ESCOLA GAÚCHA DE BIOINFORMÁTICA

17 A 21 DE JULHO DE 2017

AULAS   PALESTRAS   MINI-CURSOS   SESSÃO DE PÔSTERES

LIVRO DE RESUMOS

REALIZAÇÃO

UFRGS
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL

.Inf
INSTITUTO
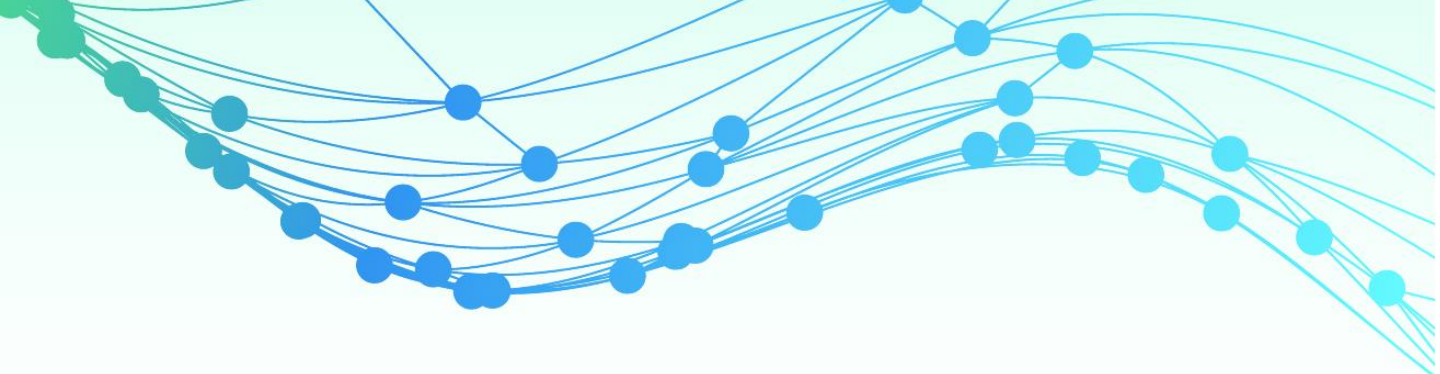DE INFORMÁTICA
UFRGS

CBiot
Centro de Biotecnologia
da UFRGS

LOCAL

Centro de Eventos do Instituto
de Informática da UFRGS
SITE › WWW.UFRGS.BR/EGB

Escola Gaúcha de Bioinformática

A Escola Gaúcha de Bioinformática, EGB, foi organizada de forma integrada pelo Instituto de Informática e pelo Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul, buscando construir um espaço de formação, integração, qualificação e desenvolvimento das atividades de pesquisa envolvendo o emprego de métodos computacionais no estudo de sistemas biológicos.

Sob uma perspectiva interdisciplinar de problemas biológicos e terapêuticos, o evento agrega tanto abordagens através de sequências (de nucleotídeos ou aminoácidos), de biologia de sistemas e de larga escala, quanto de estruturas 3D (e suas conformações) ao suporte e desenvolvimento possibilitados pela Ciência da Computação.

Comitê Organizador

Hugo Verli - Centro de Biotecnologia - UFRGS
Márcio Dorn - Instituto de Informática - UFRGS
Rodrigo Ligabue Braun - Centro de Biotecnologia – UFRGS

Comissão Científica

Ana L. C. Bazzan - INF - UFRGS
Eduardo Eizirik - FaBio - PUCRS
Laurent Dardenne - LNCC
Mario Inostroza-Ponta - DIINF - USACH

REALIZAÇÃO

UFRGS
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL

.Inf
INSTITUTO
DE INFORMÁTICA
UFRGS

CBiot
Centro de Biotecnologia
da UFRGS

LOCAL

Centro de Eventos do Instituto
de Informática da UFRGS
SITE › WWW.UFRGS.BR/EGB

DETECTION OF **LTR** RETROTRANSPOSONS IN THREE *Hamiltosporidium* lineages

ALBUQUERQUE, NATHALIA RAMMÉ MEDEIROS[1]; SILVEIRA, JULIANO DE OLIVEIRA[1]; HAAG, KAREN LUISA[1].
[1]Laboratório de Genômica Evolutiva, Departamento de Genética, Universidade Federal do Rio Grande do Sul.

Microsporidia have one of the smallest genomes among eukaryotes, due to their obligate intracellular parasitism that resulted in an extensive reduction in genome size and complexity. Transposable elements (TEs) compose a high fraction in eukaryotic genomes, the vast majority being retrotransposons, due to their "copy-paste" transposition mechanism. TEs are able to mutate genes, alter gene regulation and generate new genes. Transposition events can be deleterious when TEs integrate in active coding regions and disrupt important genes. Due to the absence of recombination, TEs may accumulate in asexual genomes and contribute to extinction of asexual organisms. However, there is no evidence showing different TE densities between sexual and asexual lineages. We assessed the genomes of three Hamiltosporidium lineages for TE content. Two genomes of H. magnivora lineages (BEOM2 and ILBN2) that reproduce sexually and are vertically transmitted, and one genome of an H. tvaerminnensis lineage (FIOER33) that reproduces asexually and is both vertically and horizontally transmitted. The RepeatMasker software was used for initial screening of TE content. A set of core protein domains encoded by LTR retrotransposons was used in BLASTp searches in the proteomes of BEOM2, ILBN2 and FIOER33. Finally, LTRharvest was used to detect full-length elements. No copies of DNA transposons larger than 80 pb long and with >80% identity to the reference elements were found in the genomes. Nevertheless, LTR retrotransposons copies were detected in all three genomes. BLASTp identified 31 core protein domains in BEOM2, 199 in ILBN2 and 15 in FIOER33 predicted proteomes with *e-values* >1e-10. LTRharvest found 40 elements in BEOM2, 41 in ILBN2 and 23 in FIOER33 genomes. However, only two potentially active LTR retrotransposons, with pair of LTRs 5' and 3', and the presence of at least 3 key protein domains (reverse transcriptase, RNAseH and integrase) were found in the genomes. Both elements are shared between all three lineages, indicating fixed insertions that occurred before the divergence of the species. The genomes of sexual lineages (BEOM2 and ILBN2) have increased TE load compared with the asexual lineage (FIOER33). These results indicate that the H. tvaerminnensis (FIOER33) were able to eliminate deleterious TEs, but the underlying molecular mechanism that eliminates TE insertions remain unknown.

Key words: Microsporidia, Transposable elements, LTR retrotransposons.

Molecular Evolution of Subtilisin-Like Serine Proteases Depicts Ongoing Diversification in Metarhizium anisopliae (Ascomycota: Hypocreales)

ANDREIS, FÁBIO CARRER[1,2]; SCHRANK, AUGUSTO[1,3]; THOMPSON, CLAUDIA ELIZABETH[1,2,4]

[1] Programa de Pós-Graduação em Biologia Celular e Molecular (PPGBCM), Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil; [2] Unidade de Biologia Teórica e Computacional (UBTEC), Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil; [3] Laboratório de Biologia Molecular de Fungos Filamentosos, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil; [4] Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, RS, Brasil;

Entomopathogenic fungi have been used for biocontrol purposes for over a century. Organisms such as *Metarhizium anisopliae* are the most frequently used worldwide mainly due to their efficient infection mechanisms. The Pr1 family of subtilisin-like serine endopeptidases has an important role in pathogenicity and virulence of *M. anisopliae*. These virulence factors allow for the penetration of the host's cuticle, a vital step in the infective process of this fungus, which possesses 11 Pr1 isoforms (Pr1A through Pr1K). This family is divided in two classes with Class II (proteinase K-like) comprising 10 isoforms further split into three subfamilies. It is believed that these isoforms act synergistically and with other virulence factors, allowing pathogenicity to multiple hosts. As virulence coevolves through reciprocal selection with hosts, positive selection may lead to the evolution of new proteases/isoforms that can withstand host defenses. This work tests this hypothesis in Class II Pr1 proteins, focusing on *M. anisopliae*, employing multiple methods for phylogenetic inference in amino acid and nucleotide datasets in multiple arrangements – grouped by isoform, subfamily, and all of Class II - for *Metarhizium* spp. and related species. Nucleotide alignments and constructed trees were analyzed regarding synonymous and non-synonymous substitutions for positive selection inference. Subfamilies comprising multiple isoforms were analyzed for pairwise Type I functional divergence based on their amino acid alignments and corresponding topologies. Phylogenies depict groups that match the taxonomy of their respective organisms with high statistical support, with minor discrepancies. Positively selected sites were identified in six out of ten Pr1 isoforms, most of them located in the proteolytic domain. Moreover, there was evidence of functional divergence in the majority of pairwise comparisons within subfamilies. These results imply the existence of differential selective pressure acting on Pr1 proteins, likely affecting host specificities, virulence, or even adapting the organism to different host-independent lifestyles. Also, evidences of Type I functional divergence further support the hypothesis that Pr1 proteins are not functionally redundant, while also pointing to a potential novel isoform. This work is the most comprehensive study on the molecular evolution of Class II subtilisin-like serine proteases in *Metarhizium* spp., unveiling patterns of protein diversification that were not addressed before on this family of virulence factors, also introducing a fully resolved phylogeny at the subfamily and isoform levels, with high statistical support, providing a more solid basis on the classification of Class II subtilisin-like Pr1 proteins, while consistent with previous knowledge on the subject. Further work is required to confirm the importance of functional divergence and positively selected sites, as to identify their effects on protein structure and function.

Keywords: *Metarhizium anisopliae*, entomopathogenic fungi, Pr1, serine endopeptidases, molecular evolution, positive selection, functional divergence.

THE NEUROPROTECTIVE POTENTIAL OF A HIGH FAT DIET IN COCKAYNE SYNDROME INDIVIDUALS

BALDISSERA[1], GABRIEL; MIYAMOTO[1], KENDI NISHINO; BONATTO[1], DIEGO.
[1]Biotechnology center, Biosciences institute, Rio Grande do Sul Federal University (UFRGS)

The Cockayne Syndrome (CS) is a rare Progeroid disease whose affected patients present symptoms of premature aging and morphological alterations (like short size and a bird-like pointed face). Besides, alterations on the myelin patterns (irregular or absent) are commonly observed, which can be a consequence of a neurodegenerative process. Moreover, the cerebellar region is one of the most affected areas in the central nervous system. Thus, symptoms associated with motor coordination are also observed in these patients. Genetically, CS has an autosomal recessive inheritance through mutations in the CSA and CSB genes, which are related to the DNA repair mechanism. These mutations lead to an alteration in the activity of the DNA damage-associated signaling protein PARP, which causes an intracellular NAD+ and ATP depletion and, in the most severe cases, cell death. Data has demonstrated the efficiency of dietary interventions on neurodegenerative diseases, as presented by studies that show the neuroprotective properties of some fatty acids, which is promising, since the CS does not have any effective treatment available so far. In this sense, studies demonstrated that CS mice fed with a high fat diet (HFD) presented an increase in their post-synaptic density length as well as better performance of their cerebellar function. However, how such diet can act in the attenuation of neurological symptoms is not biochemically understood. Therefore, the aim of this research is to use transcriptome and systems biology tools in order to evaluate how this diet can attenuate CS symptoms as well as find genes and their pathways related to the diet and to the symptoms that are relevant for a contextual understanding of HFD role on CS mice cells. Gene Expression Omnibus (GEO) database was used to obtain microarray data from the cerebellar tissue of CS mice submitted to different diets (GSE62194). The microarray data quality analysis was performed with arrayQualityMetrics R package. Transcriptome data were statistically analyzed through the limma R package, in which we compared CS mice, submitted to a HFD against those on a standard diet. The differentially expressed genes (DEGs) generated in this comparison were used as an input to prospect an interactome network on the STRING (version 10.5) and NetworkAnalyst metasearch websites. Additionally, the main lipids present on the HFD and proteins associated with their transport and metabolism were utilized for prospecting a chemical-protein interaction network on the metasearch site STITCH (version 4.0). Both networks were imported and merged on the Cytoscape (version 3.5.1) software and further analyzed with the plugins MCODE (version 3.3), BiNGO (version 3.0.3) and Centiscape (version 1.4.2) to perform clustering, gene ontology (GOs) and centralities analysis, respectively. Preliminary network data analysis showed 18 significative clusters. Using gene ontology analysis, we selected five clusters that are associated with metabolism, regulation of cell death and neurogenesis. Within these clusters, it was observed the overexpression of Etfa gene (electron carrier), as well as the Sdhb (succinate dehydrogenase) and Suclg1 (succinyl-CoA ligase) key metabolism genes. These data corroborates the hypothesis that CS cells, on a HFD, presented a metabolic tendency towards β-oxidation, which is also suggested by previous in vivo studies. Considering the energy depletion observed in CS, the maintenance of cell processes would be facilitated by the increased FA catabolism. This metabolic scenario is also corroborated by the underexpression of Srebf1, since this protein product acts as a transcription factor by regulating fatty acid synthesis metabolism. In addition, we observed the presence of many processes related to neurogenesis and myelination. The Lingo1 gene, for example, which acts negatively on the myelination process and oligodendrocyte precursor cells (OPC) maturation. This regulation occurs through the RhoA GTPase pathway activation, which inhibits the aforementioned process. Since Lingo1 is underexpressed, this activation will be affected. Besides that, we also observed the overexpression of Mbp (myelin basic protein), which is a major constituent of the myelin sheath. Previous studies demonstrated that MBP has a central role on autoimmune demyelinating disorders, since MBP is essential for myelin sheaths physical stability. Also, a previous in vitro study demonstrated an inverse correlation between MBP and LINGO1 protein expression, since overexpression of LINGO1 led to reduced MBP level, and the contrary was also observed. Finally, further steps of this project include a deeper evaluation of the myelination process and cell metabolism pathways, specially focusing on its relationship with the DNA repair pathways. In addition, a correlation network analysis will be applied in order to observe the correlation patterns among genes and check the presence of highly correlated gene clusters related to CS symptoms and the HFD.

Keywords: Cockayne; Syndrome; Fat; Diet; Neuroprotection; Cerebellum

In silico prospection and evaluation of primers for application in population genetics studies of *Schinus molle* L.

BEISE, Dalvan Carlos[1]; Castro, Renata Machado[2]; Lemos, Rafael Plá Matielo[3];
D' Oliveira-Matielo, Cristiane Barbosa[4]; Stefenon, Valdir Marcos[5].
[1]Acadêmico do curso de Ciências Biológicas Bacharelado, Bolsista do Programa de Educação Tutorial (PET Ciências Biológicas), Núcleo de Genômica e Ecologia Molecular, Universidade Federal do Pampa, São Gabriel-RS.
[2]Acadêmica do Curso de Ciências Biológicas Licenciatura, Bolsista de Iniciação a Docência (PIBID Biologia), Núcleo de Microbiologia Aplicada à Biorremediação, Universidade Federal do Pampa, São Gabriel-RS.
[3]Coorientador, Núcleo de Genômica e Ecologia Molecular, Universidade Federal do Pampa, São Gabriel-RS.
[4]Coorientador, Núcleo de Genômica e Ecologia Molecular, Universidade Federal do Pampa, São Gabriel-RS.
[5]Orientador, Núcleo de Genômica e Ecologia Molecular e Núcleo de Microbiologia Aplicada à Biorremediação, Universidade Federal do Pampa, São Gabriel-RS.

Schinus molle L., known as Peruvian Peppertree, is a widely used tree in folk medicine and by the pharmacological branch due to the presence of important secondary metabolites in its essential oil. Characterized as a pioneer species due to its adaptive plasticity, it can occur in different climates and habitats, and plays an important role in restoration of degraded environments. Its fruits are appreciated by fauna that assist in the dispersal process, thus increasing ecological restoration.

With these perspectives, the objective of this work was to verify the *in silico* quality of prospective primers, obtained through the sequencing of part of *S. molle* genome, in addition to evaluate the feasibility for use in population genetics studies. The partially sequenced samples belong to the Caatinga, Mata Atlântica and Pampa biomes and the new generation sequencing (NGS) was performed by the Ion Torrent® platform. The "De Novo" assembly was through Velvet software, for the prospection of the primers was used the software SSRLocator and finally designed with the software Primer3. These were tested by *in silico* amplification using SPCR software in a range of 90 to 500 base pairs.

Sequencing resulted in a total of 13,849 contigs. The evaluation of the primers followed some criteria for use in the laboratory: only those with a single locus between the expected region are considered optimal, those with more than one loci are not considered for in vivo application or still need adjustments. Of the 201 primers verified, 48 were classified optimally, therefore their use is indicated, 46 were disregarded because they presented more than one amplified region, finally the remaining 107 did not show amplification.

