First ISCB Latin American Student Council Symposium

Belo Horizonte, Brazil
October 27, 2014

Program Booklet

This page intentionally left blank

Contents

Table of Contents

Welcome

The ISCB Student Council (ISCB SC) is pleased to welcome you to the 1st edition of the Latin American Student Council Symposium (LA-SCS) in Belo Horizonte, Brazil. After successfully organizing the Student Council Symposium (SCS) alongside the ISMB conference for the past 10 years and a European Student Council Symposium (ESCS) alongside the ECCB conference for the past 3 years, we are thrilled to bring this opportunity to meet and interact with students and young scientists from all over the world to the Latin American Computational Biology and Bioinformatics student community. As always, our primary goals are to promote the exchange of ideas and to foster the creation and strengthening of scientific social networks.

We are honored to have **Prof. Peter F. Stadler** (Dept. of Computer Science & Interdisciplinary Center of Bioinformatics, University of Leipzig, Germany), **Prof. Vitor Leite** (Departamento do Fisica, Universidade Estadual Paulista, Brazil) and **Dr. Francisco Melo** (Molecular Bioinformatics Laboratory, Pontificia Universidad Catolica, Chile) as keynote speakers at this Symposium. Their keynotes promise to be inspiring presentations of exceptional work relevant to everyone in the field.

Throughout the day we will hear **oral presentations** from a selection of **outstanding student abstracts** spanning a wide range of research areas. In the morning and evening, the **poster sessions** will offer exciting science in various domains, and give everybody a chance to discuss their research topics in more depth. Snacks and refreshments will be offered during the poster sessions.

Everyone involved in the organization of this Symposium contributed significantly to make this event happen. Our volunteers have spent months preparing all aspects of this Symposium ranging from the invitation of keynote speakers, fund raising, advertising, organizing the peer-review process and many more technical aspect like designing the website.

We encourage you to make the most out of this opportunity and to be very active in engaging other delegates, asking questions, discussing ideas and showcasing your own research. You can make this Symposium a starting point for fruitful future collaborations and another step towards a successful career in computational biology.

We would also like to invite you to our **Social Event**! We promise you that our social events are fun. Check at the symposium for more information.

Agenda

Time	Event/Activity
08:00-08:45	Registration
08:45-09:00	Organizers Welcome
09:00-:9:50	**Students Talks Session I:** *Mendel,MD: a user-friendly online program for clinical exome analysis* *Raony Guimarães Corrêa Do Carmo Lisboa Cardenas* and Natália Duarte Linhares and Sérgio Danilo Junho Pena. Laboratory of Clinical Genomics, Brazil *Taxonomic and functional diversity of microbial community from a mining environment* **Julliane D Medeiros, Laura R Leite, Sara Cuadros-Orellana and Guilherme Oliveira. UFMG, CPqRR/ Fiocruz, Brazil**
9:50-10:40	**First Keynote:** *Protein Folding Funnels and Applications to Bioethanol Production* **Prof. Vitor Leite** **Departamento do Fisica at the Universidade Estadual Paulista, Brazil**
10:40-11:10	Coffee break
11:10-12:00	**Students Talks Session II:** *Sifter-T: A scalable framework for phylogenomic probabilistic protein domain functional annotation* Danillo C. Almeida-E-Silva and *Ricardo Z. N. Vêncio. Universidade de São Paulo, Brazil* *Exploring the membrane potential of a simple dual membrane system by using a constant electric field.* *Yerko Escalona, Jose Antonio Garate and Tomas Perez-Acle. Laboratorio de Biología Computacional (DLab), Fundación Ciencia & Vida, Chile*
12:00-14:00	Poster Session + Launch

Second Keynote:
Development of new bioinformatics tools to study the key molecular determinants that mediate protein-DNA recognition

14:00-14:50 **Dr. Francisco Melo**
Molecular Bioinformatics Laboratory at the Pontificia Universidad Catolica, Chile

Students Talks Session III:

Using coevolution to improve protein subfamily classification
<u>Franco Simonetti</u>, Martin Banchero, Ariel José Berenstein, Ariel Chernomoretz and Cristina Marino Buslje. Fundación Instituto Leloir, Universidad de Buenos Aires, Argentina

14:50-15:40

Structural and dynamical transitions in Ankyrin repeat proteins
<u>R. Gonzalo Parra</u> and Diego U. Ferreiro. Universidad de Buenos Aires, Argentina

Analyzing the effect of homogeneous frustration in protein folding
<u>Vinícius Contessoto</u>. UNESP, Brazil

15:40-16:10 Coffee break

Third Keynote:
Phylogenetics from Paralogs

16:10-17:00 **Peter F. Stadler**
Dept. of Computer Science & Interdisciplinary Center of Bioinformatics, University of Leipzig, Germany

17:00-17:20 Concluding Remarks

17:20-19:00 Poster Session + snacks and drinks

19:00 *Symposium Ends + Social Event*

Enjoy your time in Belo Horizonte!

R. Gonzalo Parra

LA-SCS Chair

Avinash Shanmugam

LA-SCS Co-Chair

Program

Note: Please check lascs2014.iscbsc.org for last minutes updates and announcements.

Keynote Speakers

Peter F. Stadler

Professor of Bioinformatics at the University of Leipzig, Germany and External Faculty Member of the Santa Fe Institute, United States.

Corresponding member abroad the Math and Science of the Austrian Academy of Sciences and External Scientific Member of the Max Planck Society

Phylogenetics from Paralogs

Sequence-based phylogenetic approaches heavily rely on initial data sets to be composed of orthologous sequences only. Paralogs are treated as a dangerous nuisance that has to be detected and removed. Recent advances in mathematical phylogenetics, however, have indicated that gene duplications can also convey meaningful phylogenetic information provided orthologs and paralogs can be distinguished with a degree of certainty.

We demonstrate that plausible phylogenetic trees can be inferred from paralogy information only. To this end, tree-free estimates of orthology, the complement of paralogy, are first corrected to conform cographs and then translated into equivalent event-labeled gene phylogenies. A certain subset of the triples displayed by these trees translates into constraints on the species trees. While the resolution is very poor for individual gene families, we observe that genome-wide data sets are sufficient to generate fully resolved phylogenetic trees of several groups of eubacteria. The novel method introduced here relies on solving three intertwined NP-hard optimization problems: the cograph editing problem, the maximum consistent triple set problem, and the least resolved tree problem. Implemented as Integer Linear Program, paralogy-based phylogenies can be computed exactly for up to some twenty species and their complete protein complements.

Vitor Leite

Leader of the Vitor Leite Research Group at the *Departamento do Fisica* at the *Universidade Estadual Paulista,* Brazil

Protein Folding Funnels and Applications to Bioethanol Production

The protein folding is a fundamental problem in Molecular Biophysics. It has been successfully understood using the energy landscape theory. The effective folding landscape topology is like a funnel, which has an energy gradient toward the nativestate region. We discuss approaches to characterize and visualize the funnel using energy landscape theory and simple models.

Several parameters - such as transition temperatures (Tf and Tg), φ-alues and measurement of frustration - can be used to characterize proteins from a theoretical-computational perspective, as well as making connection with experimental results. These parameters allow studying functional aspects, especially when associated with mechanisms involving conformational changes, such as allosteric regulation and enzymatic mechanisms. We discuss applications of these formalisms in the study and development of relevant enzymes in bioethanol production.
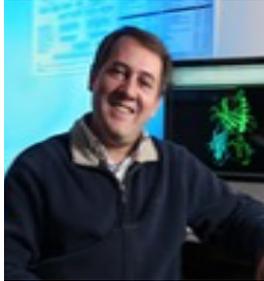
Francisco Melo

Member of the Board of Directors of the ISCB

Leader of the Molecular Bioinformatics Laboratory at the Pontificia Universidad Catolica, Chile

Development of new bioinformatics tools to study the key molecular determinants that mediate protein-DNA recognition

In this talk, a description of several bioinformatics tools recently developed in our lab to assist the study of protein-DNA interactions will be provided. This include a structural database of protein-DNA interfaces, knowledge-based potentials to describe protein-DNA interactions, a software for the fullatom 3D modeling of duplex DNA and protein-DNA complexes, Metropolis-Montecarlo simulations to search the vast sequence space of potential DNA binding sites and a PyMol plugin to visualize the binding interface of protein-DNA complexes from a different perspective.

Oral Presentations
Session I

1. Mendel,MD: a user-friendly online program for clinical exome analysis

Raony Guimarães Corrêa Do Carmo Lisboa Cardenas[1], Natália Duarte Linhares[1], Sérgio Danilo Junho Pena[1]
[1]Laboratory of Clinical Genomics

Introduction

With the advent of next-generation methodology, sequencing of the whole exome of a patient has become economically viable for clinical diagnosis of genetic diseases, including complex and rare ones. The strategy for identification of the pathogenic variant is complex, since in every exome there are 40 to 50 thousand nucleotide variants in comparison with the reference human genome. To simplify this procedure, computational filters that sequentially eliminate common and synonym variations, reducing the size of the total sample, should be used. After identifying pathogenic variants, laboratory confirmation should be carried out, for instance by traditional Sanger sequencing, to reach a definitive diagnosis.

The bioinformatics challenge is that the software has to be efficient and sophisticated from the computational point-of-view and, at the same time, simple and friendly to be used by clinicians. To address this matter, Mendel,MD was developed as a free and open-source tool that can be downloaded, installed and executed locally by any laboratory in the world with aim to analyze exomic data from their patients.

Results

After submission of a standardized file with the exome information (VCF file) into the system, annotation with different methods and tools is done, preceded by calculation of metrics with the information generated. The information about the mean of coverage and quality for all the variants of each individual is presented. Those values are used when defining thresholds for the parameters in the next implemented method which is called Filter Analysis.

Filter Analysis is a method which combines different annotations, databases and scores of pathogenicity allowing to reduce the number of variants and genes of each clinical case from thousands of candidates to only a few dozens. We claim that the final list of genes should always be investigated by doctors and researchers in the search for good candidates causing mutations taking into consideration each specific clinical case.

In order to integrate into the results the possibility of considering different models of inheritance (recessive, compound heterozygous, dominant and X-linked) the Family Analysis method was developed. It enables the search for compound heterozygous variants (the mutation which comes from both parents) and de novo variants in exomes from trios, quartets or even a larger number of individuals from a certain family.

The ultimate method developed in our tool is Pathway Analysis and it can be used to investigate variants and genes grouped by each pathway in KEGG. To test the method we used data from two different disorders Hurler Syndrome and Hunter Syndrome respectively, which, although caused by mutations in two different genes (IDUA and IDS) are both members of the same pathway category (glycosaminoglycan degradation.

The tool was validated with data from 15 different clinical cases submitted from specialized laboratories from different countries. It was consistently possible to identify a very short list of causal gene candidates, which included the correct diagnosis in all cases.

Conclusions

Mendel,MD is an efficient, secure and reliable software in exploration of variants from exome data of patients with Mendelian disorders, so-

phisticated from the bioinformatics perspective and yet simple enough to be used by doctors and scientists to quickly analyze genomic data.

2. Taxonomic and functional diversity of microbial community from a mining environment

Julliane D Medeiros[1], Laura R Leite[1], Sara Cuadros-Orellana[2], Guilherme Oliveira[2]
[1]Universidade Federal de Minas Gerais
[2]CPqRR/ Fiocruz

Biomining uses acidophilic and chemolithotrophic microbes capable of oxidizing iron and sulfur to recover metals of interest from complex minerals. One of the most critical issues in mining environments is the generation of acid mine drainage (AMD) that pollutes water and sediments with acids and metals. The Sulfate-reducing bacteria (SBR) are an alternative to bioremediate contamination of AMD. The current knowledge on the microbial diversity and the metabolic pathways involved in bioming and bioremediation is still limited. In this context, metagenomics has become a valuable tool to investigate previously uncultured microorganisms in environmental samples. The aim of this study was to assess the taxonomic and functional microbial diversity in a mining area in the Brazilian Amazon. We collected acid mining drainage water; sediment and water from the surface and 15 meters depth of a tailings dam that received ~90 million tons of chalcopyrite mining waste. The prokaryotic biomass from water samples was concentrated on filters with 0.22m pores. Metagenomic DNA was isolated and the V4 region of the 16S rRNA was amplified. Shotgun and amplicon libraries were sequenced on Ion Torrent platform. We used Qiime to cluster the sequences into OTU and we observed that the microbial diversity was higher on the sediments and AMD metagenomes (Sediment - Chao:1988; Shannon:6.05 / AMD - Chao:2375; Shannon:5.03) compared to the water (Chao:370; Shannon:4.16). MG-RAST was used to classify the reads (representative hit, identity cut-off Biomining uses acidophilic and chemolithotrophic microbes capable of oxidizing iron and sulfur to recover metals of interest from complex minerals. One of the most criti-

cal issues in mining environments is the generation of acid mine drainage (AMD) that pollutes water and sediments with acids and metals. The Sulfate-reducing bacteria (SBR) are an alternative to bioremediate contamination of AMD. The current knowledge on the microbial diversity and the metabolic pathways involved in bioming and bioremediation is still limited. In this context, metagenomics has become a valuable tool to investigate previously uncultured microorganisms in environmental samples. The aim of this study was to assess the taxonomic and functional microbial diversity in a mining area in the Brazilian Amazon. We collected acid mining drainage water; sediment and water from the surface and 15 meters depth of a tailings dam that received ~90 million tons of chalcopyrite mining waste. The prokaryotic biomass from water samples was concentrated on filters with 0.22 >=75%), and the results indicate the dominant phyla on water metagenomes were Proteobacteria (49.3%), Actinobacteria (31.6%) and Bacteroidetes (7.6%). The most abundant phyla in sediments were Proteobacteria (48.7%) Bacteroidetes (17.2%) and Firmicutes(7.5%). We analyzed OTUs from specific SRB-families (Thermodesulfovibrionaceae, Desulfobacteraceae, Syntrophaceae, Peptococcaceae, Desulfuromonadaceae, Desulfomicrobiaceae) and there was an enrichment of these families in sediment and AMD samples. Blastx against NR was used for functionally classify reads in MEGAN using the SEED and KEGG hierarchy. Despite the prevalence of housekeeping functions, we observed reads matching relevant adaptive traits, such as sulfur oxidation, ABC transporters and resistance to metal. Our data indicate the potential for biomining and bioremediation on the studied environments, however new rounds of sequencing must be done for more conclusive results.

