2nd Argentine Symposium of Young Bioinformatics Researchers (2SAJIB) organized by the ISCB-SC RSG-Argentina

Fiorella Cravero[1], Lionel Uran Landaburu[2], Nicolás N. Moreyra[3], Emilio Fenoy[2], Carla L Padilla Franzotti[4], Esteban Lanzarotti[2], R. Gonzalo Parra[5,*], Nicolas Palopoli[6,*], Alexander Miguel Monzon[6,*]

1. Process Engineering - Chemoinformatics Group, PLAPIQUI (UNS-CONICET), Bahia Blanca, Buenos Aires, Argentina

2. Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martin, San Martin, B1650HMP, Buenos Aires, Argentina.

3. Instituto de Ecología, Genética y Evolución de Buenos Aires (IEGEBA-CONICET), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina.

4. Facultad de Bioquímica, Química y Farmacia, Universidad Nacional de Tucumán, San Miguel de Tucumán, Tucumán, Argentina

5. Quantitative and Computational Biology Group, Max Planck Institute for Biophysical Chemistry, Goettingen, Germany.

6. Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, CONICET. Roque Sáenz Peña 352, Bernal, B1876BXD, Buenos Aires, Argentina.

* Corresponding authors. Emails: parra.gonzalo@gmail.com; nicopalo@gmail.com; monzon.alexander@gmail.com

Abstract

The 2nd Argentine Symposium of Young Bioinformatics Researchers (2SAJIB, according to its acronym in Spanish) took place on November 2017 in Buenos Aires, Argentina. SAJIB is the main annual student-based activity in the field of Bioinformatics and Computational Biology in Argentina and one of the largest in Latin America. Here we summarize the main activities and outcomes from the 2nd SAJIB that gathered together young students and researchers from all over the country and featured recognized Principal Investigators both from the local and international scenes.

Introduction

In 2012, a group of PhD and undergraduate students decided to establish the Regional Student Group of Argentina (RSG-Argentina) as part of the Student Council from the International Society for Computational Biology (ISCB-SC). At that moment RSG-Argentina constituted the only active RSG in Latin America and faced the challenge of reactivating the presence of the Student Council in the region. Ever since, the student community in the fields of bioinformatics and computational biology has steadily grown not only in Argentina but also throughout the Latin American continent. RSG-Brazil was reactivated and RSGs in Chile, Mexico and Colombia were also created. Starting with small events, in which many lessons were learned from the mistakes of a young and inexperienced organization [1], the Latin American branches of the ISCB-SC progressively evolved to be able to organize National Meetings [2], host the biennial Latin American Student Council Symposium [3,4] and have a major role in organizing other international events such as the Student Council Symposium [5,6] and the ISCB Latin American Conference [7]. The RSG-Argentina itself, due to their many activities and its fundamental role for orchestrating the regional efforts has become one of the most active RSGs in the ISCB Student Council. This is evidenced, for example, by its recognition as one of the spotlight groups in 2013 and 2016 [8] and by the central positions at the core organization committees occupied by former RSG-Argentina leaders. These efforts kept on growing and crystallized in 2017 with the organization of the 2SAJIB, whose highlights are summarized in this article.

Highlights from 2nd Argentine Symposium of Young Bioinformatics Researchers

The 2nd Argentine Symposium of Young Bioinformatics Researchers (2SAJIB) was held on 20-21 November 2017 at Universidad Nacional de San Martín (UNSAM) in Buenos Aires, Argentina. The symposium followed the format from previously organized events by RSG-Argentina, consisting of a first day of talks and poster presentations and a second day of practical workshops. We had 2 keynotes speakers and 9 short talks given by students in the first day, plus two workshops offered in parallel the day after. In total, there were 62 attendees from different states of the country **(Figure 1)**.