The developed and validated Primers *in silico*, demonstrated great potential of application in laboratory. It is inferred that these primers can be used in studies of distribution, diversity and population genetic structure, and also related to biotechnological field.
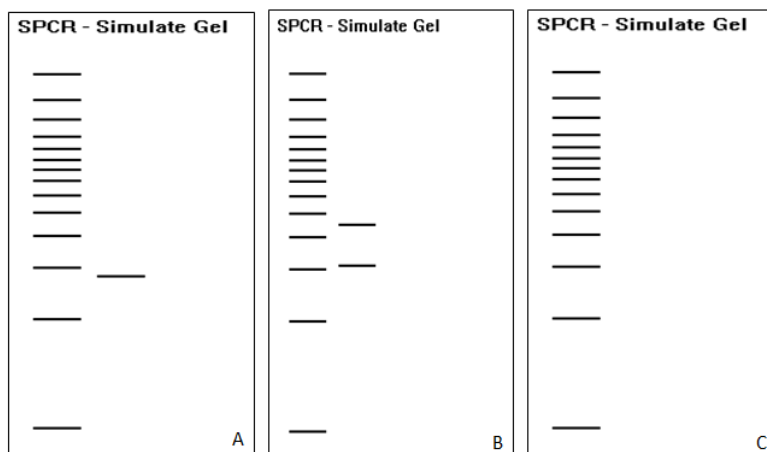

Figure 1: Amplifications obtained from the tests performed by SPCR software. A) Amplification considered optimal with application potential, B) Amplification not considered, C) There was no amplification.

SEARCHING FOR PATHOGENIC VIRUSES IN STINGLESS BEES AFFECTED BY A SEASONAL DISEASE

CAESAR, LÍLIAN[1], CIBULSKI, SAMUEL PAULO[2], CANAL, CLÁUDIO WAGECK[2], BLOCHTEIN, BETINA[3], SATTLER, ARONI[4] AND HAAG, KAREN LUISA[1]

[1] Programa de Pós-Graduação em Genética e Biologia Molecular (Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil); [2] Faculdade de Veterinária (Laboratório de Virologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil); [3] Faculdade de Biociências (Departamento de Biodiversidade e Ecologia, Pontifícia Universidade Católica do RS, Porto Alegre, RS, Brazil); [4] Faculdade de Agronomia (Laboratório de Apicultura, Departamento de Fitossanidade, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil)

In Southern Brazil the stingless bee *Melipona quadrifasciata* is no longer found in the wild and beekeepers report annual losses of their colonies due to a seasonal and acute disease of unknown causes. The symptoms are similar to those caused by certain viral infections in other bees, such as those associated with Colony Collapse Disorder (CCD). Therefore, the aim of this work is to investigate whether the annual colony collapse of *M. quadrifasciata* is related to a viral infection. For this, samples of *M. quadrifasciata* workers from symptomatic (unhealthy) and asymptomatic (healthy) colonies were collected during the outbreak periods (late summer) of 2016 and 2017. Pools of 5 (2016 sampling) and 25 (2017 sampling) workers were subjected to filtration and ultracentrifugation for the isolation of viral particles, followed by the extraction of DNA and RNA, cDNA synthesis, and, eventually, random amplification. DNA and cDNA libraries were prepared using a Nextera XT DNA sample preparation kit and sequenced using an Illumina MiSeq instrument (2x150 and 2x250 paired-end reads, insert size of 300pb and 500pb, respectively). The quality of the sequencing was inspected (FastQC) followed by the removal of reads (Sickle) with a phred quality score lower than 30. The filtered reads were then assembled into contigs using SPAdes v.3.10.1 and metaSPAdes with different parameters. Assembly statistics were generated (Quast 4.5) for comparison and choice of the best assembly based on the size and quantity of contigs. Contigs were locally blasted (e-value 1e-5) against a non-redurtant viral database extracted from the total non-redundant database (nr, NCBI) and taxonomic classification was confirmed by using blastX (e-value 1e-5) against the total nr. Alignment of reads against contigs was done with Bowtie 1.2.0 to extract with Samtools 1.3.1 the information about coverage. Four assemblies from two successive years were obtained: assembly (1) has 10399 contigs and N50 of 696pb (unhealthy bees; 2016; obtained from mixed DNA and cDNA); (2) has 2625 contigs and N50 of 1443pb (healthy bees; 2016; obtained from mixed DNA and cDNA); (3) has 16856 contigs and N50 1122pb (unhealthy bees; 2017; DNA only); and (4) has 10226 contigs and N50 872pb (unhealthy bees; 2017; cDNA only). The viromes of assemblies 1-4 contain 3, 1, 240 and 311 viral contigs respectively, with average coverages per base of 5.41, 3.67, 35.69, and 72.13x. Viromes (3) and (4), obtained with a larger sampling in 2017, contain three contigs similar to pathogenic bee viruses. Assembly (3) contains a 220bp sequence similar (46.5% identity; 1.91e-12 e-value) to the Apis mellifera Filamentous Virus (AmFV) with coverage of 1.4x. Assembly (4) also contains a 209pb sequence similar to AmFV (96% identity; 1.38e-21 e-value) with coverage of 1.4x and a 330pb sequence similar (37.8% identity, 4.26e-17 e-value) to Acute Bee Paralysis Virus (ABPV), with coverage of 3.3x. Both viruses, for which we find similar sequences in our assemblies, are known to cause symptoms, such as inability to fly, trembling and crawling, which are also observed in unhealthy *M. quadrifasciata* workers. Moreover, ABPV is known to cause the death of a worker in a few days with less than 100 viral particles, which could explain its low coverage. Whether or not these viruses, or other unknown viruses, are at the root of the collapse of *M. quadrifasciata* colonies is a matter of further investigation of our research group.

ANALYSIS OF AUTODOCK VINA EXHAUSTIVENESS PARAMETER IN THE EXECUTION OF MOLECULAR DOCKING SIMULATIONS

CECOTTI, LETÍCIA K. S.[1]; WEHRLI, ADRIANO V.[1]; MACHADO, KARINA S.[1]

[1] Computacional Biology Research Group, Universidade Federal do Rio Grande, RS

Molecular docking is a computational technique applied to predict the best position and conformation of a ligand on the binding site of a target receptor. It also estimates the binding energy between the final receptor- ligand complex. AutoDock Vina is one of the most used molecular docking tools. To perform a molecular docking simulation with AutoDock Vina the user defines the coordinates and size of a grid box where the docking algorithm will try to find the best position of the ligand on the receptor structure. Thus, during the execution of one molecular docking simulation, AutoDock Vina randomly chooses an initial position for the ligand on the grid box and starts to search for the best ligand conformation and position. The time spent on this search is varied heuristically and varies linearly with the value of Exhaustiveness. So, the objective of the work is to analyze the impact of the exhaustiveness parameter in the molecular docking results.

To achieve this, we applied redocking technique, which is defined as the process of removing the ligand from a receptor-ligand experimental complex, predicting the ligand conformation and position using molecular docking and comparing these ligands final coordinates. All the molecular docking simulations were performed using AutoDock Vina, and the protein-ligand complexes considered on the redocking experiments were obtained from the PDBbind Database, which has experimental measured affinities between these complexes.

In this work, were performed experiments considering the proteins, HIV-1 Protease (PDB ID 1A94) and HIV-1 Protease G48H (PDB ID 1A9M), using a box size of 34 x 58 x 38 = 74,936 $\mathring{A}^3$ for protein 1A94 and a box size of 40 x 44 x 54 = 95,040 $\mathring{A}^3$ for protein 1A9M. For each receptor-ligand complex, it was performed docking simulations considering Exhaustiveness parameter as 1, 2, 4, 8, 16, 32, 64 and 128. Each experiment was performed 10 times, and we calculate the value of RMS (Root Mean Square), between the original position of the ligand (from PDBbind) and its position at the end of the molecular docking simulation. Figure 1 shows the RMS averages calculated for both proteins. The errors bars represent one standard deviation from the average. Larger bars represent more fluctuation among the results and smaller bars represent agreement in the results.

Considering the results presented on Figure 1 we can notice that for 1A94 receptor, for the exhaustiveness value set to 16, accurate results are already found, while for 1A9M protein, we obtain only with the exhaustiveness value of 32. For the two proteins studied, it is noted that there are cases where defining a larger parameter for the value of exhaustiveness is not always necessary. As future work, it is intended to carry out these experiments of the same sample with an expressive amount of receptor-ligand complexes to form a database. We expect to apply Data Mining and Machine Learning techniques in this database in order to learn how to safely define the Exhaustiveness for docking with Vina.
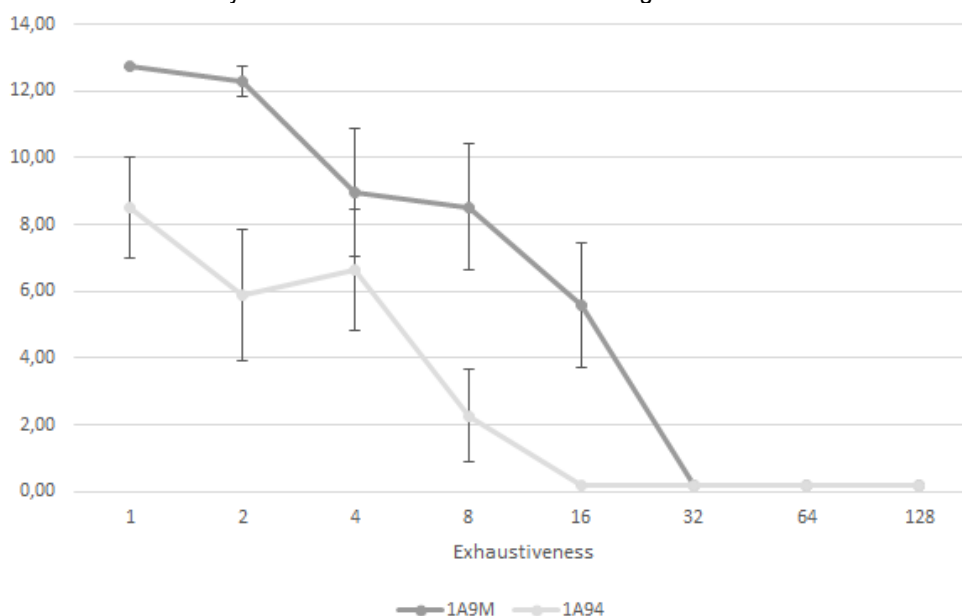

Figure 1: Average RMSD

GENE INTERACTION NETWORK ANALYSIS IN WOLF-HIRSCHHON SYNDROME

CORRÊA, THIAGO[1]; MERGENER, RAFAELLA [1]; LEITE, JÚLIO CÉSAR [2]; GALERA, MARCIAL FRANCIS [3]; MOREIRA, LILIA MARIA DE AZEVEDO [4]; VARGAS, JOSÉ EDUARDO5, RIEGEL, MARILUCE[1,2].

[1] – PPG em Genética e Biologia Molecular, UFRGS, Porto Alegre, RS, [2] – Serviço de Genética Médica, HCPA, Porto Alegre, RS; [3] – Dept de Pediatria, UFMT, Cuiabá, MT; [4] – PPG em Genética e Biodiversidade, UFBA, Salvador, BA; [5] – Centro Infant, PUCRS, Porto Alegre, RS e UPF, Passo Fundo, RS

Introduction and objectives: Deletions in the 4p16.3 region cause Wolf-Hirschhorn syndrome (WHS), a contiguous gene deletion syndrome involving variable size deletions. This study aimed to perform a gene interaction network analysis within the WHS critical region and to stablish the cytogenomic profile of the chromosome rearrangements involving the 4p16.3 region. Material and methods: 16 samples from individuals with a clinical indication of WHS were retrospectively analyzed of which 9 had a cytogenetic visible deletion and 7 a submicroscopic deletion not previously identified. Using FISH, chromosomal microarray analysis and whole exome sequencing, we define the critical breakpoints within the 4p16.3 chromosome rearrangements. Gene Multiple Association Network Integration Algorithm (GeneMANIA) version 3.1.2.8, available at http://www.genemania.org/ was used to identify protein– protein interactions (PPI). In the present study, the association data of GeneMANIA was based on the PPI databases and co-expression profiles, in which each interaction between proteins are experimentally proven. In addition to 12 classical terminal deletions, we mapped 1 interstitial deletion, 2 ring chromosomes and 1 typical translocation 4;8. The deletions sizes ranged between 3.7 and 26 Mb. We fully characterized the 4p deletions in 8 samples. An initial genes list from 343 genes a interactome network composed by 136 nodes and 750 edges was obtained. From these nodes, GO categories were identified as more significant as positive regulation of vasoconstriction and dopamine receptor signaling pathway. 4p chromosomal rearrangements associated with WHS have different mechanisms of origin, which leads to a heterogeneous spectrum of phenotype features, from very subtle or mild, to a wide range of severe abnormalities. The critical region in our study includes four candidate genes (TACC3, FGFR3, LETM1, and WHSC1) associated with seizures and microcephaly. Spaning a common region of 170 kb. This study refined the critical chromosomal susceptibility region within 4p16.3 and is further exploring the gene interaction between candidate genes related to seizures and microcephaly associated with WHS.

Altered Microbial-Microbial Connections and Absence of a Coordinated Succession Are Associated with Necrotizing Enterocolitis in Preterm Infants

Dobbler, Priscila[1]; Freitas, Anderson S.[1]; Roesch, Luiz[1]

[1] Centro Interdisciplinar de Pesquisa em Biotecnologia – CIP-Biotec, Universidade Federal do Pampa, São Gabriel, RS, Brasil

Necrotizing enterocolitis (NEC) is a major cause of complications and death in preterm infants. It is characterized by inflammation of the bowel with air accumulation leading to necrosis. NEC onset and development is thought to have a complex and multifactorial etiology. Current clinical predictors of NEC are poorly defined and nonspecific, and once this signs are diagnosed, clinical progression is fast. Although NEC affects about 15% of all preterm newborns, the diverse etiologies of NEC have to date remained largely elusive. Therefore, the field is in urgent need for the development of risk biomarkers allowing for an early diagnosis of NEC and prediction of NEC associated outcomes. So far, research has focused on the identification of a single causative pathogen, however no specific microorganism has been consistently found. Thus, instead of focusing in the detection of specific causative agents, we investigated the community assembly and its microbial relations. To address this challenge we compared microbiome successions (microbial abundance and structure) in newborns that developed Necrotizing Enterocolitis with preterm controls without NEC. 40 preterm babies were enrolled in the study, 11 in the NEC group. Average gestational age in the control group was 31.02 weeks and in the NEC group was 29.79 weeks, a Welch Two Sample t-test showed that the difference between both groups were not statistically different. All newborns were delivered in the Hospital das Clínicas de Porto Alegre –RS and admitted in the same Neonatal Intensive Care Unit (NICU). A total of 135 fecal samples ranging from first evacuation (meconium) till the fifth week of age. Following DNA extraction, the V3 region of the 16S rRNA gene was amplified from all samples, using the 515F and 806R primers and the amplicons were sequenced using the Ion Torrent-PGM technology. Quality filtering and assembling of sequences into Operational Taxonomic Units were performed according to the Brazilian Microbiome Project pipeline, available at: http://www.brmicrobiome.org. Longitudinal analysis of phyla abundance was performed in QIIME, and the co-occurrence network was calculated for the meconium samples using SparcCC, implemented in Mothur. Network analysis were performed using only phylotypes present in 70% of samples, in each group, and only strong correlations were kept, with a Pearson's correlation of ± 0.7 and a p-value <0.05. Overall, the controls presented with a shift in the most abundant phyla, from *Proteobacteria* to *Firmicutes* in the first week of life, which steadily increased until the fifth week. While in the NEC group, this shift in dominance was chaotic and not permanent, marked by the high abundance of *Proteobacteria* and low abundance of *Firmicutes* at third week of age. Moreover, in order to find the microbial correlations essential for community assembly and/or stability at birth, we applied the network analysis, and the eigenvector centrality was used to find the microbial phylotype most influential in each network. A microbial community comprised only by obligatory anaerobic, such as *Phascolarctobacterium*, which is the most influential in the controls, *Faecalibacterium prausnitzii*, *Rikenellaceae*, *Paraprevotella* and *Lachnospira*, was present in the gut microbiota of controls, in the first days of life, while entirely absent in infants with future diagnosis of NEC. Moreover, a phylotype identified as belonging to the family *Enterobacteriaceae* and another identified as *Pseudomonas* were the most influential in the network of the NEC group, however entirely absent in the controls. Premature babies were previously shown to have impaired early colonization of strictly anaerobes and subsequent transition to a more anaerobic gut environment. In our co-occurrence analysis an entire cluster of strictly anaerobes were absent in the first days of life in the gut of the preterms with future NEC development. In addition, *Faecalibacterium prausnitzii*, found in the controls, was previously shown to have a protective role in the gut mucosa. Therefore, the gut environment of preterm newborns with future development of NEC presented altered microbial connections, and a subsequent absence of a coordinated microbial succession. Thus, our data suggest that early-altered microbial network structure, during the first days of life, correlate with Necrotizing Enterocolitis risk in preterm infants. Confirmation of our findings in other NICU's might facilitate the development of a microbiota based early NEC screening approach.