Session II

3. Sifter-T: A scalable framework for phylogenomic probabilistic protein domain functional annotation

Danillo C. Almeida-E-Silva[1], Ricardo Z. N. Vêncio[1]
[1]Universidade de São Paulo

Background

In the functional annotation field, Sifter v2.0 is regarded as one of the best when it comes to annotation quality. Recently, it has been considered one of the best tools for functional annotation according to the initiative "Critical Assessment of Protein Function Annotation" (CAFA), an open collaborative experiment designed for large-scale assessment of protein function prediction tools. Sifter combines two powerful ideas: phylogenomics and bayesian graphical models. Nevertheless, it is still not widely used. This contradictory observation is probably due to issues with usability and suitability of the framework to a high throughput scale.

Although powerful in terms of approach, it can be considered prototype level in terms of software. The current Sifter version does not allow nucleotide or amino acid sequences input directly, nor accepts current standards in gene annotation formats. Moreover, several parameters are still hardcoded and difficult to be tuned by the end user. Finally, its relationship to third party dependence software is cumbersome, along with its output.

Description

In this study, we had two goals: (i) enhance the tool's usability, through local implementations or a web-based front end; and (ii) optimize the original source-code for better performance, allowing it to be used in genome-wide scale.

Among the implemented strategies we have: parallel threads; CPU load balancing; best use of disk access, memory usage and runtime; adaptation to the currently used biological databases formats; improved user accessibility; expansion of accepted input types; auto-

mation of the reconciliation process; new output format; detailed documentation; and other minor implementations.

The increased performance allowed, for example, the reannotation of 419,029 Saccharum officinarum (sugarcane) ESTs to be performed by Sifter-T in 5 days, while BLAST took 49 days in a standard bioinformatics laboratory machine.

Conclusion

This implementation result is presented as Sifter-T (Sifter Throughput-optimized), an open source tool with better usability and performance when compared to the original Sifter workflow implementation. The new Sifter-T features allow researchers to have easy and quick access to the Sifter's powerful annotation mathematical method, now with enhanced experiment customization and keeping the inference engine intact. Sifter-T, and its on-line interface, is freely available at http://labpib.fmrp.usp.br/methods/sifter-t/.

4. Exploring the membrane potential of a simple dual membrane system by using a constant electric field

Yerko Escalona[1], Jose Antonio Garate[1], Tomas Perez-Acle[1]
[1]Laboratorio de Biología Computacional (DLab), Fundación Ciencia & Vida

Connexins (Cxs) constitute Gap Junction Channels (GJCs). GJCs connect the cytoplasm of adjacent cells providing a hydrophilic path between cells that allow the movement, by passive diffusion, of water, cations and small molecules. The opening or closing of GJCs is dependent on the voltage difference between the apposed cells and/or the membrane potential. An approach to understand the voltage gating mechanisms of GJCs is to study a simplified system that can account for the basic features of a GJC. In this work, we have devised a series of simple systems bearing in mind that idea. The systems here presented are: i) a dual membrane, ii) a dual membrane with a pore on each membrane, iii) a dual membrane with a channel connecting both membranes and iv) a dual membrane with a channel having explicit charges inside.

In all cases, membrane and pore were build solely with carbon atoms. Both equilibrium and non-equilibrium MD simulations were performed in all systems. Non-equilibrium simulations were produced by applying a uniform external electric field in order to produce a potential difference across the membranes. We then performed detailed analyses of the electrostatic potential, ionic current and the potential of mean force of an ion through the system pores. This study provided important insights regarding the behavior of the electrostatic potential and ion currents inside simple dual membrane systems with or without a connecting channel, and will be useful in understanding of the voltage effects and ion transport mechanisms of GJCs.

Session III

1. Using coevolution to improve protein subfamily classification

Franco Simonetti[1], Martin Banchero[1], Ariel José Berenstein[2], Ariel Chernomoretz[2], Cristina Marino Buslje[1]
[1]Fundación Instituto Leloir, [2]Universidad de Buenos Aires

Background

The common approach for protein subfamily classification relies on grouping protein sequences according to their degree of similarity. However, there is no single sequence similarity threshold for accurately grouping sequences into isofunctional groups.

Current subfamily classification methods use bottom-up clustering to construct a cluster hierarchy, then cut the hierarchy at the most appropriate locations to obtain a single partitioning. These methods usually integrate data such as protein sequence similarity, residue conservation within groups and HMM profiles. Despite this straightforward approach, results usually predict a great number of subfamilies with few members and limited biological meaning.

The goal of this study is to identify subsets of functionally related sequences within a given superfamily. Since all proteins within a superfamily share a common ancestor, we hypothesize that functional diversity within superfamilies has arisen through a series of concerted changes that must have left an identifiable co-evolutionary signal.

Material and Methods

The challenge is to be able to separate the subfamilies coevolutionary signals and use them in the process of subfamily classification. This information can be used to guide a hierarchical clustering. Our approach uses Mutual Information to calculate covariation and commonly used clustering methods based on sequence similarity. We have defined a select group of superfamilies from the Structure Function Linkage Database as our gold standard dataset.

Results

Different approaches were considered for integrating Mutual Information data in sequence clustering. Since Mutual Information can only be calculated for a group of sequences, a preliminary sequence clustering is performed. Using solely covariation data, our method can cluster groups of sequences from the same subfamily. For a complete clustering solution, it performs almost as good as a hierarchical clustering based on sequence similarity. The next step will be to integrate both methods.

Conclusions

Automated protein classification remains an active topic of research and state of the art methods are far from predicting biologically meaningful results. Covariation data has never been used before in this context and further analysis are needed to improve the method.

2. Structural and dynamical transitions in Ankyrin repeat proteins

R. Gonzalo Parra[1], Diego U. Ferreiro[1]

[1]Protein Physiology Lab, Departamento Quimica Biologica, IQUIBICEN-CONICET, Buenos Aires University

Background

Repeat proteins are composed of tandem copies of similar motifs that can spontaneously self-organize in symmetrical structural forms. Ankyrin repeat proteins comprise a naturally abundant family present in all phyla, most frequently found enriched in eukaryotes and their pathogens. These proteins are typically constructed a variable number of copies of a ~33 residues length motif, often interspread with other non-repeating elements. How do the relatively small sequence variations change the population of the structural forms?

Description

We curated all the high-resolution structural data for ankyrin-repeat containing proteins and used a geometric approach to define repeat-units and repeat-arrays [1]. We calculated the energetics of the native-contact networks and analysed the local frustration patterns [2] of the different repeating elements, the array of repeats and the insertions or deletions with respect to canonical units. We quantified the degree of conservation of the energetic state over the canonical positions and for the interactions that these present in the natively folded forms. Coarse grained molecular simulations were performed with the AWSEM package [3] and representative ankyrin-repeat structures in order to relate the energetic heterogeneities to structural transitions.

Results

We found that natural ankyrin-repeat proteins are composed of at least three different types of repeat-units that differ in the stabilization gained by the burial from solvent upon folding. These energies are reflected in differential sequence signatures and secondary structure composition of the repeat-types. We found that natural repeat-arrays present patches of highly frustrated interactions that are typically localized in specific regions: in the insertions between repeat-units, at binding-sites for protein-protein interactions, and at the terminal re-peats. The degree of conservation of the frustration states of canonical positions is highly correlated with the variations of the sequence along the family. The positions that have high conservation of the frustrated state at the single residue level are connected by a minimally frustrated interaction network. We speculate that consensus sequences stabilize the overall fold by maximizing the energetic gap between the folded and unfolded states establishing a network of minimally frustrated interactions both within and between adjacent repeats. We will discuss how the dynamical properties of these molecules may emerge from the differential energetic distribution along an array of repeating elements.

3. Analyzing the effect of homogeneous frustration in protein folding
Vinícius Contessoto[1]
[1]Universidade Estadual Paulista

The energy landscape theory has been an invaluable theoretical framework in the understanding of biological processes such as protein folding, oligomerization, and functional transitions. According to the theory, the energy landscape of protein folding is funneled toward the native state, a conformational state that is consistent with the principle of minimal frustration. It has been accepted that real proteins are selected through natural evolution, satisfying the minimum frustration criterion. However, there is evidence that a low degree of frustration accelerates folding. We examined the interplay between topological and energetic protein frustration. We employed a $C\alpha$ structure-based model for simulations with a controlled nonspecific energetic frustration added to the potential energy function. Thermodynamics and kinetics of a group of 19 proteins are completely characterized as a function of increasing level of energetic frustration. We observed two well-separated groups of proteins: one group where a little frustration enhances folding rates to an optimal value and another where any energetic frustration slows down folding. Protein energetic frustration regimes and their mechanisms are explained by the role of non-native contact interactions in different folding scenarios. These findings strongly correlate with the protein free-energy folding

barrier and the absolute contact order parameters. These computational results are corroborated by principal component analysis and partial least square techniques. One simple theoretical model is proposed as a useful tool for experimentalists to predict the limits of improvements in real proteins.

Poster Presentations

BIOINFORMATICS OF DISEASE AND TREATMENT

1. Molecular Modeling study for discovery of new ligands for PPAR Gamma responsible for Diabetes Type 2: A Virtual Screening, Molecular Docking, ADME & Tox Study

Asif Naqvi[1], Vibhisha Vaghasia[2], Vijaya Bhargava[3], Awani Mishra[4], B. Hari Krishna[5], Preethy Paul[6], Reshma Radhakrishnan[6], Rinku Chouhan[7]
[1]BioDiscovery Group,
[2]Sardar Patel University, Rajkot,
[3]Rajiv Gandhi University of Health Sciences, Bangalore,
[4]Allahabad University, U.P,
[5]Narayna Medical College, Nellore,
[6]Vivekanandha College of Engineering for Women, Anna University, Chennai,
[7]Department of Biotechnology, MLSU, Udaipur, Rajasthan

The peroxisome proliferator-activated receptors (PPARs) are a group of nuclear receptor proteins that function as transcription factors regulating the expression of genes.PPARs play essential roles in the regulation of cellular differentiation, development, metabolism (carbohydrate, lipid, protein), and tumorigenesis of higher organisms. Peroxisome proliferator-activated receptor gamma (PPAR-γ or PPARG), also known as the glitazone receptor, or NR1C3 (nuclear receptor subfamily 1, group C, member 3) is a type II nuclear receptor that in humans is encoded by the PPARG gene. The nuclear receptor peroxisome proliferator-activated receptor gamma (PPARγ) has evolved from a relatively humble beginning as a regulator of adipocyte development (Tontonoz et al., 1994) to become a potential therapeutic target for dealing with diverse group of disorders, including type 2 diabetes, dyslipidaemia, inflammation and malignancy. A new series of PPAR-γ ligands based on 2- (2,3- dihydro- 1- benzofuran- 5- yl) - 3- (3- pyrrolidin- 1- ylpropyl) - 1,3- thiazolidin- 4- one has been designed employing virtual screening and molecular docking approach. Lamar-ckian Genetic Algorithm based docking (implemented in AutoDock 4) was performed on 3000 derivatives of 2- (2,3- dihydro- 1- benzofuran- 5- yl) - 3- (3- pyrrolidin- 1- ylpropyl) - 1,3- thiazolidin- 4- one into the active site of the PPAR-γ (PDB code 2PRG). Combining a novel algorithm for rapid binding site identification and evaluation with easy-to-use property visualization tools, the software has provided an efficient means to find and better exploit the characteristics of cavity of PPAR-γ.

The docking results showed that the binding energies were in the range of -4.26 kcal/ mol to -8.96 kcal/ mol with minimum binding energy of -8.96 kcal/ mol. 25 molecules maintained essential H-bond interaction with the active site residue, i.e. His 323. The study provides hints for the future design of new derivatives with higher potency and specificity.

