SNP discovery with fuzzy inference through computational model for decision making

Wagner Arbex[1], Marta Fonseca Martins[1], Marcos Vinícius Barbosa da Silva[1], Luís Alfredo Vidal de Carvalho[2]

[1] Empresa Brasileira de Pesquisa Agropecuária
[2] Universidade Federal do Rio de Janeiro

Abstract

SNPs discovery requires bioinformatics tools to be applied to different cases, with the ability to analyze "reads" from different sources and to establish reliable measures. These tools work with different methodologies concerning distinct attributes. However, similar results are expected, even when dealing with a same data set, but it's not unusual to yield different results, which leads to uncertainty in decision making when the results are discordant. This text describes a fuzzy inference decision model applied to assist decision making in cases when information is conflicting and also in the confirmation of coincident information.

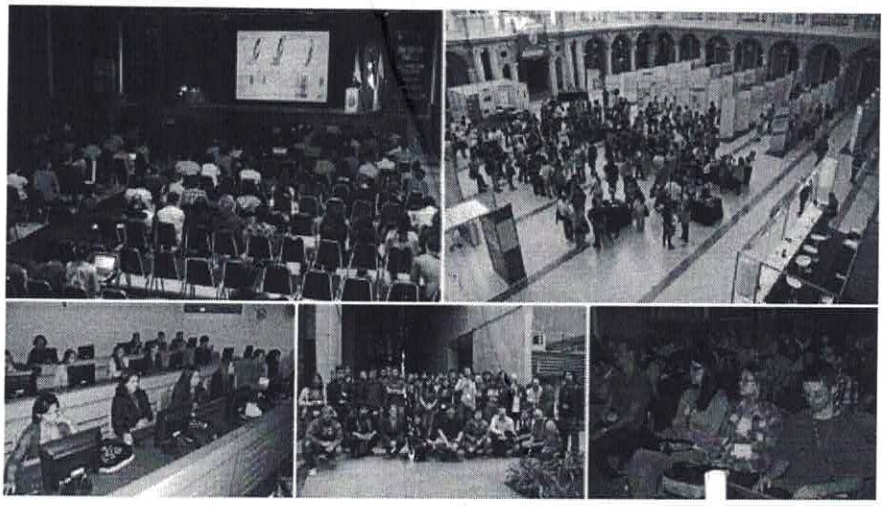
Differences between specific base pairs of different aligned sequences are the most common type of generic variability. Such differences, know as single nucleotide polymorphisms (SNPs), are important in the study of variability of species, because they may cause functional or phenotypic changes, which, by their turn, can result in evolutionary or biochemical effects on the individuals of the species. The use of computational algorithms to SNPs investigation is a widespread practice and the Polyphred and Polybayes programs stand out, because they are widely used. Thus, it is expected these programs show similar results when they are using the same data set, despite used different methods, but it isn't unusual show conflicting results. The PhD thesis "*Computational models to the identification of genomic information associate to the resistance to cattle tick*", propose a fuzzy inference model to aid in the decision process, using the fuzzyMorphic.pl, a computational tool, write in Perl language, which allow the modeling and implementation of fuzzy inference system and this text shows the computational model, inference fuzzy model specifically, proposed and developed in context of quoted thesis, to aid decision support from previous or primary results to SNPs discovery. The model proposed explores these results to aid decision support when the results are divergent or confirm them when they are similar. The computational model is based in a two steps methodological procedure, like several protocols for data mining: the data pre-processor section and the fuzzy inference section. The data pre-processor section provides extraction, integration, selection, completion and deletion procedures, which are the common procedures to pre-processing for data mining and, depending on the characteristics of research, requirements for the processing of data. The second section brings the fuzzy system itself, which includes the fuzzification, inference and defuzzification procedures. The fuzzification processes can be described by membership functions composed by fuzzy sets in a standard format, to the inference process can be used the Mamdani's or Larsen's inference models and the defuzzification process can be represented by an output function with fuzzy sets in standard format and, furthermore, using the "center of maxima" as the defuzzification method, because it takes "multiply shots" on the output

function. The Polyphred's method search for positions in sequences where were detected more than one nucleotide and the Polybayes' method look for polymorphic sites by evaluating the different nucleotides within cross-sections of a multiple alignment. However, both methods do not consider the base quality in the sequence consensus resulting of the alignment. The described model in this text combines this base quality with the previous results obtained from the Polybayes and Polyphred, setting new attributes to SNPs identification.