Application of Systems Chemical-Biology for the Study of the Association of Adenocarcinoma Sinonasal and the Chemical Risks of the Leather-Footwear Sector.

Dos Santos, Marcelo Pacheco [1]; Da Silva, Fernanda Rabaioli [1]

[1] Laboratory of Cellular and Molecular Biology, Master in Environmental Impact Assessment, University La Salle, Canoas, Brasil. E-mail: marcelopacheco.adm@gmail.com

Sinonasal adenocarcinoma is one of neoplasms that put Brazil among the countries which have a high frequency of head and neck cancer cases of Latin America. This type of cancer is known for its rapid progression and its easily in produces metastasis. Although it is recognized as an occupational disease and has its well defined etiology still a little information is known about the molecular basis of this type of cancer. Sinonasal adenocarcinoma is correlated with exposure to wood dust, leather dust, also metals like chrome, nickel and formaldehyde. Some of these substances are found in the leather-footwear sector, this is a very significant sector, in both ways, socially and economically to Brazil. The objective of this study was to predict some of the molecular pathways associated with the genesis and/or the progression of sinonasal adenocarcinoma and its effects on leather-footwear industry workers, using the tools of systems biologyFor this purpose, differentially expressed proteins in the sinonasal cancer and atmospheric contaminants from leather-footwear sector were prospected and joined to form a single network of interaction. For data acquisition and formation of the different networks, mining tools were used as the STRING 10.0 and STITCH 5.0, which link proteins and protein-protein chemical compounds. The design of binary networks were performed by Cytoscape 3.4.0 version. The cluster analysis of gene ontology interaction networks was carried out by the Complex Molecular Detection (MCODE), the Biological Network Gene Ontology (BiNGO) was used for assessment of the groups of proteins, and CentiScaPe for the analysis of centrality. A network containing 2424 node and 27600 connectors were prospected. Among those, 33 are air pollutants and 27 are proteins present in *Homo sapiens*. The clustering analysis indicated 79 modules. Gene Ontology analysis indicated an association between cluster 2 proteins and cell adhesion process and the cluster 5 protein and the mechanism of apoptosis. From our analysis it was possible to propose an action's molecular model of the air pollutants of this sector in cell adhesion and inhibition of apoptosis and the consequent formation of sinonasal cancer. In the scheme is represented the increase of oxygen reactive species (EROs) production by the occupational exposure to the complex pollutants mixture. This increase of EROs activates the expression of TGFB which induces to TGFBI [1]. This fact increases the expression and the collagen accumulation that consequently generates a greater contribution of calcium in the intracelular medium. The intracelular calcium stimulates the MMP production and this increase inhibits the cell adhesion process, as well as recruits annexin proteins that activate AKT and, consequently, block the cell apoptosis. In this model, the largest contribution of calcium in the intracellular space is the key mechanism that triggers the desregulation of cell signaling and indirectly, the blocking of apoptosis and cell adhesion.

Keywords: sinonasal Adenocarcinoma, leather footwear industry, systems biology

[1] KRSTIĆ, Jelena et al. Transforming growth factor-beta and oxidative stress interplay: implications in tumorigenesis and cancer progression. Oxidative medicine and cellular longevity, v. 2015, 2015.

ANTIBIOTIC EFFECTS IN THE STUDY OF MICRORGANISMS ASSOCIATED WITH PRETERM BIRTH BY BIOINFORMATICS TOOLS

FREITAS, ANDERSON S.[1]; DOBBLER, PRISCILA C.T.[1]; ROESCH, LUIZ F.W.[1]

[1] Centro Interdisciplinar de Pesquisas em Biotecnologia – CIP-Biotec, Universidade Federal do Pampa, Campus São Gabriel, São Gabriel, RS, Brasil

The study of microrganisms associated with several states of healthy and disease was strongly increased in the last decades. Novel Molecular Biology techniques, like Next Generation Sequencing, associated with the efforts from a large number of researches across the world, led to a proeminence of the role of microorganisms in diseases never correlated before to microbial influence, such as cardiovascular diseases, neurodegenerative disorders, authism and preterm birth. Changes in normal balance of microbial species in body environments generally are indicative of disorders. In vagina of pregnant women imbalance of *Lactobacillus* sp. levels, for example, may lead to bacterial vaginosis and urinary tract infections. Some studies have also investigated how these imbalance, not just in *Lactobacillus* levels but also in whole vaginal microbial community, are correlated with preterm birth. Although the scientific community is concentrating efforts to better understand microbial disbiosis, it is important remind that several confounding variables on such studies exists. One of them is the antibiotic usage. The use of antibiotics aim to stop potential or real infections that may cause health problems. Nonetheless, they act killing a wide group of microrganisms, creating a selective environment for most microbial species. Specifically speaking about preterm birth, antibiotics are used when there is a suspect of infection by Group B *Streptococcus* (GBS) in the vaginal cavity. This usage may affect microbial abundance and cause microbial imbalance. Based on this, the present work aimed to explore the antibiotic influence in the vaginal microbiota and its effects on abundance and diversity of the microbial community. A total of 23 vaginal swab samples were collected from mothers recruited during hospital admission for delivery at the Hospital de Clinicas de Porto Alegre (HCPA). None of the women sampled had urogenital infection in a period of three months before the labor. All mothers had preterm delivery and, among them, 15 mothers have used antibiotic (penicillin) 4 hours before labor (average gestational age = 30.97 ± 0.49) whereas 8 mothers have not used antibiotic (average gestational age = 30.20 ± 0.62). The 16S rRNA gene (V4 region) was amplified and sequenced by using the Ion PGM Platform. Sequences were processed using the pipeline from the Brazilian Microbiome Project and library comparisons were performed using the phyloseq package in R and the online tool MicrobiomeAnalyst. A marginal difference in alpha and beta diversity was found between groups (p=0.4056 and p=0.065, respectively). An analysis of phyla abundance has shown a very lower abundance of *Firmicutes* and *Actinobacteria* and increased abundance of *Proteobacteria* and *Bacteroidetes* in those groups with antibiotic. Differential abundance analysis at OTU level had shown a positive influence of the antibiotic usage to *Pseudomonas* and a negative influence to *Anaerococcus* and *Prevotella*. Antibiotic usage before labor affected microbial composition of vaginal environment, especially in most high taxonomic levels. Penicillin act mainly in gram-positive bacteria, what explain the decrease of *Firmicutes* when antibiotic was used. As the vaginal microbiome is generally dominated by *Lactobacillus*, a representative of the genera *Firmicutes*, or others fermenting bacteria, such as *Prevotella* and *Anaerococcus*, that were also affected by the antibiotic usage, not taking in account their effects is a mistake that can lead to a high incidence of false positives or false negatives during bioinformatic analysis. This way, it is possible conclude that antibiotic is an important confounding variable in studies of vaginal microbial influence in preterm birth and must be always considered in analyses of microbial abundance in vaginal microenvironment.

EVIDENCE FOR *HOVENIA DULCIS* ANTI –INFLAMMATORY ACTIONS BASED IN SYSTEM CHEMO-BIOLOGY

GODOI, RAFAEL SOUZA[1]; ALMERÃO, MAURÍCIO PEREIRA [1,2]; DA SILVA, FERNANDA RABAIOLI[1,2]
1- Laboratório de Biologia Celular e Molecular, curso de Ciências Biológicas Universidade La Salle, RS, Brasil
2-Mestrado em Avaliação de Impactos Ambientais, Universidade La Salle, Canoas, RS, Brasil

Hovenia dulcis Thunberg (Rhamnaceae), Japanese raisin tree is a native species from East Asia which was introduced in many countries, including South America, especially Brazil. It presents medicinal properties as antioxidant, antipyretic, hepatoprotective and also decreases the symptoms of alcohol intoxication. Because of these medicinal properties it has been widely used as an ingredient in natural health products and functional foods in East Asia and, increasingly, information on the medicinal properties of this plant have been published on scientific literature. Despite this, the biological mechanisms involved in these medicinal actions are poorly investigated. Flavonoids, a class of secondary plant metabolites, are chemical compounds abundant in *H. dulcis* for those many beneficial health actions have been associated: antioxidant, anti-tumor, antiallergic, epigenetic modulator and anti-inflammatory. Studies indicate that *H.dulcis* presents, among many medicinal activities, an anti-inflammatory action through a down-regulation mechanism that suppresses NO (nitric oxide) production and reduces, after a pretreatment, the phosphorylation and degradation of IkB-a and NF-k. Flavonoids such as quercetin and myricetin are known to suppress macrophage-secreted TNF-α but also to reduce levels of pro-inflammatory cytokines and chemokines such as IL-8 and IL-6 in *in vitro* experiments. Under these premises we used Systems Chemo-Biology tools to investigate how the *H. dulcis* flavonoids act on the human immune system, and which are the possible metabolic pathways involved. For this, we performed a systematic review of *H. dulcis* chemical compounds in the PubMed database. Subnetworks of proteins and chemical compounds were created using mining tools STRING 10.0 and STITCH 5.0 and were merged using Cytoscape 3.4.0. The cluster analysis and assessment of bioprocesses were performed using Complex Molecular Detection (MCODE) and the Biological Network Gene Ontology (BiNGO) tools, respectively. The centralities analysis was performed using CentiScaPe plugin. A network containing 198 nodes and 1819 connectors were prospected. Of these nodes, 11 were *H. dulcis* chemical compounds and 187 were proteins present in *Homo sapiens*. The clustering analysis indicated 7 modules and gene ontology analysis showed bioprocesses as: MAPKKK cascade, I-kappa B kinase/ NF- kappa B cascade, regulation of transcription DNA dependent, regulation of gene-specific transcription. The centrality results showed node hub-bottlenecks, which may be considered important points of the information flow in the network, and play a role in the regulation of some biological processes, like immunological pathways. In the *H. dulcis* network, quercetin, HDAC1, MAPK1 were indicated as node hub-bottlenecks. Quercetin is a flavonoid, compound that exhibits several medicinal activities, as well as epigenetic regulations. It is known that HDAC1 acts on the regulation of gene expression of eukaryotic cells, among other actions, inhibiting NF-kappaB transcription activity. In addition, MAPK1 and NF-kB play an important role in the cellular response by external signals and also playing an important role in the regulation of pro-inflammatory molecules such as cytokines and chemokines. Therefore, it seems *H. dulcis* flavonoids can have a role of epigenetic regulation on HDAC1, which would result in a lower activity of MAPK1 and NF-kB, generating a lower release of pro-inflammatory molecules such as TNF-a, IL-1, IL-6 and other cytokines.

Keywords: *Hovenia dulcis,* epigenetic regulation, anti-inflamatory activity.

DESIGN AND EXPRESSION OF RECOMBINANT CHIMERIC PROTEINS USING LEPTOSPIRAL ANTIGENS IDENTIFIED BY REVERSE AND STRUCTURAL VACCINOLOGY

HECKTHEUER, AMANDA S.[1]*; BETTIN, EVERTON[1]; CUNHA, CARLOS E. P.[1]; DELLAGOSTIN, ODIR A.[1]; GRASSMANN, ANDRÉ A.[2]; MCBRIDE; ALAN J. A.[2]

[1]*Laboratório de Vacinologia, Unidade de Biotecnologia, Centro de Desenvolvimento Tecnológico (CDTec), Universidade Federal de Pelotas (UFPel), Pelotas, RS, Brazil*
[2]*Laboratório de Pesquisa em Doenças Infecciosas, Unidade de Biotecnologia, CDTec, UFPel, Pelotas, RS, Brazil*
*hecktheuer.as@ufpel.edu.br

Leptospirosis is a neglected zoonosis of global distribution caused by diderm spirochetes belonging to the genus *Leptospira*. The global incidence is estimated to be over one million cases every year, resulting in ~60.000 deaths. There are 15 *Leptospira* spp. that cause leptospirosis, serologically classified in ~300 serovars. The current leptospiral vaccines are inactivated whole-cell preparations (bacterins), with worldwide veterinary use but limited use in humans. Bacterins confer a short term immune response that only protects against the serovars included in the vaccine and cause adverse reactions. Several recombinant proteins have been evaluated towards the development of new improved leptospiral vaccines, with moderate success. A new, effective universal vaccine against leptospirosis will likely include surface-exposed epitopes that are conserved among the pathogenic leptospires and able of generating neutralizing immune responses. Our group recently described an innovative bioinformatics approach for the discovery of leptospiral vaccine candidates based on reverse and structural vaccinology. We identified 17 β-barrel transmembrane proteins (βb-OMP) that were highly conserved among pathogenic *Leptospira* spp. Herein, we present the design and heterologous expression of five recombinant chimeric proteins that included surface-exposed regions from all 17 βb-OMPs. The three-dimensional (3D) structures of the 17 βb-OMPs were predicted by I-TASSER and quality assessed. The amino-acid sequences from these proteins were analysed using NetMHCII 2.2 to predict immunogenic T cell epitopes with high affinity to 14 HLA-DBR alleles of the human major histocompatibility complex class II (MHCII). Only epitopes predicted as strong binders (IC50 < 50 nM) were selected. The MHCII epitopes were manually mapped onto the 3D models of each βb-OMP. Surface-related MHCII epitopes were identified and further analysed using BepiPred 1.0 for the presence of B cell linear epitopes. We identified 44 regions within all 17 of the βb-OMPs that were predicted to be exposed on the leptospiral cell surface. These regions were analysed by multiple sequence alignment to confirm their presence in the orthologs in the remaining pathogenic *Leptospira* spp. Each chimera was designed to contain surface-related epitopes from 3-4 βb-OMPs. Each chimera comprised six to 11 different surface-related immunogenic regions. The DNA coding sequences for each of the chimeras were codon optimized for *E. coli*, synthetized and cloned into the expression vector pAE. The recombinant plasmids were characterized by DNA sequencing and used to transform *E. coli* BL21 (DE3) Star. Heterologous expression was induced by 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h. Cells were harvested, lysed, and the recombinant proteins were purified by nickel affinity chromatography using the AKTA Start automated system. The recombinant proteins were expressed as inclusion bodies and were solubilized in 8M urea prior to purification. Subsequently, the purified proteins were dialyzed against phosphate saline buffer (PBS) to remove the urea and imidazole. The dialyzed chimeras remained soluble in PBS after dialysis, a significant improvement compared to the production of whole, insoluble βb-OMPs that aggregate in polar, hydrophilic buffers. Precipitated proteins adopt tertiary structures dissimilar to the native protein, limiting their use as vaccine antigens. Analysis by SDS-PAGE and Western blotting, using anti-6xHis tag antibodies, showed that the recombinant purified proteins had molecular masses of 41, 38, 36, 33, and 30 kDa, as expected. Currently, these chimeric proteins are undergoing characterization with convalescent sera from human and animal leptospirosis patients. The subcellular localization of the native proteins used in the construction of the chimeras is being experimentally confirmed. Future experiments will evaluate the immune response induced by each of the chimeric proteins. Vaccine efficacy and cross-protection induced by the chimeras will be evaluated in the hamster model of acute leptospirosis using both homologous and heterologous challenge strains. In conclusion, this is a promising and innovative approach, rationally designed to avoid the problems associated with traditional vaccine discovery and, hopefully, will lead to the development of a universal vaccine against leptospirosis.

MICRORNA-16 FEEDBACK LOOP WITH P53 AND WIP1 CAN REGULATE CELL FATE DETERMINATION BETWEEN APOPTOSIS AND SENESCENCE IN DNA DAMAGE RESPONSE

ISSLER , MARIA V.[1]; MOMBACH, JOSÉ C.M.[1]
[1] Department of Physics, Universidade Federal de Santa Maria, Santa Maria, Brazil;

Cell fate regulation is an open problem whose comprehension impacts several areas of biosciences. DNA damage induces cell cycle checkpoints that activate the p53 pathway to regulate cell fate mechanisms such as apoptosis or senescence. Experiments with different cell types show that the p53 pathway regulates cell fate through a switch behavior in its dynamics. For low DNA damage the pathway presents an oscillatory pattern associated with intense DNA damage repair while for high damage there are no oscillations and either p53 concentration increases inducing apoptosis or the cell enters a senescence state. MicroRNA 16-1 (miR-16) seems to regulate the fate between senescence and apoptosis in non-Hodgkin lymphomas.

