differentiation; GO:2000738 positive regulation of stem cell differentiation; GO:0014816 satellite cell differentiation; GO:0048864 stem cell development; GO:0048865 stem cell fate commitment. By using the tool SeedServer we collected all orthologues of the manually curated entries. The three to 46 queries have been organized in two up to 30 clusters per GO term studied, the most frequent being hematopoietic stem cell differentiation, mesenchymal cell differentiation and cardioblast differentiation, with respectively 30, 25 and 20 different clusters of orthologous genes. They have grouped genes from 1 up to more than 2000 distinct organisms. The first gene of these three GO processes has appeared in cellular organisms, except for mesenchymal cell differentiation where the first gene appears only by Metazoa. The distribution of gene appearance for hematopoietic stem cell differentiation follows the appearance of the total of genes in humans, while most (72%) of the genes for mesenchymal cell differentiation are raised between the époque of Metazoa and Coelomata. For the cardioblast differentiation process, a peak of gene appearance rises in clades Eumetazoa and Bilateria (50%), with some others appearing in by the époque of Chordata and Euteleostomi. The process is accomplished with the addition of very recent genes, respectively appearing only in clades level 30, 20, 32, thus mesenchymal cell differentiation being the process which is accomplished earlier in evolution. Conclusions. Evolution of the Stem Cell Differentiation process varies according to the cell studied. The époque of origin of the comprised genes may follow the appearance of the total of genes in human genome, but it can follow specific patterns. By combining Gene Ontology process terms attributed manually with the analysis of clusters of homologous genes one can understand the evolution of the Stem Cell Differentiation processes. Supported by: Capes, FAPEMIG, FAPESP, CNPq.

Keywords: Stem cell differentiation, Gene Ontology, Evolution, SeedServer

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 198

Preliminary validation study for SNPs associated to rib eye area in Canchim beef cattle

Andressa Oliveira de Lima[1], Fabiana Barichello Mokry[2], Roberto Hiroshi Higa[3], Maurício de Alvarenga Mudadu[4], Sarah Laguna Conceição Meirelles[5], Luciana Correia de Almeida Regitano[4]

[1]UFSCar, Brazil, [2]UFSCar/CNPTIA, Brazil, [3]CNPTIA, Brazil, [4]CPPSE, Brazil, [5]UFLA, Brazil

Keywords:

Topic: Transcriptomics and Proteomics

Abstract #: 199

Analysis of RNA sequencing data from sugarcane plants submitted to microgravity

Helaine Cristiane Silva[1], Diego Gomes Texeira[1], João Paulo Matos Santos Lima[1], Katia Castanho Scortecci[1]

[1]Universidade Federal do Rio Grande do Norte, Brazil

Background: Plants are sessile organisms that depend on gravity to guide its growth and development through a phenomenon called graviotropism. Thus, microgravity has been considered as a stressor factor that it can trigger changes in the plant homeostasis like others abiotic stressors (drought, salinity, cold). The model of this study is the sugarcane plants (Saccharum spp.), that has an economic important as it is source for sugar and ethanol production. Due to the sugarcane importance in the global economy, it is important to understand how this plant can be affect by different abiotic then sugarcane plants were submitted to microgravity using the souding rocket VBS-30. Samples from leaves and roots were isolated for for RNA extraction and sequencing using the Illumina plataform. Results: First, the quality of data obtained by the Illumina sequencing was validated by NGS QC package, running the script IlluQC.pl using phred> 20 (quality value associated with the likelihood of one in a hundred of base being misnamed). Then, the reads filtered contained in the fastq file were grouped into contigs using the Velvet package through three different values of k-mers: 35, 41 and 47. The output files from different k-mers were combined according to the pipeline ANGUS 2.0, which includes the following programs: CD-HIT, Minimus2 and AMOS, generating the final contigs used in the annotation. The IlluQC generated graphs that demonstrate the excellent quality of the RNA sequences from all samples. It was observed that the filtered high quality sequences have an average phred score over 23. It was noted a substantial reduction in the size of sequences for each library during each assembly stage. Conclusions: The sequences obtained will be used for annotation and for calculating the differential expression from the libraries in order to identify the molecular responses related to microgravity, as well as the possible signaling pathways related to abiotic stresses. Supported by: AEB, FINEP, FAPERN, INCT-INEspaco/CNPq.

Keywords: RNA sequencing, sugarcane, microgravity

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 200

Intragenomic Non-homologous Isofunctional Enzymes (NISE) in Leishmania amazonensis

Leandro Pereira[1], Monete Gomes[1], Diogo Tschoeke[1], Jeremy Mottram[2], Alberto Dávila[1], Antonio Miranda[1]

[1]Oswaldo Cruz Institute, Oswaldo Cruz Foundation, Rio de Janeiro, RJ, Brazil., Brazil, [2]University of Glasgow, United Kingdom

Background: Leishmaniasis is a disease caused by a complex of more than 20 protozoan parasites of the genus Leishmania. Among these species, L. amazonensis is one of the main pathogens of such disease in Brazil. The severity of the disease ranges from cutaneous forms to fatal visceral forms. So far there are no available vaccines and treatments are based on chemotherapies that are highly toxic and not specific to the parasite. In this respect, desirable characteristics for a candidate drug that minimize the risks of resistance include specificity (lack of the target in the vertebrate host) and the identification of isoforms or unrelated enzyme forms involved in the same metabolic steps or in alternative biochemical routes. Therefore, the aim of this study is the identification of Non-Homologous Isofunctional Enzymes (NISE) in L. amazonensis proteome. NISE display the same functional activity and no significant similarity is found either between their primary or between their tertiary structures. Being the result of independent evolutionary events, NISE may be involved in alternative metabolic pathways.

Keywords: Non-homologous isofunctional enzymes, Analogous enzymes

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 201

Non-homologous Isofunctional Enzymes (NISE) between Leishmania amazonensis and Homo sapiens: a source for potential drug targets

Monete Gomes[1], Leandro Pereira[1], Diogo Tschoeke[1], Jeremy Mottram[2], Alberto Dávila[1], Antonio Miranda[1]

[1]Fiocruz, Brazil, [2]University of Glasgow, United Kingdom

Background: Leishmaniasis is a disease caused by a complex of protozoan parasites of the genus Leishmania. It is mainly a rural disease, prevalent in 88 countries, with an estimated number of 1.6 million new cases annually, causing considerable mortality which implies in economic and health burdens. In humans, leishmaniasis is usually divided in four different clinical forms: cutaneous, diffuse cutaneous leishmaniasis, mucocutaneous and visceral. In the New World, L. amazonensis is the main causative agent of the cutaneous form. Presently, there are no commercially available vaccines for such disease, and there is no effective treatment against it, since current drugs present many side effects and resistance. This study aims the identification of Non-Homologous Isofunctional Enzymes (NISE) between L. amazonensis and Homo sapiens, since these enzymes do not have structural similarities, and therefore they have the potential to be studied as candidate therapeutical targets. Results: Initially we identified 25 potential cases of NISEs when we compared the proteome of L. amazonensis with the proteome of H. sapiens with the clustering pipeline AnEnPi (Analogous Enzyme Pipeline, www.dbbm.fiocruz.br/AnEnPi/). The folding patterns and superfamily classification of these 25 enzymes were checked against the SUPERFAMILY database. They were considered NISEs only if they had totally different folds under the same EC number (Enzyme Commission number). For those cases where there were no significant hits against the SUPERFAMILY database we gave the classification as “Predicted NISE”. 15 NISE cases were confirmed and 1 was considered “Predicted NISE”. Finally, we searched for the confirmed NISEs in drug targets databases (TDR targets, TTD and DrugBank). Conclusions: NISE identification between L. amazonensis and H. sapiens produced a list of 15 potential drug targets, some of them already under study as therapeutical targets against other parasites, which suggests that they may also be suitable drug targets for L. amazonensis as well.

Keywords: Non-homologous Isofunctional Enzymes, analogous enzymes, Leishmania amazonensis

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 202

Identification of inhibitors of snake venom class P-I metalloproteases using molecular modeling techniques.

Raoni Souza[1], Mariana Alves[1], Rodrigo Ferreira[1], Adriano Pimenta[1], Ronaldo Nagem[1], Eladio Sanchez[2], Rafaela Ferreira[1]

[1]Universidade Federal de Minas Gerais, Brazil, [2]Fundação Ezequiel Dias, Brazil

Approximately 90% of accidents with venomous snakes in Brazil are caused by Bothrops species. Their venom induce various local and systemic effects, characterized by a high proteolytic and hemorrhagic activity mainly caused by 'snake venoms metalloproteases' (SVMPs), a group of toxins with a highly conserved catalytic site containing zinc. Drug-like inhibitors of these toxins could therefore represent a complementary procedure to serum therapy and may help to develop more effective treatment for the local effects. The search for such inhibitors can be optimized through molecular modeling techniques.

In this work we used the tridimensional structures of two P-I SVMPs, the hemorrhagic toxin atroxlysin-I, of Bothrops atrox, and the non-hemorrhagic toxin leucurolysin-a, of Bothrops leucurus, as models to find SVMPs inhibitors. We used the x-ray structure of leucurolysin-a and produced a three-dimensional structure model of atroxlysin-I by comparative modeling to perform virtual screening and molecular dynamics simulations. This model was built using Modeller and four SVMP toxin templates selected from PDB with sequence similarity bigger than 50% and structure resolution better then 2Å. The model structures with higher DOPE-score values were used for further analysis through quality evaluation programs. The best model was then selected for the dynamics simulations. The templates were also evaluated with these programs aiming to compare the scores of all the structures.

To search for inhibitors of these toxins 1.4 million compounds from the ZINC database were virtually screened using Autodock Vina. The quality of results obtained in this step could be further improved if used in conjunction with molecular dynamics simulations for the production of protein conformations and free energy calculation. To obtain these conformations, molecular dynamics simulations were performed with the structures of both toxins using NAMD on NPT conditions. After the minimization and equilibration of the system, verified by measuring the variation of total energy and RMSD values, 20 ns of production runs were performed, extracting protein structures at every 50 ps. These structures were subsequently clustered with Chimera using the RMSD values of aminoacid residues near the active site. Representative structures of each of the topten most populated clusters were then selected for a second virtual screening round with compounds selected in the first screening round.

