



## Master project 2024-2025

### Personal Information

<b>Supervisor</b>	Julio Rozas
<b>Email</b>	jrozas@ub.edu
<b>Institution</b>	Departament de Genètica, Microbiologia i Estadística (Universitat de Barcelona)
<b>Website</b>	<a href="https://www.ub.edu/molevol/julio/">https://www.ub.edu/molevol/julio/</a>
<b>Group</b>	Evolutionary Genomics and Bioinformatics

### Project

## Computational genomics

### Project Title:

Comparative and evolutionary analysis of multi-gene families in spider genomes

### Keywords:

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

### Summary:

Understanding the origin, amplification and functional role of repetitive sequences in eukaryotic genomes is a central question in Evolutionary Biology. Despite that modern high-throughput sequencing 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 is changing thanks to the irruption of the so called third-generation sequencing technologies (i.e., long-read sequencing). Currently, our research group generates genomic and transcriptomic data (mainly of spiders) to obtain chromosome-level assemblies using long reads DNA sequence data and Omni-C based scaffolding. The objective of this project is to perform a comparative genomic study of 1) the major gene families involved in the chemosensory system (olfactory and gustatory), or 2) those encoding venoms and toxins or 3) the repetitive elements (transposable or other types of repetitive sequences) in chelicerates and, by extension, in arthropods. This research has very relevant biological characteristics with many applications, beyond evolutionary biology. For the analysis, we are using comparative genomics and transcriptomics approaches, under the theoretical framework of molecular evolutionary genetics to identify the genomic regions and gene functions driving diversification. We applied powerful bioinformatics tools, some of them generated in our research group, to perform the comparative genomic analyses (in coding and non-coding regions, in gene copy number, and in gene expression levels). The student will participate in the assembly, annotation and analysis in several spiders (and chelicerates) species. For that, he/she will use high quality genome/transcriptome 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 visualize 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 across other genomic elements, such as protein-coding genes. Many of these analyses will be carried out in our high performance computer cluster.

### References:

- Escuer, P., Pisarenco, V. A., Fernández-Ruiz, A. A., Vizueta, A. J., Sánchez-Herrero, J. F., Arnedo, M. A., Sánchez-Gracia, A., Rozas, J. 2022. The chromosome-scale assembly of the Canary Islands endemic spider *Dysdera silvatica* (Arachnida, Araneae) sheds light on the origin and genome structure of chemoreceptor gene families in chelicerates. *Mol. Ecol. Res.* 22: 375-390. doi: 10.1111/1755-0998.13471.
- Frias-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.
- Rispe, C. et al. 2020. The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC Biol.* 18: 90. doi: 10.1186/s12915-020-00820-5.
- 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., Escuer, P., Frías-López, C., Guirao-Rico, S., Hering, L., Mayer, G., Rozas, J., Sánchez-Gracia, A. 2020. Evolutionary history of major chemosensory gene families across Panarthropoda. *Mol. Biol. Evol.* 37: 3601-3615. doi: 10.1093/molbev/msaa197. • Vizueta, J., Sánchez-Gracia, A., Rozas, J. 2020. BITACORA: A comprehensive tool for the identification and annotation of gene families in genome assemblies. *Mol. Ecol. Res.* 20:1445-1452. doi: 10.1111/1755-0998.13202. • Vizueta, J., Rozas, J., Sánchez-Gracia, A. 2018. Comparative Genomics Reveals Thousands of Novel Chemosensory Genes and Massive Changes in Chemoreceptor Repertoires across Chelicerates Genome *Biol. Evol.* 10: 1221-1236. doi:10.1093/gbe/evy081. Research Group References: (<https://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 programming languages commonly used in bioinformatics (e.g., Python, R).

**Possibility of funding:**

To be discussed

**Possible continuity with PhD:**

To be discussed