To investigate the regulation of cell fate we developed a logical model of the G1/S checkpoint in DNA damage response that takes in account different levels of damage and contemplates the influence of miR-16 through its positive feedback loop formed with p53 and Wip1. In a logical model the proteins have discrete state values and the interactions are represented by the logical operators AND, OR and NOT. The input of the model is the level of DNA damage: no damage(S=0), low or reparable damage (S=1) and high or irreparable damage (S=2). Simulations of the model were generated using the tool GINsim 2.9.5.
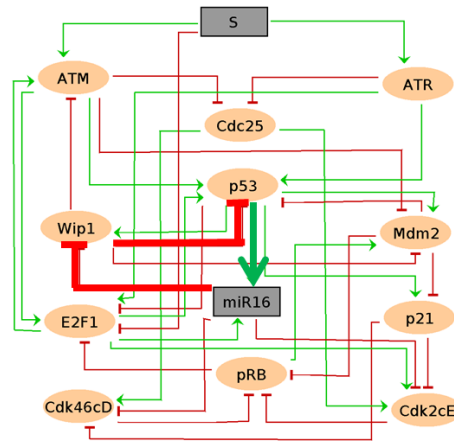

Figure 1: Regulatory network for the G1/S checkpoint pathway. Elliptic nodes represent proteins and rectangular nodes inputs or microRNA. The input node in grey at the top of the network denote DNA damage level. Green lines represent activations and red lines, inhibitions.

The model reproduces the observed cellular phenotypes in experiments: oscillatory (for low DNA damage) regulated by negative feedback loops involving mainly p53 and Mdm2 and apoptotic or senescent (for high DNA damage) regulated by the positive p53-Wip1-miR-16 feedback loop. We find good agreement between the level of DNA damage and the probability of the phenotype produced according to experiments. We also find that this positive feedback makes senescent and apoptotic phenotypes to be determined stochastically (bistable), however controlling the expression level of miR-16 allows the deterministic control of cell fate determination as observed experimentally.

In this work we proposed a model for cell fate regulation in DNA damage response involving the p53 pathway and miR-16. The model shows that under low damage (S= 1) the experimentally observed oscillatory phenotype is regulated by negative circuits involving mainly p53 and Mdm2. The knockdown of Mdm2 (Loss of function) or p53 nodes abrogates the oscillatory phenotype. So, our results present fair agreement with experimental works and we conclude that after p53, Mdm2 is the main regulator of the oscillatory dynamics. In addition, we found that this dynamic is robust against several perturbations of the model elements including those of node miR16. For high damage the positive circuit Wip1-p53-miR16 makes the fate determination between apoptosis and senescence. Perturbations of miR-16 destroy the bistable dynamics pushing the system to decay deterministically in a specific phenotype that can be changed according to miR16 expression level and DNA damage. For S= 2 it presents a phenotype change from senescence to apoptosis with increasing miR16 expression levels. Finally, we determined that the bistability is robust to several LoF and GoF perturbations of all network nodes showing that it is a general solution of our model and not a particular case. We also found that Wip1 is the main regulator of the bistability since its perturbations destroy the bistable dynamics.

Keywords: cell fate, logical modeling, G1/S checkpoint, cell cycle regulation.

Bioinformatics: A case study to distinguish between coding and non-coding RNAs through complex networks

Ito, Eric Augusto; Katahira, Isaque; Lopes, Fabricio Martins

Universidade Tecnologica Federal do Parana - Campus Cornelio Procopio,

Avenida Alberto Carazzai, 1640 - Centro, Cornelio Procopio- PR, Brasil

{eric.ito,isaque.katahira,fabricio.lopes,}@utfpr.br

http://www.utfpr.br

With the emergence of Next Generation Sequencing (NGS) technologies, a large volume of biological data was quickly sequenced at relatively lower costs. The sequencers perform the extraction of data from thousands of base pairs simultaneously, producing a massive and complex data volume to be analyzed. In this sense, computational tools are increasingly needed to understand the functioning of organisms at various levels of detail, through the genome, transcriptome, and proteome to the phenotype. Given this need for understanding, the objective of the present research is to represent the biological sequences through complex networks and to extract measures that characterize them. Complex networks are chosen because of their efficiency in representing real systems in which biological systems are embedded. The study is justified by the fact that it contributes to the understanding of many important biological processes, such as: protein coding, regulation of gene expression and cell cycles, often associated with certain mutations and diseases. To demonstrate the applicability of this approach, it was initially performed the extraction of metrics of 1000 mRNAs and 1000 lcnRNAs from GENCODE GRCh37.p11 - Homo sapiens. In order to do this, the metrics of centrality and path length were chosen, since the observation of the number of connections made by the vertices, as well as the distance between the interactions, allowed to characterize the analyzed sequences. The configuration of the network was made from the parameters of the step size and word size, with the increasing elimination of less connected vertices and the extraction of measurements from each of these thresholds, in order to compose a matrix of characteristics. In this matrix, the supervised classification algorithms Naive Bayes, Random Forest, J48 and SVM are applied, both with cross validation of 10 folds. The results obtained allowed to observe 81.25% of the correct answers using the classifier Naive Bayes; 86.65% with Random Forest; 83.55% with J48; e, 84.5% with SVM. Following the methodology was applied to Pan Troglodytes organism, with 1906 mRNAs and 1166 ncRNAs, using the Random Forest classifier with cross validation of 10 folds, with an accuracy of 98.0% of the mRNAs and 99.1% of the ncRNAs. The predictions of the mRNAs were able to overcome two other prediction tools: CNCI with 90.2% and PLEK with 87.1%. In the case of prediction of ncRNAs, the proposed method was able to correctly predict 99.1% of the ncRNAs, a percentage very close to the other two tools: CNCI with 100% accuracy of ncRNAs and PLEK with 99.5% accuracy of ncRNAs . It is noteworthy that the proposed methodology uses open source tools and can be implemented on a personal computer. These results justify the feasibility of the applied methodology, since the correct values demonstrate the representativeness of the sample for a greater amount of data . Thus, it opens the way for the identification of biological sequences in other organisms.

Keywords: Bioinformatics. Supervised classification. Network Modeling. Pattern Recognition. Biological Sequences.

Transcriptome assembly and functional annotation of the Antarctic green algae *Prasiola crispa*

Maciel, Lucas F[1]; Echeverria, Pablo M[1]; Dezordi, Filipe[1]; Carvalho, Evelise L[1]; Wallau, Gabriel L[2]; Pinto, Paulo M[1];

[1] Laboratório de Proteômica Aplicada, Universidade Federal do Pampa, São Gabriel, RS, Brasil;
[2] Departamento de Entomologia Centro de Pesquisas Aggeu Magalhães, Fiocruz, Recife, PB, Brasil;

Prasiola crispa (Lightfoot) is a green macroalga belonging to *Trebouxiophyceae* class, first collected in Scotland in 1777, with a wide distribution from Arctic to Antarctica. In the Antarctic territory are between the most important primary producer, where Prasiola crispa needs to tolerate extreme environments in the course of the seasons, such as repeated freeze and thaw cycles, physiological drought, limitation of nutrients, salinity stress and high levels of UV radiation. Due to its ability to colonize such an inhospitable and extreme environment, it must have adaptive mechanisms naturally selected by evolution. The genes and biomolecules behind these mechanisms have a great interest and potential in Biotechnology field. Therefore, in order to access this potential biotechnological source and better understand the genetic and metabolism of this species, we sequenced the transcriptome of *P. crispa.* The RNA-seq was chosen because of its advantages of being cost-effective, highly sensitive and accurate. Besides, it has been widely used for gene discovery and functional and comparative genomics in non-model species. The aim of this work was the transcriptome assembly and functional annotation of *P. crispa* in order to search transcripts behind the important adaptative mechanisms.

Prasiola crispa was collected in areas near Arctowski Polish Station Region, Admiralty Bay, King George Island (61°50' – 62°15' S and 57°30' – 59°00'W), Antarctic. Total RNAs from samples were extracted using RNAqueous Micro Total RNA Isolation Kit (Thermo Fisher Scientific Inc., USA). The transcriptome was sequenced by Macrogen Service using the Solexa-Illumina HiSeq 2500 next-generation sequencing platform. A paired-end reads with a read size of ~100 bp separated by insert size of 300 bp was employed. Raw reads from data sets were filtered to remove the adapter sequences and low-quality reads with Fastx-toolkit and Trimmomatic. After that, we used Trinity as Bruijn graph assembler with 25 kmer size. Due to the sequencing of a complex sample extracted from the Antarctic soil we expected some amount of bacterial and fungi contamination. In order to clean our transcriptome, we performed a tblastx with default parameters search against all the NCBI nucleotide non-redundant database and recovered all contigs in which the best blast hit occurred with algae and plant sequences. Then we used Bowtie 2 with default parameters to recover only the reads that mapped against those *P. crispa* contigs. The assembled unigenes were searched against the NCBI protein non-redundant (NR) database with the BLASTX algorithm, the E-value cut-off was set at 10^{-10}. Genes were tentatively identified based on the best hits against known sequences. Blast2GO was used to predict the functions of the sequences, to assign gene ontology (GO) terms and to predict the metabolic pathways in KEGG databases.

A total of 42,978,976 raw reads were obtained from the sequencing. After a stringent filtering process to remove contamination and sequencing errors 5,233,428 valid reads were recovered. Using the processed reads the transcriptome was assembled into 17,205 unigenes, ranging from 200 to 12,802 bp and an N50 of 1,000 bp. Of the 17,205 unigenes, 3,973 (23.09 %) were >1,000 bp long. *P. crispa* when compared with others species from the Trebouxiophyceae class with transcriptome already sequenced, *Chlorella minutissima* with 14,905 unigenes and *Trebouxia gelatinosa* with 19,601, has an intermediate size. The BLASTX result demonstrates that a total of 8,980 (52.19 %) unigenes had at least one hit on NCBI protein NR database. Thus, 47.81 % unigenes had no match to any of the sequences in this database. This was expected, taking into account that there is very little information of sequences from close species, indicating that they may contain novel sequences.

A total of 7,009 (40.73%) unigenes had at least one GO term assigned and these terms were utilized to annotate the sequences. Using the annotation and GO terms we searched sequences that could be putatively involved with important adaptative mechanisms. Sequences associated with response to UV radiation (GO:0070914; GO:0010224; GO:0034644; GO:0009411), nutrient transport and reservoir activity (GO:0006810; GO:0045735), cold response (GO:0009409) and osmotic stress (GO:0047484) were found. Besides, results from KEGG database show several different metabolic pathways actives on *P. crispa*. Among them, the thiamine metabolism, which in plants is a cofactor in response to abiotic stress, can be highlighted with 767 unigenes associated.

In conclusion, our results demonstrate that our transcriptome assembly and functional annotation was satisfactory. Besides, there are lots of unigenes that need to be analyzed individually in order to confirm their true biological function and possible biotechnological application. The perspectives for future works are the analyses of these unigenes evaluating the presence of ORFs, protein domain, family signature conservation, three-dimensional structure and a possible heterologous expression.

Key words: RNA-seq; Biotechnology; Stress response;

BACE1 inhibitory activity and molecular docking analysis of 1-(7-chloroquinolin-4-yl)-N-(4-methoxybenzyl)-5-methyl-1H-1,2,3-triazole-4-carboxamide

MARTINS FERREIRA, MARIA CLARA[1]; GALLIO FRONZA, MARIANA[1]; GOLDANI, BRUNA[2]; ALVES, DIEGO[2]; SAVEGNAGO, LUCIELLI[1]

[1]Grupo de Pesquisa em Neurobiotecnologia (GPN) – Universidade Federal de Pelotas; [2]Laboratório de Síntese Orgânica Limpa (LASOL) – Universidade Federal de Pelotas

Alzheimer's disease (AD), the most common neurodegenerative pathology, is a complex disease characterized by an accumulation of beta-amyloid (Aβ), neurofibrillary tangle formation and progressive memory loss. It has been estimated that AD affected more than 37 million individuals worldwide. In the amyloidogenic pathway, Aβ is generated by the consecutive proteolytic cleavage of amyloid precursor protein (APP) by two proteases, β and γ-secretase. This peptide is neurotoxic in several modes, it induces oxidative stress, lipid peroxidation, inflammatory responses and mitochondrial dysfunction, causing neuronal death. However, a cure for this disease is currently unavailable although extensive research has been focused on the development of therapeutic approaches. In this way, investigations on new drug discovery and development for AD are extremely necessary and bioinformatics tools appear as a novel approach to screening for new potential drugs. In this scenario, β-secretase (BACE1) has emerged as a promising target for the prevention and/or treatment of AD. Thus, the aim of our study was providing a rational design, molecular docking and pharmacokinetic prediction of 1-(7-chloroquinolin-4-yl)-N-(4-methoxybenzyl)-5-methyl-1H-1,2,3-triazole-4-carboxamide (QTC-4-MeOBnA) as a potential anti-AD drug. To explore the inhibitory mechanism, we choose BACE-1 as a target to perform the virtual screening using the ZINC database (http://zinc.docking.org/), where approximately 35 million of different molecular structures are deposited. We set chemical parameters based on drug-like properties, such as low molecular weight (<500kDa), number of hydrogen donors (<5) and acceptors (<10). These molecular parameters are extremely important, since are able to predict the compound druggability. After this initial virtual screening, 79 promising molecules were drawn using ChemDraw and their geometry optimized using the Avogadro 0.9.4 software. The molecular docking simulation was used as a second screening for these molecules utilizing Autodock Vina software. In the study, we used different crystallographic structures of BACE (PDB: 1SGZ and PDB: 2ZHU) from Protein Data Bank (PDB) (http://www.pdb.org/). After obtaining the ligands and enzymes, their structures were prepared using the Auto Dock Tools 1.5.4 program, in which all the rotatable bonds of ligands were allowed to rotate freely, and the receptors were considered rigid. For docking studies, we used the Auto Dock Vina (version 1.1.1), utilizing a grid box centered on the active site of the enzyme with high resolution, allowing the program to search for additional places of probable interactions. The compound with the best score was QTC-4-MeOBnA possessing a binding free energy of -8.6 kcal/mol. The compound interacts with catalytic aspartate dyad that is, Asp32 and Asp228. Further, QTC-4-MeOBnA interact with residues belonging to large hydrophobic pockets (S1, S3, S2), which are key regions to BACE inhibition. Our results showed interactions with residues that comprises the hydrophobic cleft of S1 and S3: Trp115, Leu30, Ile 110 and Phe 108. With the main chain of S3: Gln12, Gly11 Gly230 and Thr232 as well with residues of S2 pocket: Val69 and Ile126. Besides that, we also evaluated pharmacokinects properties, such as absorption and toxicity using QuikProp 4.4 and admetSAR online tool (http://lmmd.ecust.edu.cn:8000). ADMET properties reveal that QTC-4-MeOBnA cross the blood brain barrier and had 75% of predictive activity in central nervous system (CNS). Our results show that QTC-4-MeOBnA is probably completely absorbed in human intestine and have low oral toxicity, probably with a DL-50 of 500-5000 mg/kg. In conclusion, the compound demonstrated high affinity with BACE-1, probably being capable of modulate the amyloidogenic pathway, as low predictive toxicity. These characteristics indicates QTC-4-MeOBnA as a potential therapeutic drug for Alzheimer's diseases, although further studies are necessary to elucidate the complete mechanism of action.

Keywords: Alzheimer's disease, molecular docking, BACE1, pharmacokinetic

MINISATELLITES IN *HERVES-LIKE* ELEMENTS AS PUTATIVE GENERATORS OF MIRNAS AND THE DISTRIBUTION OF THESE ELEMENTS IN DROSOPHILA

MOMBACH, Daniela M.[1]; BERNARDO, Larissa P.[2]; LORETO, Élgion L.S.[3]

[1] Graduanda de Ciências Biológicas em Universidade Federal de Santa Maria - UFSM; [2] Doutoranda em Biodiversidade Animal em UFSM; [3] Departamento de Biologia Molecular e Bioquímica UFSM

Transposable elements (TEs) are nucleotide sequences found in most genomes. These elements are highly variable in copy number, molecular structure and transposition strategies. They can move within and between genomes, are also highly diverse and can promote genetic variability, acting as modulators of structure, function and evolution of host genomes. Their move between genomes its termed "horizontal transfer", which can even happen across organisms from vastly distant taxa. TEs are sorted in two classes by their transposition intermediate: RNA (class I or retrotransposons) or DNA (class II or DNA transposons). Both classes present autonomous elements able to produce their own enzyme for transposition, and non-autonomous elements, the degenerate copies of TEs that use the autonomous elements enzyme for transposition. A special group of TE is classified as miniature inverted repeat transposable elements (MITEs) which are considered as truncated derivatives of autonomous DNA transposons that are non-autonomous elements.

