Research proposal

Jesús A. Ballesteros

Genomic sciences have brought many of the hitherto understudied organism into the limelight. Technological advances and lower sequencing costs are now providing the basis for more powerful comparative evolutionary and functional genomics. In the case of arthropods, the continuing production of reference genomes is if finally enabling the study of the genetic machinery and its evolution beyond traditional model systems such as the fruit-fly *Drosophila melanogaster*. My research goals include the continue development of bioinformatic tools and methods to study the evolution of arthropod genomes and the empirical test of these methods in understanding the evolution of chelicerate diversity. My research leverages on emerging genomic resources in arthropods and arachnids in particular. In addition to the computational component, my research plans involve empirical study of evolutionary dynamics based of genomic data, combined with functional assays to unravel the elusive relation between the genotype and phenotype in one of the most diverse groups of animals.

The research I have planed for the upcoming years (3-5 years) is structured around three major projects that are currently partially funded or are competing for internationally funded resources.

**Gene homology, the age of gene duplicates and phylostratigraphy**

Homology is one of the cornerstone concepts of comparative and evolutionary biology. For genomic disciplines, the process of gene duplication further complicate homology assessment by the necessity of distinguishing between orthologous genes, originated trough speciation, and paralogous genes derived from gene duplications (Fitch, 1970). My work on the development of bioinformatic tools to assess orthology (UPhO Ballesteros and Hormiga, 2016) allows the incorporation of gene genealogy into the orthology assessment, instead of relying on overall similarity approaches (such as reciprocal BLAST hits).

While the relevance of filtering orthologous genes for its use in phylogenetics is well documented, this distinction is of equal relevance for the study of gene identity and function. After duplication, additional gene copies follow one of three alternative paths; (1) they disappear via mutations, becoming pseudogenes and ultimately removed, (2) they maintain the and dived the function of the ancestral copy or (3) one of the copies is released of selective pressures and allowed to attain new functions (Lynch and Conery, 2000). Gene duplication is also a major source of genetic novelty a promoter of phenotypic diversification. A prevailing conjecture in functional genomics is that orthologues in different species are more likely to perform similar functions than paralogous copies. This hypothesis however, still lacks a rigorous test.

It is generally accepted that new genes are produced by duplication events, and at its most extreme the duplication involved the whole genome. Whole genome duplications (WGD) are considered rare events in animal evolution. Whole genome duplications have been demonstrated more notably in the vertebrates, where two rounds of WGD in the ancestor leading to crown vertebrates are though to be major driver for the diversification of this group of organisms (Dehal and Boore, 2005).
Figure 1: Phylostratographic profiles of developmental transcriptomes uniquely found in different appendages of the bark scorpion Centruroides sculpturatus.

In the case of arthropods, and thanks to recently sequenced genomes, WGD have been documented in spiders, scorpions and horseshoe crabs (Schwager et al., 2017; Sharma et al., 2014). Recent studies based on transcriptomes also suggest that these genes expansion, many of genomic scale may be more common than currently thought (Li et al., 2018). One of the most intriguing aspects relates to the study of macroevolutionary consequences of WGD events. In the case of chelicerates, the main question centers in the relation of this genome wide changes with the origin of iconic adaptations, such as terrestrialization, the diversity of venom components, silk and the effects of these in promoting and sustaining the observed diversity.

Complex phenotypes, often arise from equally complex coordination of several genes during development. Therefore, even when armed with genomes and expression data, it is hard to establish de novo the genetic mechanism of specific phenotypes. Of particular interest is the role of new genes in the origin of new structures or if novel structures arise through co-option of older genes to attain new functions. The first step towards unraveling this mystery is to establish the relative age of origin of genes.

I am currently developing a phylostratigraphy approach to elucidate the relative origin of “genes” across the chelicerate tree of life. The idea, analogous to geological stratigraphy, postulates the origin of a gene (orthologous, or homologous) to the most recent common ancestor of the species that share a given gene. This method was first described in 2007 (Domazet-Lošo et al., 2007) and has been applied to study phenotypic novelty in a variety of groups, for example in mollusks (Hilgers et al., 2018). A main hurdle in the broader use of this method is the lack of standard bioinformatic tools able to handle diverse datasets.

