

Orthology & phylostratigraphy unravel the contribution of new genes in the development of chelicerates



WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

*

Jesús A. Ballesteros* & Prashant P. Sharma

✉ ballesterosc@wisc.edu

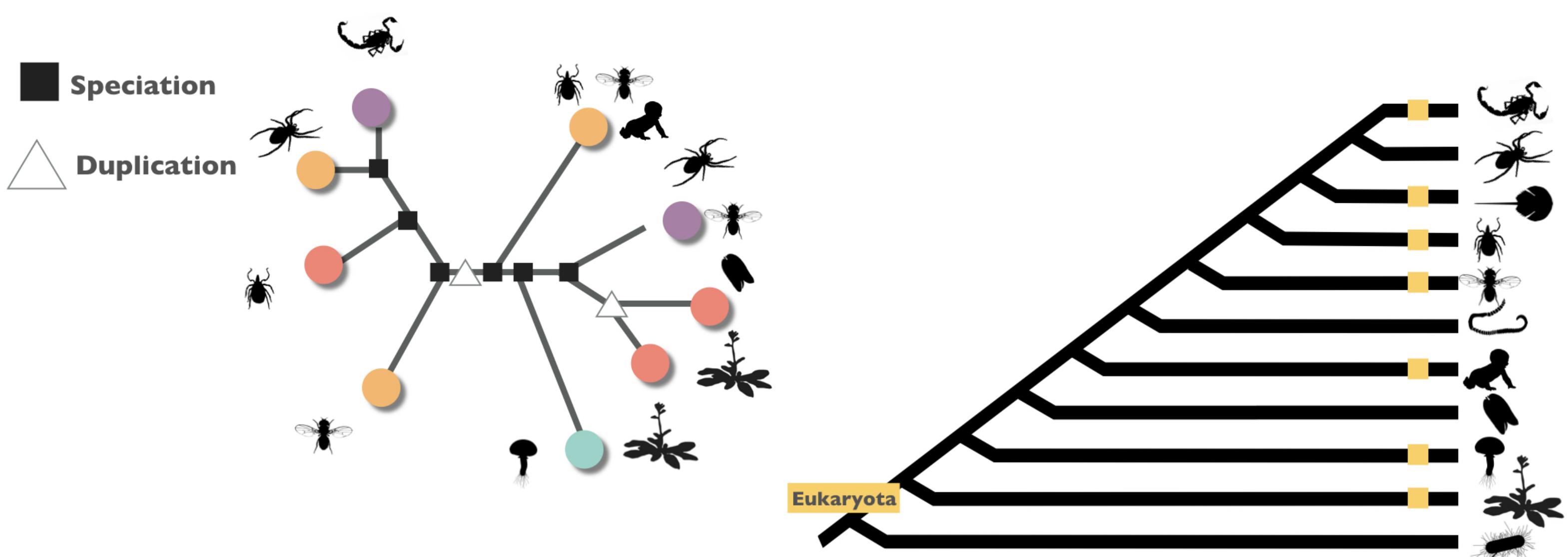
🌐 <https://github.com/ballesterus>



Introduction

Understanding the evolutionary origin of genes is of particular interest for discerning the relative contribution of "new" and "old" genes underpinning the development of morphological innovations. Phylostratigraphy is a comparative approach aimed to elucidate the evolutionary origin (relative age) of genes of interest [1]. In its original formulation, this method searches for similar sequences (BLAST) of genes of interest to reference genomic data from organisms spanning the whole tree of life. The minimum age of the gene of interest is inferred to equal the age of the most recent common ancestor of the focal species and the most distant relative that exhibits a significant hit to the queried gene. Unfortunately, there is no standard implementation for performing phylostratigraphic analyses and the use of similarity as the homology criterion can obscure the role of gene duplications in producing new genes. This distinction may be more relevant in lineages with widespread gene duplicates consistent with whole genome duplication events, as is the case for some chelicerates [2]. Here, we combined an evolutionary orthology criterion (UPhO [3]) with a custom implementation of phylostratigraphy to investigate patterns of gene age and gene expression level at (a) different developmental stages and (b) different appendage types in two chelicerate species.