The *hAT* superfamily of class II TEs are a widely distributed group with representatives in animals, plants and fungi. In this superfamily there is a group of elements termed *herves-like*, composed of *herves* (*Anopheles gambiae*), *hAT-2_AG* (*Aedes aegypti*), *hAT-1_DP* (*Drosophila pseudoobscura and D. persimilis*), *Howilli1* (*D. willistoni*), *Homo2* (*D. mojavensis*) and *hAT-1_DSi* (*D. simulans* and *D. sechelia*) elements. *hAT-1_DP* and *Howilli1* contain minisatellites (MnRs) in their 5'UTR region, short repetitions of 18 nucleotides that can be highly polymorphic in number of repeats. Recently, it was found that minisatellites of *Howilli1* are transcripts and produce micro RNAs (miRNAs) in *D. willistoni*. These miRNAs have important genes of the host genome as targets as it was verified in *D. melanogaster*. Therefore, we analyzed the distribution of *herves-like* elements that contains MnRs in Drosophila species with sequenced genome. These MnRs could be using these elements in order to expand their own copy numbers in genomes, since their miRNAs might be important in the genetic function of host genomes.

The search tool used was Blastn (NCBI) with default parameters. The sequences used as query were *hAT-1_DP*; *hAT-1_Dsi*; *Howilli1* and *Homo2*. *hAT-1_DSi* and *Homo2* were also used as query because we found MnRs in their structure. Sequences of putative active elements and MITEs were analyzed and annotated with UGENE software and target site duplications (TSDs) with WebLogo construction. We found the following sequences similar to *hAT-1_DP* element: 23 MITEs, 8 putative non-autonomous elements (PNA) and 1 putative autonomous element (PA) in *D. miranda*, 13 MITEs and 3 PNAs in *D. takahashii*, 2 MITEs in *D. rhopaloa* and 1 MITE in *D. arizonae*. Similar to *hAT-1_DSi* element we found 31 PNAs in *D. melanogaster*. Regarding the MnRs, they were highly polymorphic, specially in MITEs, since there were more copies in the genomes. We found in *D. miranda* from 3 to 9 MnRs in MITEs, 5 in PA and from 1 to 2 in PNAs, in *D. takahashii* from 3 to 9 MnRs in MITEs and from 5 to 7 in PNAs, in *D. arizonae* 9 MnRs in MITEs, in *D. rhopaloa* 4 MnRs in MITEs and in *D. melanogaster* 5 to 6 MnRs in PNAs. The TSDs had conserved domains in the second and seventh nucleotide: 5'-nTnnnnAn-3'.

The element *hAT-1_DSi* is restricted to *D. melanogaster* species, however recently studies discovered horizontal transfer to *Zaprianus* genus. Our results suggest that *hAT-1_DP*, found in *D. miranda*, *D. persimilis* and *D. pseudoobscura*, is also being horizontaly transferred to *D. takahashii*, *D. rhopaloa* and *D. arizonae*. Analyzing *hAT-1_DP* between the restricted species and the ones where we found the element, the similarity and the high time of divergence of the host species led us to suggest the occurrence of horizontal transfer of this element. Transposable elements are recognized as an important source of miRNAs. MITEs are important in this process due to their structural characteristics which, due the proximity of terminal inverted repeats (TIRs), allow the generated transcripts to produce the hairpins necessary for miRNA processing. The following steps in this research are to analyze the MITES flanking regions in Drosophila without sequenced genome. It is possible that they are inserting themselves in important genes, or near to them, since they might be relevant in the genetic function of these genomes.

GENETIC DIVERSITY AND POPULATION STRUCTURE OF BRAZILIAN POPULATIONS OF *DROSOPHILA INCOMPTA* (DIPTERA, DROSOPHILIDAE)

MOREIRA, HENRIQUE DA ROCHA[1]; CENZI DE RÉ, FRANCINE[2]; LORETO, ELGION LUCIO DA SILVA[2]; ROBE, LIZANDRA JAQUELINE[1,2]

[1]Laboratório de Genética, Universidade Federal do Rio Grande; [2]Programa de Pós-Graduação em Biodiversidade Animal, Universidade Federal de Santa Maria

Phylogeography focus on the contemporary distribution of gene lineages with an emphasis on its historical perspective, based on DNA sequences. Its main purpose is to explain the present biodiversity distribution patterns, being used in several studies related to conservation biology and speciation. To accomplish the microevolutionary inferences, several bioinformatic analyses are used, and the most prominent are based on Bayesian inferences and, mainly, coalescent theories. Drosophilidae is considered one of the most diverse and widely distributed families of Diptera and due to its wide niche range, this family has been an important model in phylogeographical studies. *Drosophila incompta* belongs to the *flavopilosa* group of *Drosophila* and like the other species of this group uses *Cestrum* flowers as unique substrate for both feeding and breeding. Glacial and interglacial Quaternary periods changed the climate and the vegetation of the Neotropical region, and may have impacted *D. incompta* in direct and indirect ways. After all, in both evolutionary and ecological perspectives, the distribution of the flies is associated with the distribution of their host, which typically show discontinuous flowering patterns throughout the year. This raises a question about drosophilids' survival on periods of resource scarcity. Here we evaluate the genetic diversity and structure of Brazilian populations of *D. incompta*, trying to understand the influence of historical and ecological factor on the evolutionary history of this species. For this, the individuals were collected within the species range, in the states of Rio Grande do Sul and Paraná. Total DNA was extracted of each individual using the NucleoSpin Tissue XS kit. Fragments of DNA were amplified using the primers TYJ1460 and C1N2329 for the mitochondrial gene *Cytochrome oxidase subunit I* (*COI*), and HB106F and HB903R for the nuclear gene *Hunchback* (*HB*). Sequencing was performed on a MegaBACE 500 automatic sequencer using the DYEnamic ET kit (Armesham), according to the protocol provided by the manufacturer. Bioinformatics analysis based on *COI* and *HB* sequences were executed, including Bayesian phylogenetic analysis performed on MrBayes 3.2, Median-joining network carried out Network software, FST index and measures of genetic diversity were performed on Arlequin v.3.5 and DNAsp 5, respectively. The total number of sequences obtained for COI and HB were 75 and 39, respectively. Despite their independence, both sets of markers recovered a similar signal for *D. incompta* evolutionary history, indicating that this species underwent a population expansion event. The network recovered by the COI marker revealed a starlike pattern, with a high number of exclusive haplotypes (15) and 3 main shared haplotypes, suggesting no geographical structure. Thus, the phylogenetic trees also suggest high levels of gene flow, even though the FST values presented signals of a medium to high differentiation when populations of Itaara/RS and Frederico Westphalen/RS are concerned. In regard to the *HB* marker, the network also revealed a starlike pattern, a high number of exclusive haplotypes (28) and only one main shared haplotype, reinforcing the shallow geographical structure recovered by the *COI* marker. In this case, nevertheless, the FST values indicate significantly low genetic differences when comparing populations of Curitiba/PR with Frederico Westphalen/RS and Montenegro/RS. In relation to diversity indices, the two markers presented high haplotype diversity (0.821 and 0.938 for *COI* and *HB*, respectively). Additionally, the Tajima's D, Fu and Li's D and F neutrality tests for both genes had negative results, congruent with the other previously mentioned results, indicating that the populations of *D. incompta* present signs of population expansion. The high levels of genetic diversity, but not defined geographical structure is compatible with a strategy of recurrent but scattered gene flow along alternate host flourishing seasons as a mean of overcoming periods of scarcity of resources. However, additional analyses are needed to access the paleogeographic period, in which this population expansion event occurred.

Keywords: *Drosophila flavopilosa* group; phylogeography; population expansion; restricted ecology

On the mobilome of *Drosophila incompta* (Diptera: Drosophilidae), a specialist model for evolutionary studies

Moura, Rafael D.[1]; Fonseca, Pedro M.[2]; Loreto, Elgion L.S.[3]

[1] Undergraduate Student at Ciências Biológicas - Bacharelado – Universidade Federal de Santa Maria (UFSM)
[2] PhD Student at Genética e Biologia Molecular – Universidade Federal do Rio Grande do Sul (UFRGS)
[3] Professor at Departamento de Bioquímica e Biologia Molecular – Universidade Federal de Santa Maria (UFSM)

Flies from several species in the genus *Drosophila* (Diptera: Drosophilidae) have been extensively used for over a century as model organisms in the field of genetics and, more recently, genomics and molecular biology in general. This is due to their genetic similarities with humans and the ease to cultivate and maintain them. A particular species, *Drosophila incompta*, offers an opportunity to investigate the genetic mechanisms that allow a population to specialize and adapt to a strict ecological niche, since it, along with other members of the flavopilosa group, feed, develop and reproduce exclusively on *Cestrum* (Solanaceae) flowers.

Transposable elements (TEs) are traditionally defined as DNA sequences capable of changing their position within a host genome or moving between genomes, a phenomenon called Horizontal Transfer, which can even happen across organisms from vastly distant taxa. TEs can either move using an RNA intermediate to replicate (and are called, in that case, retrotransposons or Class I TEs) or move directly to a different site, with no RNA intermediate (called transposons or Class II TEs). The entire mobile portion of an organism's genome is referred to as its mobilome, which accounts for a great part of the total genome in several species, *e.g. Drosophila melanogaster* (12-22%), *Homo sapiens* (about 50%) and maize (more than 80%). Though often deleterious for the host, TEs may greatly increase the plasticity of genomes, ultimately leading to genetic variation and biological innovation, and may, therefore, have strong correlations with the specialization of populations to their niches. We aim to investigate the existence of such correlations regarding *D. incompta* by assembling its mobilome based on next generation sequencing data and standard computational tools for TE detection.

Paired-end Illumina sequences of *D. incompta* genome were uploaded to the Galaxy platform, as two separate files in FastQ format. The sequences had their quality evaluated, were filtered based on their quality scores and both files were joined, using the FastQ Groomer, Filter by quality and FastQ Interlacer tools respectively, all of which are available within Galaxy and were run on default parameters. As a result of filtering, 16.08% of reads were discarded in total. The resulting file was then converted into the Fasta format and, using RepeatExplorer's Clustering tool, similar sequences were assembled into clusters, which were compared to repetitive sequences from multiple online databases. Reclustering followed, using Cluster Merge (RepeatExplorer) to group sequences that matched the same TEs. Once the new clusters had been assembled, we were able to evaluate which TEs are likely to be present, as well as their percentage in the genome.

The resulting data have shown that around 24% of the *D. incompta* genome is composed of repetitive sequences. The most prevalent were *Helitron* transposons, accounting for around 20.40% of that amount, followed by LINE (~16.24%) and *Gypsy* (~12.05%) retrotransposons. 44.48% of the total repetitive sequences were simple repeats, low complexity reads, and most were classified as "unknown" repeats, which means that many non-described TEs could be present in the *D. incompta* mobilome. The percentage of *Helitron* and other transposons among repetitive sequences is similar to that of *Drosophila virilis*, the most closely related species to *D. incompta* that has had its whole genome sequenced. The next analyses to be performed in this work are to assemble the clusters into contiguous sequences (contigs) using the CAP3 program, search for potential coding regions with ORF Finder and search for characteristic domains. We have done that to some extent, having found a potential active *Helitron* that matches a characterized one from *D. virilis* (83.5% similarity). We shall review that preliminary result and continue by finishing the assembly and manually annotating significant TEs found. Following that, a de novo approach will be used to automatically identify, analyze and characterize TEs within the genome, which will be done with the RepeatModeler, TEDNA and REPET packages.

AN EVALUATION OF THE GENOMIC RELATIONSHIP BETWEEN *ECHINOCOCCUS* AND THEIR HOSTS AS AN OPPORTUNITY TO HORIZONTAL TRANSPOSON TRANSFER

OLIVEIRA, PâMELA S.[1,2]; JARDIM, SINARA S.[2]; D'ÁVILA, MARíCIA F.[2]; GRAICHEN, DANIEL A. S.[2]

[1] Curso de Ciências Biológicas, Universidade Federal de Santa Maria, Campus Palmeira das Missões – RS – Brazil,
[2] Laboratório de Genética Evolutiva, Universidade Federal de Santa Maria, Campus Palmeira das Missões – RS - Brazil

Echinococcus worms are parasites belonging to the class cestoda of phylum Plathyhelmintes These parasites are distributed worldwide and are responsible for many hydatid diseases among humans and domestic and wild mammals. Due to the lack of comprehensive studies concerning host-parasite relationship, the genomic approach through data sequencing may be useful to clarify these issues. The close relationship between host-parasite is usually pointed in scientific literature as being able to increase the amount the horizontal gene transfer (HT) events. HT is defined as the transference of genetic material between genomes by means other than parent-to-offspring inheritance. Those events occur with transposable elements (TE), which are sequences of DNA, due to their ability to mobilize within and between different genomics. The Retrotransposon BovB belongs to the long LINE family of TEs and is the dominant retrotransposon type in both ancestor and extant mammal species. The BovB was firstly described in *Bovidae* (bovids) and has been found in several mammals and reptiles' genomes and, in divergent species, it has a mosaic distribution, thus suggesting the existence of HT events. The goal of this work is to evaluate the presence of host transposable elements within parasite genomes. We utilized the sequencing projects of many Platyhelminthes species (34 genomes) to look for host sequences inside parasite genomes. In order to get this, we chose the BovB element from *Bos taurus* (bovids) due to its promiscuous features and then we performed BLASTn searches over all available Platyhelminthes genomes. The BLASTn searches were performed using BovB consensus sequence from Repbase with default parameters. To rule out host DNA contamination inside the parasite genomic sequences, we performed BLASTn using B. taurus COI sequence as bait. The search of BovB sequences among Platyhelminthes showed positive results only in *Echinococcus canadensis*. The element BovB founded in *E. canadensis* had 1698 pb of length and 97% of identity with the BovB consensus. Furthermore, the host DNA contamination was not found in the *E. canadensis*, and the result confirms the genome integrity. Also others species of *Echinoccocus* genus were searched by BLASTn and that element was not present. This is the first case of BovB element in this genus, thus inducing us into arguing about a possible case of HT in *E. canadensis*, since the element was not distributed in all genus as in vertical transfer and due the high similarity of the sequence of BovB between *E. canadensis* and *B. Taurus.*

Bayesian network applied to laboratorial diagnosis of meningitis caused by *Cryptococcus neoformans*

PIGOZZO, DANIELA FERNANDA[1]; PUCCI, CAROLINE DO AMARAL FETZNER[2]; STAHNKE, FERNANDO RAFAEL[3]

[1]Affiliation (CA- Applied Computation, Line of Research: Applied Informatics to Health, Universidade Feevale, Novo Hamburgo-RS); [2]Affiliation (CA- Applied Computation, Line of Research: Applied Informatics to Health, Universidade Feevale, Novo Hamburgo-RS); [3]Affiliation (CA- Applied Computation, Line of Research: Applied Informatics to Health, Universidade Feevale, Novo Hamburgo-RS)

Introduction: Bayesian networks (BN) can constitute a virtual learning environment, where they assist in teaching both in face and distance modalities. They are formed by acyclic graphs, such as symptoms, examinations, treatment, which represent the variables of the systems, that is, the qualitative part and by their probabilities, which constitute the quantitative part within the network. They enable the representation of uncertain knowledge, and must be developed with one or more specialists in the area in which it is desired to apply them. Objective: Construction of a Bayesian network aiming at the clinical and laboratory diagnosis of meningitis caused by *Cryptococcus neoformans*, and formalization of probabilities related to each variable. This network will be part of a clinical case simulator, called the Health Simulator, which is being developed and can be applied to the teaching of the Biomedicine course. Methodology: The structural model of the network was made by students of the Biomedicine course using the Bayes Editor and Hugin Graphical User Interface applications, which allow the construction of the Bayesian network. The variables used were symptoms such as sore throat, prolonged fever, excessive sweating, weight loss, acute eye pain, AIDS patient and exams such as cerebrospinal fluid analysis using Chinese ink, fungal blood culture, latex test, microscopic analysis Of the presence of encapsulated yeasts, cultivation of the fungus in Sabourad Agar, presence of mucoid colonies or smooth cream-colored colonies, producing urea and non-fermenting. Subsequently, the determination of the initial probabilities of each variable was done, extracting the percentages previously established in articles and comparing the statistical proximity with other works and also based on the expertise of the domain specialist. Some of the variables did not present close percentages, and for them the probability was determined by the arithmetic mean and by the expert. The article search was done in electronic databases, such as Scielo, PubMed, Lilacs and Google Scholar, using articles from the last ten years and the following keywords: Cryptococcus neoformans Meningitis; Symptoms and Prevalence; Cryptococcus neoformans Diagnosis; Cryptococcal meningitis Epidemiology. Discussion and Conclusion: Today, the large amount of information generated constantly, associated with long periods of teaching in the classroom overwhelms the students, making them unable to reconcile all the knowledge acquired with the practice, proving to be a big problem. In view of this problem, the Bayesian Network is an interesting tool to be inserted in the educational environment, since it formalizes theoretical, scientific and uncertain knowledge in probabilities, in this case, about meningitis caused by *Cryptococcus neoformans*. The *Cryptococcus neoformans* is a spherical yeast surrounded by a capsule, responsible for numerous cases of meningitis, especially in patients with HIV. Therefore, the RB methodology allows students and professionals to develop autonomy, confidence, clinical reasoning and diagnosis for decision making, without generating real risks for the patient. Another advantage of using this methodology is the possibility for the teacher to follow the evolution of the student and to evaluate the educational needs to be reviewed.