Studying this phenomenon requires whole sequenced genomes as the starting point and only recently we have achieved the dense sampling of chelicerate genomes that will allow this investigations.

To bridge this gap I have recently developed a new pipeline that facilitates these phylostratigraphic analyses in a flexible a reproducible manner. I am currently testing its utility in genomic datasets of the scorpion Centruroides sculpturatus and the spider Parasteatoda tepidariorum. In the case of the scorpion, first I identified uniquely expressed transcripts from diverse appendages dissected from scorpion embryos during morphogenesis. This produced a list of transcripts that are commonly expressed across all tissue types and those that are uniquely expressed in the development. By comparing the
phylostratigraphic profile of these tissues specific gene collections we can identify lineage specific genes that can help us understand the what genes are involved in the development of these structure and the relative contribution of old and new genes in the development of different structures. In the case of the spider, comparisons will be made between developmental stages, which are well documented in this model organism, instead of structures. We expect the phylostratigraphic profiles to illuminate the origin of different genes cascades through the developmental series.

Although, this research is only in its first stages, preliminary results are very promising and show distinct phylostratigraphic profiles between genes of specific structures (with higher proportion of younger genes), and the genes those found across all tissues (with higher proportion of genes of older phylogenetic origin).

My plans to continue the development of this pipeline and its application to empirical dataset includes:

1. Introducing models of gene loss to output numerical estimates (likelihood) of observing the presence/absence of gene copies in members of the stratum.
2. Identifying orphan genes, genes without homologues or orthologues, and ghost genes or genes not found in all members of the phylostratum.
3. Direct comparison of strata derived from homologues and orthologues.

**Evolution of blindness in cave Arachnids of Israel**

Despite the antiquity of eastern Mediterranean settlement and the legacy of natural history surveys in this region, there is much room for biodiversity discovery and exploration, with cave systems representing one of the most promising habitats for new discoveries. The region of the Levant, today’s Israel and Palestine territories, is scattered with a multitude of caverns that harbor a unique fauna composition. Cave dwelling organisms are one of the most fascinating examples of the prowess of evolutionary adaptation. Such organisms vary in their degree of association to the cave environment. Some are frequent but not obligate cave visitors (troglophiles) recorded inside and outside of the caves; others are found exclusively in the darkest depths of the caverns (troglobites) and those are the ones showing the most complete and dramatic morphological cave adaptations (Barr Jr and Holsinger, 1985).

Several animal groups have repeatedly and independently colonized cave environments. Due to natural selection, these groups have converged on traits better adapted for life in the depths. One of the most dramatic examples is loss of eyes. Surprisingly, as clearly exemplified by the eyeless cave fish Astyanax mexicanus (De Filippi, 1853), acquiring these dramatic adaptations does not imply genetic break up with the parent surface-dwelling (epigean) eyed population 2–4. In the case of the many arthropods, the assumption of genetic break up and speciation is far less documented.

Reduction or complete loss of eyes is a common trait of animals adapted to the cave environments. Recent investigations of eye loss in the cave fish Astyanax mexicanus and the isopod Asellus aquaticus have elucidated the genetics of eye loss in cave dwelling populations of these species (Protas et al., 2011; Hinaux et al., 2013). Nevertheless, the genetic and developmental mechanisms leading to this phenotype are poorly understood for arachnids. This knowledge gap is attributable to two different considerations. Firstly, arachnids have two types of eyes (median and lateral eyes) which differ in both morphology and function. The homology of the two arachnid eye types to the faceted eyes of the isopod A. aquaticus is unclear, to say nothing of homology to the eyes of the vertebrate A. mexicanus. Second, with respect to developmental genetics, little is known about how spider eyes are patterned
Figure 2: A. Geographic location of caves sampled during the 2018 expedition to Israel funded by the National Geographic Society. B. Close up of the eye region of the funnel web spider *Tegenaria pagana*. C. Close up of the eye region of an eyeless species *Tegenaria*. D. Habitus of the blind whipspider *Charinus ioanniticus*.