The next step is to select compounds ranked in virtual screening by visual inspection to perform free energy calculation. Due to the difficulties of choosing parameters to be used in this calculation by molecular dynamics, experimental binding affinity of known inhibitors can be used to guide this procedure. Hence broad-spectrum inhibitors were selected to obtain experimental data by inhibition of proteolytic activity, enzyme kinetics and calorimetry. These include inhibitors of other SVMPs or of ADAM and MMP, which are metaloproteases with very similar active sites to SVMPs.

Keywords: snake venoms metalloprotease, docking, molecular dynamics, inhibitors

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 203

Whole genomes obtained of Dengue, Yellow Fever and distinct groups of arboviruses and zoonotic RNA viruses using the semiconductor sequencing method

Layanna Freitas de Oliveira[1], Janaina Mota de Vasconcelos[1], Daisy Elaine Andrade da Silva[2], Jedson Ferreira Cardoso[1], João Lídio da Silva Gonçalves Vianez Júnior[2] Pedro Fernando da Costa Vasconcelos[2], Márcio Roberto Teixeira Nunes[2]

[1]Universidade Federal do Pará, Brazil, [2]Instituto Evandro Chagas, Brazil

Background: In the microbial word, the needs of improvement for genetic characterization of full-length genomes are evident in terms of high coverage, accuracy and protocols for RNA viral whole genome sequencing. To date, obtaining of whole genome data for different organisms has been increased exponentially with Next Generation Sequencing methods development. Among the current platforms, Ion Torrent PGM, has simply and automated process, with only two days of benchtop preparation. There is no published protocol described for RNA virus sequencing on Ion PGM, only for DNA virus and, for that, it was used specific primers. We describe by the first time a new approach for RNA virus genome sequencing on Ion Torrent PGM using viral samples belonging to distinct viral families (Bunyaviridae, Flaviviridae and Rhabdoviridae). This method uses an initial RT and PCR approaches for double-strand cDNA synthesis with random primers, followed by library preparation on Library Builder System. Both RNA shearing methods (sonication and enzymatic cleavage) were applied. Emulsion PCR was performed on OneTouch2 System and Ion Spheres with templates loaded on 318v2 chip. Results: The sequencing has generated 164 M bases in average. Polyclonal data (27%) and 31% low quality reads (Phred score based) were removed, obtaining 4,109,732 usable reads, which resulted on a 15-fold coverage. Using 200bp chemistry, a mean read size of 160 bp was produced. There were no differences of RNA shearing regarding the genomes assembly. Complete genomes, including the terminal 5' and 3' non coding regions, were assembled by the De Novo method using the MIRA 4 algorithmic. For viral species with genome sequences available at the GenBank database, the alignment was made against reference genomes which reinforced the assembly obtained by this method. Sequences generated by the Ion PGM were longer than those observed for the reference genomes deposited in the GenBank. Conclusions: The current work demonstrated the feasibility of complete viral RNA genomes by the Ion Torrent PGM sequencer, using a combination of cDNA synthesis, library preparation and massive DNA sequencing through the semiconductor technology.

Keywords: Ion Torrent PGM; RNA virus; sequencing; genome assembly

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 204

Molecular basis for the allosterism in the dimeric retinoid X receptor's DBD/DNA interaction: A molecular dynamics study

Leonardo Henrique França de Lima[1], Leandro Martínez[2]

[1]Universidade Federal de São João Del-Rei, Brazil, [2]Universidade Estadual de Campinas, Brazil

Molecular basis for the allosterism in the dimeric retinoid X receptor's DBD/DNA interaction: A molecular dynamics study LH Lima1, L Martínez 1Campos de Sete Lagoas, Universidade Federal de São João Del-Rei, Sete Lagoas - MG 2Instituto de Química, Universidade Estadual de Campinas, Campinas – SP Background: Nuclear receptors (NRs) are ligand sensitive transcription factors, representing the last link between system physiology and cell transcriptional control. Despite a number of structural and biophysical data, the detailed mechanisms of the cooperative interaction of their dimeric DNA binding domains (DBDs) with the DNA and how it could be propagated to the other domains remains still poorly understood. Results: Here we probe the molecular basis for the allosterism in the interaction of the homodimeric DBD of the retinoid X receptor (RXR) with its specific hormone response element (HRE) in DNA through a systematic study by molecular dynamics (MD) simulations. Simulation analysis indicates a central hole of the unfolding of the C-terminal extension (CTE) from the DBD, both for an initial readjust of the domain global dynamics, as for an induced protein:DNA and (in less extension) protein:protein fitting. The allocation of the CTEs of the two respective domains at the DNA minor-groove stimulates a local DNA bending that, in turn, optimizes the specific protein:DNA contacts at the major-grooves. Such major groove contacts favored by the CTE insertion at the minor one involve residues at the DBD corn apparently correlated along the NR evolution. Also, and surprisingly, this protein:DNA induced fit promotes a global DBD structural and dynamic rearrangement that contributes for water exclusion along the domain hydrophobic surfaces. Such rearrangement occurs in regions distant from the protein:DNA interface and, according recent structural data, involved in contacts between the DBD and the ligand binding domain (LBD). Conclusions: Taking all together, these results shed new lights on the mechanisms of cooperativity in the dimmeric interaction of NR-DBDs with specific HREs, correlating some features that have been individually suggested to be important in such process by literature experimental data – i.e, CTE local unfolding and insertion at the minor-groove, protein:DNA induced fit, water exclusion from hydrophobic surfaces and long range allosteric modulation of the NR surface by the HRE interaction – but, up to so, without a complete view encompassing all of them and their mutual inter-relationship. Supported by: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Conselho Nacional de Pesquisa e Desenvolvimento Fundação de Amparo à Pesquisa do Estado de São Paulo

Keywords: Nuclear Receptors, DNA Binding Domain, Molecular Dynamics Simulation, Statistic Coupling Analysis, Retinoid X Receptor

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 205

In silico evaluation of 5 'and 3' UTRs gamma globin gene regions and identification of regulatory elements motifs associated with erythropoiesis and globin genes regulation

Larissa Paola Rodrigues Venancio[1], Gisele Cristine Souza Carrocini[1], Claudia Regina Bonini-Domingos[1]

[1]UNESP - Universidade Estadual Paulista, Brazil

Background: Sickle cell disease and beta thalassemia are among the most common genetic disorders and higher impact on mortality and overall health. One of the most important factors associated with improvement in the clinical picture is the production of fetal hemoglobin (Hb F), regulated by multiple genetic factors. The Hb F regulation process is influenced by transcription factors (TF), acting in the regulation of globin genes and changing the expression of HbF by Hb A. Untranslated regions (UTR) are typically rich in motifs FT that influence the stability, localization and translation of mRNA. The aim of this study was to evaluate the 5' and 3' UTRs gamma globin gene in order to understand the genetic determinants that act in modulating the levels of Hb F. For this analysis, we used the program MAPPER (Multi-Genome Analysis of Positions and Patterns of Elements of Regulation) which evaluates large-scale transcriptional regulation, based on hidden Markov chain models for building alignments of known regions. For analysis, we aligned the HBG1 and HBG2 Homo sapiens gene sequences against those from Mus musculus and Rattus norvegicus. The tool identifies TF recognition sites by combining the TRANSFAC and JAS–PAR. Results: The results indicated the occurrence of 8 motifs of transcription factors in 5'UTR and 27 3'UTR . We highlight motifs that are related to erythropoiesis and globin genes regulation, such as RUNX1 (haematological stem cells differentiation into mature blood cells), PBX (repression of globin genes expression), both in the 5'UTR; HAT5 (acts on acetylation of erythroid Krüppel-like factor, which are associated with the regulation of the expression of gamma globin), SOX5 (beta globin activation), BCL6 (related to cooperation with the BCL11A transcription factor in silencing of gamma globin gene expression, affecting the levels of Hb F), RXRA (associated with erythropoietin expression, which promotes proliferation of erythropoietic progenitor cells during the fetal stage), and SP1 (works with GATA - 1, the main erythroid transcription factor in exchange for the adult and fetal hemoglobin), all these in the 3'UTR . Conclusions: The data suggest that these regions may be related not only to the transcriptional regulation and on the production of fetal and adult hemoglobin, but with the regulation of other genes in the cluster beta globin. Knowledge of regulatory elements globin gene may help to develop therapeutic strategies that affect improvement of the clinical condition of patients with beta- hemoglobinopathies.