Distinguishing orthologous and homologous relationships affect phylostratigraphy



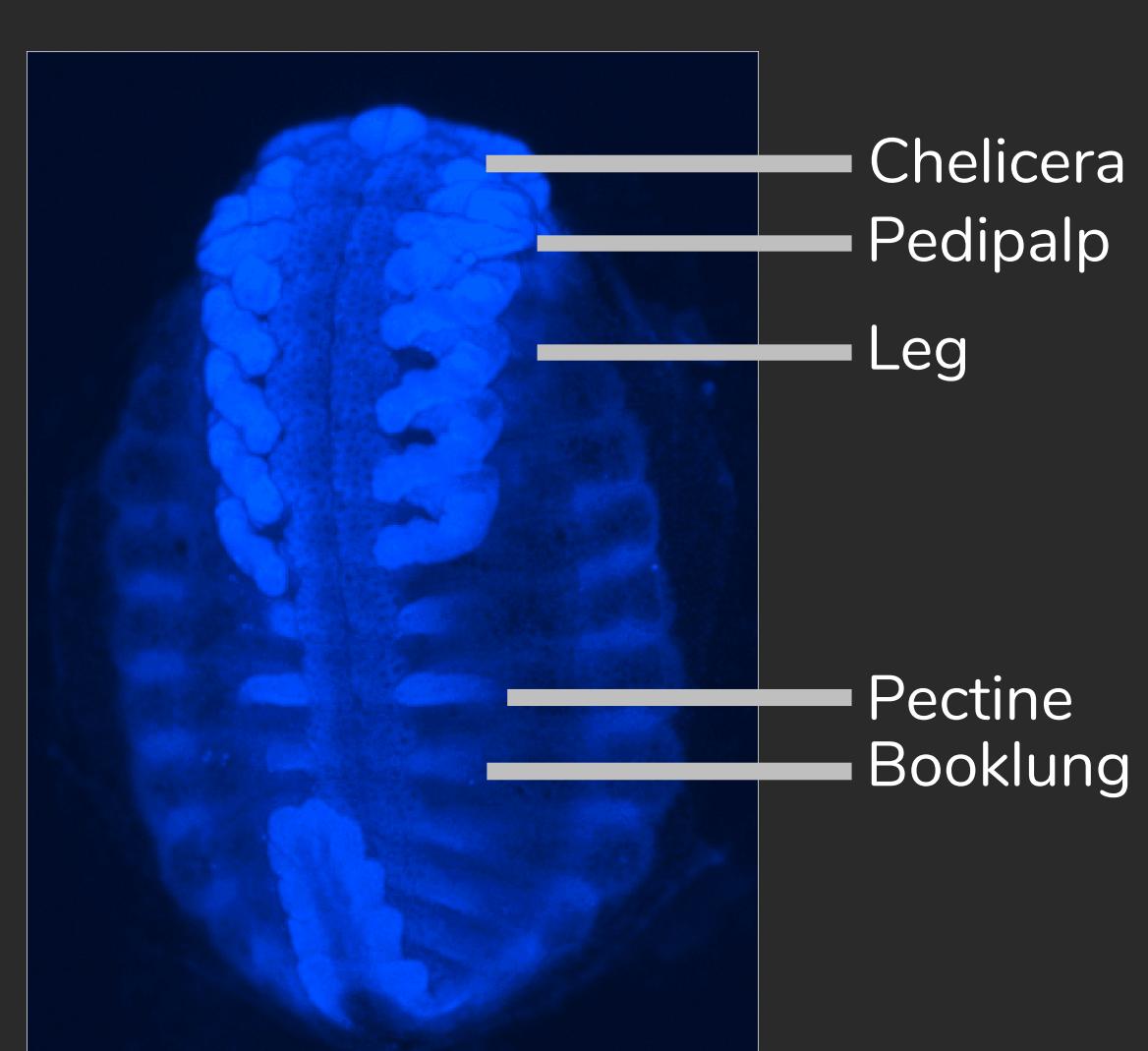
Images from phylostratigraphy.org

Pipeline overview

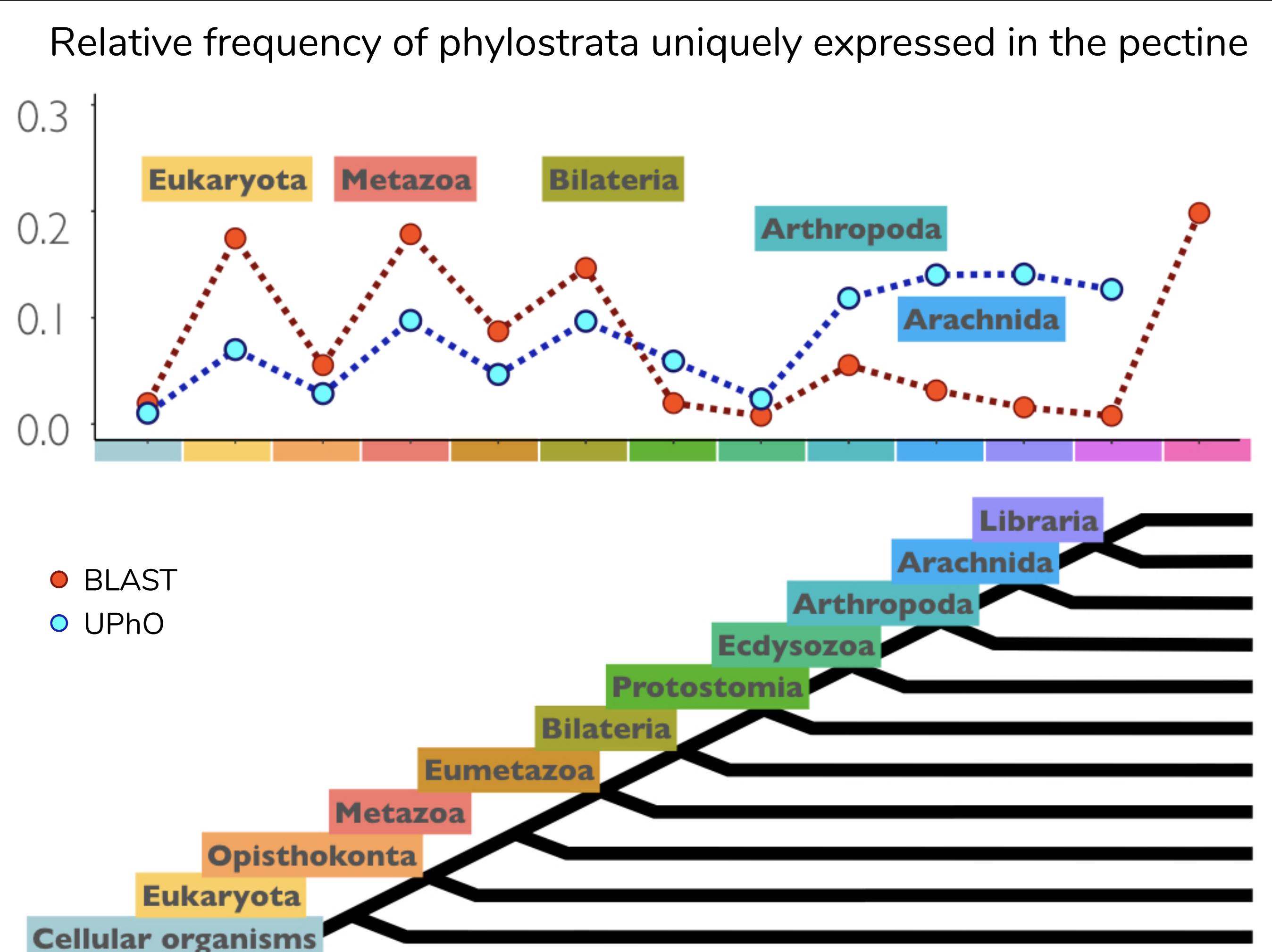
- Protein coding regions of 27 reference genomes spanning the Tree of Life were combined with the focal chelicerate species, the spider *Parasteatoda tepidariorum* and the scorpion *Centruroides sculpturatus*.
- Initial similarity of transcripts of interest used BLAST, producing clusters of "homologous" genes. A phylogenetic criterion (UPhO) was separately used to identify orthologous clusters.
- The strata of similar (BLAST) or orthologous (UPhO) clusters of sequences including the gene of interest was identified with a custom python script.