2. Cancer missense mutations in β-2-microglobulin result in aggregation

José Soto Lambert[1], Daniel Almonacid Coronado[1], Felipe Venegas Faundez[1]
[1]Universidad Andrés Bello

The massive sequencing of the human genome has resulted in large databases with the genetic variations of the human population in health and in sickness. This information is useful to understand better the mutations related with diseases such as cancer. As the databases have grown the statistical methods to study them have grown too, allowing us to better understand the statistic significance of mutations that lead to cancer. The new statistical methods include, for instance, normalization by the size of the gene, the replication time of the gene, the mutational spectra from each cancer type, and the expression levels of the gene, among others.

In developed countries cancer is the first cause of death. This disease is defined as a deregulation of cellular processes such as: replication, growth and differentiation. The organs that are more affect by cancer are: the prostate, breasts and lungs.

It is now well known that significantly mutated proteins in cancer form amyloid aggregates with TP53 infective behavior. By using phylogenomics as well as bioinformatics approaches we identify another protein mutated in cancer that is capable of making amyloid fibbers: Betha-2-microglobulin (β2m). β2m is a soluble factor of the major histocompatiblity class I, thus, it is involved in the presentation of antigens to the immune system. β2m is a 99 aa protein, that has a highly conserved sequence when compared to β2m proteins present in other species. β2m has the typical immunoglobulin family folding. Its amyloidal behavior was reported in chronic hemodialysis patients, making fibbers in bone junctions, but the implication of clinical cancer mutations affecting β2m is still unknown.

The aggregation process begins with a destabilization of the protein structure, exposing hydrophobic sequences that overlap with themselves to make seeds that propagate the fibber, acting as a catalytic unit of aggregation. The causes of protein destabilization are principally the ionic strength, the concentration and mutations.

In one hand, using MutsigCV and Mutsig0.9 to analyze cancer mutations from exomes from over 4000 patients identified 23 significant mutations for β2m in 5 different cancer types. The structural impact was calculated using the FoldX force field, DUO machine learning and evolutionary database and Polyphen, predicting that 5 mutations should destabilize the structure. In the other hand, aggregation propensity was calculated after a mutation using TANGO, Zyggregator and FoldAmyloid algorithms, corroborating the putative importance of these mutations in the formation of amyloids. These results are now being tested in vitro using thermal stability assays to identify mutations that make β2m prone to aggregate.

3. Assessing the functional impact of cancer mutations in DNA binding proteins through bioinformatics approaches

Daniel Pizarro[1], Daniel Almonacid[1]

[1]Universidad Andres Bello

Cancer is defined as a number of diseases in which abnormal cells divide without control and are able to invade other tissues. Genetic studies have been informative, but they fail to provide information for uncharacterized genes and lack to incorporate evolutionary aspects of genes. To overcome these problems, mutations in genes can be considered in the context of other homologous genes, improving the relevance of seemingly non-significant data and presenting new work sets of study.

We obtained a set of genes mutated in tumor exomes from 4327 patients affected with one of 19 different types of cancer from The Cancer Genome Atlas. We filtered out insertions, deletions and synonymous mutations. This subset was then subjected to statistical analysis using MutSigCV and MutSig 0.9 obtaining a total of 352 genes significantly mutated in Cancer. MutSig correlates the mutation rate with a set of events that can change the probability of a gene being mutated by chance. It includes gene length, identity of mutations in certain cancer types, expression rate, replication timing and chromatin status. Among the 352 genes we found many families to be highly mutated. 64 of these 352 genes were identified as producing sequence-specific DNA binding proteins by the Gene Ontology, and many of these seemed to be of relevance to cancer.

Transcription factors (TF) are proteins that bind specific sequences of DNA, activating or repressing the target gene's transcription. They are in charge of controlling many metabolic and developmental pathways, making them an important target for cancer mutations. Mutations in TFs can mainly have three types of effect: loss of function, in which case the TF fails to recognize the target sequence; gain of function, when TF binds DNA permanently interfering with its normal behavior, and finally a switch in the DNA recognition, when it binds DNA with the same affinity, but binds a different sequence. Of the 64 DNA binding proteins, we identified those that were TFs only, with crystallized structures

under 2.2 A˚ of resolution, and that were bound to double stranded DNA. Finally we filtered the data to only have proteins with mutations in their DNA binding domain, which lead to a final dataset of 5 proteins and 16 mutations. Mutant models were created with Maestro, and then minimized with the FoldX force field. The resultant models were subjected to a FoldX DNA specificity protocol to create sequence logos. Logos were also created by using web tools 3DTF and PiDNA. In all three cases the software showed a switch in the DNA recognition motif of the selected proteins, which together with energy binding calculations between DNA and the proteins suggested that cancer mutations in TFs are not producing a loss or gain of function, but instead switching their binding affinity for another DNA sequence.

To sum up, in this work we analyzed the full set of TFs significantly mutated in 19 types of cancer, proposing that some of these mutations cause a switch in the recognized DNA sequence, which corresponds to a new type of effect of mutations in TFs, having implications in the role of TFs in cancer genesis and progression.

4. Finding new, more efficient beta-glucuronidases in the glycosyl hydrolase family 2 for hydrolysis of morphine-6-glucuronide and codeine-6-glucuronide

Catalina Valdivia[1], Paullett Arriagada[1], Manuel Rozas[2], Daniel Almonacid[1]
[1]Universidad Andrés Bello, [2]Kura Biotec

Background

The human body metabolizes drugs mainly by modification and conjugation to signal that they must be eliminated from the organism. Among the conjugating enzymes, those that attach glucuronic acid to drugs are common. In the market exists a high demand for enzymatic methods for hydrolysis of drug glucuronides, an important step for their detection when analyzing biological samples. This step is mainly performed using Escherichia coli betaglucuronidase, but this enzyme is not efficient with all glucuronide conjugates, and while several studies to increase efficiency of the existing methods have been performed, it is clear that they are not enough to meet the demands of the market.

Description

In this work we studied the universe of functionally uncharacterized enzymes belonging to the Glycosyl Hydrolase 2 (GH2) family, to which E. Coli beta-glucuronidase belongs, with the purpose of finding a protein that has greater affinity to certain drugs of interest than the enzymes currently used. We generated sequence similarity networks of all glycosyl hydrolase 2 family members, which yielded two clusters containing known beta-glucuronidases. These clusters were further analyzed filtering by sequence length and presence of three Pfam domains characteristic to beta-glucuronidases: PF00703, PF02836 and PF02837. However, beta-galactosidases also present those domains, in addition to other domains, such as PF02929 and PF2449. Due to this and other similarities between the enzymes we searched for sequence differences, because of the possibility that some sequences were beta-galactosidase fragments. We performed a multiple sequence alignment of reviewed beta-galactosidases and beta-glucuronidases using PROMALS3D, finding that the distance between the catalytic residues was higher in beta-glucuronidases. Using this knowledge, from the two networks created we selected 100 sequences for homology modeling. We used crystal structures of beta-glucuronidase from E.coli and human available from the Protein Data Bank (PDB) as templates. We then performed molecular docking of the models with the drugs of interest in order to characterize the binding affinities for each of the ligands.

Conclusions

We are now identifying a set of possible beta-glucuronidases with better affinity to morphine-6-glucuronide and codeine-6-glucuronide, among others, and elucidating the causes for the different interactions between the drug glucuronides and the active sites of the enzymes. By doing this, we hope to solve the problem of finding an enzyme capable of hydrolyzing the aforementioned conjugates, which can't be efficiently hydrolyzed with the currently commer-

16

cially available enzymes. The enzymes identified in silico will be then tested in vitro and hopefully reach the analytical toxicology labs worldwide.

5. Identification and Rational Design of Peptide Ligands for Cancer Screening.

Claudia Pareja[1], Sebastián Cruz[2], Diana Gaete[2], Álvaro Lladser[2], Tomás Pérez-Acle[1]
[1]Computational Biology Lab, Fundación Ciencia & Vida,
[2]Laboratory of Gene Immunotherapy, Fundación Ciencia & Vida

Phage display is a laboratory technique used for finding new ligands such as enzyme inhibitors, receptor agonists and antagonists, to target protein. In this case the target is human carcinoembryonic antigen (CEA), a protein that has been found to be associated with various types of cancer, particularly colorectal carcinoma, and developed to be a molecular target for cancer diagnosis and therapy.

In the present study, we designed a series of potential CEA binding peptides for early detection of cancer. In general, we applied phage display technique in combination with computational modeling to identify high-affinity peptides for CEA. For this, we used molecular modeling to design the target affinity-selected peptide, experimentally determined by phage display technique, and then we performed a molecular docking with the target to do an analysis of the CEA-peptide complex. In this analysis we considered mainly the hydrogen bonds, hydrophobic interactions, and van der Waals forces to evaluate the connections. In the next stage, a serie of peptides was designed in reference to the binding site on target, and finally the conformational affinity of the CEA-modified peptide complexes was calculated from where the designed peptides that showed the complexes of lower energy were selected.

In conclusion, here we study the use of phage display in combination with molecular modeling to identify high-affinity ligands for CEA. These techniques have broad application to the design of novel ligand peptides and can be extrapolated to other targets.

6. In silico structure analysis and epitope prediction of E3 CR1-beta protein of Human Adenovirus E for vaccine design

Noman Ibna Amin Patwary[1], Shah Md. Shahik[1], Md. Saiful Islam[1], Md. Sohel[1] and Mohd. Omar Faruk Sikder[1].
[1]Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chittagong-4331, Bangladesh

Background

The Human adenovirus E (HAdVs-E) genome is a linear, double-stranded DNA containing 38 protein-coding genes. Wild-type adenoviruses type E, are linked to a number of slight illnesses. The E3 CR1 beta protein controls the host immune response and viral attachment.

Description

We use numerous bio-informatics and immuno-informatics implements comprising sequence and construction tools for construction of 3D model and epitope prediction. The 3D structure of E3 CR1-beta protein was generated and total of ten antigenic B cell epitopes, 6 MHC class I and 11 MHC class II binding peptides were predicted.

Conclusions

The study was carried out to predict antigenic determinants/epitopes of the E3 CR1-beta protein of Human adenovirus type 4 along with the 3D protein modeling. The study revealed potential T-cell and B-cell epitopes that can raise the desired immune response against E3 CR1-beta protein and useful in developing effective vaccines against HadVs-E.

7. Comparative genomics and in-silco subtractive genomics approach for targeting multidrug resistant Neisseria

gonorrhoeae with medicinal-plant-derived antimicrobial compounds

Syed Babar Jamal[1], Sandeep Tiwari[1], Leticia C Oliveira[1], Syed Shah Hassan[1], Sintia Almeida[1], Vinicious Ac Abreu[1] and Vasco Azevedo[1]

[1]UFMG - Universidade Federal de Minas Gerais

Neisseria gonorrhoeae is responsible for causing gonorrhea, one of the most common sexually transmitted diseases prevailing globally. To control the transmission of the disease and to develop drug(s) against the pathogen, researches are in progress but, till date no effective vaccine or specific drug could be developed and only antibiotic treatment is in use. The rapid emergence of antimicrobial resistance in Neisseria gonorrhoeae to a range of antibiotics with no new therapeutic agents being produced. There are several drug resistant strains have been found. In the present study, we tried to find candidate drug and vaccine targets against the multidrug resistant strain of N. gonorrhoeae using comparative genomics and an in silico subtractive genomics approach. A data-set of active natural antimicrobial compounds derived from different natural resources were retrieved and a library was made and used in this analyses. Molecular properties and prediction of bio-activity of all the compounds were determined. Various drug relevant properties; like mutagenic, tumorigenic, irritant and reproductive effect of the compounds following Lipinski Rule of five, were also checked. The compounds fulfilling the criteria of having these molecular properties and drug-like ability were separated and used in the docking analyses. Compounds demonstrated good docking results with all target proteins were considered to be used as potential therapeutic candidates for the treatment of Pharyngeal diphtheria, owing to experimental validations. The proposed putative target of Neisseria gonorrhoeae play a critical role in the pathogenesis and help bacteria in multidrug resistant. Compounds derived from medicinal plants are better enough for their accessibility and low price. This will definitely change the trend in the near future as new inhibitors are discovered and optimized, and the 3D structures of inhibitor complexes are determined. Structural studies of inhibitors, in complex with putative targets should also further facilitate the application of virtual screening approaches, giving more insight into the inhibitory mechanism of the target proteins.