Keywords: UTRs, transcription factor, MAPPER, erytropoiesis, gamma globin regulation

Topic: Transcriptomics and Proteomics,Human Pathophysiology; Animal Models of Disease

Abstract #: 206

Musashi1 Overexpression: Global Impact on Intestinal Cells

Bruna R. S. Correa[1], Michela Plateroti[2], Luiz O. F. Penalva[3], Pedro A. F. Galante[1]

[1]Hospital Sírio-Libanês - Centro de Oncologia Molecular, Brazil, [2]University Claude Bernard Lyon 1, France, [3]University of Texas Health Science Center at San Antonio, United States

Background: RNA-binding proteins (RBPs) are critical post-transcriptional gene expression regulators. Alterations in their expression can impact the cellular proteome and lead to diseases states, like tumorigenesis. Musashi1 (Msi1) is a highly conserved RBP that has multiple functions mediating several post-transcriptional processes. Musashi1 participates in gene networks related to cell cycle, proliferation, differentiation and apoptosis. Furthermore, Msi1 is an adult stem cell marker in several tissues and play a crucial role in the balance between self-renewal and differentiation. Over-expression of Msi1 has been associated with tumor development besides less differentiated and more aggressive phenotype in multiple tumors, such as colon, breast, melanoma, medulloblastoma and glioblastoma. In colon cancer, high expression of Msi1 is related to higher clinical stages, poorer overall prognosis and survival. Methods and Results: To understand the impact of Msi1 overexpression, we explored the global gene expression of an intestinal cell population from murine models that overexpress Msi1 (Msi1-Over), with wild-type mice (WT). We used RNA-Seq to sequencing the transcriptome of the both conditions. Then we performed the sequences mapping against the mouse reference genome (mm10) using TopHat. Next the reliable alignments were selected and the gene expression quantified. Afterward we use three methodologies to identify differentially expressed genes: Cuffdiff2, DESeq and EdgeR, and only those genes identified by at least two methods were selected. We found 1,018 down-regulated genes and 1,365 up-regulated genes in Msi1-Over. Then we mapped all differentially expressed genes in Gene Ontology and KEGG Pathways databases, for functional enrichment analysis. Among the enriched GO terms and pathways, those related to cell cycle appeared, consistent with the Msi1 role. Moreover, we observed the up-regulation of genes from Notch and Wnt pathways. Conclusions: The up-regulation of genes of these pathways is consistent with the overexpression of Msi1 in intestinal cells, supporting that Msi1 can increase cell proliferation and induce tumorigenesis through the activation of both the Notch and Wnt pathway. Supported by: FAPESP.

Keywords: RBPs, Musashi1, Gene Expression, RNA-Seq, Intestinal Cells

Topic: Transcriptomics and Proteomics

Abstract #: 207

TRANSCRIPTOME ASSEMBLY OF THE GIANT SUGARCANE BORER (TELCHIN LICUS LICUS)

Lucas Pimenta[1], Sérgio Alencar[1], Maria Fátima Grossi-de-Sá[2]

[1]Universidade Católica de Brasília, Brazil, [2]Embrapa Recursos Genéticos e Biotecnologia, Brazil

Sugarcane (Saccharum spp. hybrid) is produced in more than 100 countries, of which Brazil is the largest producer. The expansion of the production of sugar and alcohol derived from sugarcane and the growth of mechanized harvesting are changing the entomofauna within sugarcane fields. Insect pests are an important biotic stress that can affect sugarcane production. One of the main threats to sugarcane is the giant sugarcane borer, Telchin licus licus Drury, 1773 (Lepidoptera: Castniidae), which has been a pest problem only in Brazilian northeastern states but was also detected in the state of São Paulo in 2008. The larva can reach up to 8cm and causes severe damage to crops, reducing its biomass, therefore reducing the sugar and alcohol production. To evaluate methods to control the giant sugarcane borer, a transcriptome of its midgut was assembled. RNA was extracted to make a cDNA library and sequencing was performed with the 454 GS FLX platform, resulting in 653988 reads. The quality of raw data from pyrosequencig was analyzed with FASTQC V0.10.1. Adaptors were removed and low quality sequences were trimmed with Trimmomatic-0.30, and the homopolymers were removed with NGSQCToolkit_v2.3. The output data with a total of 192822 reads was used to perform a de novo assembly using MIRA-4.0 that assembled into 8358 contigs with an average length of 382bp, without singlets. Further analysis will be performed with the transcriptome.

Keywords: moth, economic lost, biological control

Topic: Transcriptomics and Proteomics

Abstract #: 208

Detection of differentially expressed SNVs in miRNA target binding sites: preliminary results

Natasha Andressa Nogueira Jorge[1], Carlos Gil Ferreira[1], Fabio Passetti[1]

[1]Instituto Nacional de Câncer (INCA), Brazil

Background MicroRNAs are small non coding RNAs capable of controlling gene expression by pairing with target sites in the 3' UTR region of messenger RNAs. Allelic imbalance is the specific expression of each allele of an individual. This expression difference has several causes and is present in different pathologies, including cancer. Using High-throughput sequencing technology, one may quantify haplotypes and identify new single nucleotide variations (SNVs), therefore it can be applied to the quantification of allelic imbalance in tumors. Our objective is to identify differentially expressed SNVs that may affect known and putative microRNAs target binding sites using public RNA and microRNA high-throughput sequencing data. Results We used the data published by Kim and collaborators (2013) in which the authors sequenced the mRNAs and microRNAs of six non-smokers women with lung cancer to start the development of a methodology to detect differentially expressed SNV in microRNA target biding sites. In our preliminary analysis using the microRNA-Seq samples, we have detected 95 different SNVs in common between normal and tumor samples for at least 4 patients and 65 SNVs in common for all six patients. We were also able to detect 4 SNVs present only in normal samples for at least 4 patients and only 1 SNV in common for all patients. 14 SNVs present only in tumor samples and only 1 SNV for all patients were also detected. Conclusions This initial approach shows that there are several SNVs in both tumor and normal samples, but only a few are present for more than 4 patients. This analysis will be repeated for other datasets and extended to the identification of putative microRNA biding sites and differentially expressed SNVs.

Keywords: allelic imbalance, microRNA target, single nucleotide variations

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 209

Combination of ab initio folding and homology methods to generate a full-length model of Histidine Kinase PhoR Sensor Protein of Corynebacterium pseudotuberculosis

Gleiciane Leal Moraes[1], Alberto Monteiro dos Santos[1], Cinthia Cunha Maradei Pereira[1], Jerônimo Lameira Silva[1]

[1]UFPA, Brazil

Corynebacterium pseudotuberculosis is the causative agent of several veterinary diseases in a broad range of economically important hosts. This pathogen has a two-component regulatory system known as PhoPR, which consists of a sensory histidine kinase protein (PhoR) and an intracellular response regulator protein (PhoP). This regulatory system is involved in the regulation of proteins present in various processes, including virulence by phosphoryl group transfer from a conserved histidine residue of the PhoR to a conserved aspartate residue of the PhoP. Here we describe the first 3D structure of the sensor histidine kinase PhoR of C. pseudotuberculosis obtained through molecular modeling using a combined homology and ab initio approach. The model generated provides structural information for a homodimeric membrane receptors full-length histidine kinase class I (I HKs). Each monomer contains a short N-terminal cytoplasmic domain followed by a transmembrane α helix (TM1) and a periplasmic sensory domain that is connected via a second transmembrane α helix (TM2) to the cytoplasmic. A conserved helix-turn-helix motif HAMP domain connects TM2 to the cytoplasmic catalytic core HK. The catalytic core consist of two-helix dimerization and histidine-containing phosphotransfer (DHp) domain, connecting to a ATPbinding (CA) globular domain C-terminus that phosphorylates the histidine. The sensory domain shows extremely low sequence similarity among related two-component sensory systems, therefore the portion N-terminal was modeled by ab initio relax protocol using Rosetta Software Suite. 50.000 structures were generated, selection of the 100 lowest energy structures by cluster-based score rosetta and clustering with maxcluster. 48.000 models generated for refinement of the sensor. Models were joined using the Modeller program. Ligand binding to the periplasmic sensory domain of sensor HKs constitutes the trigger for subsequent transmembrane signaling events, leading to quaternary structural changes within the dimer and ultimately autophosphorylation in cytoplasmic kinase domain. The HAMP domain and catalytic core HK (DHp and CA domains) were modeled by molecular modeling by homology in Modeller program using as templates 2LFR (39% identity) and 2C2A (31% identity), respectively. The templates were selected from Protein Data Bank (PDB) based in resolution and sequence similarity. Models were joined using the Modeller program. The final homodimeric structural obtained contains 1002 amino acids and showed 96.9% of the residues in the most favored regions of the Ramachandran plot. In addition, to mimic the membrane of gram-positive bacteria, we use a 3:1 mixture of palmitoyloleoylphosphatidylglycerol (POPG) with palmitoyloleoylphosphatidylethanolamine (POPE) and it was build a protein-membrane complex in CHARMM-GUI website (http://www.charmm-gui.org). The 3D obtained for the full-model containing all the features of PhoR show structural compatibility with the experimental data described in the literature. This structure will be used as a model for further studies using molecular dynamics, docking, and drug development for Histidine Kinase of Corynebacterium pseudotuberculosis.

Keywords: Ab initio, Molecular Homology, Histidine Kinase PhoR, Protein-Membrane Complex

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics,Human Pathophysiology; Animal Models of Disease

Abstract #: 210

Molecular Modeling of New triazole Inhibitors of α -glucosidases with Therapeutic Potential in Type II Diabetes

Paulo Vinícius Sanches Daltro de Carvalho Daltro[1], Mário R. Senger Senger[1], Sabrina B. Ferreira Ferreira[2], Carlos R. Kaiser Kaiser[2], Vitor F. Ferreira Ferreira[3], Floriano Paes Silva Junior Silva[1]

[1]Fiocruz - RJ, Brazil, [2]Universidade Federal do Rio de Janeiro, Brazil, [3]Universidade Federal Fluminense, Brazil

Recent numbers of cases of Type II diabetes demonstrate the urgency to develop new drugs for the treatment of this pathology. Our group has synthesized glycoconjugate triazole compounds (GTCs) with 20X greater inhibitory activity against yeast MAL12 (Saccharomyces cerevisiae maltase) than acarbose (a hypoglycemiant α-glucosidase inhibitor in clinical use). Studies on the kinetic mechanism of inhibition of these compounds showed that all inhibitors act non-competitively both on yeast maltase and porcine (Sus scrofa) pancreatic α-amylase (PPA). Based on the kinetic data, a structural mechanism of inhibition for both enzymes by GTCs has been hypothesized by our group where these compounds are bound in the +2 and +3 subsites, adjacent to the substrate (4-nitrophenyl-α-D-maltoglucoside - pNPG for Mal12 and 2-chloro-4-nitrophenyl-α-D-maltotrioside - CNPG3 in the case of PPA). Since there is a lack of experimental data on the structural aspects of the inhibition process of the GTCs, we are currently conducting a comparative analysis of the active site topology of three enzyme belonging to the GH13 family: MAL12, PPA and human pancreatic α-amylase (HPA). In agreement with the results of non-competitive inhibition, we used both free enzyme (E) and enzyme - substrate (ES) complex as receptors in docking simulations of the GTC ligands using the induced-fit docking (IFD) protocol with help of the Glide and Prime softwares (Schrödinger, LLC). Preliminary results with yeast maltase confirm the hypothesis raised by our group where the GTCs are preferentially bound in subsites adjacent to the substrate. The results will allow us to rationalize the experimental inhibition results on MAL12 and PPA and to determine the possible differences in enzyme-inhibitor interactions within the human enzyme in order to obtain more efficient inhibitors in the treatment of type II Diabetes.