Keywords: Cryptococcus neoformans Meningitis; Symptoms and Prevalence; Cryptococcus neoformans Diagnosis; Cryptococcal meningitis Epidemiology.

Bayesian network applied to the clinical condition of tuberculosis and its diagnosis

PUCCI, CAROLINE DO AMARAL FETZNER[1]; PIGOZZO, DANIELA FERNANDA[2]; STAHNKE, FERNANDO RAFAEL[3]

[1]Affiliation (CA- Applied Computation, Line of Research: Applied Informatics to Health, Universidade Feevale, Novo Hamburgo-RS); [2]Affiliation (CA- Applied Computation, Line of Research: Applied Informatics to Health, Universidade Feevale, Novo Hamburgo-RS); [3]Affiliation (CA- Applied Computation, Line of Research: Applied Informatics to Health, Universidade Feevale, Novo Hamburgo-RS)

Introduction: Tuberculosis is a pneumopathy that generates millions of deaths throughout the year around the world. It is caused by Mycobacterium tuberculosis, also called Koch bacillus, is the second infectious agent that kills the most in the world. Therefore, professionals trained to identify you are strictly necessary to reduce cases and treat patients earlier. In view of this, Bayesian Networks can be an excellent support tool for training and academic construction of health students. They are techniques of artificial intelligence based on acyclic graphs, variables and probabilities. They are made up of a qualitative part that is represented by variables, where symptoms and diagnostic techniques can be attached, and by a quantitative part where the probabilities of each variable are inserted. They can be inserted into the educational environment as a complementary teaching methodology, because they represent the probabilistic knowledge of a certain subject, in this case of tuberculosis, stimulating students to think in a more rational and fast way. Objective: To formalize the knowledge about the clinical picture of Tuberculosis through the use of Bayesian networks, which can be used throughout the Biomedicine course and be part of a clinical case simulator, called Health Simulator. This simulator uses a gamut environment and is based on the uncertain knowledge represented by RB. Methodology: The structural model of the network was made by students of the Biomedicine course through the Bayes Editor and Hugin Graphical User Interface applications, which allow the construction of Bayesian networks. The students' knowledge about the pathology and information contained in books and academic articles were used to establish the important variables that lead to the diagnosis of the problem. Excessive cough, bloody sputum, weight loss, night fever, heavy afternoon sweating, chest pain and exams such as chest radiography, gram negative examination, culture, staining with Ziehl Neelsen, Presence of alcohol-acid bacillus constituted the other part of the variables. After the structural elaboration of the Bayesian network, the probabilities of the symptoms and the tests were determined. The initial probabilities were collected in articles that reported the percentages of symptoms related to their prevalence and the tests related to their sensitivity. Not all variables presented similar prevalences in the literature, so the arithmetic mean was used to establish the initial statistical probability of these variables, together with the experiences of specialists in the area. The search mechanisms for the academic articles were electronic databases such as Scielo, Lilacs, PubMed and Google Scholar, using works referring to the last decade, with the following keywords: Tuberculosis, Diagnosis, Sensitivity of Exams, Prevalence of Symptoms. Discussion and Conclusion: Thus, as shown, RB can represent theoretical, scientific and uncertain knowledge in probabilities, allowing students and professionals to develop confidence, clinical reasoning and diagnosis for decision making, without generating real risks for the patient.

Keywords: Tuberculosis; Diagnosis; Sensitivity of Exams; Prevalence of Symptoms.

IN SILICO SCREENING OF SELECTIVE INHIBITORS OF 11-BETA HYDROXYSTEROIDENESIDROGENASE TYPE 1 (11β-HSD1) WITH PERSPECTIVE OF PREVENTION OF SKIN DAMAGE

RICHARDT, JOVELINE; ALBRECHT, DANIELLE; AMORIM, HERMES
Laboratório de Bioinformática Estrutural, ULBRA, Canoas/RS

The skin has several fundamental functions in our body. Thus, their integrity is essential to maintaining health and well-being. In the skin, the excess of cortisol may limit tissue repair and negatively regulating wound healing and skin integrity, as well as its production is associated with the skin aging process. Cortisol is systematically released in response to various stressors. In tissues, its concentration is modulated by the isoenzymes of the 11-beta hydroxysteroidenehydrogenase (11β-HSD) family, responsible for activation and inactivation of glucorticoids (CGs). 11β-Hydroxysteroid dehydrogenase type 1 (11β-HSD1) is a bidirectional GC-activating enzyme that works primarily as a reductase, driven by the supply of its cosubstrate NADPH by hexose-6-phosphate dehydrogenase (H6PDH). The fact that 11β-HSD1 has a key role in the production of cortisol has encouraged the planning and development of selective inhibitors for topical use, in order to combat undesirable effects of the high cortisol concentration. This work describes an *in silico* approach to prospect and select potentially inhibitors of 11β-HSD1 from two compound libraries: a) compounds from Traditional Chinese Medicine (TCM) and b) compouds approved in the world for clinical use. TCM ligands were filtered by ADMET rules to select compounds with suitable pharmacological characteristics. The affinity of the selected ligands for two 11β-HSD enzymes (11β-HSD1 and 11β-HSD2) was then evaluated by virtual screening (VS). Subsequently, the complexes with the best ligands were submitted to molecular dynamics (MD) simulations aiming to obtain better quality information on affinity and selectivity. Since 11β-HSD2 isoform does not have 3D structure available it was necessary to model its structure by homology modeling. Eleven compounds were selected from virtual screening for further analysis by the MM-PBSA (Molecular Mechanics – Poisson Boltzmann Surface Area) method combined with MD simulations. Eplerenone (ZINC3985982) and Hamanasic Acid (ZINC13377292) were selected as potential selective inhibitors to be considered in a subsequent development project.

Keywords: 11β-HSD1. Cortisol. Selective inhibitors. Virtual Screening. MD simulation. MM-PBSA.

WRKY FACTORS IN PLANT RESPONSES TO ABIOTIC STRESS: A META-ANALYSIS PREVIEW

SANTOS, Railson Schreinert dos[1]; VIANA, Vivian Ebeling[1]; ARAUJO JR, Artur Teixeira de[2]; COSTA DE OLIVEIRA, Antonio[1]

[1] Plant Genomics and Breeding Center (CGF), Universidade Federal de Pelotas; Pelotas, RS, Brazil.; [2] Centro de Biotecnologia (CBiot), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

WRKY genes encode DNA binding proteins that recognize the conserved nucleotide sequence (T)(T)TGAC(C/T) (W box) found in promoters of plant defense genes. The importance and the large number of members contained in this family, calls the attention of the scientific community which aims to identify the role of some of these genes in stress tolerance. With many members present in most plant genomes, an analysis of the sequences and of the transcriptional behaviors of these genes in different species, allowing the better understanding of their role in plant evolution, is of great importance. Considering such a need for information, the first plant model for studies in molecular genetics to be customarily thought of is *Arabidopsis thaliana* L., a plant of the Brassicaceae family, with five chromosomes and a relatively small genome of 125 million base pairs sequenced and published in 2000 [1]. Given the large availability of transcriptional data which has recently been made available, a broad analysis covering several plant species and environmental conditions is about to be made in our lab. Here we report the first part of this meta-analysis, where we use AT-00120 access, from the ATGenExpress database, through Genevestigator [2]. Young *A. thaliana* rosetes have been examined for conditions of cold, drought, salt and osmotic stresses, where we searched specifically 59 of the 60 genes analyzed by Wang et al. [3]. The conditions of greater impacts on the responsiveness of these genes were osmotic (13), cold (12), salt (10) and drought (9), as shown in Figure 1A. The Log2-ratio values of differentially expressed WRKY ($p \leq 0.05$) are shown in Figure 1A. The number of differentially expressed genes shared between conditions can be seen in Figure 1C. WRKY46 was the gene that was differentially expressed in the greatest number of different conditions, being altered in plants under cold, drought and osmotic stresses. These results represent a first step aiming to identify factors commonly altered in certain conditions throughout the evolution of plants, with special attention to model species. This information should allow us to better understand essential factors involved in the response to adverse environmental conditions, helping plant breeding to obtaining more productive and resilient crops.
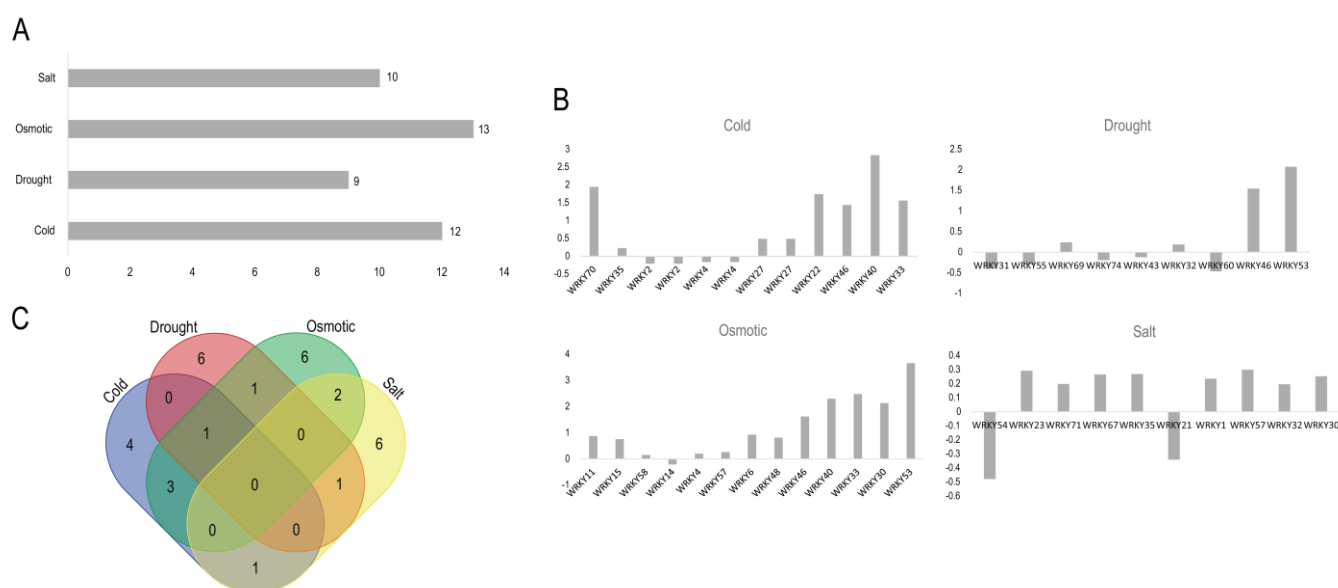

Figure 1. **Expression of WRKYs from *Arabidopsis thaliana* L. under conditions of cold, drought, osmotic and salt stress ($p \leq 0.05$). A.** Total number of WRKYs differentially expressed in each condition; **B.** Log2-ratio values of WRKYs transcriptional expression in each stress; **C.** Venn diagram showing how many genes are differentially expressed in each single or combined situations.

[1] The Arabidopsis Genome Initiative. Nature, 796-815, 2000; [2] Zimmermann, Philip et al. Mol Plant., 851-857, 2008; [3] Wang, Qishan et al. Genetica, 973–983, 2011.

BIOINFORMATICS ANALYSIS IN GENE EXPRESSION DATA FROM PUBLICLY AVAILABLE DATABASES

SARTOR, IVAINE TAÍS SAUTHIER[1,2]; RECAMONDE-MENDOZA, MARIANA[3]; ASHTON-PROLLA, PATRICIA[1-4]

[1] Laboratório de Medicina Genômica, Serviço de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre, LMG-SPE-HCPA; [2] Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, PPGBM-UFRGS; [3] Departamento de Informática Aplicada, Instituto de Informática, Universidade Federal do Rio Grande do Sul, INF-UFRGS; [4] Departamento de Genética, Universidade Federal do Rio Grande do Sul, DEGEN-UFRGS.

The gastrointestinal (GI) cancers account for 20% of estimated new cancer cases and 15% of estimated death worldwide. Among GI neoplasias, we especially highlight: i) the esophageal cancer (EC), which comprises two subtypes: the squamous cell carcinoma (SCC) and adenocarcinoma (ADC), the last one can progress from a pre-malignant lesion, known as Barrett's esophagus; ii) the gastric cancer (GC), which can progress from pre-malignant lesions such as chronic gastritis (ChG) and intestinal metaplasia (IM) and iii) the colorectal cancer (CRC), which can be stablished from colorectal adenoma lesion. GI malignancies are aggressive and heterogeneous diseases with poor survival. Further knowledge about the molecular pathogenesis and biological features of GI cancers is necessary to enable the identification and characterization of novel molecular biomarkers and therapeutic targets. In a previous study, which used a computational approach, was identified the transcription factor *TULP3* as a master regulator of carcinogenesis in pancreatic ductal adenocarcinoma (PDAC). The authors observed a poor prognosis in patients with higher *TULP3* expression in PDAC. Considering that pancreas and other gastrointestinal organs (such as esophagus, stomach and intestine) have the same embryonic origin, we investigated the profile of *TULP3* gene expression in GI tissues hypothesizing that it may have a role in these diseases. Therefore, we performed bioinformatics analysis to compare *TULP3* expression in GI tissues and to analyze patient survival. Gene expression data from patient biopsies were obtained from GEOdatabase and TCGA public repositories. GEOdatasets were downloaded under accession numbers: GSE26886 (GPL570) and GSE1420 (GPL96) for EC; GSE79973 (GPL570), GSE33335 (GPL5175) and GSE2669 (GPL2048) for GC; and GSE21510 (GPL570) and GSE24514 (GPL96) for CRC. From TCGA we obtained the RNASeq data of the following studies: ESCA (Esophageal Carcinoma), STAD (Stomach Adenocarcinoma) and READ (Rectum Adenocarcinoma). Preprocessed microarray data from COAD-TCGA (Colon Adenocarcinoma) study was obtained as provided by the authors. Principal component analysis was performed in each study to filter possibly biased samples. Gene expression raw data were normalized using *affy BioConductor* R-package for GPL570 microarrays and *oligo BioConductor* R-package for GPL5175 and GPL2048 microarrays. Raw counts from RNASeq data were normalized using *limma BioConductor* R-package. To select a single probe to represent a gene, we used the JetSet score for Affymetrix GPL570 and GPL96 microarrays. Data normality assumptions were verified and the appropriate statistical tests were chosen. Survival analysis was performed using *survival* R-package, the Kaplan-Meier method was used to estimate survival curves and LogRank test was used to compare the curves. TULP3 gene expression comparison between groups in ESCA-TCGA (p-value=3.21e-06), GSE26886 (p-value=2.03e-06) and GSE1420 (p-value=0.01) could differentiate the esophageal lesions. Despite *TULP3* showed significant statistical differences, the esophageal lesions analyzed and the expression trend observed in all datasets were not the same. Nevertheless, the survival analysis in TCGA Esophageal ADC associated a poor prognosis in patients with higher *TULP3* expression, ranging from ($\log_2 4.45$, $\log_2 6.16$], with a p-value=0.03, HR=2.11(1.05-4.21), while in TCGA Esophageal SCC, an unfavorable prognosis was associated with lower *TULP3* expression ($\log_2 3.62$, $\log_2 5.34$], p-value=0.04, HR=0.46(0.22-0.94). Considering GC, *TULP3* analysis in STAD-TCGA (p-value=0.02), GSE33335 (p-value=4.45e-07) and GSE2669 (p-value=3.44e-05) presented higher expression in gastric cancer samples in comparison with adjacent non-tumoral mucosa (non-GC) and ChG and IM. When we analyzed survival probability according the gender of patients in STAD-TCGA study we observed a worse prognosis in females with higher *TULP3* levels, ranging from ($\log_2 5.07$, $\log_2 7$], with a p-value=3.77e-3, HR=2.44(1.30-4.44). In male patients no difference was observed. In addition, increased *TULP3* expression in diffuse-type GC was also associated to an unfavorable prognosis ($\log_2 4.91$, $\log_2 6.25$], p-value=0.04, HR=2.93(1.00-8.54), but the same trend was not observed in the group of patients with intestinal-type GC. In patients diagnosed with gastric cancer not otherwise specified (NOS), higher *TULP3* expression ($\log_2 5.05$, $\log_2 7$], was also associated with worse prognosis, p-value=2.29e-04, HR=3.46 (1.71-6.98). The dichotomized TULP3 expression presented significant difference in univariate analysis in diffuse-type (p-value=0.03; HR=2.93(1.00-8.54)) and NOS (p-value=3.34e-04; HR=3.46(1.71-6.98)). Although the *TULP3* gene expression analysis in EC samples showed significant differences, the lesions analyzed and the observed trend of its expression in all datasets were not the same. In addition, the prognostic value associated to esophageal ADC and SCC, despite statistical significance, should be exploited in future works to comprehend biological process involved in EC. Considering GC, higher *TULP3* gene expression was observed in GC groups in STAD and GSE33335 studies, and a worse prognosis was associated with higher *TULP3* expression. Finally, in CRC higher *TULP3* gene expression was observed in CRC group in all studies and poor prognosis was assigned in patients with lower expression. Indeed, it is possible that *TULP3* has a role as a biomarker, and more studies are needed to confirm these *in silico* findings.