FUNCTIONAL GENOMICS

8. Analysis of the Trypanosoma cruzi transcriptome in response to gamma radiation

Michele A. Pereira[1], Mariana Boroni[1], André L. M. Reis[1], Carolina F. Castro[1], Helaine G. S. Vieira[1], Priscila Grynberg[2], Glória R. Franco[1]

[1]Universidade Federal de Minas Gerais,
[2]Embrapa Recursos Genéticos e Biotecnologia

Background Trypanosoma cruzi, the etiologic agent of Chagas disease, is a kinetoplastid organism highly resistant to DNA damage caused by ionizing radiation. After a dose of 500 Gy of gamma rays, the genomic DNA is fragmented and the parasite can restore the chromosomal bands pattern in less than 48 hours. To understand changes in gene expression during cell recovery, previous studies were already performed. Microarrays analysis showed that gamma rays affect T. cruzi gene expression in a time-dependent manner. This study aims to compare the gamma radiation effect to the T. cruzi transcriptional profile using high-throughput RNA sequencing (RNA-seq). Descriptions Epimastigote cells from CL Brener strain were exposed to a dose of 500 Gy in a cobalt (60Co) irradiator. Total RNA was extracted from non-irradiated cells (control sample) and irradiated cells (4, 24 and 96 hours post-irradiation). For RNA-Seq analysis, Illumina HiSeq was used to perfom paired-end sequencing. Approximately 13 millions of paired-end reads were obtained for each library and their quality was evaluated using FastQC. Reads were mapped to the CL Brener reference genome with Bowtie2 and BWA-MEM aligners and the alignments were inspected through the Integrative Genomics Viewer. The uniquely mapped reads were filtered using samtools and submitted to

HTSeq for gene counting. The CL Brener genome has approximately 14000 annotated genes, however the CL Brener .gff file used contains only 3997 annotated genes. The remaining genes will be used in future analysis. The obtained gene counts were ploted in a Venn diagram to analyze the relation between the four time points. Results showed that 35, 45, 83 and 27 genes were specific for the control, 4h, 24h and 96h post-irradiation samples, respectively. 2433 genes are shared by all time points and 480 were absent in all of them. Reads will further be mapped to the Esmeraldo-Like and Non-Esmeraldo-Like CL Brener genomes, which present approximately 10600 coding genes and the differential gene expression in the time course will be analyzed. Conclusions This study will help to understand how the parasite can handle such a harmful stress and which genes are up and down regulated to maintain its survival.

9. In silico Identification of Non-coding RNAs in the Crenarchaea Sulfolobus acidocaldarius Genome: an RNA secondary structure-based approach

Victor Aliaga-Tobar[1], Raúl Arias-Carrasco[1], Vinicius Maracaja-Coutinho[1], Alvaro Orell[1]
[1]Universidad Mayor

Non-coding RNA refers to RNA molecules that do not encode proteins. The functions of non-coding RNAs are diverse and associated to important biological tasks. For instance, in eukaryotes small nuclear RNAs are required for splicing, while microRNAs are involved in gene expression regulation. In bacteria small regulatory non-coding RNAs range in size from about 50 to 500 nucleotides and are often required in stress response. In contrast, non-coding RNAs identification is rather scarce for members from Archaea. In this regard, most studies have focused on determining a large number of small RNAs, though other types of non-coding RNAs, as long non-coding RNAs have not been considered. Here, we have performed an exhaustive in silico prediction of non-coding RNAs molecules in the genome sequence of the crenarchaea Sulfolobus acidocaldarius. By using our in-house "structRNAfinder" bioinformatic tool, which is based in covariance models to analyze non-coding RNAs secondary structure conservation, we were able to identify 137 non-coding RNAs present in Sulfolobus acidocaldarius genome, thereof excluding stable non-coding RNAs (tRNA and rRNA). Consequently, out of the 99 previously annotated non-codings RNAs we extended to 38 newly predicted molecules. In general, 71 RNA molecules correspond to small RNAs, 27 to CRISPR system associated-RNAs, 18 to transfer-messenger RNA, 4 to miRNAs, 4 to long non-coding RNAs and 10 unclassified non-coding RNAs. Moreover, we found 26 non-coding RNAs with an antisense orientation that matched annotated protein-coding ORFs, 20 of them were not hitherto predicted as non-coding RNAs. Because anti-sense-non coding RNAs might act regulating gene expression we are currently i) extending our non-coding RNAs identification within RNA-seq transcriptomic dataset, as well as ii) performing in vivo functional studies so as to evaluate their contribution in S. acidocaldarius physiology.

10. Analysis of gene classification associated to differential expression using geometric approach in data coming from RNA-Seq

Tiago Tambonis[1], Marcelo Boareto[2], Vitor Barbanti Pereira Leite[1], Nestor Felipe Caticha Alfonso[2]
[1]São Paulo state University, UNESP/IBILCE, Dept. Physics – São Jose do Rio Preto – São Paulo (Brazil),
[2]University of São Paulo - Sao Paulo (Brazil)

Analysis of differential expression is studied through of microarray and RNA sequencing (RNA-Seq). Despite of the fact that microarray be largely used by experimentalists, it is a more limited technique compared with RNA-seq. The use of RNA sequencing technology produces thousands of small sequences that later will be aligned to a reference genome or transcriptome (usually). The execution of these steps will produce a table which relates the amount of mRNA sequences aligned to a gene or transcript in different experimental conditions. The presence of errors in one of these steps can be propagated, producing noises that hinder the inference of differential

expression. From this context, the method used for analysis of differential expression must be robust so that all information provided is extracted of concise and precise way. Thus, in this work, it is proposed a statistical geometric method to study differential expression in data sequencing coming from RNA-seq. The geometric approach defines a metric and an objective function that is minimized such that elements in the same group remain close and elements from different groups remain distant. This method determines the relevant genes if differential expression is detected between the samples. This approach was analyzed using a benchmark dataset called SEQC, which includes replicated samples of the human whole body reference RNA and human brain reference RNA along with RNA spike in controls. These sample are part of the MAQC project and include close to 1000 genes that were validated by TaqMan qPCR. The evaluation was performed using ROC analyses in curves of sensibility vs specificity with the R tool pROC comparing with another packages (edgeR, DESeq, baySeq) commonly used by researchers. Using the graphs is possible to conclude that the results of the geometric approach are comparable to another packages in some cases and better in other cases. Furthermore, this method not need any assumption about the distribution associated with the count table, it is not parametric, it is of simple understanding and not necessary high computational resources.

11. *Aeromonas salmonicida Transcriptomics Hunting for Loss of Virulence Mediated by Culture Temperature*

Cristopher Segovia[1], Vinicius Maracaja-Coutinho[1], Javier Santander[1]
[1]Universidad Mayor

Background

Aeromonas salmonicida is one of the oldest known fish pathogens and the causative agent of furunculosis. It has broad host range a nearly worldwide distribution, causing significant mortality to wild and farm fish. A. salmonicida can be cultivated at temperatures as high as 34.5°C. At temperatures over 22°C its chromosome and virulence plasmid undergo recombination. The effects of this recombination are a faster growth and lost of virulence, both due to genetic recombination of the chromosome and virulence plasmid. Despite the documented relationship between loss of virulence and increase of growth rate the mechanisms involved in this genetic rearrangements remain unknown. Here, we used phenotypical test, transcriptomics and genomics to determine the possible mechanism involved in A. salmonicida temperature inducible recombination.

Description

It is known that vapA, an important virulence factor responsible for the synthesis of the A-layer, is lost due to insertion sequence ISAS1 and ISAS2 endogenous recombination. A. salmonicida A-layer is a VapA protein array that covers most of the LPS O-antigens and it's necessary for complement resistance and congo red A+ phenotype (red colonies). Using the A+ and A- (withe colonies) on congo red agar we determined the frequecy of chromosomal recombiantion. Also, we isolate total mRNA from an unstable strain grown at 15°C and 28°C. We found that the rRNA of A. salmonicida lost significant molecular weight at 28ºC compared to 15ºC. This result was coincident with RNA-seq analysis, where the 5S subunits of the rrn genes were missing at 28°C. Also, the mRNA G+C composition shifted after recombination. Finally, we determine the up/down-regulated genes that might be related to the recombination machinery, such as the transposases IS630, IS30 and IS3.

Conclusions

We conclude that all A. salmonicida strains lost vapA gene associated to A-layer when they were grow at 28 ºC. Ribosomal mRNA, 16S and 23S, lost significant molecular weight suggesting recombination at rrn operons. This is coincident to the loss of 5S rrn genes at 28 ºC. Three trasposases, IS630, IS30 and IS3, seems to play a mayor role during recombination. These results support the idea that A. salmonicida is under accelerated evolution.

12. Gene expression profiling of the parous mouse oviduct. Possible implications in ovarian cancer

José Carlos Márquez[1], Carlos Chacon[1], Sandra Ampuero[2], Ulises Urzúa[1,3]

[1]Laboratorio de Genomica Aplicada, ICBM, Facultad de Medicina, Universidad de Chile

[2]Programa de Virología, ICBM de Medicina, Universidad de Chile,

[3]Centro de Investigacion y Tratamiento del Cancer, Fac de Medicina, Universidad de Chile

Background

Ovarian cancer (OC) is among the major lethal gynecological cancers worldwide. The facts that prophylactic salpingectomy reduces OC risk plus recent genomic analyses uncovering an oviductal epithelium signature in the most frequent OC variant, led to the notion that OC would actually originate from oviductal cells. In addition, periods of interrupted ovulation such as pregnancy and oral contraception have largely known to decrease OC risk through a protective hormonal effect. As the oviduct might also be responsive to pregnancy hormones, the aim of this study was to obtain transcriptional profiles of multiparous C57BL6 female mice compared to age-matched virgin animals aiming to detect altered expression of genes implicated in OC protection by pregnancy and pre-neoplastic transformation by repetitive ovulation.

Description

Gene expression profiling was conducted in one-channel Illumina MouseRef-8 v2 beadchips (Illumina, Inc., San Diego, CA) for two divergent reproductive conditions (multiparous vs virgins; n=4/group). Regarding the multiparous/virgin comparison, 159 genes were upregulated and 336 genes were downregulated. Importantly, 26 of these genes were described as tumor suppressors in TSG database, in particular 4 upregulated (Dido1, Igfbp5, Eif2ak2 and Isg15) and 22 downregulated. Functional enrichment analyses (GO, KEGG and WiliPathways) of upregulated genes showed heat shock protein binding (e.g. Cdc37, Pdxp, innate immune response (e.g. Irf9, Stat), cholesterol biosynthesis (e.g Mvd) and mRNA processing, for instance Rbm4. In contrast, major downregulated genes functions included vascular morphogenesis, negative regulation of inflamatory response, tight junctions and focal adhesions (eg ITGB, TGFBR1), regulation of actin cytoskeleton, cell cycle, signaling of the Wnt, MAPK and insulin pathways and mRNA procesing.

Conclusion

Parity promotes oviductal expression of genes related to inmunologycal response and cell stability suggesting a preparation to stress. In addition, the repression of those genes that promote cell migration, anti-apoptosis and proliferation suggests that pregnancy condition inhibiting or decrease likelihood of pre-neoplastic lesions or cell migration toward ovary reducing the risk of ovarian cancer. This event may drive by progesterone and estradiol, the same pregnancy hormones that may protect the ovary. The possible mechanism would be to subject the oviduct to a constant and mild stress that it will be prepared for a worse one, for instance, neoplastic process.

GENETIC VARIATION ANALYSIS

13. Selecting potential functional candidate genes for temperament of Guzerá cattle in a region of chromosome 1

Carolina Matosinho[1], Fernanda Santos[1], Maria Gabriela Peixoto[2], Pablo Fonseca[1], Maria De Fátima Pires[2], Izinara Rosse[1], Frank Bruneli[2], Glaucyana Santos[2], Maria Raquel Carvalho[1]

[1]Universidade Federal de Minas Gerais, [2]Embrapa Gado de Leite

Background

The Guzerá cattle presents heat tolerance and resistance; and the ability to survive under limited food resources during dry periods of the year, due to it, Guzerá cattle is one of the most common zebu breeds in northwest semi-arid region of Brazil. Zebu breeds was known popularly as "brave" breeds, due to having more aggressive temperament. When compared with calm animals, the angry ones produce meat of poorer quality, are more susceptible to

diseases, have lower weight gains, lower reproductive efficiency and lower milk production. Genome-Wide Association Studies (GWAS) use genotyping technologies for analyzing several single-nucleotide polymorphisms (SNPs) and relate them to clinical conditions and measurable traits. This kind of study usually identify SNPs in linkage disequilibrium with the real variant that contributes to the phenotype measured. A region around SNPs found to be associated with reactivity in a previous GWAS performed by our research group was delimited and investigated by a literature data-mining in the present study. The region was delimited by the rising of p-values around the associated marker found in chromosome 1. A list with all the genes located inside the region previously delimited, was constructed. The software Gene Relationships Across Implicated Loci (GRAIL) was used to perform the data-mining. This software uses our text-based definition of relatedness to identify a subset of genes, and it is based on two key methods: a text-based similarity measure that scores two genes for relatedness to each other based on text in PubMed abstracts and a novel statistical framework that assesses the significance of relatedness between genes in disease regions. The input file for GRAIL analyzes was made using the Gene name and the Gene ID of the homologous gene in Human genome.

Results

Initially, the input file consisted for 39 genes. After the analyses we were able reduce considerable the number of genes that can be associated with temperament in Guzerá cattle. The keywords related to the phenotype evaluated were temperament, behavioural and stress. These results show the genes in chromosome 1 that may be associated with behavioural traits and temperament in Guzerá cattle.

Conclusions

The bovine temperament is an important characteristic in the animal's well-being, and it affects the animals productivity. Given these facts, and the importance of this study for the breeding selection, the results shown in the present work are very important for the temperament genetics from Guzerá and milk production in Brazil.