Keynote speakers presentations

Our first keynote speaker was Dr. Toby Gibson, who is a Group Leader at the European Molecular Biology Laboratory in Heidelberg (EMBL), Germany. He's an expert in sequence analysis and has been leading the field of research in protein-protein interactions mediated by Short Linear Motifs (SLiMs) for the last 15 years. His team develops and maintains the Eukaryotic Linear Motif resource [9], the reference online database of manually-curated SLiMs. Among his major achievements are his role in the early sequencing of the Epstein-Barr by the Sanger method [10] and his contribution to the development of Clustal [11], a breakthrough sequence alignment method and one of the top-ten most cited publications ever [12]. Such a productive and diverse background was reflected on an engaging talk by Dr. Gibson that covered the origins of Bioinformatics, his career path and perspectives on the future of the field. Dr. Gustavo Parisi delivered the afternoon keynote talk. He leads the Structural Bioinformatics Group at Universidad Nacional de Quilmes in Buenos Aires, Argentina and is the current president of the Argentine Association of Bioinformatics and Computational Biology (A2B2C). Dr. Parisi has produced top research on protein structure, function and evolution for the last two decades, and is currently focused on increasing the knowledge and consideration of protein conformational diversity. He delivered his talk, with the provocative title 'How to avoid doing Confusional Biology', as a call of action to encourage a critical view on research topics and practices. His talk was followed by an intense and participative discussion with students.

Student and early-career researchers' presentations

2SAJIB featured 9 student oral presentations along with 11 posters presented by undergraduate students, PhD candidates or postdoctoral researchers. A special talk about teaching bioinformatics in high schools was also presented, extending the reach of our community to intermediate education. A2B2C sponsored the awards to the best talks and posters at the symposium closing ceremony.

In the first talk, titled "Splicing site recognition and regulation", Maximiliano Beckel presented a study about the biological role of splice site sequence variability and their connection with PRMT5 (Protein arginine methyltransferase 5)-mediated regulation. In previous works with *Arabidopsis thaliana*, they proposed the negative effect of mutations on the prmt5 gene on weak

5' splice sites (5ss) recognition. They have now developed a maximum entropy model, based on correlation of up to two sites, that allowed them to established an energy scale related with site strength. From the interaction analysis between the effect of prmt5 mutation and changes in the 5ss strength, they suggest that PRMT5 could help the splicing machinery to recognize weak splice sites.

Ariel Aptekmann presented his work entitled "Natural DNA motif exploration of sequence space". By applying theorems originally developed for protein linear motifs, he generated regular expressions to describe DNA motifs from knowledge databases, which were then inspected for an answer to many fundamental questions such as how many linear motifs can coexist in the same genome and how different do they need to be.

Alejandro Ricci presented "APRANK (Antigenic Peptide/Protein Ranker), a bioinformatic tool for genome-wide prioritization of candidate antigens of human pathogens". APRANK ranks the antigenicity of a protein by using a number of features predicted or measured from protein sequences, and thus helps prioritize candidate antigens (and candidate antigenic peptides) from a given proteome.

The second talk session was started by Julia Marchetti with "Ensembles from ordered and disordered proteins reveal similar structural constraints". Intrinsically disordered protein regions (IDRs) or whole proteins (IDPs) lack a well defined three-dimensional structure under physiological conditions. They found that disordered proteins have 20-30% of their sites evolving under structural constraints, and that these constraints were similar to those observed in ordered proteins. These results contrast with the generally accepted hypothesis that the high evolutionary rate of IDPs is due to the lack of these constraints.

Patricia Araceli Quispe presented her work "Computer-assisted search of inhibitors of the surviving protein with potential antitumor effect and in tumoral cell models". They found a new binding region in the protein, corresponding to a hydrophobic cavity, in which the insertion of a molecule can affect its function through the distortion of the structure. Based on these computational studies, they were able to propose a model of a specific protein-compound interaction.

The last work presented in the session was "Bioinformatic pipeline for Protein-Carbohydrate complexes structure prediction" by Carlos Modenutti, on the prediction of sugar binding modes in proteins. He showed the results from an integrated approach on conformational-space sampling of receptor structures built from a wide range sequence identity templates. They developed and used the Water-Site Biased Docking Method (WSBDM), an Autodock4 docking protocol with a special FF (Force-Field) that showed an improvement over a conventional docking method.