Keywords: Glycosidases; Triazoles; Molecular Modeling; Type II diabetes.

Topic: Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 211

A relational database for the investigation of functional SNPs from the 1000 Genomes Project

Sérgio de Alencar[1], Gabriel Fernandes[1]

[1]Universidade Católica de Brasília, Brazil

The rapid progress in high-throughput sequencing platforms has led to a huge increase in Single Nucleotide Polymorphism (SNP) discovery. Currently, these markers are widely used in plants and animals for Marker Assisted Selection (MAS), and in humans for Genome Wide Association Studies (GWAS). Resequencing studies, which are carried out when there is a reference genome available, can be done using many individuals in order to obtain a SNP frequency for a whole population, or whole group of individuals of interest. Recently, the public "1000 Genomes Project" carried out the resequencing of 1,092 human genomes obtained from individuals from different populations around the world, providing an invaluable source of SNP frequency information. In this study, we report the construction of a database containing results from in silico functional prediction tools (PolyPhen-2, SIFT, SNPEff and CONDEL) which were used in order to evaluate the potential impact of coding and non-coding SNPs identified in the "1000 Genomes Project". Using dbSNP accession numbers as common identifiers, this database also integrates information from several other databases relevant to SNP functional studies, including OMIM, Ensembl, GeneCards and GWASdb. This allows not only the filtration of potentially damaging SNPs using various sources of evidence, but also to obtain allele frequencies determined by the "1000 Genomes Project". Future projects include the creation of a website containing various search options which can be used to access all information contained in this database, including the option to search rare SNPs, which have been recently suggested to be more frequently associated with diseases than common SNPs.

Keywords: SNP; functional impact

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 212

Small-angle X-ray scattering and structural modeling of full-length cellobiohydrolase I from Trichoderma harzianum

Leonardo Henrique França de Lima[1], Viviane Serpa[2], Flávio Rosseto[2], Mário Oliveira Neto[3], Leandro Martínez[4], Igor Polikarpov[2]

[1]Universidade Federal de São João Del-Rei, Brazil, [2]Universidade de São Paulo, Brazil, [3]Universidade Estadual Paulista, Brazil, [4]Universidade Estadual de Campinas, Brazil

Small-angle X-ray scattering and structural modeling of full-length cellobiohydrolase I from Trichoderma harzianum LHF Lima1, VI Serpa2 , FR Rosseto2, MO Neto3, L Martínez4, I Polikarpov2 1Campos Sete Lagoas, Universidade Federal de São João Del-Rei, Sete Lagoas – MG, Brazil 2Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos – SP, Brazil 3Instituto de Biociências de Botucatu, Universidade Estadual Paulista, Botucatu – SP, Brazil 4Instituto de Química, Universidade Estadual de Campinas, Campinas – SP, Brazil Background: Cellobiohydrolases hydrolyze cellulose releasing cellobiose units. They are very important for a number of biotechnological applications, such as, for example, production of cellulosic ethanol and cotton fiber processing. The Trichoderma cellobiohydrolase I (CBH1 or Cel7A) is industrially important exocellulase. It exhibits a typical two domain architecture, with a small C-terminal cellulose-binding domain and a large N-terminal catalytic core domain, connected by an O-glycosylated linker peptide. The mechanism by which the linker mediates the concerted action of the two domains remains a conundrum. Results: Here, we probe the protein shape and domain organization of the CBH1 of Trichoderma harzianum (ThCel7A) by small angle X-ray scattering (SAXS) and structural modeling. Our SAXS data shows that ThCel7A linker is partially-extended in solution. Structural modeling suggests that this linker conformation is stabilized by inter- and intra-molecular interactions involving the linker peptide and its O-glycosylations. Conclusion: The union of the experimental low resolution data with the structural high resolution modeling approach here presented allowed first insights about the domain organization, structure and dynamics of this industrially importante class of enzymes. Such insights are significant for the comprehension of the domain concerted cellulase action, being particularly promising for future studies of enzyme engineering. Supported by: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Conselho Nacional de Desenvolvimento Científico Fundação de Amparo à Pesquisa do Estado de São Paulo

Keywords: Cellulase, Celullose, Cellobiohydrolase I, Small Angle X-Ray Scattering, Structural Modeling, Molecular Dynamics Simulation, Simulated Annealing

Topic: Transcriptomics and Proteomics

Abstract #: 214

SpliceProt: a protein sequence repository of predicted human splice variants

Raphael Tavares[1], Nicole Scherer[1], Bianca Alves Pauletti[2], Elói Araújo[3], Edson Luiz Folador[1], Gabriel Espindola[1], Carlos Gil Ferreira[1], Adriana Franco Paes Leme[2], Paulo Sérgio Oliveira[2], Fabio Passetti[1]

[1]INCA/LBBC, Brazil, [2]LNBio, Brazil, [3]Universidade Federal do Mato Grosso do Sul, Brazil

Background: The mechanism of alternative splicing in the transcriptome may increase the proteome diversity in eukaryotes. Recent Bioinformatics analysis using RNA-Seq experiments showed that approximately 90% of human genes produce more than one transcript due to alternative splicing events. The impact of alternative splicing in the human proteome has been the focus of Bioinformatics approaches and it is expected that most of these alternative transcripts can alter the polypeptide chain produced after its translation. Due to its importance, many studies have been developed focused on the identification of alternative splicing events based on cDNA data and also in mass spectrometry experiments using protein repositories. Nonetheless, even high quality spectrums are not identified in these experiments frequently. In proteomics, several studies aim to use protein sequence repositories to annotate mass spectrometry experiments or to detect differentially expressed proteins. However, the available protein sequence repositories are not designed to fully detect protein isoforms derived from mRNA splice variants. To foster knowledge for this field, here we introduce SpliceProt, a new protein sequence repository of putative splice variants based on the human transcriptome. Results: Current version of SpliceProt contains 159,719 non-redundant putative polypeptide sequences. The assessment of the potential of SpliceProt in detecting new protein isoforms resulting from alternative splicing was performed by using publicly available proteomics data. We detected 173 peptides hypothetically derived from splice variants, which 54 of them are not present in UniprotKB/TrEMBL sequence repository. In comparison to other protein sequence repositories, SpliceProt contains a greater number of unique peptides and it is able to detect more splice variants. Conclusions: SpliceProt provides a solution for the annotation of proteomics experiments regarding splice isofoms.

Keywords: Alternative splicing, mass spectrometry, proteomics

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Transcriptomics and Proteomics

Abstract #: 215

Identification of proteins involved in pathways of RNA mediated gene silencing small in flatworms

Santiago Fontenla[1], Lucas Maldonado[2], Nicolas Dell'Oca[3], Mara Rosenvit[2], Pablo Smircich[4], Laura Kamenetzky[2], José Tort[3]

[1]Departamento de Genética, Facultad de Medicina (UDELAR), Uruguay, [2]Instituto de Microbiología y Parasitología Medica (IMPaM), CONICET, Facultad de Medicina - Universidad de Buenos Aires, Argentina, [3]Departamento de Genética, Facultad de Medicina (UDELAR), Uruguay, [4]Departamento de Genética, Facultad de Medicina (UDELAR); Laboratorio de Interacciones Moleculares, Instituto de Biología, Facultad de Ciencias (UDELAR), Uruguay

Various Small RNA regulated gene expression pathways orginally discovered in model organisms, are widely conserved in metazoans, revealing a common cellular machinery. miRNAs generated in the nucleus specifically block translation of some mRNA showing an endogenous mechanism of regulation. Other pathways associated with the presence of exogenous RNAs, that presumably emerged in response to viral infections, leads to the specific degradation of messengers (siRNA), while other small nuclear RNA (piRNA) participate in transposon silencing. The mediators and pathways underlying these processes seem to be strongly interrelated. In the model nematode Caenorhabditis elegans 66 different proteins have been identified as related to these small RNA mediated pathways. We used them as baits to find orthologs in the available genomes and transcriptomes of flatworms, consisting of 4 cestodes genomes, 3 genomes and 4 transcriptomes of trematodes and 2 partial genomes and transcriptomes of free living turbellarians. In relation to C.elegans we found a reduced set of orthologous proteins, consisting of 33 major mediators. These include central characters as Argonauta , Drosha , Pasha and Dicer. Interestingly a novel group of flatworm specific Argonaute proteins was identified. While the RNAi enhancer gene ERI seems to be reduced or absent in cestodes, is present in trematodes. A potential ortholog of the RNA dependent RNA polymerase (RdRP) was detected in the turbelarian Macrostomum lignano, a remarkable finding since this secondary route of amplificiation through synthesis of small RNAs has been found only in C. elegans and plants so far. We advanced in the characterization of the pathway in the parasitic trematode Fasciola hepatica confirming by RT-qPCR, the expression of four key proteins of the pathway: Dicer, Argonaute, ERI and transporte protein SID relevant for RNAi uptake.