14. SNP clustering on DBSCAN and sensitivity analysis

Vinicius C. Brum[1], Itamar L. Oliveira[1], Marcos Vinícius Silva[2], Wagner Arbex[1,2]
[1]Universidade Federal de Juiz de Fora, PGCC – UFJF
[2]Empresa Brasileira de Pesquisa Agropecuária – Embrapa

The aims of this work are the knowledge acquisition from genotype data of 56K single nucleotide polymorphism (SNP) chip in genome of three cattle breeds through the algorithm DBSCAN (Density-Based Spatial Clustering of Applications with Noises) and its sensitivity analysis with respect to changes on its initial conditions in this database. The database used has 2467 samples distributed in three cattle breeds. Two breeds are taurine cattle and have 56947 markers: Holstein (577) and Jersey (1024); and one breed is zebu cattle and have 54000 markers: Nellore (866). Through this work was passible to find markers that do not discriminate samples of these breeds, since they remain constant independently of their breeds. In addition, it was possible to analyze the algorithm's sensibility and show that it presents relative stability in this scenario, since the clustering results converged with no difficulty. Finally, this work demonstrates the possibility to build a model able to generalize and predict the unknown breed samples of cattle efficiently.

GENOME ASSEMBLY AND ANNOTATION

15. CGIIS: an automated tool to Close Gaps Inter and Intra Scaffolds

Raul Arias-Carrasco[1], Vinicius Maracaja-Coutinho[1]
[1]Universidad Mayor

Background

One of the main issues in whole genome sequencing is to fill the gaps on the final assembled genomic sequence, linking the different

scaffolds reconstructed. A common strategy to deal with this problem is generating additional sequencing runs, using: (i) mate-pair/paired-end strategies or (ii) long reads sequencing technologies (i.e. Sanger, PacBio). However, its mapping and quality evaluation is not a trivial task. Currently computational procedures are still not completely automated and time consuming.

Description

We developed an automated toolkit of Perl scripts to solve this problem. It uses sequences obtained from additional sequencing runs, other assemblies, or a taxonomic related organism as a reference. It uses this information, in order to fill the missing fragments between the extremity of each consecutive (intra) contigs on the same scaffold, or in different scaffolds (inter) from the assembled draft genome. Based on user defined identity and coverage cutoffs, CGIIS aligns the extremity of each contig on the scaffold, against the reference sequences generating a final reconstructed super-scaffold. Finally, original raw reads are mapped against the reconstructed region, and a confidence score is calculated based on the re-mapping coverage. The algorithm was applied on different public available draft genome sequences form Bacteria, using new sequencing runs, related organisms and different strains as a reference. It reduced by around 50% the number of contigs on the final super-scaffolds reconstructed.

Conclusions

The implementation of CGIIS allowed us to close different Bacteria draft genome sequences in an automated way, and in less time compared to current methodologies. Moreover, the tool can be used to join different assemblies of the same organism produced by distinct assembly tools.

16. Study and Comparison of Heuristics for the Problem of Genome Rearrangement by Unsigned Reversals
Eduardo Fernandes[1], Eloi Araujo[1]
[1]Universidade Federal de Mato Grosso do Sul

Genome rearrangement is a problem in computational molecular biology that can be described as a succession of changes on a chromosome, specifically in the order of blocks of gene in a sequence of the chromosome. There are lots of variants for the same problem, and one of them is known as genome rearrangement by unsigned reversals. The main goal of this work is to study this specific kind of genome rearrangement.

A chromosome can be represented by a permutation, that is a sequence of genes in which each gene is indicated by integers. The problem of genome rearrangement by reversals consists in sorting the genes of a sequence using the smallest amount of operations known as reversals. An unsigned reversal, or simply reversal, is defined as an operation that reverses the order of genes in a range of the sequence. After a reversal is applied, the first element before the reversal becomes to last, the second one becomes the penultimate, and so on.

Genome rearrangement problem is difficult because there is no polynomial algorithms for them unless $P = NP$.

This work consists of: the theoretical study of a sorting algorithm by reversing previously described; the proposal of some new heuristics; and the comparison among the implemented algorithms.

Basically, there are four heuristics that were studied in this work, the first one is from literature and the other three were proposed by us. The first heuristic (proposed by Watterson et al., 1982) puts the k-th element in the correct position in the iteration k, using one reversal. As well a famous sorting algorithm Quick-Sort, the second heuristic, using reversals, separates into two parts the permutation, one with the smallest elements and the other with the greatest elements and then sorts recursively the two parts. The third one, for each step of ordering, searches for reversal whose sum of distances between the current position of the element and the position that it must be, for each element, is minimum. The fourth heuristic works as follows: for each step of the ordering, it searches for reversal whose sum of dis-

tances, for each element (call it n) of the permutation, between the current position and the position of n-1 and n+1 elements is minimum.

All the algorithms were implemented using C++ language and executed in the supercomputer of the Technological Center of Electronics and Computer (Centro Tecnológico de Eletrônica e Informática – CTEI), at the Federal University of Mato Grosso do Sul (UFMS). For this work, it was used a set of random entries with different sizes.

17. SIMBA: a simple way to make complete assemblies of bacterial genomes

Diego C. B. Mariano[1], Letícia C. Oliveira[1], Edson L. Folador[1], Edgar L. Aguiar[1], Leandro Benevides[1], Felipe L. Pereira[1], Marcus Canário[1], Thiago J. Sousa[1], Rommel T. J. Ramos2, Vasco A. C. Azevedo[1]
[1]UFMG,
[2]UFPA

Background

The evolution of large-scale sequencing platforms has reduced the time and spent cost on the process of DNA fingerprinting. However, sequencers still have limitations, such as the capacity to read the maximum size of DNA fragments, leading to the need to fragment the DNA into small pieces before sequencing. Through this approach, it is necessary, after this step, to rearrange the fragments read (reads), for then can be possible to represent the original genome. This process is known as genome assembly. The process is very complex and dependent of limitations of sequencers, so several computer programs are necessary to work with the data. Nowadays, a lot of strategies for genome assembly have been proposed, but there isn't consensus on the best approach yet. One of the problems detected was the large amount of programs that should be implemented in assembly processes, requiring a large computational domain from the bioinformaticians. The adoption of a tool with good usability could reduce the spent time on hand labor training. In this context we present SIMBA (Simple Manager for bacterial genomes assembly), a Web tool designed to manage strategies for a hybrid pipeline that aims to facilitate the implementation of the processes of genome assembly.

Results

SIMBA was developed using PHP and SQLite database. The software allows data processing and conversion extensions through various scripts. To de novo assembly, SIMBA allows to use four separated softwares, which three are based on overlap- layout-consensus algorithm (Mira3, Mira4, and Newbler), and one are based on De Bruijn graph (Minia) algorithm. To finish the installation, SIMBA utilizes two approaches: the first one is based on ordering contigs by a reference genome and the second one is based on reports of optical mapping generated by Opgen MapSolver. Finally, SIMBA allows the download of the results generated and the manual curation of the data with some other software.

Conclusions

The friendly interface of SIMBA allows an easiest performing of the genome assembly process and also facilitates the ordering of the contigs and closing gaps. Any bioinformatician without a great specific knowledge of hardware and software can do all the steps for assemble a genome. The application all the developed scripts and the source code were made available in http://github.com/dcbmariano/simba.

18. The complete genome sequence of Streptococcus agalactiae strain GBS85147

Edgar Lacerda de Aguiar[1], Diego C. B. Mariano[1], Leticia Castro Oliveira[1], Lucas Amorim Gonçalves[1], Alberto F. Oliveira[1], Marcus Canario[1], Flávia de Souza Rocha[1], Felipe Luiz Pereira[1], Siomar de Castro Soares[1], Fernanda Alves Dorella[1], Carlos Augusto Gomes Leal[1], Henrique Cesar Pereira Figueiredo[1], Vasco Ariston de Carvalho Azevedo[1]
[1]Universidade Federal de Minas Gerais

Streptococcus agalactiae (Lancefield group B, GBS), is a gram-positive and cocci, bacterial patho-gen. This species can causes diseases in humans, cattles and fishes. In humans, it is associated with neonatal sepsis and meningi-

tis, as early-onset or late-onset diseases (EOD, LOD). It can also affect immunocompromised adults, although also being a common colonizer of the gastrointestinal and genitourinary tracts. In dairy cattle, GBS is an im¬portant pathogen of clinical and subclinical mastitis, affecting quality and production of milk. In fish, S. agalactiae is an emerging pathogen that causes septice¬mia and meningoencephalitis with high mortality in wild and cultured species worldwide. In this work, the genome was sequenced with PGM Ion Torrent, using fragment library approaches, and 200 bp sequencing kit, according to manufacturer's recommendations. A coverage approximately of 246x was obtained, with 578,082,183 bp in 2,973,022 reads with mean length of 203 bp and Phred quality greater than or equal to 20 in 91.25% of bases. The data was assembled with Mira Assembler version 3.9.18 (N50 length of 104.996 bp) using the recommended parameters. One hundred and four contigs amounting 2.032.890 bp were mapped over reference (S. agalactiae GD201008) using CONTiguator 2.0, and 34 contigs had similarity. There were micro contigs remaining, less than 600bp. The overlaps contigs were removed using "in-house scripts" and the last 8 gaps were filled manualy using extract consensus of reads map over S. agalactiae GD201008 and S. agalactiae 09mas018883. Through CLC Genomics Workbench 7 version it was possible to curate the frameshifts, which 254 regions of pseudogenes were observed. To validate the corrections, Blastn tool was used in NCBI database. For prediction of rRNA and tRNA, it was used RNAmer and tRNA Scan, respectly, through prokaryote parameters. For gene prediction, the data was performed in FgenesB, where using Streptococcus agalactiae 09mas018883 as reference. The genome was in Artemis, using the protein blast of Uniprot database. The Interproscan 5 was used for re-analysis of hypothetical proteins that were found in the genome. Streptococcus agalactiae strain GBS 85147 is comprised by a circular chromosome with 1996163 bp. There are 1925 coding sequences, 18 rRNA genes, 63 rRNA genes and G+C content of 35,98%. The bacterium has a high genetic diversity and considerable amount of hosts, which highlights the importance of the sequencing of new samples for classification as to the pathogenesis, characteristics, etiology and target host. The elucidation of these aspects is essential for the development of improved therapeutic strategies and new genomic comparative analysis between samples, minimizing the socioeconomic impact of bacteria in the society.

PATHOGEN INFORMATICS

19. Metabolic advancement of multidrug resistant E.coli: A Comparative Metabolomics analysis of Multidrug resistant and Pathogenic Isolates E. coli Strain with Commensal E. coli Strain

Sandeep Tiwari[1], Syed Babar Jamal[1], Leticia C Oliveira[1], Syed Shah Hassan[1], Sintia Almeida[1], Vinicious Ac Abreu[1], Vasco Azevedo[1]
[1]UFMG - Universidade Federal de Minas Gerais

Host-microbiota is interaction symbiotic relation, on the other hand interaction between Host-pathogen results in host damage. To grow and replicate bacteria need energy, carbon and nitrogen for this Animal tissue are reach source of nutrients for bacteria. But the report of whole microbiome project there are tenfold more bacteria than human cell within the body. Thus, the pathogens have to invade niches that already occupied by perfectly adapted bacteria. For compete with the resident microbiota pathogen modulate metabolic pathways with in the host tissues. These modulation of pathogens help them to coordinately regulate virulence factor expressions. Escherichia coli is common microflora found in the human gastrointestinal tract. However some isolates of E.coli can causes illness. E. coli strains cause diverse intestinal and extraintestinal diseases by means of virulence factors that affect a wide range of cellular processes. The rapid emergence of antimicrobial resistance in E.coli to a range of antibiotics with no new therapeutic agents being produced. Here, we compared the all metabolic pathways of commensal E.coli, with the multidrug resistant and pathogenic isolates. This comparative

study provide the role of metabolic pathways on the pathogenicity and virulence of bacterial pathogen. Since pathogenic microbes rapidly develop resistance against existing antibiotics, new anti-infective strategies are needed to keep ahead of the inevitable resistance that accompanies antimicrobial use. Since metabolism is an essential for virulence, such pathways could potentially be good targets for anti-microbial therapies. Bacterial in vivo metabolism is one of the most fundamental aspects of virulence of pathogenic bacteria yet our understanding of it is relatively limited. Although comparative genomics should identify new pathways unique to specific pathogens or associated with virulence genes, new strategies for the analysis of the importance of specific metabolic pathways in host tissue must be developed. This study help us in understating the pathogenic behavior of pathovar of E.coli at metabolic level.

POPULATION GENETICS, VARIATION AND EVOLUTION

20. Unraveling evolutionary pathways for genomic and phenotypic diversification and its predicted impact on the ecophysiology of biomining Acidithiobacilli

Carolina González[1,2,] Darwin Guzmán[2], Paz Tapia[2], Jorge Valdés[2] and David S. Holmes[1]
[1]Center for Bioinformatics and Genome Biology, Fundación Ciencia & Vida y Facultad de Ciencias Biológicas, Universidad Andrés Bello, Santiago, Chile.
[2]Bio-Computing & Applied Genetics Division, Fraunhofer Chile Research Foundation-Center for Systems Biotechnology, Santiago, Chile.