Gaston Rozadilla presented "Metatranscriptome and transcriptome of the digestive tract of *Spodoptera frugiperda*, a plague of agricultural interest". They integrated the gene expression data of intestinal tract *S. frugiperda* with its associated metatranscriptome, to identify genes and microorganisms with potential biopesticides. This study revealed potential candidates for biocontrol of this pest.

Paula Nizovoy in "Cleaning 'Nwanted regions: A novel approach to getting rid of N content by assemblies combination" presented a strategy to reduce assemblies $(N)n$ tracts by merging overlapping scaffolds obtained from different assemblers. Previously undefined regions were resolved in Naganishia and Dioszegia assemblies. The improvements achieved in draft genomes assemblies enable a more confident genomic study of these basidiomycetous yeast species taking full advantage of different assembler's strengths.

To finish this session, Dario Elias presented his work "Study of the microbial communities of Puna lagoons and other environments based on the abundance and characteristics of rhodopsins". The authors aimed to determine if sunlight is of vital importance for the microbial communities of Puna lagoons, based on the comparison between the characteristics and abundances of rhodopsins in the metagenomes of these communities and in those of other environments.

Ana Julia Velez Rueda presented a novel project in Argentina called "Bioinformatics goes to school: A way to rethink the teaching of biology". She coordinated a pilot project by the Structural Bioinformatics Group at Universidad Nacional de Quilmes to promote the usage of bioinformatics as a teaching and learning tool for biology in public and private schools. The project was sponsored by A2B2C and had the participation of 100 high school students (and

their teachers), aged from 16 to 19 years old, 87% of whom had no prior knowledge about programming or bioinformatics. The workshops taught the basics of programming, using the Python language, and proposed different biological questions to be solved using simple programming routines that were run on computers, online terminals or smartphone applications. The final stage of the project was a Bioinformatics programming competition among groups of students representing each school, with eight participants presenting their solutions at the poster session and receiving the recognition and awards from A2B2C.

Workshops

As in previous editions of our symposium, in 2SAJIB we offered two workshops of very interesting and current topics in the bioinformatics community. 17 undergraduate or PhD students from five institutions took part in these workshops, taught by experienced PhD candidates from our community. Computers were provided and refreshments were served on both morning and afternoon sessions.

The advancements in Machine Learning techniques over the last few years have permeated not only most fields in science but society as a whole, and offer the most exciting perspectives of development for the near future. BSc/MSc. Emilio Fenoy led the Machine Learning Workshop, which was a pleasant and enriching experience both for its teacher and students. In a companionship environment, they studied the basics behind training and evaluation of predictors based on different artificial neural networks architectures. For the practical part, a friendly competition was proposed. The students were asked to train their own model looking for the best possible performance while avoiding overfitting, and then analyze the differences between the best and worst models. Throughout the day, we discussed how the students would like to apply what they were learning, sharing ideas and suggestions among all participants.

The means through which bioinformatic tools are made available to scientific community is increasingly becoming an important matter to consider in academia. Even for those not involved in tool development, there is much a web server savvy user can do to exploit web serviced applications, in comparison to the user proscribed to typically low-throughput user interfaces. The Web Servers Workshop led by Dr. Esteban Lanzarotti and BSc/MSc. Lionel Uran Landaburu, was conducted as a full-day, hands-on tutorial on web development and server deployment. Workshop participants learned about client and server side request architecture, plus basics on web interface development (in HTML, Javascript and CSS). They made their own applications using a Python framework to serve PDB structures (using PDB API) and a Perl

model-view-controller framework to assess a simple CRUD drug-target knowledge database. All participants showed at least a minimum programming knowledge, which stimulated a very good environment for learning web server skills without having to address programming limitations.