Keywords: RNA interference pathways, Orthologs, Flatworms

Topic: Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 216

PORTING A DESKTOP BIOINFORMATIC TOOL TO THE WEB - THE BENEFITS OF USING WEB SERVICES

Eduardo Langowski[1], Ricardo Vialle[2], Alessandro Brawerman[1], José Miguel Ortega[2], , , , , , , , , , , , AU

[1]Federal University of Paraná, Brazil, [2]Federal University of Minas Gerais, Brazil

Background. Web services are commonly used to integrate several systems regardless their infrastructure. They act like a middleware connecting these systems, which can be either local, web or mobile. The advantage is that the system becomes global and not only users can exploit it via a web interface, but also programs can consume the available services. In this work, we describe the steps necessary to port a bioinformatics desktop tool to the Web, making it available to other systems and users via web services. Results. The first step to ensure functional web services is to think about the server environment. It is necessary to have at least a Web server, like Apache, installed and running. Sometimes, depending on the programming language, an application server, like Tomcat, is also needed. Afterwards, the developer has to decide if he/she will use SOAP or REST as the web service specification and XML or JSON to code the data into objects. Nowadays, it is easier and simpler to implement RESTfull web services (REST over HTTP) with JSON encoding and decoding the information from one system to another. This combination has been slowly becoming the de facto pattern applied over the Web, especially when involving mobile devices as services consumers. The next step includes getting the desktop tool and developing the web service to access it. We developed a tool in MatLab, called getSOV, to calculate the segment overlap measure (SOV) of two distinct proteins. The web services developed allowed the getSOV tool to choose between two secondary structure predictors (PSI-Pred and SSPro4) and two alignment modes (Prank and Needleman), which were installed in the Web server. The use of those algorithms is completely transparent to the developer, i.e., there is no need to program those functions or to import them as an internal library in his/her system. Using the web services available, the getSOV tool only needs to send a Fasta file as an input and choose the predictors and alignment modes, the processing is done in background in our server, totally transparent to the user. It is also a good strategy to have another web service to report the job status. In this fashion, the user or program easily knows if a job is been loaded, running or finished. Finally, a web service to return the result set to the user should be implemented. In our example, the getSOV tool created multiple requests using RESTfull web services with JSON encoding to carry the data out. Our Web server received the requests and while them were being processed, another web service made the job status available for the user (human or program). Finally, after execution completion, the last web service reported the response back to the origin. The getSOV tool sent a total of 1000 pairwise comparisons to our server, which analyzed the deviation of SOV measurements as a function of evolutionary distance and returned the measurement back to the user. Conclusions. Porting a desktop bioinformatics tool to the Web, besides being simple, raises several benefits; mainly, other users (human or programs) can consume the services provided by the tool; there is no need to redo what has been previously done; and in a plus side, the tool author gains notoriety.

Keywords: web services, rest, json, sov

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 217

Transcription Factor abundance of cowpea bean (Vigna unguiculata) an important insight in abiotic stress responses

João Pacifico Bezerra Neto[1], Lidiane Lindinalva Barbosa Amorim[1], Valesca Pandolfi[1], Ana Maria Benko Iseppon[1]

[1]Universidade Federal de Pernambuco, Brazil

Due to their sessile condition, plants are often exposed to a wide range of both biotic and abiotic stresses. Therefore, they have developed intricate mechanisms to detect precise environmental changes, allowing optimal responses to adverse conditions. Understanding the principles of abiotic and biotic stress responses, tolerance and adaptation remains important in plant physiology research to develop better varieties of crop plants. Transcription factors (TFs) are proteins that act together with other transcriptional regulators promoting over-expression or suppression of genes that may improve the plant's tolerance. These transcription factors often belong to large gene families, which in some cases are unique to plants, like DREB (Dehydration Responsive Element Binding Factor), ERF (Ethylene Responsive Factor), NAC and WRKY. The present work aimed to perform a data mining-based identification of eight important transcription factors families in the cowpea transcriptome (NordEST database) obtained from two varieties (susceptible and resistant to drought), using well known sequences involved in the abiotic stresses response as probe/template. The sequences were classified as a TF if they presented a significant match (e-value <e-05) with seed sequences. The searches showed a great abundance and diversity of transcription factors in cowpea data based on EST and RNAseq libraries, with the identification of more than 900 probable orthologs in cowpea bean. Among the results, prevalent genes were NAC (28%), MYB (27%), DREB/ERF (16%) and WRKY (16%). This result was expected, since these represent families that present a large number of members. The RNAseq data counting enabled a determination of expression pattern profile under drought condition in different times from roots and leaves. Considering the transcription factors as whole, more expression was noted in roots, unlike the bZIP family that present more expression level at leaf tissue. The more abundant families in root expression levels with respect to absolute counting were NAC, ZFP and MYC, with more than 279.000 sequences of RNAseq all three. It is important to note that transcription factor expression could be observed in all tissues, in both susceptible and resistant varieties, point to the fact that these genes probably have basal expression under normal conditions and are up-regulated or repressed during stresses conditions. The identified transcription factors of cowpea regard interesting candidates for improvement of legume resistance to abiotic stress. Evaluations including their identification, diversity, and expression may help in their manipulation and biotechnological breeding programs.

Keywords: Transcription Factor, Fabaceae, Transcriptome, Drought, Stress

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 218

Analysis of Impact of Multiple Tautomeric Representations of Chemical Compounds in 2D and 3D Virtual Screening Studies

Fábio Mendes dos Santos[1], Julio César Dias Lopes[1]

[1]Universidade Federal de Minas Gerais, Brazil

In this work we investigated the impact of multiple tautomeric forms of chemical compounds in 2D and 3D ligand-based virtual screening studies. We used as a benchmark the DUD database (dud.docking.org) which contains datasets for 40 biological targets, each one with a set of actives and decoys carefully picked to avoid trivial classifications. In this study we compared the performance of 22 different 2D and 3D ligand-based approaches with Screen Suite available from Chemaxon (www.chemaxon.com) and ROCS software from OpenEye (www.eyesopen.com). Five molecular 2D descriptors (CF, PF, BCUT, ECFP and FCFP) using ScreenMD software were tested. The Screen3D Chemaxon software was used for 3D studies using "shape" (volume overlap) and "match" (atomic distances) methods in 12 different combinations. For 3D studies we also used the ROCS software that uses a smooth Gaussian function to represent the molecular volume and pharmacophores in five different combinations of metrics. The compounds used as targets (DUD actives and decoys) and queries (PDB ligands) were downloaded in mol2 format from DUD database and were subjected to tautomeric and protonation states treatment at pH=7 using Calculating Plugin software from Chemaxon. The "dominant" option produces different tautomers found at the given pH and the "major" option produces only the most significant tautomer at the specified pH. The three sets (original, "major" and "dominant") were used to perform all virtual screening simulations. The area under de ROC curve (AUCROC) was used as a figure-of-merit to compare the results obtained using compound's similarity against the respective references. Results for 2D studies showed that tautomer treatment bring no improvement in the AUC, except for CF descriptor where the "major" tautomers reaches slightly better AUC (average AUC = 0.745). However, 3D studies suffer high impact of the tautomers treatment which depends markedly on the software used. For Screen3D the "major" tautomers increased very much the AUC for all methods used (best average AUC = 0.757). For ROCS the effect on the average AUC was the reverse of that obtained for Screen3D and "major" tautomers presented the worst results (best "major" AUC = 0.755, best "dominant" AUC = 0.774 and best original tautomers AUC = 0.768). It is worth of noting that our results are better than previously reported for 2D (CF) e 3D (ROCS) methods even using original tautomers. As concluding remarks we can assert that the proper treatments of the tautomeric and protonation states have a great impact on ligand-based virtual screening studies and their effects cannot be underestimated.

Keywords: virtual screening, ligand, tautomer, 2D, 3D, AUC, DUD

Topic: Transcriptomics and Proteomics

Abstract #: 219

Analysis of the Triatoma brasiliensis anterior midgut transcriptome by Expressed Sequence Tags

Michele Araújo Pereira[1], Mariana Boroni[1], Ceres Luciana Alves[1], Lissandra Souza Dalsecco[2], Marcela Gonçalves Drummond[2], Marcos Horácio Pereira[3], Carlos Renato Machado[1], Andrea Mara Macedo[1], Marina Moraes Mourão[4], Ricardo Nascimento Araújo[3], Glória Regina Franco[1]

[1]Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, Brazil, [2]Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, Brazil, [3]Departamento de Parasitologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, Brazil, [4]Grupo de Genômica e Biologia Computacional, Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Minas Gerais, Brazil, Brazil

Triatoma brasiliensis is a hematophagous arthropod and the main vector responsible for the transmission of the protozoa Trypanosoma cruzi in the semiarid areas of the Northeast region of Brazil. Since T. cruzi survival is influenced by molecules expressed in the digestive tract of the insect, is of utmost importance to understand the factors responsible for the insect feeding performance. In this work we report the characterization of the anterior midgut transcriptome of T. brasiliensis using Expressed Sequence Tags (ESTs). We have isolated the messenger RNA from the anterior midgut of 101 fifth-instar nymphs and generated a cDNA library. A total of 768 clones were randomly selected and sequenced with M13 forward and reverse primers. ESTs were edited in silico to remove low quality regions and sequences of vector, adapters and poli(A) tail using DNA Baser and SeqClean software. Sequences were also clustered using the CAP3 software and submitted to similarity searches, functional annotation and classification with blastn, blastx and blast2GO. We have obtained 311 uniques (138 contigs e 173 singletons). Some transcripts were considered as contaminants and further removed from analysis. From the remaining transcripts, 180 uniques did not present similarity with sequences deposited in databases and are, possibly, T. brasiliensis specific genes. The library exhibited a high expression level of transcripts of secreted proteins and transcripts related to energetic metabolism, protein synthesis and modification machinery, immunity and viral transcripts. We observed transcripts coding for defensin, lysozyme, brasiliensin and digestive enzymes. Five transcripts were submitted to RT-qPCR analysis. The expression of three transcripts tends to be influenced by fasting and the other two, by blood feeding. The characterization of the T. brasiliensis midgut transcriptome will provide useful information to develop new strategies for vector control and to understand the processes involved in vector/T. cruzi interaction.