Members of the Acidithiobacillus genus are characterized for their ability to derive energy from inorganic sources under extreme acidic conditions and high concentrations of heavy metals. They are typically found in acid mine drainage as part of a microbial consortia, constituting a powerful biotechnological tool for the industrial recovery of metals.

Genome sequences from several Acidithiobacillus representatives have been generated, annotated and compared to assess evolutionary processes potentially associated with gene gain/loss and its impact on functional and properties for each species. Conserved, dispensable and unique genes were identified in order to decipher evolutionary paths responsible for shaping the current genomic structure and phenotypic properties of each representative. Comparative genomic analysis identified more than a thousand conserved protein families that constitute the genus' core functionalities, anticipated to be key for life in extreme acidic environments.

Gene gain/loss events seems to be a major explanatory mechanisms for driving diversification within this genus and could explain the adaptation to their extreme environment conditions, including the ability to survive using a chemolithoautotrophic metabolism and taking advantage alternative and complimentary mechanism for iron and sulfur oxidation, allowing the exploitation of mineral sources for the extraction of precious metals.

The present analysis provides a comprehensive comparative genome analysis to aid the understanding of marked functional similarities and points to a common origin in this bacterial genus, delineating species boundaries between Acidithiobacillus representatives and proving an information platform for the continuous classification and analysis of current and future genomic and metagenomic data from extremely acidic environments.

21. Evolution of 3'UTR-associated RNAs

Jan Engelhardt[1], Peter F. Stadler[1]
[1]Bioinformatics Group, Department of Computer Science, University of Leipzig

Background

Despite their abundance, unspliced EST data have received little attention as a source of information on non-coding RNAs. Very little is know, therefore, about the genomic distribution of unspliced non-coding transcripts and their relationship with the much better studied regularly spliced products. In particular, their evolution has remained virtually unstudied. A subclass of the unspliced EST cluster consists of

so called 3'UTR-derived RNAs (uaRNAs). They can have functions in cis and trans, independent from the harboring gene. uaRNAs can be detected by combining EST data with predicted transcription start sites.

Results

We systematically study the evidence on unspliced transcripts available in EST annotation tracks for human and mouse, comprising 104,980 and 66,109 unspliced EST clusters, respectively. 15-20\% of the unspliced EST cluster are conserved between human and mouse. More than 7,000 human and 6,000 mouse unspliced EST cluster overlap the 3'UTR of a RefSeq gene or are located within 5kb downstream of the 3' end. Using TSS predicted by chromatin data we identify a total of 1,547 bona fide uaRNA candidates in human. Integrating only the public available CAGE data by the FANTOM5 consortium we predict a total of 1,891 uaRNA candidates in human and 2,477 candidates in mouse. We also give a first glimpse on the sequence and structure conservation of these uaRNA candidates.

Conclusions

Expressed sequence tag data combined with experimentally predicted promoter data, e.g. CAGE, is a powerful tool to identify candidate uaRNAs. This combination of data sets could also be applied to non-model organisms without a sequenced genome. uaRNAs are a quite new class of non-coding RNAs and have not been extensively analysed yet. We present a catalog of candidates which are excellent targets for experimental verification. Increasing evidence hints to additional regulatory functions of 3'UTRs independent from the processing of the corresponding gene. It is very likely that this mechanism is not only present in human and mouse but also other eukaryotes. Using public data is a great way to get a glimpse on the uaRNAome of the respective species.

PROTEOMICS

22. Visual strategies to reveal patterns of protein-ligand interactions

Alexandre Victor Fassio[1], Sabrina Azevedo Silveira[1], Raquel Cardoso de Melo Minardi[1]
[1]Universidade Federal de Minas Gerais

Molecular recognition plays an important role in biological systems and is a phenomenon of organization very difficult to predict or design even for small molecules. Due to its remarkable importance molecular recognition was studied under different perspectives in Bioinformatics. Several studies focused on seeking patterns of molecular recognition on datasets consisting of a specific receptor and multiple ligands or vice versa by using varied data mining techniques. The analytical process is extremely toilsome as an expert has to assay each of the patterns carefully and they can be very voluminous. Data visualization has in recent years become an established area of study in academia and is increasingly being used in biological data visualization. Therefore, we propose visual and interactive strategies to depict the types of interactions established between a protein and its ligands and materialized these strategies into a tool called nAPOLI (Analysis of PrOtein Ligand Interactions). nAPOLI is a quite different approach which aims at being more easy and intuitive. Here, we focus on a case study of an important family of enzymes: the Cyclin Dependent Kinases II (CDK2).

23. Analysis of the active and inactive states of EGFR kinase domain by pockets and cavities structural properties comparison

Marcia Anahí Hasenahuer[1], Yanina Powazniak[2], Guillermo Bramuglia[2], Gustavo Parisi[1], Maria Silvina Fornasari[3]
[1]Universidad Nacional de Quilmes, [2]Fundación Investigar - Argenomics,
[3]Centro de Estudios e Investigaciones, Universidad Nacional de Quilmes, Buenos Aires, Argentina

Background

EGFR (Epidermal Growth Factor Receptor) is a tumor marker in many cancer types. Several single aminoacid substitutions (SASs) in this protein are present in different cancers. Most of these SASs are characterized as "activating", through the stabilization of the conformer required to drive the phosphorylation (active form). EGFR is a trans-membrane protein, with extracellular and cytoplasmic regions. The latter has a juxtamembrane, a Tyr-kinase and a C-terminal intrinsically disordered tail (C-tail)

regions. Auto-phosphorylation on different C-tail tyrosine sites triggers signaling pathways for cell growth and proliferation. Most interaction sites of proteins with their ligands and substrates are located in cavities or pockets. The goal of this project is to understand the structural and physicochemical characteristics of EGFR kinase pockets that differentiate the active and inactive conformations and to try to elucidate the effect of SASs in those pockets, that could trigger the unregulated kinase activity. Particularly, the effect of not previously reported SASs observed in Argentine cancer affected patients is a central objective of this work.

Description

Pockets and cavities calculations were performed on conformers of human EGFR Tyr-kinase region with fPocket [http://fpocket.-sourceforge.net/]. Different conformers coordinates were taken from PDB [http://pdb.org/pdb/home/home.do] and CoDNAS [http://www.codnas.com.ar/about.php]. The characteristics of the pockets were analyzed using per-site RMSD, volume, polarity of surface area, charge and hydrophobicity, also considering the presence of mutations in conformers. Clustering methods were applied to compare this information using statistical packages of R [http://www.r-project.org/].

Further, SASs from Argentine patients and others compiled from COSMIC database [http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/] were mapped in the structures, to analyze their relationship with pockets.

Conclusions

The main pockets related to kinase active site that we found, either contain or are in close contact with 70% of the 195 positions with SASs related to cancer in EGFR cytoplasmic region. Reorganization of pockets could favor the binding of the C-tail to be phosphorylated and could affect the affinity for ATP or anti-cancer drugs . Disease related SASs could affect the dynamics and shape of pockets, promoting a deregulated EGFR activity. Co-localization of most of cancer related sites in pockets could be important to develop predictive computational tools.

24. Characterizing the Physicochemical

Properties of SiO2-Supported Membranes: A Molecular Dynamics Study
Sebastian E. Gutierrez-Maldonado[1], Maria Jose Retamal[2], Marcelo Cisternas[2], Ulrich G. Volkmann[2], Tomas Perez-Acle[1]
[1]Computational Biology Lab (DLab), Fundacion Ciencia para la Vida,
[2]Laboratorio de Superficies (SurfLab UC), Instituto de Fisica, Pontificia Universidad Catolica de Chile

Various biotechnological applications have arisen from the manufacturing of synthetic membranes. These applications range from modelling the function of cell membranes to developing biosensors with the incorporation of membrane proteins. Synthetic membranes, composed mainly by phospholipid bilayers supported on solid surfaces, have several advantages over similar systems, as they are stable, easy to manufacture and they exhibit lipid mobilities similar to those found in vesicles. However, upon inclusion of membrane proteins, the proximity to the solid support becomes a liability in terms of protein functionality, also limiting lateral displacements of the protein. Several approaches have been employed to solve this problem, with the incorporation of polymers in the form of cushions being one of the most widely used. In previous work carried out by our group, a biocompatible polysaccharide (chitosan) and a lipid bilayer (dipalmitoylphosphatidylcholine, DPPC) were deposited from the gas phase in high vacuum onto a solid surface covered with its native SiO2 layer, providing a hydrated support. This system was characterized by Raman spectroscopy, very high resolution ellipsometry and atomic force microscopy.
In order to provide a complementary description to experimental data, we have built an in silico model of this synthetic membrane, which will allow us to describe its physicochemical properties by means of molecular dynamics simulations. Characterizing properties such as area per lipid, order parameter, lateral diffusion, among other properties, will give us an insight of the interactions occurring at the atomic and molecular level. These simulations are being performed using an United Atom forcefield description and the GROMACS simulation package. Our results indicate that the interaction between chitosan and DPPC affects the ordering of the alkane tails of the bilayer, although leaving the area per lipid unmodified. Further description of the system is needed as the incorporation of the Si/SiO2 solid support will introduce more variations to the ones already described in the present work.

25. Rational discovery of new capsaicin analogues as TRPV1 activators

Javier Cáceres[1], Romina Sepúlveda[1], Camila Navas[1], Ramón Latorre[2], Fernando González-Nilo[1]
[1]Center for Bioinformatics and Integrative Biology (CBIB),
[2]Centro Interdisciplinario de Neurociencia de Valparaíso (CINV)

The ability to interpret environmental signals is a fundamental feature to ensure the integrity and survival of living organisms. The nociceptors are a group of sensory terminals that can detect a variety of noxious stimuli such as thermal or chemical, and are related with the generation of the pain signal and inflammation. These sensors have the ability to respond to capsaicin, the pungent agent of chilli peppers, which allowed the identification of TRPV1, a polymodal non-selective cation channel, tightly related with the generation of acute and neurogenic pain. TRPV1 is a member of the Transient Potential Receptor (TRP) family. It can be activated by noxious temperature (>42°C), low pH and ligands like capsaicin and resiniferatoxin. Even when the physiology of the channel has been characterized in some detail, the structural events that the channel undergoes during gating are still unknown. Recently the TRPV1 structure was resolved by electron cryo-microscopy in a fully closed conformation and in complex with activators such as capsaicin , allowing us to get insight of the distinct conformations of TRPV1. Consequently, the understanding of the structural background of the channel gating in a ligand-dependent manner, provide exciting opportunities for pharmacological intervention.

The aim of this research is to identify novel activators of TRPV1 in a rational framework, taking the characterization of capsaicin as a model for structure-ligand interaction and to unravel the structure-function relationship involved in channel gating via molecular dynamics (MD) simulations. We have analyzed the vanilloid binding site within the third and fourth transmembrane segments, revealing a hydrophobic pocket with a marked presence of polar residues on the intracellular edge of the membrane. Furthermore, our MD assays has been shown that the capsaicin-dependent activation of the channel involves a shift in the curvature of the 'S6-TRP domain' segment disrupting the pore domain, presumably providing the first step for the ligand-dependent channel gating and thus revealing the great relevance of the vanilloid binding site to the channel opening. Molecular docking assays indicates the influence of the polar residues in the orientation of capsaicin in the binding pocket showing the vanilloid ring facing the polar residues and the aliphatic ramification pointing to the center of the membrane. These findings will allow us to perform a rational search strategy of novel TRPV1 activators, testing a database of 112.935 molecules with a massive molecular docking strategy. As a first approach we have selected a group of 10 molecules using the binding affinity and the hydrophobicity relative to capsaicin as selection criteria. Our goal is to contribute to the development of novel analgesic drugs and to increase the knowledge of the gating of this class of ion channels.

26. Generation of a β-glucuronidase with enhanced affinity for codeine-6-glucuronide

Nicole Garcia[1], Daniel Almonacid[1]
[1]Universidad Andres Bello

The goal of the project is to improve the binding affinity of β-glucuronidase from Escherichia coli for codeine-6-glucuronide, to reduce the time of hydrolysis of the conjugate, as it represents an important sample preparation step for doping tests. The enzyme from E. coli has been structurally characterized and its active site thoroughly studied, being known that its catalytic residues correspond to: E413, Y468 and E504, and that each of these amino acids are found in the conserved motifs IANEP, RYYGW and TEYG.

From the crystals available in the Protein Data Bank (PDB), we selected the best template for our in silico mutation study by performing a docking analysis of the drugs of interest with all the chains of all the crystals of E. coli β-glucuronidase available. To evaluate the best protein-ligand complexes we used both i. the distance between an oxygen of the nucleophilic glutamic acid and the electrophilic carbon forming the glycosidic bond in the glucuronide conjugate, and ii. the best binding energy reported by the docking algorithm. We considered the distance between the two mentioned

atoms important because of the retention mechanism used by the enzyme to catalyze the reaction, and because we were looking for a conformation in which the catalytic residues were interacting closely with the substrate.