While gene expression surveys have been conducted for the homologs of eye-patterning genes identified in the fruit fly, no functional work accompanies these data. On top of that, the presence of gene duplicates in the spiders further complicate our understanding of the genetic mechanism of eye morphogenesis in arachnids.

Understanding of how blindness has evolved in cave arachnids also requires investigating the evolutionary dynamics between putatively isolated cave populations and their closest epigean relatives. The most common assumption is that cave dwelling populations represent independent (genetically isolated) lineages from surface dwelling counterparts. However, as exemplified by the *A. mexicanus* case, acquiring these dramatic adaptations does not require genetic breakup with the parent surface-dwelling (epigean) eye-bearing population. Thus, this hypothesis must be tested by assessing the evidence for historical gene flow, via a modern, genome-scale approach to population genetics.

The region of the Levant, today’s Israel and Palestine territories, is scattered with a multitude of caverns that harbor a unique fauna. As part of a collaboration with Hebrew University, we are describing several species of blind, cave-dwelling arachnids that will constitute the study systems for my proposed work. To facilitate fieldwork for this investigation, I successfully applied for funding from the National Geographic Society for a collecting campaign of several weeks. During the first six-week
expedition that I led in Israel, we explored 41 caves (92 collection events) and secured specimens from our target species complexes. The differences in population size, geographic structure and habitat preferences of different arachnid species represent an ideal opportunity to explore complex evolutionary dynamics of independent cases of eye reduction/loss in a similar geographic setting.

My study system consists of four cases of putative sister species, with one species bearing eyes and the other lacking them, as follows:

**Case 1:** Funnel web spiders (Araneae: Agelenidae: Tegenaria sp.). These spiders (Figures 2B, 2C) were the most widespread and abundant in our collection. One of the most exciting aspects of the spider system is that within the same cave, there are eye-bearing and blind specimens. Hundreds of specimens from dozens of populations have already been collected and are available for sequencing.

**Case 2:** Woodlouse hunter spiders (Araneae: Dysderidae: Gen. spp.). These wandering spiders are not confined to their webs. We found blind dysderids in 12 cave sites.

**Case 3:** Whip-spiders (order Amblypygi). The widespread species in the region is *Charinus ioannticus*, which is known to occur in caves and in man-made underground cavities from Turkey to Israel. A second species (*Charinus israelensis*) was described for two geographically distant caves in Israel. These two whip spiders are very similar morphologically but with a key difference: *C. israelensis* shows extreme reduction of its eyes (Figure 2D).

**Case 4:** The harvestman (Opiliones: Phalangodidae). The genus *Haasus* has a single eye-bearing species (*Haasus judaeus*) known from several Israel localities. A new eyeless species that we are describing was recently found in the deep chambers of a single cave in the Judean desert, isolated from the known epigean populations.

I am implementing a RAD-Seq approach to characterize the population genetics of these species, assess the evidence for gene flow between the two phenotypes, and generate estimates for the duration of reproductive isolation. These four cases represent natural replicates of cave invasion and provide a unique opportunity to study the genetics associated to cave adaptations. The comparative framework of this study, encompassing four divergent and understudied arachnid groups, will provide a unique setting for (a) identifying the common evolutionary processes leading to the reduction or loss of eyes in caves, (b) isolating the genetic mechanisms of ecological adaptation, and (c) assessing the conservation status of these endemic lineages.

To understand whether the mechanism underlying the evolution of blindness is the same in all four species, I procured embryos in different developmental stages of funnel-web spiders and whip spiders (Figure A) that were either blind or bore eyes. I generated RNA-Seq data for this material to be used for differential gene expression analyses. I will then compare which genes are being differentially expressed in surface and cave dwelling arachnid embryos, across species pairs, to infer whether blindness always evolves through a common mechanism.

To ground the patterns we observe in natural populations with the investigative arsenal of a model organism, I will leverage the findings of the differential expression analyses and develop a functional assay, using the model spider system *Parasteatoda tepidariorum*. A reference genome is available for this species, from which homologs of the differentially expressed candidates will be identified. I will use parental RNAi to induce gene knockdowns and assess how the four pairs of *P. tepidariorum*’s eyes are affected by these manipulations. These data will constitute the first functional data unraveling how spider eyes are formed and provide strong experimental evidence for the genetic mechanism leading to eye loss in these cave dwelling arachnids.