Keywords: gastrointestinal cancers, gene expression, bioinformatics

PHYLOGENETIC STUDY OF *COMPANION OF CELLULOSE SYNTHASE* (CC) GENES AND POSSIBLE RELATION WITH TOLERANCE TO ENVIRONMENTAL STRESSES

SCHETINGER, ISABELA CHITOLINA[1]; ARAUJO JR, ARTUR TEIXEIRA[1]; OLIVEIRA, BEN-HUR NEVES[1]; RICACHENEVSKY, FELIPE KLEIN[2]; SPEROTTO, RAUL ANTONIO[3]; FETT, JANETTE PALMA[1]

[1]Centro de Biotecnologia e Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS; [2]Departamento de Biologia, Universidade Federal de Santa Maria, Santa Maria, RS; [3]PPGBiotec, Centro Universitário UNIVATES, Lajeado, RS

Key words: plant cell walls, abiotic stress, cellulose synthesis

The cell wall is a key component in plant cell morphology, important for cell growth, development and stress responses. Cellulose is the most abundant component of the primary cell wall. Recent studies of our group, using RNA-seq data from germinative and vegetative stages, indicate that cold tolerance in rice (*Oryza sativa* ssp. *indica*) may be related to modifications in the expression of genes responsible for cellulose synthesis, including genes that encode the proteins *Companion of Cellulose Synthase* (*CC*). Due to the lack of functional information about this protein family, this study also aimed to perform phylogenetic analyses of CC genes, and to investigate *in silico* their expression under stress conditions. The known Arabidopsis sequences were used to search for CC homologue genes in other plant genomes using *BLAST* search in the *Phytozome* database. The protein and nucleotide sequences were used for phylogenetic analyses performed by MEGA 6, based in the *Neighbor Joining* method. The motifs found in common between rice and Arabidopsis in both the protein sequences and the promoter regions were analyzed using the database *MEME* and *PlantPan* 2.0, respectively. The expression analysis of the rice CC genes was investigated using the platform *Genevestigator*. The phylogenetic tree was reconstructed with 258 putative CC sequences, from 46 species. The tree was divided in four distinct subgroups. Within three subgroups, the sequences were separated in subdivisions of monocot and eudicot. The fourth subgroup consisted of specific sequences of green algae, bryophyte and lycophytes. Monocots harbor a larger number of CC genes in their genomes than eudicots. The CC rice paralogues were named OsCC1 to OsCC5 and may have arisen as result of two duplication events. Analysis of the exon/intron structure suggests, in Arabidopsis, large similarity between AtCC1 and AtCC2, as well as between AtCC3 and AtCC4, while in rice all sequences seem largely similar. The rice protein sequences contain a transmembrane domain and two motifs related to the *Late Embryogenesis Abundant 2* (LEA 2) domain. The sequence comparison of promoter regions showed four types of conserved sites potentially required for transcription factor binding (FAR1, AP2, MYB e WRKY) in all Arabidopsis CC sequences, while in rice there are six (NAC, C2H2, MYB, B3, bZIP e GATA). These transcription factors play important roles in processes related to development and responses to biotic and abiotic stresses. *In silico* analyses indicated that the gene expression of three rice CC genes is up-regulated in plants submitted to salt, drought and cold stresses, while the expression of two CC genes is down-regulated by drought and salt treatments. The results suggest that CC genes may have different evolutionary histories in monocot and eudicot lineages, and that these genes may be important in tolerance to abiotic stresses. Additional studies are necessary to confirm these hypotheses.

Transmission dynamics of HIV-1 Subtype C in the South America

Schwingel, Gustavo Brum[1]; de Medeiros, Rubia Marília[2]; Gräf, Tiago[2,5]; Almeida, Sabrina Esteves de Matos[2,3,4]; Junqueira, Dennis Maletich[1]

[1] Centro Universitário Ritter dos Reis - UniRitter. Rua Orfanotrófio, 555 - Alto Teresópolis, Porto Alegre, RS, Brazil; [2] Centro de Desenvolvimento Científico e Tecnológico (CDCT), Fundação Estadual de Produção e Pesquisa em Saúde (FEPPS). Avenida Ipiranga, 5400 - Jd Botânico, Porto Alegre, RS, Brazil; [3] Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul (UFRGS). Avenida Bento Gonçalves, 9800 - Agronomia, Porto Alegre, RS, Brazil; [4] Instituto de Ciências da Saúde, Universidade FEEVALE. Rodovia RS 239, 2755 - Vila Nova, Novo Hamburgo, RS, Brazil; [5] College of Health Sciences, University of KwaZulu-Natal (UKZN), Durban, South Africa.

The subtype C (HIV-1C) is the predominant genetic variant worldwide representing half of the infections in the HIV/AIDS epidemic. In Brazil, despite the massive predominance of subtype B, HIV-1C is the major causative agent of HIV infections in the southern states of the country, especially Rio Grande do Sul and Santa Catarina. In this region, around 50% of all infections are caused by subtype C. In the State of Paraná, also located in the southern region of Brazil, this subtype accounts for between 20 to 50% of the infections. Although the much lower prevalence of HIV-1C in Paraná, several studies pointed Curitiba, Paraná's capital, as the main hub of initial dissemination of the virus in Brazil. In a recent study, however, Gräf et al. (2015) using a new phylogeographic approach and by calculating several phylogeny-trait association measures consistently showed that Porto Alegre (capital city of Rio Grande do Sul state) seemed to have a central role in the initial dissemination of HIV-1C in Brazil. After its introduction in the early 1970s, HIV-1C was disseminated from Porto Alegre to closer localities and progressively to further locations. The objective of the present study was to characterize the transmission dynamics of HIV-1C in Brazil and in the South American countries where this lineage circulates.

Partial sequences of the *Pol* region from 741 South American individuals were selected from GenBank. In order to understand the influence of the South American HIV-1C epidemic in a global context, we used ViroBlast to identify the top fifty related sequences to each one of the selected South American HIV-1C isolates, excluding identical sequences. This set only included sequences isolated outside of South America. Through maximum likelihood phylogenies, transmission links were inferred using ClusterPicker. Bayesian analyses were applied to estimate the median maximum inter-transmission interval.

The 741 sequences of 901 bp analyzed in our study included samples from Brazil (718 sequences representing 28 cities in 13 States), Argentina (N=18), Venezuela (N=3), Peru (N=1) and Uruguay (N=1). We found 212 sequences in 95 transmission clusters identified in a SH-aLRT support threshold of 90 and a maximum genetic distance of 4.5%. Six clusters linking two individuals included only sequences from non-South American countries (5 including only Zambian sequences and 1 linking isolates from South Africa) and none transmission cluster were observed linking South American to non-South American sequences. Around 58% of the transmissions clusters identified here are linking sequences from the same city. Twenty percent of the clusters are linking sequences from patients within different cities in the same state and only 17% are linking individuals from different states inside South America. None of the clusters linked sequences from different countries. The Bayesian analysis suggest that the HIV-1C transmission between individuals in South America takes in average 8.3 years (95% CI 7.52- 9.09) to occur.

These results suggest that subtype C in South America seems to be predominantly transmitted locally (in the same city), not nationally or continentally, maintaining a restricted area of dissemination. A previous study on the HIV-1 subtype B circulating in South America showed that 71% of the transmissions clustered sequences from the same geographical region. Although subtype C presents a lower proportion of local transmissions in comparison to the subtype B, this difference is not statistically significant (P=0.816). However, residual analysis test points to a higher number of transmission clusters involving samples from different states for subtype C (P=0.04). The answer for this difference may rely on the time since the initial dissemination of these viruses in the continent. Subtype B was first introduced in the Americas around 1964 (1950-1967) while the subtype C epidemic was initiated in the 1970s [6,30]. As the subtype C epidemic is newer and more recently experienced an exponential spread [31], it is easier to detect higher rates of interstate links in comparison with subtype B. In addition, the lack of significance when comparing proportion of local transmissions between subtypes B and C suggests that the HIV-1 epidemic in South America is following a pattern only dependent on the host, regardless the HIV variant. In summary, our results suggest that the subtype C is disseminating in small clusters, including few patients and that the transmission between individuals might take several years to occur. These results seem to explain the apparent slow dissemination of the HIV-1C in South America since its introduction in the 70's and corroborates previous studies about the importance of the Southern region of Brazil for the epidemic.

PHILOGENY OF PLATYHELMINTHES CONSIDERING MITOGENOMIC DATA

TAMIOSSO , RAQUEL T USI [1]; ROSA , M ARCOS TRINDADE [1]; LORETO , ELGION LÚCIO DA SILVA [3]

[1] GRADUANDA EM CIÊNCIAS BIOLÓGICAS BACHARELADO (LABDROS, UNIVERSIDADE FEDERAL DE SANTA MARIA); [2] MESTRANDO EM BIODIVERSIDADE ANIMAL (LABDROS, UNIVERSIDADE FEDERAL DE SANTA MARIA); [3] DEPARTAMENTO DE BIOLOGIA MOLECULAR E BIOQUÍMICA UFSM

Platyhelminthes are a group of invertebrates dorsoventrally flattened, non-segmented, having soft-body and simple anatomy. The taxonomy of the group consists in two clades: Rhadbitophora and Catenulida. The Rhadbitophora are divided into Neodermatas (covering parasitic organisms like Acoela, Cestoda and Trematoda) and Tricladida (representing free-living organisms). Catenulida are also composed by free-living microturbaries.

Any traits that have a genetic base and have variation between the factors involved can be assessed by similarity and help us rebuild their evolutionary history. With the advancement of molecular tools, data availability and subsequent analysis of these, new proposals about the phylogeny arose. With respect to phylum Platyhelminthes, recent works suggesting that the Catenulida clade is the most basal of this group. Through the mitogenomic data analysis available in GenBank, the present work sought to verify the evolutionary relations of different clades that compose the phylum Platyhelminthes.

For the analyses, fourteen Platyhelminthes mitogenomas were downloaded by GenBank to determine the phylogenetic relationships among the main clades of the group. Of these, nine are parasitic flatworms Neodermata (belonging to three different orders Monogenea, Cestoda and Trematoda), three of Tricladida, one of Catenulida (spp. Stenostomum leucops) and one of the external group Rotifera (spp. Brachionus plicatilis) that was used as outgroup. The amino acid sequence of 12 coding genes were aligned using MUSCLE with default parameters, implemented in MEGA 6.0. After alignment of the genes, they were concatenated and gap-missing data were treated as complete deletions using Gblocks. The phylogenetic trees were constructed using the Maximum Likelihood method performed by MEGA 6.0 and the Bayesian method performed by BEAST program.

The phylogenetic tree shows a clade with nine mitogenome corresponding to parasitic flatworms Neodermata, as well a clade grouping with four mitogenome of free-living organisms (three Tricladidas and one Catenulida). The mitogenome of Catenulida represented by the species Stenostumun leucops was in the most basal position. As expected, all mitogenomes fit into phylum Platyhelminthes, sustaining the monophyly of the group.

Recent studies using molecular methods have shown that Platyhelminthes are formed by Catenulida and Rhabditophora, with Catenulida being the most basal clade. According to the results obtained in the present work, the phylogenetic comparisons made through the analyzes using the available mitogenomes of phylum corroborate with this theory. The description of new mitochondrial genomes referring to other Platyhelminthes species will certainly enable future phylogenetic analyzes, enhancing the reliability of the data and / or promoting the discovery of new relationships in their evolutionary history.

Complete genome sequences of two Torque Teno Sus Virus recovered from serum of sows with stillbirths in Southern Brazil.

Tochetto, Caroline[1]; Lima, Diane A.[1]; Varela, Ana Paula.[1]; Paim, William Pinto.[1]; Loiko, Márcia Regina.[1]; Herpich, Juliana[1]; Schmitd, Candice[1]; Cerva, Cristine[3]; Cibulski, Samuel Paulo.[2]; Santos, Anne Carolina.[1]; Mayer, Fabiana Quoos.[3]; Roehe, Paulo Michel.[1]

[1]Laboratório de Virologia, Departamento de Microbiologia, Imunologia e Parasitologia, Instituto de Ciências Básica da Saúde, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil.
[2]Faculdade de Veterinária, Laboratório de Virologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
[3]Laboratório de Biologia Molecular, Instituto de Pesquisas Veterinárias Desidério Finamor (IPVDF), Eldorado do Sul, Rio Grande do Sul, Brazil.

Torque teno viruses (TTVs) are circular, non-enveloped, single-stranded DNA viruses, classified in the Anelloviridae family. In swine, two genetically distinct species, torque teno sus virus 1a (TTSuV1a) and 1b (TTSuV1b) are currently grouped into the genus Iotatorquevirus, whereas torque teno sus virus k2a (TTSuVk2a) and torque teno sus virus k2b (TTSuVk2b) are included in the genus Kappatorquevirus. TTSuV infection in pigs is distributed worldwide, and is characterized by a persistent viremia. TTSuVs appear to spread by vertical and horizontal transmission routes and viral loads increase by age of the animals. Interestingly, the pathogenicity of TTV remains unclear.

Here, we report two complete genome sequences of TTSuV 1a identified in serum samples of sows with stillbirths from one farm in Southern Brazil. After filtered (0,22 µm) and ultracentrifuged, the samples were treated with nucleases and viral DNA was extracted using phenol protocol. The DNA obtained was prepared with Nextera and sequenced using the MiSeq reagent kit v2 300 cycle Illumina®. A total of 613,632 paired-end reads were generated. After trimmed by Prinseq, 599,286 reads were *de novo* assembled by the metaSPAdes genome assembler (v 3.9.1). The retrieved contigs were compared with sequences from BaseSpace Cloud databases using BLASTx. All assemblies were confirmed by mapping reads to contigs generated by SPAdes using Geneious version 8.1.7 software.

The two TTSuV1a genomes, strain TTSuV1a-RS12A and TTSuV1a-RS12B have 2,868 nt and 2,910 nt in length, respectively, containing four major open reading frames (ORFs): ORF 1, ORF 2, ORF 1/1 and ORF 2/2. ORF1/1 and 2/2 suffer splicing before translation. ORF 1 encodes a putative capsid protein and associated replicase. ORF 2 encodes a protein similar to a phosphatase tyrosine and the ORF 2/2 a non-structural protein with no determined function. The untranslated region (UTR) is CG-rich and included a potential stem-loop structure (stem-loop motif TAGTATTAC) and TATA box (ATATAA) domains.

In order to verify the classification of the two genomes identified in this study, a phylogenetic analyses was performed based on TTSuV genomes used on the last TTSuV classification by International Committee on Taxonomy of Viruses (ICTV). Both genomes clustered with sequences corresponding to TTSuV 1a species. After that, fifty genome sequences of TTSuV 1a from NCBI database were aligned with the two sequences identified in this study using MUSCLE version 3.5. The whole genome sequence of TTSuV1a-RS12A shares between 62.5% and 98.9% of nucleotide identity with those genomes; TTSuV1a-RS12B shares between 68.7 and 93.9% of nucleotide identity with them. TTSuV 1a-RS12A and TTSuV1a-RS12B showed a 68% sequence identity between each other. Phylogenetic analysis of whole genome sequence was performed using Maximum Likelihood method. TTSuV1a-RS12A clustered together with TTSuV 1 isolate TTV1Gx3-1 (HM633253) from China, while TTSuV1a-RS12B clustered closest to TTSuV 1 isolate TTV1Bj3, also from China (HM633245).