To select the residues that would be mutated in the active site of the enzyme, we identified those residues in direct contact with the ligands up to 4 Angstrom away. Each of the identified amino acids was mutated to the other 19 possible residues using an in-house script that uses Modeller. We then performed docking of the generated mutants with morphine-3-glucuronide (a substrate that is easily turned over by the enzyme) and codeine-6-glucuronide (a substrate that is known to be a bad substrate by the enzyme), and selected those with the smaller distances between the nucleophile and electrophile atoms and having the lowest binding energies. Finally, only 7 single mutants were selected, using physicochemical residue similarity to reduce the possible mutation space. Then, using the combinatory of all single mutants designed, we generated double and triple mutants, and selected the best two double mutants and the best triple mutant.

Currently, the 10 mutants generated are being experimentally characterized in order to confirm if they fulfill their purpose and possess a better affinity for the ligand codeine-6-glucuronide, according to what we predicted with our computational analysis.

27. Orienting the dipole in a molecular dynamics

Alejandro Bernardin[1], Jose Garate[2], Tomas Perez-Acle[3]
[1]CINV, Fundacion Ciencia & Vida,
[2]Fundacion Ciencia & Vida,
[3]Computational Biology Lab. Fundación Ciencia para la Vida

Nowadays, Molecular dynamics (MD) has become a standard method for the inspections of molecular systems at the atomic scale. Several MD studies have explored the influence of externally applied electric fields (EF) unraveling a number of interesting phenomena such as voltage dependent gating, electro-freezing, enhance single diffusion etc. In general, non-thermal effects of EF can be divided into translational and orientational, where the former are present for entities where a net charge is present whereas the latter are a consequence of the presence of a net dipole. In MD, orientational effects have been addressed by directly measuring dipolar kinetics and orientation in the form of relaxation times, dipolar alignments rotational diffusivity among others. Nonetheless, the orientational thermodynamics int the presence of externally applied EF's has not been thoroughly addressed and, up to our knowledge, only Potential of Mean force (PMF) calculations employing the direct counting method have been carried out. In this work, we explore the possibility of performing PMF calculations on a reduced space or collective variable geometrically defined by the dipole orientation. In the current implementation, harmonic and adaptative biases employing the Local-Elevation method can be performed. In practice, this is equivalent to the application of a "Local" EF, which is shown in this work. Several test were carried out in toy systems, ranging from free molecules in vacuo to confined fluids. We believe that this new collective variable could have potential usage in the study of more complex system such as conexins were orientational EF effects are poorly described at an atomic level.

28. The ion permeation process into K-channels using non-equilibrium molecular dynamics.

Romina Sepúlveda[1], Felipe Bravo-Moraga[1], Ignacio Diaz-Franulic[2], Daniel Aguayo, David Naranjo[2], F. Danilo Gonzalez-Nilo[1]
[1]Center for Bioinformatics and Integrative Biology, Universidad Andrés Bello,
[2]Centro Interdisciplinario de Neurociencia de Valparaíso

Potassium channels have been recognized as a family of membrane proteins, which controls the selective and fast flow of K+ across the membrane. The highly conserved signature sequence "TVGYGD" forms the selectivity filter of the channel and is responsible of both ion selectivity and high rates of ion transport (Heginbotham et al 1994). Nowadays, potassium channels are known for exhibiting a wide conductance rate from 8.2 pS to 250 pS despite that selectivity filter is largely conserved (Carvacho et al, 2008).
Some aspects about the relation between the charges distribution along the pore and the conductance rate have been studied. In Shaker channel, a low conductanc K-channel,

one single point mutation of the P475D increases the 4-fold the maximal unitary conductance leading to propose that a new binding site for K+ is created by negatively charged residue addition (Moscoso et al 2012). In BK channels charged residues at the internal entrance does not modifies the maximal ion transport rate and the main difference between small and large conductance K-channels could be the size of the channel pore In order to understand at the molecular level the conduction process in both K+ channel classes we performed molecular dynamics simulations under external electric field to compare the behavior of BK and Shaker channels under the application of different voltages (+300 mV, +600 mV and +800mV). The differences between the ion permeation process in both ion channels, the ions distribution, ions occupation into the selectivity filter, electric field effect into the structures and the water role are described in order to explain the factors that disturb the conductance process.

SEQUENCE ANALYSIS

29. Computational approach for amino acids visualization according to codon usage
Katia De Paiva Lopes[1], Ricardo Assunção Vialle[1], Verônica R M Costa[1], J Miguel Ortega[1]
[1]Universidade Federal de Minas Gerais

Background

Codon usage bias stands for when specific codons are used more often than others during translation of genes and it is known that it results from selection for efficient and accurate translation of highly expressed genes. The higher demand for common codons leads to a faster and more accurate translation, while the non-optimal codon usage shows a tendency for presence in turns, loops and domains linkers. It means that slowly translated regions can represent separated folding events. These preferential use of codons is a phenomenon extensively studied for several reasons, like speed control of transcription and evolutionary relationships.

Results

We developed an algorithm to fast identify and visualize the amino acids codon usage bias of a Protein Data Bank - PDB structure file. It was developed using MATLAB platform and MySQL database. First, the tool searches for the protein sequence against the translated nucleotide NCBI database using TBLASTN. Then, it uses as reference the Kazusa Codon Usage Database to find the specific codon usage of an organism. Lately, it shows in Jmol molecule viewer the amino acids colored by their codon usage rarity according to calculation of median; a blue to red color scale represents from rare to more abundant codon usage, respectively. The PDB files 3OQ3, 2IAK, 1LZ1 e 1A4X were used for parameterization of the software due to different conformations, different number of chains and CDS translation. For example, the bifunctional protein PyrR (1A4X), a Bacillus subitilis protein which regulates transcriptional attenuation of the pyrimidine nucleotide (pyr) operon by binding in a uridine-dependent manner to specific sites on pyr mRNA, corroborates the idea of codon usage showing rare codons in domains linkers (arginine).

Conclusions

Some proteins were analyzed supporting the studies that high frequency usage codons are mainly associated with structural elements, such as alpha helices, while lower frequency usage codons are related with beta-strands, coils and structural domains boundaries. Thereby, the tool shows to be useful to assist studies of protein structure related with codon usage bias.

30. Using sequence similarity networks to understand the distribution of the β-glucuronidase activity in the GH-A clan of the glycosyl hydrolase superfamily
Nelly Arriagada[1], Daniel Almonacid[1]
[1]Universidad Andres Bello

The enzyme market has increased greatly over the years, especially in pharmacology where enzymes are used in drug identification. Our aim is to identify enzymes of interest to industry, and in this particular case, β-glucuronidases capable of hydrolyzing codeíne-6-glucuronide, a chemical reaction of interest to toxicology laboratories performing drug tests.

CAZy is an authoritative database that groups a large number of glycosyl hydrolases (GH) in families according to sequence and structure similarities. In this database all β-glucuronidases appear associated to clan GH-A, also called by other authors the 4/7 superfamily, given that all members share an (β/α)8 barrel that contains an acid/base catalytic residue in β-sheet 4 of the barrel (glutamic acid 413 in β-glucuronidase from E. coli and 451 in β-glucuronidase from H. sapiens), and a catalytic nucleophile in β-sheet 7 of the barrel (glutamic acid 504 in β-glucuronidase from E. coli and 540 in β-glucuronidase from H. sapiens).

Clan GH-A (or superfamily 4/7) is composed of 19 families, each showing a different spread of catalytic activities. Beginning with the functionally characterized members of each of these families, we created a sequence similarity network (SSN) with the aim of identifying β-glucuronidases in organisms from all domains of life present in NCBI's nr database. SSNs are a phylogenomics approach based on the fast computing of all-against-all sequence similarity using BLAST, which is then depicted as a network, with nodes representing protein sequences, and edges corresponding to similarities between the connected nodes.

The network was then mapped with functional annotation from UniprotKB and CAZy. The resulting network identifies enzymes with β-glucuronidase activity in four families: GH1, GH2, GH30 and GH79. However, a literature search identifies differences between the preferred substrates for each of those families:

* Family GH1 contains only one enzyme with β-glucuronidase activity. This is the protein Klotho, whose efficiency is very low, and with a substrate specificity limited to steroid glucuronides. The physiological activity of Klotho seems to be part of the FGF23 receptor.

* Family GH2 contains 24 enzymes with β-glucuronidase activity. These proteins present specificity for the glucuronide conjugates of several drugs, including morphine. There are also two structurally characterized enzymes: those from E. coli y H. sapiens.

* Family GH30 contains one β-glucuronidase. The substrate specificity of this enzyme is exclusive for baicalin-glucuronide.

* Family GH79 contains five functionally characterized β-glucuronidases. One of them presents specificity for baicalin, and other three for β-glucuronosyl and 4-O-methyl- β-glucuronosil residues in arabinogalactan-protein complexes. There is also a structurally characterized enzyme from Acidobacterium capsulatum.

We are now beginning to study each of the families in order to underpin the structure-function linkage paradigm for the β-glucuronidases from each cluster. What is immediately interesting is the homoplasy that we have identified for the β-glucuronidase activity, i.e. this activity has emerged at least in four occasions within the context of the GH-A clan. We also note that the distribution of the families within the SSN for the GH-A clan is very varied, with the GH2 family being the most diverse, and thus perhaps not well defined as one family.

SYSTEMS BIOLOGY AND NETWORKS

31. A Local Method for the Evaluation of Gene Regulatory Network Inference based on Three Nodes Graphlets.

Alberto J.M. Martin[1], Calixto Dominguez[2], Alejandro Bernardin[1], Tomas Perez-Acle[1]
[1]Computational Biology Lab (DLab) Fundacion Ciencia & Vida. Centro Interdisciplinario de Neurociencia de Valparaiso,
[2]Bioinformatics and Genome Biology, Fundacion Ciencia & Vida. Computational Biology Lab (DLab) Fundacion Ciencia & Vida, Computational Biology Lab (DLab) Fundacion Ciencia & Vida.

Background

Networks are mathematical representations widely used in biology to depict complex

systems as graphs constituted by nodes and edges. Edges connect pairs of nodes if some relationship between them can be established. In Gene Regulatory Networks (GRNs), nodes represent genes and the connections among them depict the existence or not of a regulatory interaction of their expression. GRNs are composed of different basic building units, small induced subgraphs called graphlets. These graphlets form local interconnectivity patterns that describe how the elements represented in the network perform specific functions. Interestingly, the performance of methods for the inference of GRNs inference, is commonly estimated without considering these basic building blocks. The prevalent approach solely relays on the existence of single edges and disregards the network functional units represented by the graphlets.

Description

We have created a new methodology that focuses on the comparison of graphlets formed by the same nodes in the gold standard and in inferred GRNs. In our approach, the existence or not in an inferred network of all graphlets present in the gold standard is treated as binary classification problem for which we provide several comparison measurements. We also provide a quantitative measure of how well are reconstructed the graphlets from the gold standard in the inferred GRN.

Conclusions

We have applied our method to the three whole genome GRNs used to assess the performance of a consensus inference method. This method is a linear combination of all the individual participants in the fifth edition of the Dialogue on Reverse Engineering Assessment and Methods experiment (DREAM5) an it consistently outperforms each of the individual inference approaches. Our data proves that the evaluation of network inference methods should include the reconstruction of network graphlets as a key element to be assessed, since the prevalent performance measures tends to over-estimate the quality of inferred networks. Our approach

also has applications to perform quantitative network differential analysis based on the local topology depicted by the graphlets.

32. Use of a text-mining approach for describing a Gene Ontology poor annotated process as a case study: the blood-brain barrier

Carlos Alberto Xavier Gonçalves[1], José Miguel Ortega[1]
[1]Universidade Federal de Minas Gerais

Background

Gene Ontology (GO) is a project which developed three ontologies of terms to describe gene products with respect to their associations with cellular components, molecular functions or biological processes. By collecting all the taxonomy IDs of the sequences to which a given GO term of biological process was associated, it is possible to reveal the history of that process, allowing us to find those that are very recent in human evolution. However, this analysis revealed several terms associated to a single gene product – such as the term "GO:0090210 - Regulation of establishment of the blood-brain barrier", whose sole human associated protein is the orphan G-protein Coupled Receptor GPR124. The blood-brain barrier is a complex structure whose function is to maintain homeostasis on the CNS and control the passage of metabolites into and out of the brain. Its establishment can be divided in multiple steps, all of which require several regulatory gene products that could be associated with the GO:00909210 term.

Description

We used a text-mining approach to make a broad characterization of this process. Web application Medline Ranker was used to rank a background set of abstracts comprising the Pubmed query "Blood-Brain Barrier" by comparing them with a restricted set of manually selected relevant abstracts, such as those that mentioned many genes participating on the process; the best 1,000 were then submitted as input to another application, PESCADOR, which marks up co-occurrences of genes and associated forms of biointeractions (i.e. "upregulates", "represses"). Relevant interactions

between genes were used to create a visual pathway, using the software PathVisio. This resulted on the coupling of 35 gene products other than GPR124. Using other softwares developed by our group, we were able to attribute ancestrality levels to all these proteins, some of them appearing in very recent moments of human evolution.