A natural extension of this project invites the exploration and contrast of similar natural systems found in Mexico, that is known to harbor a diverse fauna of troglobitic arachnids and also harbors a
representatives of the same some taxa I am studying in Israel.

**Evo-devo, differential gene expression and CRISPR**

Negative results from RNAi are difficult to interpret (i.e., the absence of a gene function cannot be disentangled from poor penetrance or efficiency of RNAi) and may confound the outcomes of the genetic screen I plan to pursue. One additional project that I participate in is the effort to achieve efficient transgenesis in *P. tepidariorum*. Due to technical limitations, single-cell injections are not feasible in this spider, and thus transgenesis has eluded the grasp of the spider community. As a workaround, the Sharma Laboratory plans to modify an approach recently implemented in the mosquito with high efficiency, wherein a yolk precursor protein is fused to Cas9 and injected into an adult female (Chaverra-Rodriguez et al., 2018). This mediates higher transfer of the Cas9 ribonucleoprotein complex into developing oocytes; in the mosquito, this procedure increased the success of transgenesis by more than 100-fold. I am trialing this procedure with three homologs of yolk precursor proteins identified in the *P. tepidariorum* genome, in order to complement my functional genetic work on arachnid eye development. As a proof of concepts we aim to generate trasngenic spider with an easy phenotype to track by green incorporating green fluorescent protein to silk genes. Once the system is tested, the technique can be easily adapted to produce functional assays with virtually any gene knockouts.

Evolutionary biology is a changing field. Nowadays the acquisition of large scale genetic data is not longer a limitation. The pressing challenges in the field are to develop the methods and tools needed to sieve gargantuan piles of raw data and, to fill up the gaps in exploring and discovering the prodigious amount of life forms from which we ignore the most basic facts. My research program aims to contribute to these needs which will ultimately improve our understanding of the evolutionary process affecting of living forms, leading to the organization of their genome, their phenotype and the macro-evolutionary consequences.

**Research infrastructure**

The emerging results of these projects are being used to support our applications for two federally funded research grants that will allow me and my colleagues to expand our sequencing effort, propel our international collaborative effort with Hebrew University, and advance the capabilities of *P. tepidariorum* as a model system.

Finally, although centered on chelicerates, the products of my research program are not bound to any taxonomic group in particular. The universality of the genomic phenomena I study and the methods to investigate them are fertile fields of opportunity for establishing collaborative efforts in diverse organisms. My research plan will also support and foster the research interests of highly motivated undergraduate and graduate students promoting intra and inter institutional collaborations and the formation of new generation of scientist.

The projects outlined in this research proposal require modest computational and laboratory resources. The majority of the genomic resources (reference genomes) are already publicly available or in advance stage of production. The new genomic data I would need to produce, require simple molecular laboratory facilities for DNA and RNA extraction and quantification, as well as access to thermocyclers and reagents for DNA amplification and storage of tissues and genomic samples. More complex laboratory procedures, such as library preparation and high troughput sequencing will be outsourced as required.
Design and selection of gRNA sequences for CRISPR-Cas9 applications require the selection orthologous sequences in the organisms of interest. A, Gene family trees are built to identify target sequences, using orthology inference under the algorithm I designed and authored (UPhO; Ballesteros and Hormiga, 2016). B, A set of candidate gRNA are designed using dedicated bioinformatic pipeline and candidate sequences mapped back to the reference. C, These techniques will allow us to study the genetic mechanism in emergent model species such as the house spider Parasteatoda tepidariorum.

Access to computing resources are also not a limitation for my research. Computation intensive tasks can be accessed via UNAM’s supercomputing facilities (such as Mitzli or KanBalam), public high-performance clusters our through my network of collaborators.

I believe my research proposal complements the strengths of CCG and integrates aspects of computational biology, functional and evolutionary genomics. In line with the goals of the CGG I hope to contribute to the advancement genomic sciences and resources of Mexico in an integrative and collaborative framework.

References


Filippo De Filippi. *Nouvelles espèces de poissons*. 1853.