ISCB Latin America 2012 #iscb_latam
Santiago de Chile

Join the conversation

Welcome to ISCB Latin America 2012 Conference on Bioinformatics



The second International Society for Computational Biology Latin American regional meeting (ISCB-Latin America) took place in March 17th-21st 2012 in Santiago, Chile. More than 250 people attended, primarily from countries in Latin America. As a scientific conference with an international audience, all oral and poster presentations, as well as all printed materials, were in English.

The major aim of ISCB-Latin America 2012 was to deeply motivate and inspire young Latin American students and post-docs to conduct the best research possible in the areas of Bioinformatics and Computational Biology.

The first two days of the meeting (March 17-18) were dedicated to hands-on practical tutorials and workshops covering different topics of interest.

The main conference took place March 19th-21st and featured the following six topic sessions:

Session I. Comparative Genomics and Evolution
Session II. Genomics, Proteomics, Metagenomics and Metabolomics
Session III. Macromolecule Structure/Function Prediction
Session IV. Computer Aided Drug Design and Docking Simulations
Session V. Biomedicine and Immunoinformatics
Session VI. Functional Genomics and Systems Biology

Each session had two keynote speakers and six oral presentations. There were also two poster sessions.

ISCB - Latin America
Santiago, Chile
March 17-21, 2012

An Official Conference of the International Society for Computational Biology

About ISCB-Latin America 2012

The International Society for Computational Biology (ISCB) is holding the ISCB Latin America Conference on Bioinformatics in Santiago, Chile, in March 17-21, 2012. This meeting constitutes the second regional ISCB Latin America meeting, with the first held in Montevideo, Uruguay, in March 2010.

Conferences are key to the development and exchange of new ideas in science. Over the years many thousands of participants have attended ISCB's annual ISMB conferences. As the majority of those attendees do their research in North America and Europe, the ISCB Regional Conferences aim to break the barrier imposed by high cost travel. How? By bringing a high quality conference, including lectures delivered by world-renowned scientists, to the regions of Latin America, Africa and Asia.

Toward this aim, in 2009 ISCB began organizing a newer series of smaller, regionally-based meetings as part of its mission to advance the science through world-wide education and training activities. These regional meetings have included ISCB-Africa (Mali 2009, South Africa 2011), ISCB-Latin America (Uruguay 2010) and ISCB-Asia (Malaysia 2011).

ISCB also aims to provide more students the opportunity to discuss and participate in the latest developments in bioinformatics and computational biology by bringing these meetings closer to home. We hope the ISCB-Latin America 2012 will be the second of many conferences that will contribute to the growth of the field within this region and support scientific innovation across Latin America.

ISCB - Latin America
Santiago, Chile
March 17-21, 2012

An Official Conference of the International Society for Computational Biology

Selected Posters

Remember that the acceptance of your poster presentation was notified by e-mail to the address you wrote when submitting your work.

Poster Presentations

Odd Numbered posters:

Monday, March 19, 5:30 PM to 7:00 PM.

Even Numbered posters:

Tuesday, March 20, 5:30 PM to 7:00 PM.

NOTE: Regardless of the day of presentation, all posters must be on display **both** days.

Poster Display Size: When preparing accepted posters please note that your poster should not exceed the following dimensions: 95 cm (37.4 inches) wide by 150 cm (59.0 inches) high.

List of accepted posters

Please note that each poster was assigned a unique ID (eg. PX, where X is a serial number). Note also that the posters in the list were grouped by topic session.

Comparative Genomics and Evolution
Genomics, Proteomics, Metagenomics and Metabolomics
Macromolecule Structure/Function Prediction
Computer Aided Drug Design and Docking/Molecular Dynamics Simulations
Biomedicine and Immunoinformatics
Functional Genomics and Systems Biology

Topic Session: *Comparative Genomics and Evolution*

P1 » Identification of tranecription regulation associated proteins in plants and stramenopiles

Presenting Author: Diego Mauricio Riaño Pachón (Universidad de los Andes)

Co-authors: Francisco J. Ballrago-Florez (Universidad de los Andes, Biological Sciences); Silvia Restrepo-Restrepo (Universidad de los Andes, Biological Sciences); Bernd Mueller-Roeber (Universität Potsdam, Institute of Biochemistry and Biology)

Short Abstract: The generation of biodiversity is tied to the evolution and re-wiring of gene regulatory networks (GRNs). One component of these GRN are transcription factors and other transcriptional regulators. We have devised a pipeline for the identification of TFs and TRs, exploiting the domain architecture of these proteins. Currently we have a set of rules, representing 136 proteins families, that we have applied to the identification of ~20 different plant species and several species of Stramenopiles, where important plant pathogens are found. Results for plant species are available at http://plntfdb.uniandes.edu.co/; we are now developing a newer interface for Stramenopiles.