Conclusion

We propose the use of text-mining tools as an easy form of batch annotation of Gene Ontology terms of biological processes, with great perspectives for the enrichment of poorly annotated entries. The ancestrality analysis depicted the last added components, revealing how the system has adopted its final functionalities

33. Study of correlations among genes of the cluster nif by means of vector algebra analysis in nitrogen fixing organisms

Nilson Coimbra[1], Vinicius Weiss, Dieval Guizelini[1], Alexandre Lejambre[1], Maria Berenice Steffens[1], Roberto Raittz[1]
[1]Universidade Federal do Paraná

The metabolic capacity for nitrogen fixation is present in several prokaryotic species. A presence of a minimum set of six genes, nifHDK and nifENB, is actually used for prediction of nitrogen fixation organism. Concepts of linear algebra such singular value decomposition, vectorial algebra and principal components analysis, show a natural and powerful interpretations in analysis of microarray data, comparison of proteins and data/text-mining. We did not find any study or application of such approaches in the analysis of the nif cluster and its genes, correlating rules in the various nitrogen fixing organisms. We used vector algebra, a technique of linear algebra, with cluster of orthologous of nif genes to create organism-cluster vector which explore evidences of correlation among 82 diazotrophs genomes and orthologs groups. We created a MATLAB routine with treatment of data of 82 genomes, available on NCBI, results of in silico analysis of SILA, OrthoMCL, BBC and Inparanoid programs, to annotate and classify in orthologs groups. The analysis of organism-cluster vector identified 126 orthologs groups by Or-

thoMCL, BBC and Inparanoid softwares. Some of these are with the same gene product, whereas, 22 orthologs groups are identified as hypothetical protein. The pearson correlation coefficient of the nifH gene presented highest positive correlation with nifK, nifE, nifB and nifN with values 0.3616, 0.1789, 0.1488, and 0.1061 respectively, as indicating a high co-occurrence of these genes. The nifH gene presented negative correlation with serine O-acetyltransferase, cystathionine, beta-lyase and cysteine synthase with values -0.4344, -0.7021, and -0.7021, respectively, that may indicates a distance dependence with these groups. Corroboration for these results was found in literature on Dos Santos et al study, which defined a minimum set of nif cluster in 149 diazotrophs organisms. We expect to find evidences to discover possible new rules involving the 22 cluster of orthologs containing hypothetical proteins. We explored vectorial algebra and correlation analysis techniques applied directly on genomic studies and intent to expand further analysis to infer about functional linkage between genes or proteins of other gene clusters.

34. Molecular Design of Nucleic Acids Carriers based on Bioinformatics Analysis of
Protein-DNA Interactions

Valeria Marquez-Miranda[1,2], María Belén Camarada[1,2], Ingrid Araya[1,2], Daniel Almonacid[1], Fernando González-Nilo[1,2].
[1]Universidad Andrés Bello, Center for Bioinformatics and Integrative Biology, Av República 239, Santiago, Chile.
[2]Fraunhofer Chile Research Foundation, Av. Mariano Sanchez Fontecilla 310, Santiago, Chile.

Background

Dendrimers have gained prominence as efficient non-viral gene delivery carriers, due to their unique features as well-defined size and shape, monodispersity and variable end-groups. Several efforts have been devoted to design a nanoparticle which can associate strongly enough to nucleic acids so that it remains intact during binding and entry into the cell. However, one of the issues that must

be improved is how to modulate dendrimer - DNA interaction in order to promote the unpacking of the complex inside the cells, allowing the release of the cargo. Thus, we decided to employ the knowledge about how protein and nucleic acids interact in Nature with the goal of identifying the functional groups involved in these interactions.

Description

To gain insight into this matter, a bioinformatics strategy has been developed. By analyzing the Nucleic Acid Database, we have detected certain patterns in the interactions between proteins and nucleic acids. Using this platform, we have implemented a molecular design of new gene carriers, called Synthetic Protein Based on Dendrimers (SPBD), which consists

on a dendrimer-based nanoparticle with its surface conjugated with two or more amino acidic groups, which can act as a customizable gene carrier. These functionalized dendrimers are referred to as synthetic proteins due to their precise and biologically active structure, mirroring natural proteins. By adjusting the type of amino acids, flexibility of the terminal groups and charge distribution, we can modulate the nucleic-acid binding properties of synthetic proteins. Thus, we describe Molecular Dynamics studies to characterize SPBD-oligonucleotide complexes, design new prototypes of SPBD, and improve their affinity for specific therapeutic oligonucleotides.

Conclusion

Here we described bioinformatics and Molecular Dynamics studies to develop a new nucleic-acid carrier. We determined that the incorporation of aminoacids as a functional group of a dendrimer could improve the affinity for nucleic acids. Our results shown that a dendrimer functionalized with an aminoacid such as Arginine, with a high affinity for DNA, as bioinformatics analysis of Nucleic Acid

Database have demonstrated, could also improve the affinity of the dendrimer for DNA.

Social Event

One of the most important aspects in symposiums and conferences is the interaction among the attendants. For that reason the Student Council organizes social events at the end of each of the meeting it offers and the LA-SCS will not be the exception.

Check at the venue for the information of our LA-SCS social event, we will be promoting it along the symposium. Join us to enjoy some drinks and food in a cool and relaxed environment. You would be surprised to know how many collaborations have started with a distended talk with some snacks and beverages! The first rounds of drinks are on us, so don't be late! Join us!

Acknowledgements

The success of this event depends on the commitment of many. We would like to thank everyone involved in the organization this year for their contribution, be it a 15-minute job or months of work. For some efforts we are extraordinarily grateful and they deserve to be mentioned explicitly:

Without the logistical support and invaluable advice of ISCB Executive Director **Diane E. Kovats** the 1st Latin American ISCB Student Council Symposium would not have been possible. We deeply appreciate her continued support of the ISCB Student Council.

We are also greatly indebted to LA-ISCB 2014 conference chair **Dr. Guilherme Oliveira,** for giving us the opportunity to have the 1st Latin American ISCB Student Council Symposium in Belo Horizonte. Further, we would like to acknowledge the support of the ISCB Board of Directors and their trust in our vision. The Student Council would also like to thank our keynote speakers **Prof. Peter F. Stadler, Prof. Vitor Leite** and **Prof. Francisco Melo**. They are volunteering their priceless time to contribute to the success of this Symposium and to promote the next generation of computational biologists.

Furthermore, we would like to thank everyone on the program committee, without them, there would be no Symposium! All of our reviewers did a fantastic job and it's due to them that we stayed within our set deadlines.

We are extremely grateful for the financial support that we received from our sponsors. Without their help organization costs, travel fellowships for students and presentations awards that we offer at the 1st Latin American ISCB Student Council Symposium would not have been possible.

Thank you all!

Sponsors

We thank our sponsors **Nucleic Acids Research** and **SBV IMPROVER (Systems Biology Verification)** for sharing our vision and helping to make the 1st Latin American ISCB Student Council Symposium a success.

Best Oral Presentation and Poster Awards

Outstanding poster and oral presentations of the 1st ISCB Latin American Student Council Symposium will be recognized and awarded with the support of our sponsor **Nucleic Acids Research**. The Best Poster and Best Oral Presentation awards will be chosen through voting by the delegates at the Symposium and by the opinions of a jury of ISCB Student Council Leaders.

Don't miss the awards, to be held during the Concluding Remarks on October 27, (see the schedule table).

Travel fellowships and symposium expenses

The ISCB Student Council has teamed up with **SBV IMPROVER (Systems Biology Verification)** to give 3 students the opportunity to attend the 1st ISCB Latin American Student Council Symposium in Belo Horizonte, Brazil. Also we used funding from this sponsor to cover several expenses from this event.

We thank to SBV **IMPROVER** for believing in our symposium and deciding to support us!

Regional Student Groups Initiative

The ISCB Student Council (SC) has always strived to reach out to Students of Computational Biology and Bioinformatics around the world and promote communication between them to create a vibrant global network of peers. To accomplish this more effectively, in 2006 the SC conceptualized the setting up of Regional Student Groups (RSGs). Regional Student Groups work to fulfill the broad mission of the SC at their regional level by organizing events and initiatives tailored to the requirements of the local student community.

The RSGs initiative has turned out to be an extremely popular and successful initiative. In the past six years, the RSG network has grown to include twenty RSGs from all over the world. Our active RSG network has seen RSGs organize symposia, conduct workshops and contests, initiate discussion groups and even work with each other on trans-national collaborative student projects. As supra-institutional organizations, RSGs are perfectly placed to foster inter-institutional contacts and collaborations in their region and where possible, even serve as a link between students and the local industry. Most RSGs have also formed their own network of members using mailing lists, discussion forums or other means to ensure quick and efficient dissemination of useful information within the community.

The minimal leadership team required to run an RSG are a President and a Secretary working under the guidance of a Faculty Advisor. Since the RSGs are affiliated to the SC membership to an RSG is free. Only the President, Secretary and the Faculty Advisors are required to hold an ISCB membership. Individual RSGs are of course free to put in place a more elaborate administration team if needed. This uncomplicated administrative structure and low operating costs associated with the RSGs has made it feasible for students in many developing countries to begin and develop RSGs in their countries.

As recognition of the importance of the RSGs to the Student Council's overall mission, the RSGs funding program was initiated in July 2010, thanks to funding support by the ISCB. As a part of this program, RSGs are invited to submit proposal for events and initiatives they plan to organize and after a peer review process some of those proposals are selected to be funded by the SC. So far, RSGs have utilized these funds to organize workshops, hackathons, discussion

groups and more. Visit http://iscbsc.org/node/65/rsg-funding for more details about the funding program.

The success of the RSGs initiative is due only to the enthusiasm and commitment shown by the RSG leaders and the support that they have received from faculty advisors and other interested professors. And with these motivated students leading our RSGs, we only expect to see this initiative grow from strength to strength in the coming days.

If you would like to find out more about the RSGs initiative or find out how you too can get involved in this, please visit http://iscbsc.org/content/regional-student-groups or send an email to rsg@iscbsc.org

1st Latin American Student Council Symposium Organizing Committee

SCS-LA Chair

R. Gonzalo Parra (BSc/MSc in Bioinformatics. PhD Student at Protein Physiology Lab. Universidad de Buenos Aires. Argentina)

SCS-LA Co-Chair

Avinash Shanmugam (PhD Candidate. Nesvizhskii lab. Department of Computational Medicine and Bioinformatics. University of Michigan. Ann Arbor. United States)

General Committee
The chairs want to thank all our close collaborators,. They have been working hard for several months in many different objectives and tasks to make this event a reality.

Leonardo Lopez (MSc/BSc in Bioinformatics. Phd student in Engineering. Master student in Computer Science. Research Center for Signals, Systems and Computational Intelligence(sinc(i)). Universidad Nacional del Litoral. Santa Fe. Argentina)

Franco Lucio Simonetti (BSc/MSc in Bioinformatics, PhD Student at Fundacion Instituto Leloir, Universidad de Buenos Aires, Argentina)

Izinara Rosse Cruz (PhD student. CEBio. Centro de Excelência em Bioinformática. FIOCRUZ. Minas LGHM. Laboratório de Genética Humana e Médica. UFMG)

Juliana Assis Geraldo (Master student in Bioinformatics Lab: CEBio – Centro de Excelência em Bioinformática –FIOCRUZ-Minas–Belo Horizonte – Brazil)

Marcia Anahí Hasenahuer (MSc/BSc in Bioinformatics. PhD student at Structural Bioinformatics Group. Universidad Nacional de Quilmes. Buenos Aires. Argentína)

Nina Verstraete (Post-doc at Protein Physiology Laboratory. Department of Biological Chemistry. IQUIBICEN – CONICET. Universidad de Buenos Aires, Argentina.)

Gabriel Olguín Orellana (Bioinformatic Engineering student. Escuela de Ingeniería en Bioinformática. Universidad de Talca. Talca. Chile)

Luis Emilio Fenoy (BSc/MSc in Bioinformatics. Facultad de Ingeniería. Universdidad Nacional de Entre Ríos. Entre Ríos. Argentina)

Vinicius Contessoto. Physics departament, São Paulo State University UNESP, São Jose do Rio Preto, Brazil.

Francislon Silva (CEBio. Centro de Excelência em Bioinformática. FIOCRUZ. Minas LGHM. Laboratório de Genética Humana e Médica. UFMG)

Inti Anabela Pagnuco (MSc/BSc in Bioinformatics. PhD student. Universidad Nacional de Mar del Plata. Buenos Aires. Argentína)

Nicolas Palopoli. (RSG Argentina assessor. Research Fellow, University of Southampton, United Kingdom)

Margherita Francescatto (PhD Student at VU University Medical Center, Amsterdam)

Yesid Cuesta Astroz (PhD student in Bioinformatics. CEBio. Centro de Excelência em Bioinformática. FIOCRUZ. Minas. Belo Horizonte. Brazil)

Dan DeBlasio (PhD Student at Department of Computer Science University of Arizona, USA)

Pieter Meysman (PhD Student at Department of Mathematics and Computer Science University of Antwerp, Belgium)

Anupama Jigisha (Graduate student in Bioinformatics at the University of Geneva and Swiss Institute of Bioinformatics in Geneva, Switzerland)

Disclaimer

The ISCB Student Council has made all efforts to provide accurate information but does not guarantee the correctness of any information provided in this booklet. The ISCB Student Council is a committee of the International Society for Computational Biology (ISCB), which is incorporated as a 501(c)(3) non-profit corporation in the United States.

Copyright