Keywords: Triatoma brasiliensis, anterior midgut, ESTs

Topic: Transcriptomics and Proteomics

Abstract #: 222

Implementation of computational approaches to identify long noncoding RNAs involved in neuronal differentiation

Gabriel Zaniboni[1], Ana Carolina Oliveira[1], Alan Durham[1], Eduardo Reis[1]

[1]Universidade de São Paulo, Brazil

Recent studies have revealed that long noncoding RNAs (> 200 nt, lncRNAs) may act by distinct mechanisms to exert regulatory functions to control gene expression in eukaryotes. It is known that lncRNAs may affect gene expression patterns during cellular differentiation in a way similar to the observed for well-studied regulatory proteins. In fact, some lncRNAs have been shown to be essential to the differentiation process. We implemented computational methodologies and resources to the analysis of global gene expression data collected from murine undifferentiated embryonary carcinoma cells (P19 cells) and P19 differentiated into neuron or glia cells. The gene expression data was searched for the identification of groups of trans co-regulated lncRNAs and mRNAs during cell differentiation, as well as for expression changes in specific gene categories and molecular pathways. The gene category analysis of the co-regulated mRNAs showed gene expression enrichment in categories directly related to neurogenesis and nervous system development, leading to the identification of two lncRNA candidates for functional characterization (annotated as Gm2694 and 2610017I09Rik). The data obtained so far shows that the expressed lncRNAs during P19 cell differentiation might play a regulatory role in the neurogenesis process, both into the nucleus and the cytoplasm.

Keywords: genic expression, celular differentiation, neurogenesis, DNA microarray, noncoding RNAs

Topic: Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 223

Development of a Database for Recording and Comparison of Molecules Isolated from Plants of The Brazilian Semiarid – SAM Database

Tiago Nery de Oliveira[1], Elisson Barros de Oliveira[1], Maicon Vinícius Araújo Santos[1], Alex Ferreira dos Santos[1], Wagner Rodrigues de Assis Soares[1], Bruno Silva Andrade[1]

[1]Universidade Estadual do Sudoeste da Bahia (UESB), Brazil

Databases of molecules has become a great tool for scientists and researchers working in drug development because they provide information of various types of molecules, which becomes crucial when trying to find a structure with pharmaceutical activity. Several plants of semi-arid of Bahia have been studied in recent years in an attempt to find new therapeutic targets for a variety of pathologies. However, despite several efforts our knowledge is still very small about its mechanism of action of these natural compounds. Then, the aim of this work was to develop a database which allow the storage of information and comparison of molecules isolated from plants of Brazilian Semiarid Region, because its great diversity of plant species, and find compounds that can be useful for applied drug development against human diseases. The development of this database used Java programming language in its standard EE (Enterprise Edition). Java technology is used for enterprise applications that can be on the Internet or Intranet. Some features, such as creating the structure of molecules, calculating the similarity between molecules and MOL files generation were obtained from the Open Source library Indigo, developed by GGA Software Services. The generated data is stored in a relational database, used as System Manager Database (DBMS) MySQL Server 5.5.28.

Keywords: Molecules, Database, Semiarid, Comparison, Drug

Topic: Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 224

An in-silico semiflexible study of acetylcholinesterase ligand interaction: apparent higher plasticity in the human active site compared to Torpedo californica, influence of the "Backdoor opening" conformation and potential for new expanded ligands

Leonardo Henrique França de Lima[1], Rafaela Ferreira[2], Marcelo Santoro[2]

[1]Universidade Federal de São João Del-Rei, Brazil, [2]Universidade Federal de Minas Gerais, Brazil

Background: Human acetylcholinesterase (hAChE) is a significant target for therapeutic drugs, being particularly promising for anti-Alzheimer treatment. During many years, the acetylcholinesterase from the eletric fish Torpedo californica (TcAChE) has been used as a molecular model for the rational design of specific inhibitors. However, comparison of multiple hAChE and TcAChE crystallographic structures shows subtles changes between the half-sites of these two enzymes with apparent implications to ligand affinity and selectivity. Beyond this, the significantly large active site of AChE, the existence of three potential interaction sites (i.e., a periferic site, an anionic "gorge" and the catalytic site itself), as well the apparent existence of a transient "backdoor opening" conformation for the same encourage computational studies comparing the plasticity for diferente ligand interaction at diferente active site configurations for this enzyme. Results: Here, we use semi-flexible docking companied from principal component analysis for the study of the plasticity of interaction of galantamine (a competitive inhibitor), aflatoxin (a periferic ligand) and 3-O-Acetylleucotamine (a galantamine derivative with an aliphatic diesterasic expansion) witht the hAChE and the TcAChE active sites. For TcAChE, such in silico studies were carried for crystallographic structures of the enzyme at the native and the "backdoor opening" conformations. The analysis suggests that the substitution of a Phe to a Tyr at the 337 position between the TcAChE and hAChE terminate to allows a higher plasticity of interaction modes at the last one. Also, the "backdoor mechanism" seems to enhance the probability of binding at marginal sites (out of the catalytic region), due to a narrowing of the active site. Finally the aliphatic expansion at the 3-O-Acelylleucotamine allows a multiple site interaction for this ligand, encompassing both the catalytic region as the periferic site, this last being pointed, in multiple studies, as a promissing site for the inhibition of amyloid fibers formation Conclusion: The studies presented here are significant for new drug design, aiming to optimizate the interaction with this pharmacologically importante enzime.

Keywords: Acetylcholinesterase, docking, rational drug design

Topic: Databases & Data Integration; Text Mining & Information Extraction

Abstract #: 225

Using semantic annotation and the concepts of protein druggability and gene essentiality to prioritize drug targets in protozoa parasites

Kele Belloze[1], Maria Claudia Cavalcanti[2], Floriano Silva-Jr[1]

[1]Instituto Oswaldo Cruz - Fiocruz-RJ, Brazil, [2]Instituto Militar de Engenharia - RJ, Brazil

Neglected diseases are infectious diseases that primarily affect the poorest people in the world. The existing drugs to fight these diseases cause many side effects to patients and are insufficient or inaccessible. Another problem is that there still is drug resistance. Accordingly, it is very important to identify new targets for drugs. This study proposes a methodology to support the prioritization of targets to combat neglected tropical diseases caused by five protozoan Entamoeba histolytica, Leishmania major, Plasmodium falciparum, Trypanosoma brucei and Trypanosoma cruzi, based on the concepts of protein essentiality and druggability. The methodology takes advantage of the large amount of data and information publicly available on genomic, biochemical and pharmacological databases, and also the biomedical literature to seek and integrate data and information from model organisms and drug target proteins. Our final goal is to suggest essential and druggable candidates (target proteins) for protozoa, raising the possibility of targets for future studies and experiments. To obtain these data we used the approach of ontology-based semantic annotation to extract data from scientific articles and the concepts of homology and orthology between protein sequences stored in semi-structured databases, in order to raise essential and druggable candidates. The semantic annotation showed that terms involving the protozoa T. brucei and the model organism Saccharomyces cerevisiae as well as the technique of RNA Interference were annotated in the highest number of articles. It was also possible to classify the annotated proteins in 16 different categories. In relation to essential protein candidates, protozoans presented varying numbers of orthologs to essential proteins from model organisms in the following decreasing order: S. cerevisiae, Mus musculus, Danio rerio, Drosphila melanogaster, Escherichia coli, Caenorhabditis elegans and Arabidopsis thaliana. Regarding druggable protein candidates, it was noted that a single protozoan protein could present homology with several proteins from BindingDB, DrugBank and Therapeutic Target Database. The data about essential and druggable protein candidates, integrated with data obtained in the semantic annotation, are a starting point to prioritize proteins of protozoa, which can be tested with experimental approaches in the process of drug discovery.

Keywords: Protozoa, targets, semantic annotation, homology, orthology

Topic: Transcriptomics and Proteomics

Abstract #: 227

Integrating SNPs annotations into CLIP-seq data analysis

Paula H. Reyes-Herrera[1], Carlos A. Sierra[2]

[1]School of Electronic and Biomedical Engineering, Universidad Antonio Nariño, Bogotá 110311,, Colombia, [2]School of Computer Engineering. Universidad Antonio Nariño, Bogotá 110311, Colombia

Research on post-transcriptional gene regulation is key in order to understand the rules that govern gene expression regulation and epigenetic changes. RNA Binding Proteins (RBPs) and their systemic action are at the core of post-transcriptional gene regulation. In order to interpret the RNA Binding Proteins function, it is fundamental to identify the RNA regions where these proteins bind. Until a few years ago, experimental techniques obtained a reduced set of sample sequences containing the binding regions, but nowadays the joint action of CrossLinking ImmunoPrecipitation and High Throughput Sequencing (known as CLIP-seq) makes it possible to recover a transcriptome-wide set of sequences which contain the interaction regions for a particular protein. Nevertheless, the specific interaction region must be identified from the transcriptome-wide set of sequences recovered. In order to facilitate the identification of the specific interaction regions and improve the signal to noise ratio several protocols introduce experimental variations to CLIP-seq experimental techniques. The experimental variations introduced frequently result on mutations close to the interaction site or inside it. Therefore in order to find the specific binding region, computational approaches rank the sequences based on the induced mutations and as a result these approaches give a higher weight to sequences with mutations. Finally, these computational approaches use a motif extraction algorithm to provide a motif candidate as the RNA region recognized by the RNA Binding Proteins. We notice that nucleotide mutations are caused by experimental introduced variations, sequencing errors and Single Nucleotide Polymorphisms (SNPs). These SNPs are not induced by the experimental protocol but are existing mutations that occur at the DNA level. We propose to take into account the annotated SNPs which are common genetic variations, to consider differently mutations experimentally induced from mutations which are probably not induced. Consequently we applied a motif extraction algorithm and to evaluate the performance we quantified the number of interaction sequences that contain the extracted motif compared to the total number of sequences provided by the experimental protocol. The results guide us to conclude that SNPs should be taken into account in order to reduce the noise present in the experimental data. Considering annotated SNPs for the motif extraction algorithms is a way of taking into account pre-existing mutations.