[Full Abstract]

P2 » Novel Classes of Eukaryotic Aspartic Proteases and the Identification of their Specificity Determining Residues

Presenting Author: María Revuelta (Universidad Nacional de Mar del Plata)

P48 » SNP discovery with fuzzy inference through computational model for decision making

Presenting Author: Wagner Arbex (Empresa Brasileira de Pesquisa Agropecuária)

Co-authors: Marta Martins (Empresa Brasileira de Pesquisa Agropecuária, Centro Nacional de Pesquisa de Gado de Leite); Marcos Vinicius Silva (Empresa Brasileira de Pesquisa Agropecuária, Centro Nacional de Pesquisa de Gado de Leite); Luis Alfredo Carvalho (Universidade Federal do Rio de Janeiro, Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia)

Short Abstract: SNPs discovery requires bioinformatics tools to be applied to different cases, with the ability to analyze "reads" from different sources and to establish reliable measures. These tools work with different methodologies concerning distinct attributes. However, similar results are expected, even when dealing with a same data set, but it's not unusual to yield different results, which leads to uncertainty in decision making when the results are discordant. This text describes a fuzzy inference decision model applied to assist decision making in cases when information is conflicting and also in the confirmation of coincident information.

[Full Abstract]

Topic Session: *Macromolecule Structure/Function Prediction*

P50 » Modeling proteins using supersecondary structure library and NMR chemical shifts

Presenting Author: Andras Fiser (Albert Einstein College of Medicine)

Co-authors: Vilas Menon (Albert Einstein College of Medicine, Systems and Computational Biology); Joseph Dybas (Albert Einstein College of Medicine, Systems and Computational Biology)

Short Abstract: A new method is presented to generate theoretical protein structures for such remote homology modeling cases where sequence signal is not useful any more. The approach uses NMR chemical shift data and an exhaustive library of protein structure building blocks, Smotifs. The current modeling approach does not require any sequence information for modeling.

[Full Abstract]

P51 » Using correlation data and network decomposition to obtain sub-class determinants in protein families

Presenting Author: Lucas Bleicher (Universidade Federal de Minas Gerais)

Co-authors: Richard Garratt (Universidade de Sao Paulo, Instituto de Fisica de Sao Carlos); Ney Lemke (Universidade Estadual de Sao Paulo, Departamento de Fisica e Biofisica, Botucatu)

Short Abstract: Correlated mutation has been used mostly to detect structural contact pairs in protein families. Based on previous observations that it can also provide functional insights, we provide a framework devoted to this purpose, based on an amino acid specific correlation measure, used to build networks summarizing correlation and anti-correlation patterns in a protein family. Network decomposition results in subsets that can be further assessed by parameters and procedures, proposed for this methodology, having useful applications in protein family analysis. This framework is applied to the family of FeMn superoxide dismutases, highlighting its potential use in protein characterization and gene annotation.

[Full Abstract]

P52 » Bacterial RNAs and host human cells interactions

Presenting Author: Amir Shmaryahu (Fundación Ciencia para la Vida)

Co-authors: Pablo DT Valenzuela (Fundación Ciencia para la Vida, Bioinformatics)

Short Abstract: We have developed a bioinformatic pipeline for possible miRNAs discovery generated from bacterial RNAs that may have potential to regulate gene expression of the host human cell, in case of infection. Given their versatile roles in transcriptional and translational control of gene expression and in quality control of macromolecular products, it is suggested that the study of these predicted miRNAs will yield important clues in the future as to how the host human fine tune cell processes possible human diseases like as cancer in response to changing bacterial environments.

[Full Abstract]

P53 » AFAL: a web service for profiling amino acids surrounding ligands in Protein Data Bank crystallographic data

Presenting Author: Mauricio Arenas (Universidad de Talca)

Co-authors: Samuel Ortega (Centro de Bioinformática y Simulación Molecular, Universidad de Talca,