These findings provide useful information for further studies on TTSuVs. The presence of two complete genomes in serum samples from sows with stillbirths indicate a viremia processes. Although TTSuV has not showed any pathogenicity alone, it has an important role at co-infection with other pathogens, as porcine circovirus type 2 (PCV2). To the best of our knowledge, this is the first complete genome sequence of TTSuV 1a recovered from serum of sows in Brazil. Further investigation is necessary to better characterize these virus.

Keywords: TTV, swine, NGS, virome.

The Use of GPUs for the Acceleration of Molecular Dynamics Simulations With NAMD

TORREZIN, Aline[1]; GONZÁLEZ-DURRUTHY, Michael[2]; MONSERRAT, José Maria[2]; WERHLI, Adriano V.[1]; MACHADO, Karina S.[1]

[1] Centro de Ciências Computacionais, Universidade Federal do Rio Grande; [2] Instituto de Ciências Biológicas, Universidade Federal do Rio Grande

The Graphics Processing Unit GPU, maximizes the use of a computer together with a Central Processing Unit CPU, to accelerate computation intensive applications such as Deep Learning, Engineering applications, Computational Biology and so on. The main difference between GPU and CPU processing is that GPU uses a parallel architecture with multiples cores instead of only a few cores, like the CPU.

In Computational Biology, both GPU and CPU are used to perform molecular dynamics (MD) simulations, because many applications using these simulations demand high performance computing [1]. The MD is a computer simulation method for studying the physical movements of atoms of biomolecules as proteins [2]. In this work we propose to study and compare protein MD simulations' time, using the software NAMD [3] in CPU and GPU to verify which architecture is more appropriate to perform this type of experiment. The NAMD is a parallel molecular dynamics software designed for high-performance simulation of large biomolecular systems. So, in order to analyze more complex proteins, it is necessary to improve the efficiency of NAMD, because of the high computational requirements [4].

To achieve this we performed a case study in one workstation with GPU. In this case study we executed a MD simulation of the protein Cytochrome C Oxidase (EC 1.9.3.1). This protein has as a function to accept electrons from Cytochrome C's molecules and to transfer them to the molecular oxygen, where it's immediately reduced and converted into water. Its pathology it's usually associated to severe consequences to the human beings due the failure in the main source of energy of our cells. Leigh disease, Ischemia, Sensorineural hearing loss are a few examples of diseases that tend to occur because of these failure, usually appearing in the first stage of life and compromising the brain or the heart.

Two MD simulations were run with the software NAMD. The first simulation used the CPU, and the second one was performed in the GPU. In our first experiment, it was used only the CPU of our workstation, an Intel Xeon Series 5600 (2.8GHz, 6 cores and 12 threads). In the second experiment, the Tesla K40C GPU was employed. To compare the results, the output files with the simulation information were manipulated to obtain the time that was spent in these simulations. Based on the simulation times, it was noticed the CPU simulation was slower than GPU-simulation. The Intel Xeon Series 5600 took about 32 seconds in 500 time steps, while the Tesla K40C took about 4 seconds. In total, the protein simulation took 5 minutes by CPU and only 48 seconds by GPU. With this experiment, we can conclude the utilization of GPUs like Tesla K40C in MD simulation can greatly benefit this type of research project. We observe that the simulation time decreases approximately 7 times and, therefore, we can analyze more proteins in a same workstation in a shorter period, or run longer MD simulations.
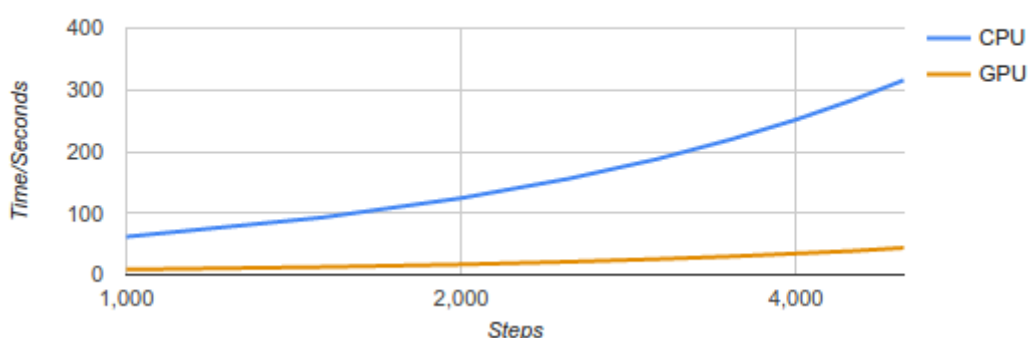

Fig 1. Simulation times in seconds of GPU and CPU.

[1] Zheyong Fan et al. Efficient molecular dynamics simulations with many-body potentials on graphics processing units. Computer Physics Communications 218, 10–16, 2017; [2] Karplus, M. and McCammon, A. (2002). Molecular Dynamics Simulation of Biomolecules. Nat. Struct. Biol., 9:646-652; [3] NAMD. **Scalable Molecular Dynamics.** Available on: <http://www.ks.uiuc.edu/Research/namd/>; [4] Pratas, Frederico et al. Fine-grain parallelism using multi-core, Cell/BE, and GPU Systems. Parallel Computing 38, 365–390, 2012

DESCRIPTION OF TRANSPOSABLE ELEMENTS IN *STENOSTOMUM LEUCOPS (PLATYHELMINTHES, CATENULIDA)*

VARGAS, VITOR HUGO SCHÜNEMANN[1]; ROSA, MARCOS TRINDADE[2]; Loreto, Elgion Lúcio da Silva[3]

[1] Graduando em Ciências Biológicas, Bacharelado - Univesidade Federal de Santa Maria; [2] Mestrando em Biodiversidade Animal – Universidade Federal de Santa Maria; [3] Dep. de Bioquímica e Biologia Molecular, Universidade Federal de Santa Maria

Keywords: Small flatworms; Mobilome; Next Generation Sequencing (NGS).

In this study, we used the species *Stenostonum leucops*, which is an aquatic small free-living flatworm, they feed mainly debris and microorganisms. It has around 0.5 to 2 mm in length, belonging to the Catenulida class, which has a basal position within the Platyhelminthes. Another striking feature of this organism is their immense regenerative capacity.

In the last decade a real revolution in nucleic acid sequencing technologies have emerged, able to generated a greater number of sequences by each run. These technologies are called Next-Generation Sequences, or NGS. Through these methods, the *S. leucops* genome of cultivar SL01_SM01 was sequenced using an Illumina MySeq sequencer.

Stenostomum leucops may become the first species of the Catenulida class to be made a comprehensive annotation of Transposable Elements. Transposable elements (TE) are described as discrete segments of DNA capable of "moving" from one locus to another within a host genome or between genomes, by means of horizontal transfer. They are highly important for several factors, including the potential to interrupt sequences, later serving as a substrate for homologous recombination causing rearrangements. Also, they are agent for mutations and genetic diseases. In the other hand, there are several evidences that they can influence positively the genome evolution.

To search TEs into the *S. leucops* genome the method by sequences homology was used, where RepeatMasker tools were applied jointly with RepBase database.

Firstly, the raw paired reads were sent to the RepeatExplorer and using FastQ Groomer tool, 34% of the reads with poor-quality were eliminated. After, we used the FastQ Interlacer tool to join both FastQ files into a single one, afterwards executing the FastQ to Fasta tool. Next, the clustering process was started creating clusters with sequences showing similarities. All tools were run using default parameters. Among the reads that went through the filtering, the tool pointed out the existence of 44% of repetitive sequences. After manual analysis, clusters with a number greater than 100 hits were selected, resulting in a percentage of 13.62% of reads presenting similarity to Transposable Elements already described in the RepBase.

Among those that presented similarities with described TEs, 18 families of available elements were found distributed among 8 families of retrotransposons, representing 8.11% of the genome, and 10 families of transposons, with a total of 5.5%. From class II, *mariner* family correspond to 1.98% of the genome, and *maverick* family to 1.94%.

In other Platyhelminthes that had Transposable Elements annotated in their genomes using search by homology, such as the worm *Macrostomum lignano*, was detected a total of 7.7%, which is much less when compared to the 44% found in *S. leucops* using the same methodology. In *M. lignano*, 0.06% and 0.11% were annotated for Classes I and II respectively. In this aspect the results from our researches pointed to values of 8.11% and 5.5% respectively, suggesting the genome of *S. leucops* is significantly occupied by a transposable elements.

POLYADENYLATION REGULATORY SEQUENCES AND FREQUENCY OF ALTERNATIVE POLYADENYLATION SITES IN A COMPREHENSIVE SET OF CANCER PREDISPOSITION GENES

VIEIRA, IGOR ARAUJO[1,2]; RECAMONDE-MENDOZA, MARIANA[3]; SILVA, VANDECLECIO LIRA DA[4,5]; LEÃO, DELVA PEREIRA[1,6]; SCHEID, MARINA ROBERTA[2]; SOUZA, SANDRO JOSÉ DE[4,5]; ASHTON-PROLLA, PATRICIA[1,2,6,7]

[1] Programa de Pós-graduação em Genética e Biologia Molecular (Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, Rio Grande do Sul, Brazil); [2] Laboratório de Medicina Genômica (Serviço de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre - HCPA, Porto Alegre, Rio Grande do Sul, Brazil); [3] Instituto de Informática (UFRGS, Porto Alegre, Rio Grande do Sul, Brazil); [4] Programa de Pós-Graduação em Bioinformática (Universidade Federal do Rio Grande do Norte - UFRN, Natal, Rio Grande do Norte, Brazil); [5] Instituto do Cérebro (UFRN, Natal, Rio Grande do Norte, Brazil); [6] Serviço de Genética Médica (HCPA, Porto Alegre, Rio Grande do Sul, Brazil); [7] Departamento de Genética (UFRGS, Porto Alegre, Rio Grande do Sul, Brazil).

Almost all eukaryotic mRNAs acquire a poly(A) tail at their 3' ends in a process termed polyadenylation. Two core polydenylation elements (CPE) located in the 3' untranslated region (3'UTR) of pre-mRNAs play an essential role in this process: the polyadenylation signal (PAS), a highly conserved hexamer AAUAAA or its close variants; and the actual cleavage site (CS), preferentially a CA dinucleotide 10-30 nucleotides downstream of the PAS. In human genes, PAS sequences include 12 functional hexamer variants, because some positions are tolerant to point mutations. In addition, alternative polyadenylation (APA) is defined as use of more than one functional PAS/CS, allowing a single gene to encode multiple mRNA transcripts with variable 3'UTR. In the present study, we characterized PAS and CS sequences in a comprehensive set of cancer predisposition genes (CPGs), besides exploring the occurrence of APA events in the same group of genes. NCBI (reference source) and APA databases (APADB and APASdb) were queried to characterize CPE sequences in the selected CPGs (n=117), including 81 tumor suppressor genes and 17 oncogenes. Regarding CPGs with no PAS described in the NCBI database, we developed a computational method using in-house Perl scripts to identify the 3'-most hexamers that may function as PAS (putative PAS) in the full sequence of corresponding transcripts. Based on NCBI analysis, we did not find an established PAS in 21 of the 117 CPGs (~18%), and most PAS already described (74.4%) had the canonical sequence AAUAAA, while 24.1% contained the variant AUUAAA. Our computational strategy was able to detect putative PAS sequences for 17/21 CPGs with no established PAS in NCBI database, and putative PAS were not identified for the *ERCC4*, *FH*, *MUTYH* and *SHOC2* genes. Interestingly, we found the AA dinucleotide in most CS sequences associated with this set of CPGs. CA dinucleotide was only the fourth most frequent CS in our gene set, indicating that certain estimates provided by long-standing polyadenylation studies do not apply to all human transcripts. Moreover, an integrative analysis of the data obtained through the NCBI, APADB and APASdb databases allowed to identify 105 CPGs (~90%) with APA sites among their transcript variants, while a previous estimate indicated that it occurs in about 54% of human genes, suggesting a greater complexity in the regulation of polyadenylation in transcripts derived from CPGs. The strongest evidence of APA arose from the *PTEN* transcript which has 61 APA sites differentially used in its processing among different normal human tissues according to APASdb analysis. Overall, our study generated a landscape of polyadenylation regulatory sequences in CPGs that may be useful in the development of molecular analyses covering these often neglected regulatory elements of 3' end processing in human genes. These findings reinforce the relevance of establishing updated methods and/or databases to detect PAS and CS sequences. Furthermore, the computational strategies used here could be easily applied to similar situations with additional genes outside the CPG context. This is the first study focused on the comprehensive characterization of CPE sequences in CPGs.

Keywords: polyadenylation, cancer predisposition genes, alternative polyadenylation.

IN SILICO PREDICTION OF TRIDIMENSIONAL STRUCTURE AND T- AND B-CELL EPITOPES OF MOUSE MAMMARY TUMOR VIRUS env (GP52) PROTEIN

VITIELLO, Glauco Akelinghton Freire[1]; Amarante, Marla Karine[1]; Pereira, Nathália de Sousa[1]; Banin-Hirata, Bruna[1]; Motoori-Fernandes, Carolina Yukari[1]; Rosa, Marcos Henrique[1]; Losi-Guembarovski, Roberta[1]; Watanabe, Maria Angelica Ehara[1]

[1] Laboratory of DNA Polymorphisms and Immunology, Department of Pathological Sciences, Biological Sciences Center, Londrina State University, Londrina, Parana, Brazil.

Mouse Mammary Tumor Virus (MMTV) has been recognized as an etiological agent for breast cancer (BC) in mice, and its' influence in human BC is a field for intense debate in recent literature, with groups reporting convincing epidemiological data associating this (or a 98% homologue) virus with this tumor in different human populations; however, there is still a profound lack of knowledge regarding possible mechanisms of infection and carcinogenesis of this virus in humans: currently, there is no putative receptor described for it in human cells and immunological response to the virus is incompletely understood. Thus, the present project aims to predict tridimensional (3D) structure and identify putative T- and B-cell epitopes of the gp52 subunit of MMTV env protein, paving the way to understand molecular and immunological mechanisms of human infection by MMTV. Ginzu domain predictor was used to predict protein domains which were then structurally predicted through Rosetta's *de novo* method. ProTSAV scoring and ranking function was used to elect the best model for each domain, which were then assembled through *Ab initio* Domain Assembly (AIDA) server. Potential 9 AA T-cell epitopes were predicted by ProPred (for class-II MHC) and ProPred-I (for class-I MHC) while B-cell epitopes were predicted through BCPREDS (for linear epitopes with 20 AAs) and DiscoTope2.0 (for conformational epitope residues) servers. Obtained model had 3 identified domains on the 376 AA protein, and showed 81% of residues on the most favored regions, 18.7% in additional allowed regions, 0.3% in generously allowed regions and no residue in disallowed regions in Ramachandran plot and there were no unusual G-factors (< -0.05) for dihedral angles and main-chain covalent forces in the model in PROCHECK analysis. Analysis in VaxiJen server revealed env protein as a probable antigen using default threshold score (0.4). ProPred-I identified 16 peptides with potential to bind to 41 of the 47 class-I MHC alleles queried (threshold = 4%), while ProPred identified 60 peptides that putatively bind to all 51 class-II MHC alleles available in the server (threshold = 3%). Of these, 7 peptides showed potential to bind both class-I and -II MHC alleles, of which the two most promiscuous were FVAAIILGI, which was predicted to bind to 12 class-I and 41 class-II different MHC alleles (VaxiJen score = 0.77), and YPYAILLGL, predicted to bind 29 class-I and 2 class-II alleles (VaxiJen score = 1.12). BCPREDS analysis predicted 8 potential B-cell linear epitopes (threshold = 75%; score range: 0.8 – 1.0), while DiscoTope revealed 91 B-cell epitope residues. Of the 8 predicted linear B-cell epitopes, only 1 did not shown any sequence overlap with conformational B-cell epitope residue. In conclusion, the present study reveals that MMTV gp52 env subunit may elicit both humoral and cellular adaptive immune response in humans and reveal putative epitopes for T- and B-cell responses. These data may be useful in screening for individuals at risk for MMTV infection and for vaccine and immunotherapy design, which may be of great importance given the compelling data associating MMTV-like sequences with human BC. Furthermore, prediction of gp52 may pave the way for further *in silico* and experimental studies investigating mechanisms of MMTV internalization by human cells, a crucial step for viral infection.

REALIZAÇÃO

UFRGS
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL

.Inf
INSTITUTO
DE INFORMÁTICA
UFRGS

CBiot
Centro de Biotecnologia
da UFRGS

LOCAL

Centro de Eventos do Instituto
de Informática da UFRGS
SITE › WWW.UFRGS.BR/EGB