Keywords: RNA Binding Proteins, CLIP-seq, RNA-Protein Interactions Prediction. SNPs

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 229

WHOLE EXOME SEQUENCING TO IDENTIFY NEW GENES RELATED WITH HEREDITARY BREAST AND OVARY CANCER

Jessica Rodrigues Plaça[1], Jorge Estefano S. de Souza[2], Wilson Silva Junior[2]

[1]Tecnological Development Center, Federal University of Pelotas, Brazil, [2]Center for Medical Genomics, Hospital of the Ribeirao Preto Medical School, University of Sao Paulo, Brazil

Background: Hereditary Breast and Ovary Cancer (HBOC) is characterized by the presence of breast adenocarcinoma in ductal or lobular origin and epithelial ovarian carcinoma, with a prevalence of 1 in 800 or even 2500 women in the general population. It is estimated that about 10% of all cases of breast and ovarian cancer are hereditary and those are associated with mutations in BRACA1/BRCA2 genes. The Next Generation Sequencing technologies have been extensively applied to study the genetic machinery that drives the tumorigenesis and to find potential gene targets for clinical therapy. Exome is a part of genome formed by protein-coding regions of all genes and the whole-exome sequencing information can reveal germline and somatic mutations that depict the causal relationship between the mutations and phenotypes. Seen this characteristics, the study aimed to analyze the exome of 15 patients with HBOC to know the frequency of mutation in 197 genes related with breast cancer and repair mechanisms widely studied in the literature. After the DNA extraction from peripheral blood, it was kept in solution and the exome enrichment was done using the Nextera Exome Enrichment kit according to the manufacturer's instructions. The enriched DNA samples were 2×100 paired-end sequenced using the Illumina Genome Analyser IIx. Sequence reads were aligned to the human reference genome (hg19) using Burrows–Wheeler Aligner with default parameters and variants were identified using Genome Analysis Toolkit (GATK). Reads that matched exonic regions, including exon–intron-boundaries were analyzed. The variant detection frequency was set at a minimum of 20% of the reads covering any mutation. A minimum coverage of 10 reads was set as the threshold for any variant to be considered a real mutation. Results: A total of 127.783.772 reads were obtained after filtering with a Phred-like Quality Score (Q score) greater than 20. The result of the mutation analysis identified 7641 single nucleotide variations (SNVs), with 510 SNVs occurring in coding regions, 1122 in intron regions, 94 in 5' Flanking regions and 90 3' Flanking regions. The total number of nucleotide transitions were 5065 and the total number of transvertions were 2136. Were found 1537 non-synonyms SNVs and in stop-codon regions were found 17 SNVs. Looking for SNVs at specific genes we found 11 SNVs in introns, 2 SNVs in 5' UTR, 8 SNVs in 3' UTR, being 11 SNVs non-synonyms. It was found one missense SNV at position 30672353 of chromosome 6 corresponding to the Mediator of DNA Damage Checkpoint Protein 1 (MDC1) gene. Discussion: The point mutation types are heterogeneous and includes different regions of the exome. The data shows new candidate genes involved in tumoral progression and maintenance. Further analyses would be important to perform the molecular characterization of the mutated genes, like the MDCI.

Keywords: Exome sequencing, Hereditary Breast and Ovary Cancer, mutagenesis, phenotype related mutatations

Topic: Transcriptomics and Proteomics

Abstract #: 230

An automated generic pipeline for de novo identification of long and small non-coding transcripts on RNAseq raw data

Marcelo Rojas-Herrera[1], Raul Arias-Carrasco[1], Juan Azalte[2], Nicole Trefault[1], Victor Polanco[1], Vinicius Maracaja-Coutinho[1]

[1]Centro de Genómica y Bioinformática, Instituto de Biotecnología, Universidad Mayor, Chile, [2]Centro Nacional de Secuenciación Genómica, Universidad de Antioquia, Colombia

Background: Non-coding RNAs (ncRNAs) transcribed from intronic, intergenic and antisense genomic regions are important players in a wide variety of critical Eukaryotic cellular processes. Despite the existence of a number of different algorithms for ncRNAs identification, there are no straightforward tools for large-scale identification of novel transcripts on organisms without an available reference genome sequence. Currently solutions require the management of a plethora of input/output formats and conventions, resulting in a non-trivial task for researchers with low programming skills. Results: Here, we describe a generic, flexible and modular automated pipeline for de novo identification of long and small ncRNAs based on RNAseq raw data. Our pipeline provides different automated Perl modules for each important step for ncRNAs identification (reads quality control, assembling, identification and annotation of coding/non-coding RNAs) in datasets obtained from common used sequencing platforms (Illumina and 454-Roche), obtained from single- or paired-end libraries. We applied it on different 454-Roche and Illumina RNA sequecing runs from the Dinoflagelate Alexandrium catenella and the plant Aristotelia chilensis (Maqui or Chilean Wineberry), respectively. Our analysis resulted on the identification of a total of 3,372 transcriptional fragments in A. catenella transcriptome, of which 2,019 (61.63 %) are protein-coding genes, and 1,294 (38.37 %) non-coding RNAs. In A. chilensis, we identified a number of 64,924 transcriptional fragments, of which 36,019 (55.48 %) are protein-coding genes, and 28,906 (44.52 %) non-coding RNAs. Conclusions: Our pipeline emerges as a new tool for identification of ncRNAs on RNAseq projects related to organisms without a genome available. Also, it is the largest exploration of the non-coding transcriptional activity developed so far for Alexandrium catenella and Aristotelia chilensis, evidencing an exhaustive non-coding activity in both organisms.

Keywords: Non-coding RNAs, transcriptome, transcription, RNAseq, pipeline

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 232

Inter-domain residue correlations in hormone nuclear receptors

Leonardo Lima[1], Marcelo Afonso[2], Lucas Bleicher[2]

[1]UFSJ, Brazil, [2]UFMG, Brazil

Nuclear receptors are transcription factors composed of a modular architecture, virtually all of them containing at least a DNA binding domain (DBD) and a ligand binding domain (LBD) connected by a flexible hinge. We have previously studied those domains by decomposition of correlated residues networks, detecting that evolution of simultaneously conserved residues in the LBD seemed to be mostly related to hormone selectivity and co-activator binding, and that strong correlation among residues in the DBD is related to a clade-specific P-box which suggests the existence of an alternative DNA hormone response element for nematodes. Here, we apply this methodology to detect and discuss residue specific correlation between the DBDs and LBDs, which can now be analyzed in the light of the recent DNA-DBD-LBD complex x-ray structures. The correlated residues found are related to an interaction pair in the LBD with varying attraction intensities depending on residue identity, which relates to different nuclear receptor subtypes. Those residues also present correlation with DBD positions which are important for dimerization, which may explain the differences in complex formations for nuclear receptor sub-classes.

Keywords: correlated mutations, nuclear receptors, decomposition of residue correlation networks

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 233

Comparative Genomics of Two Newly Sequenced Brazilian Isolates of Vibrio parahaemolyticus

Leandro Santos[1], Luísa Hoffmann[2], Rosane Silva[2], Turán Ürményi[2], Paulo Bisch[1], Wanda von Krüger[1]

[1]Laboratório de Física-Biológica - IBCCF / UFRJ, Brazil, [2]IBBCF/ UFRJ - INCT-INPeTAm/CNPq/MCT, Brazil

Background: Vibrio parahaemolyticus, a Gram-negative bacterium, is a worldwide cause of food-borne gastroenteritis found in aquatic environments free- living, attached to abiotic and biotic (e. g., plankton and shellfish) surfaces, or associated with marine animal hosts. Since the complete genome sequence of clinical isolate RIMD2210633 was published, draft genome sequences of several others V. parahaemolyticus strains have been announced, but no one is from Brazil. Our group recently, obtained draft genome sequences of two V. parahaemolyticus strains isolated in Brazil: they are, Cascavel (O3:K6 – same serotype of the pandemic clone), a clinical isolate from a gastroenteritis outbreak in the city of Cascavel, CE, in 2002, and the strain 20173, isolated from coastal waters in the Northeast of the country. We did a comparative analysis of the genome sequences of Cascavel and 20173 with those of two reference strains of V. parahaemolyticus, namely, RIMD2210633 (O3:K6 serotype), isolated in 1996 from a patient with travellers´ diarrhea (Osaka, Japan), and BB22OP (O4:K8 serotype), a Bangladesh environmental isolate from early 1980s. The contig sequences obtained for Cascavel and 20173 were aligned with the genome of the reference strains and annotated using RAST, Blastx, and Blast2Go. We also, used the CLC Genomics Workbench to do the comparative genomic analysis of the annotated genes in the genomes of the four V. parahaemolyticus strains mentioned above and V. cholera strain N16961. Results: Most of the genes annotated so far for the V. parahaemolyticus isolates from Brazil share high sequence similarity with orthologs from the Asian isolates and V. cholera N16961. Those genes encode essential functions such as amino acids and carbohydrate metabolism, DNA replication, transcription, and translation. A small percentage of the annotated genes is V. parahaemolyticus specific, including the genes found in pathogenic islands, tdh (thermolabile hemolysin), ure (urease), and T3SS (type three secretion system) genes. An even smaller percentage of the annotated genes seems to be strain specific. Conclusion: Comparative genomics approach proved to be an extremely powerful tool to determine similarities and differences between isolates from different parts of the world, where the environmental conditions are quite distinct, allowing a better understanding of the mechanisms behind the population dynamics.

