Universitat Pompeu Fabra Barcelona | Master in Bioinformatics for Health Sciences

Master project 2020-2021

Personal Information

Supervisor	Julio Rozas
Email	jrozas@ub.edi
Institution	Universitat de Barcelona
Website	http://www.ub.edu/molevol/EGB/
Group	Evolutionary Genomics & Bioinformatics

Project

Computational genomics

Project Title:

Comparative and evolutionary analysis of repetitive elements in spider genomes

Keywords:

Comparative genomics; Transposable elements; Repetitive elements; phylogenomics; Adaptive genomics; genome annotation

Summary:

Understanding the origin, amplification and functional role of repetitive sequences in eucaryotic genomes is a central question in Evolutionary Biology. Despite that modern high-throughput sequencing (HTS) technologies are currently accessible for many labs, the accurate identification and annotation of gene family is one of the major challenges in the field. This scenario will change in the near future thanks to the irruption of the so called third-generation sequencing technologies (i.e., long-read sequencing). In this sense, our research group is generating new high quality genomic data from a group of Canary Island endemic spiders (Chelicerata) using long-read sequencing technologies but also chromosome-scale assembly techniques, such as Hi-C and Chicago libraries. The objective of this project is to study the abundance, distribution and evolution of repetitive elements in chelicerates and, by extension, in arthropods, including transposable elements (TEs) and other types of repetitive sequences. A large body of evidence suggest that these elements have structural functional significance. TEs, for instance, can generate variability by movement and insertion, are responsible of defining centromeric regions, or can activate gene expression under stress conditions. Our bioinformatic study in a comparative context will enable understanding of the nature and behavior of this important genomic components. The student will participate in the identification, annotation and/or analysis of repetitive elements in complete genomes of several spiders (and chelicerates) species. For that, he/she will use high quality genome sequences (data generated by our group based on third generation sequencing technologies, and sequences already available in databases), bioinformatics tools (software and scripts to manipulate and visualizte sequences and genomic annotations, to identify repetitive elements, to conduct evolutionary genetics analyses). The basic work-flow will consist in the identification and primary annotation of repeats, the determination of families, types and classes, the estimation of gene turnover rates, or the characterization of the distribution of these repetitive sequences across chromosomes or with respect to other genomic elements, such as protein-coding genes . Many of these analyses will be carried out in our high performance computer cluster.

References:

References from our researhch group • Frías-López, C., Sánchez-Herrero, J. F., Guirao-Rico, S., Mora, E., Arnedo, M. A., Sánchez-Gracia, A. and Rozas, J. 2016.DOMINO: Development of informative molecular markers for phylogenetic and genome-wide

population genetic studies in non-model organisms. Bioinformatics 32: 3753-3759. doi:10.1093/bioinformatics/btw534. • Rendón-Anaya, M. et al. 2019.The Avocado Genome Informs Deep Angiosperm Phylogeny, Highlights Introgressive Hybridization, and Reveals Pathogen-Influenced Gene Space Adaptation. Proc. Natl. Acad. Sci. USA. 116: 17081-17089. doi: 10.1101/654285. • Sánchez-Herrero, J. F., Frías-López, C., Escuer, P., Hinojosa-Alvarez, S., Arnedo, M. A., Sánchez-Gracia, A., Rozas, J. 2019.The draft genome sequence of the spider Dysdera silvatica (Araneae, Dysderidae): A valuable resource for functional and evolutionary genomic studies in chelicerates. GigaScience 8: 1-9. doi: 10.1093/gigascience/giz099. • Vizueta, J., Macías-Hernández, N., Arnedo, M. A., Rozas, J. Sánchez-Gracia, A. 2019.Chance and predictability in evolution: the genomic basis of convergent dietary specializations in an adaptive radiation. Mol. Ecol. 28: 4028-4045. doi: 10.1111/mec.15199. • Vizueta, J., Sánchez-Gracia, A., Rozas, J. 2019.BITACORA: A comprehensive tool for the identification and annotation of gene families in genome assemblies. bioRxiv XX:. doi: 10.1101/593889. • Vizueta, J., Rozas, J., Sánchez-Gracia, A. 2018.Comparative Genomics Reveals Thousands of Novel Chemosensory Genes and Massive Changes in Chemoreceptor Repertories across Chelicerates Genome Biol. Evol. 10: 1221-1236. doi:10.1093/gbe/evy081. Research Group References: (http://www.ub.edu/molevol/julio/SelPublications.html)

Expected skills::

Basic knowledge on NGS data handling and analysis, especially in genome assembly and annotation, notions of comparative genomics and transcriptomics approaches and phylogenetic methods, and experience with Linux operating systems and some of the high level programing languages commonly used in bioinformatics (Perl, Python, R).

Possibility of funding::

To be discussed

Possible continuity with PhD: :

To be discussed