Conclusions and future directions

The 2nd edition of our annual meeting SAJIB constituted another successful event organized by RSG Argentina. It provides clear evidence of the growth of our student organization over the last years, accompanying the expansion of the local community in computational biology and bioinformatics. SAJIB has established itself as a centralized forum in Buenos Aires, home of most students and directly accessible from other locations, that complements the annual conference organized by the Argentine Society for Bioinformatics and Computational Biology (A2B2C) in different regions of the country. SAJIB provides the best opportunities for engaging new members in the RSG and discovering potential leaders for the near future. New workshops are organized every year and added to a growing list of more than ten workshops that RSG-Argentina is ready to offer at any institute, university or school along the country. We are confident that SAJIB will keep growing, uniting our community and fostering collaborations with other organizations, helping to make Bioinformatics and Computational Biology so widely recognized that even grandmothers will talk about it [13] over dinner!

Grant information

Acknowledgements

About the authors

All authors are part of or collaborate with the RSG Argentina. Fiorella Cravero is the RSG Argentina President. Nicolas N. Moreyra is the RSG Argentina Secretary. Nicolás Palopoli is the

RSG Argentina Faculty Advisor. Alexander Miguel Monzon is the RSG Vice Chair for Latin America. R. Gonzalo Parra is the ISCB Student Council Executive Committee Secretary.

Figure 1: Delegates who attended to the 2SAJIB. Photo credit: Alexander Miguel Monzon.

References

1. Mishra T, Parra RG, Abeel T. The upside of failure: how regional student groups learn from their mistakes. PLoS Comput Biol. 2014;10: e1003768.
2. Parra RG, Defelipe LA, Brenda Guzovsky A, Monzon AM, Cravero F, Mancini E, et al. Highlights of the 1st Argentine Symposium of Young Bioinformatics Researchers (1SAJIB) organized by the ISCB RSG-Argentina [Internet]. PeerJ Preprints; 2016 Oct. Report No.: e2494v1. doi:10.7287/peerj.preprints.2494v1
3. Monzon AM, Hasenahuer MA, Mancini E, Coimbra N, Cravero F, Cáceres-Molina J, et al. Second ISCB Latin American Student Council Symposium (LA-SCS) 2016. F1000Res. 2017;6. doi:10.12688/f1000research.12321.1
4. Parra RG, Simonetti FL, Hasenahuer MA, Olguin-Orellana GJ, Shanmugam AK. Highlights from the 1st ISCB Latin American Student Council Symposium 2014. BMC Bioinformatics. 2015;16: A1.
5. Rahman F, Farmer R, Das S, Vayani F, Hassan M. Highlights of the 1st Student Symposium of the ISCB RSG UK. F1000Res. 2015;4: 154.
6. Rahman F, Wilkins K, Jacobsen A, Junge A, Vicedo E, DeBlasio D, et al. Highlights from the tenth ISCB Student Council Symposium 2014. BMC Bioinformatics. 2015;16 Suppl 2: A1–10.
7. Palopoli N, Monzon AM, Parisi G, Chernomoretz A, Agüero F. A report on the "International Society for Computational Biology - Latin America (ISCB-LA)" Bioinformatics Conference 2016. EMBnet.journal. 2017;23: e883.
8. Shome S, Meysman P, Parra RG, Monzon AM, Palopoli N, White B, et al. ISCB-Student Council Narratives: Strategical development of the ISCB-Regional Student Groups in 2016. F1000Res. 2016;5. doi:10.12688/f1000research.10420.1
9. Gouw M, Michael S, Sámano-Sánchez H, Kumar M, Zeke A, Lang B, et al. The eukaryotic linear motif resource - 2018 update. Nucleic Acids Res. 2017; doi:10.1093/nar/gkx1077
10. Baer R, Bankier AT, Biggin MD, Deininger PL, Farrell PJ, Gibson TJ, et al. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature. 1984;310: 207–211.
11. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22: 4673–4680.
12. Van Noorden R, Maher B, Nuzzo R. The top 100 papers. Nature. 2014;514: 550–553.
13. Bernard V, Michaut M. Explain Bioinformatics to Your Grandmother! PLoS Comput Biol. 2013;9: e1003305.