Keywords: Comparative Genomics; Annotation; Vibrio parahaemolyticus; Brazilian Isolates

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny,Structural Bioinformatics; Molecular and Supramolecular Dynamics

Abstract #: 234

NEW SMALL RNA BIOGENESIS PROTEINS IN PARASITE PLATYHELMINTHS

Lucas Maldonado[1], Adolfo Fox[1], Natalia Macchiaroli[1], Santiago Fontela[2], Marcela Cucher[1], Jose Tort[2], Mara Rosenzvit[1], Laura Kamenetzky[1]

[1]IMPAM, Argentina, [2]Universidad de la República, Uruguay., Uruguay

Background: small RNAs including microRNAs (miRNAs), small interfering-RNAs (siRNAs) and piwi-RNAs (piRNAs) control several pathways such as developmental timing, hematopoiesis, organogenesis, apoptosis, cell proliferation and tumorigenesis. Canonical animal miRNAs are generated from long hairpin precursors which are processed by three principal proteins: Dicer, DGCR8 (Pasha) and Drosha. The resulting mature miRNA is loaded onto the effector protein Argonaute (Ago). The miRNA-Ago complex is guided by the loaded miRNA to specifically interact with the target mRNA. This interaction produces the inhibition of the target protein expression by 'slicing', destabilization or translation inhibition of the target mRNA. Argonaute proteins are evolutionarily conserved and can be phylogenetically subdivided into the Ago and Piwi subfamily. Ago proteins are ubiquitously expressed and bind to siRNAs or miRNAs but Piwi proteins expression is mostly restricted to germ line and associate with piRNAs to facilitate silencing of mobile genetic elements. With the aim to identify the main small RNA pathway proteins in Platyhelminth parasites we searched for Ago, Piwi, Dicer, Pasha and Drosha proteins in recently generated Echinococcus genomes and transcriptomes.

Keywords: Platyhelminths, parasites, Argonauts, miRNA.

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 237

Sequencing and Annotation of the Brazilian Vibrio parahaemolyticus isolate 20173 Genome

Mateus Perissé[1], Leandro Santos[1], Luísa Hoffmann[2], Rosane Silva[2], Turán Ürményi[2], Paulo Bisch[1], Wanda von Krüger[1]

[1]Laboratório de Física-Biológica - IBCCF / UFRJ, Brazil, [2]IBBCF/ UFRJ - INCT-INPeTAm/CNPq/MCT, Brazil

Background: Vibrio parahaemolyticus, a Gram-negative bacterium, is a worldwide cause of food-borne gastroenteritis. It can be found in marine environments free-living, attached to abiotic and biotic (e. g., plankton and shellfish) surfaces, or associated with marine animal hosts. The organism is phylogenetically close to V. cholerae, the causative agent of cholera. Since the complete genome sequence of the clinical isolate RIMD2210633 was published, draft genome sequences of several others V. parahaemolyticus strains have been announced, but no one was isolated in Brazil. In the present work, we aimed to sequence and annotate the whole genome of the Brazilian environmental strain V. parahaemolyticus 20173. The bacteria was cultured in LB medium (+ 1.2% NaCl), at 37oC, stirring over-night, and the DNA was extracted and sequenced according to the protocol of PGM platform (Ion Torrent, Life Technology) to obtain the reads. The FastQC program was used to access the quality of the reads. Using the CLC Genomics Workbench the reads were trimmed by quality (Phred > 25), by length (> 15bp), removing the ends (40 bp), and assembled with de novo methodology in contigs. These were aligned with two the genomic sequences of the V. parahaemolyticus references strains RIMD2210633 and BB22OP and then annotated using RAST, Blastx, and Blast2Go. Results: Using this sequencing methodology we obtained 1,830,779 reads, with average length of 142.3 bp. After the trimming step, we ended up with 1,606,830 reads (87.77% left) with average length of 84.3 bp. These reads we assembled in 1,068 contigs (N50 = 20,365 bp; maximum length = 76,577 bp; average length = 4,768). From all the contigs, 877 (82.12%) were mapped onto the RIMD2210633 reference genome, covering 80.5% of chromosome 1 and 79% of chromosome 2. 883 contigs (82.68%) were mapped onto the BB22OP reference genome, covering 80.5% of chromosome 1 and 84.5% of chromosome 2. About 4,600 genes were annotated, including tdh (thermolabile hemolysin) and T3SS (type three secretion system) genes, important virulence factors on V. parahaemolyticus. This is an interesting result, since strain 20173 was isolated from coastal waters in Brazil. Conclusions: The methodology used in this study proved to be a useful way to characterize the genome and rapidly annotate important genes, such as virulence factors or hypothetical ones. In the near future, we intent to do another run on PGM platform to increase genome coverage and use paired-end sequencing technology for the scaffolding of pre-assembled contigs.

Keywords: Next-Generation Sequencing; Annotation; Vibrio parahaemolyticus; Brazilian Isolates

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 238

Sequencing and Annotation of the Genome of the Brazilian Vibrio parahaemolyticus Strain Cascavel

Cristóvão Lanna[1], Leandro Santos[1], Luísa Hoffmann[2], Rosane Silva[2], Turán Ürményi[2], Paulo Bisch[1], Wanda von Krüger[1]

[1]Laboratório de Física-Biológica - IBCCF / UFRJ, Brazil, [2]IBBCF/ UFRJ - INCT-INPeTAm/CNPq/MCT, Brazil

Background: Vibrio parahaemolyticus is a rod-shaped, Gram-negative bacterial pathogen found in marine and estuarine environments, both as a free-living organism and associated with aquatic eukaryotes. It is a pathogen of worldwide relevance, being responsible for gastroenteritis cases in Japan and in the United States, where it is the main contaminant of seafood. V. parahaemolyticus can also cause skin infections, which can lead to necrosis and sepsis. Even though V. parahaemolyticus is a relevant pathogen, its virulence factors and pathogenicity are not completely understood. There are only two completely assembled genomes from V. parahaemolyticus in the GenBank. They are from strains RIMD 2210633 and BB22OP. There are also some drafts available, but no one from a Brazilian isolate. Many Brazilian V. parahaemolyticus strains have been recently isolated from environmental (including contaminated seafood) and clinical samples; however, none has been sequenced so far. Our aim in this study was to sequence and annotate the complete genome of V. parahaemolyticus Cascavel, a clinical strain. To this end, the bacterium was cultured in LB medium (+ 1.2% NaCl), at 37oC, stirring over-night. The DNA was extracted and used for sequencing according to the protocol of PGM platform (Ion Torrent, Life Technology) to obtain the reads. The FastQC program was used to assess the quality of the reads. Using the CLC Genomics Workbench the reads were: trimmed by quality (Phred > 25), by length (> 15bp), removing the ends (40 bp), and assembled with de novo methodology in contigs, which were aligned with two V. parahaemolyticus reference strains (RIMD2210633 and BB22OP) genomic sequences and annotated using RAST, Blastx, and Blast2Go. Results: We obtained 1,191,343 reads, with average length of 141.2 bp. After the trimming step, we ended up with 1,063,979 reads (89.31% left) with average length of 85 bp. With those reads we assembled 1,493 contigs (N50 = 11,093 bp; maximum length = 73,014 bp; average length = 3,375). From all those contigs, 1,233 (82.59%) were mapped onto the RIMD2210633 reference genome, covering 87.98% of chromosome 1 and 84.96% of chromosome 2. About 1,200 contigs (83.12%) were mapped onto the BB22OP reference genome, covering 86.28% of chromosome 1 and 86.83% of chromosome 2. About 4,658 genes were annotated, including tdh (thermolabile hemolysin) and T3SS (type three secretion system) genes, important virulence factors on many V. parahaemolyticus strains. Conclusions: This methodology proved to be a useful way to characterize the genome and rapidly annotate important genes, such as those involved in virulence factors and hypothetical ones. In the near future, we intent to do another run on PGM platform to increase genome coverage and use paired-end sequencing technology for the scaffolding of pre-assembled contigs.

Keywords: Next-Generation Sequencing; Annotation; Vibrio parahaemolyticus; Brazilian Isolates

Topic: Genomics; Sequence Analysis; Evolution and Phylogeny

Abstract #: 239

Identification and regulatory role prediction of a miR398 homologue from Arabidopsis in superoxide dismutase expression in Cowpea

Diego Teixeira[1], Taffarel Melo Torres[2], Heverton Valdevino[1], João Paulo Matos Santos Lima[1]

[1]Universidade Federal do Rio Grande do Norte, Brazil, [2]Universidade Federal Rural do Semi-árido, Brazil

Background: Plants, like other organisms, are subject to several fluctuations of the environmental conditions. However, due to its sessile nature, their survival under unfavorable conditions depends on an orchestration of a series of biochemical and physiological processes. One of the abiotic stress acclimation consequences is the high production of reactive oxygen species (ROS). Though these species possess defined role in cell signaling, they can also cause oxidative damage to proteins, lipids and DNA. Plants have a complex and redundant antioxidant system, both enzymatic and non-enzymatic, which control ROS accumulation and limit their formation The first enzymatic defense against excessive ROS accumulation, on cellular compartments, is the superoxide dismutase (SOD) enzyme which performs the dismutation of O2●─ in H2O2 e O2. The Cu/Zn SOD isoforms are coded by two genes (SOD1 and SOD2), which are strongly regulated in the post-transcriptional levels, mainly by small interference RNAs, in special by microRNA-398. The aim of this study was to identify, in public databases, miR398 homologues and SOD isoforms sequences in Vigna unguiculata, and also infer evolutionary relationships related to the regulatory role of the miRNAs in plant redox homeostasis control. Results: Using BLASTn tool with mature miRNA dataset as input against V. unguiculata EST database and the RNAz WebServer to investigate the pre-miRNA structure, two ESTs entry with both sequence and secondary structure similarity to miR398 were identified, as well as the putative sequences for V. unguiculata SODs isoforms. The phylogeny inferred by Bayesian methods, showed a correlation of the putatives miRNAs inside the Rosides taxon. Conclusion: This study provides the first indications of homologous sequences for miR398 in V. unguiculata, including a phylogenetics interpretation on its regulatory mechanism in the oxidative stress response. The results guide us to a better experimental studies interpretation and comprehension about the post-transcriptional regulatory mechanisms involving ROS protection genes in plant cells.

Keywords: Cowpea, ROS, Cu/Zn Superoxide Dismutase, miRNA