Case 1: Appendage formation in scorpions

Arizona bark scorpion (*Centruroides sculpturatus*)

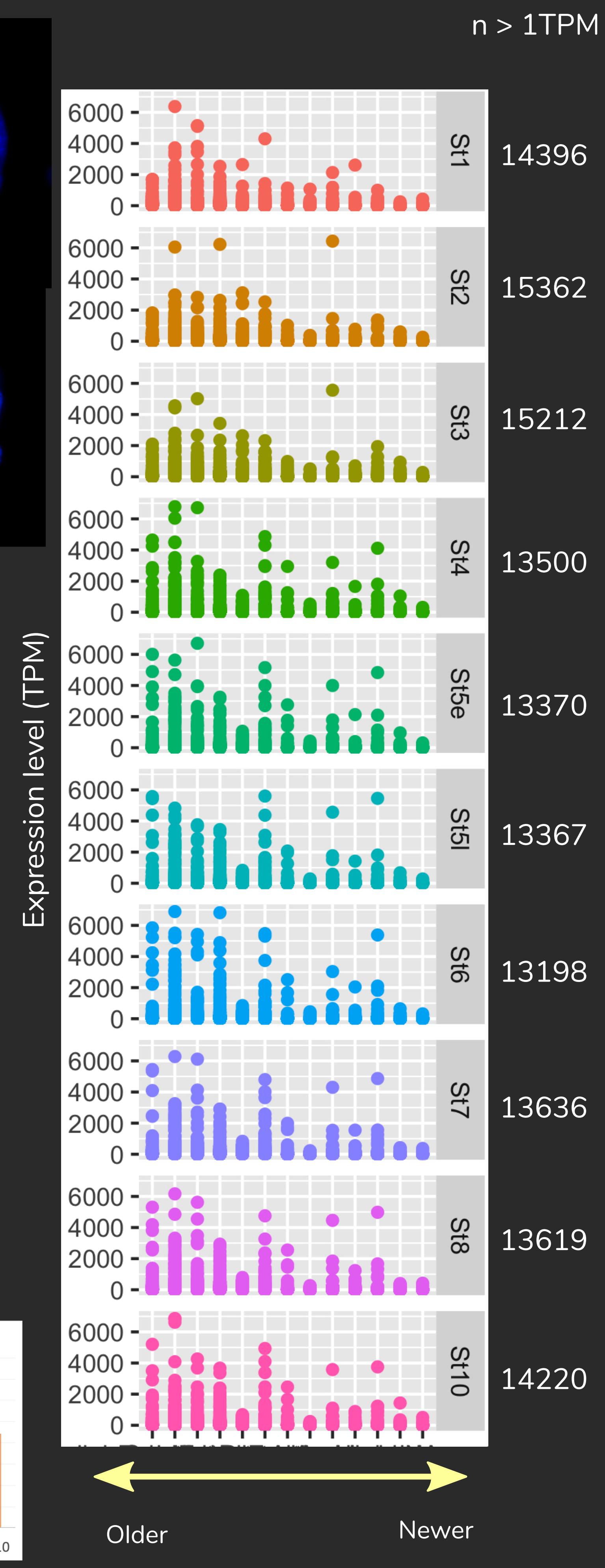
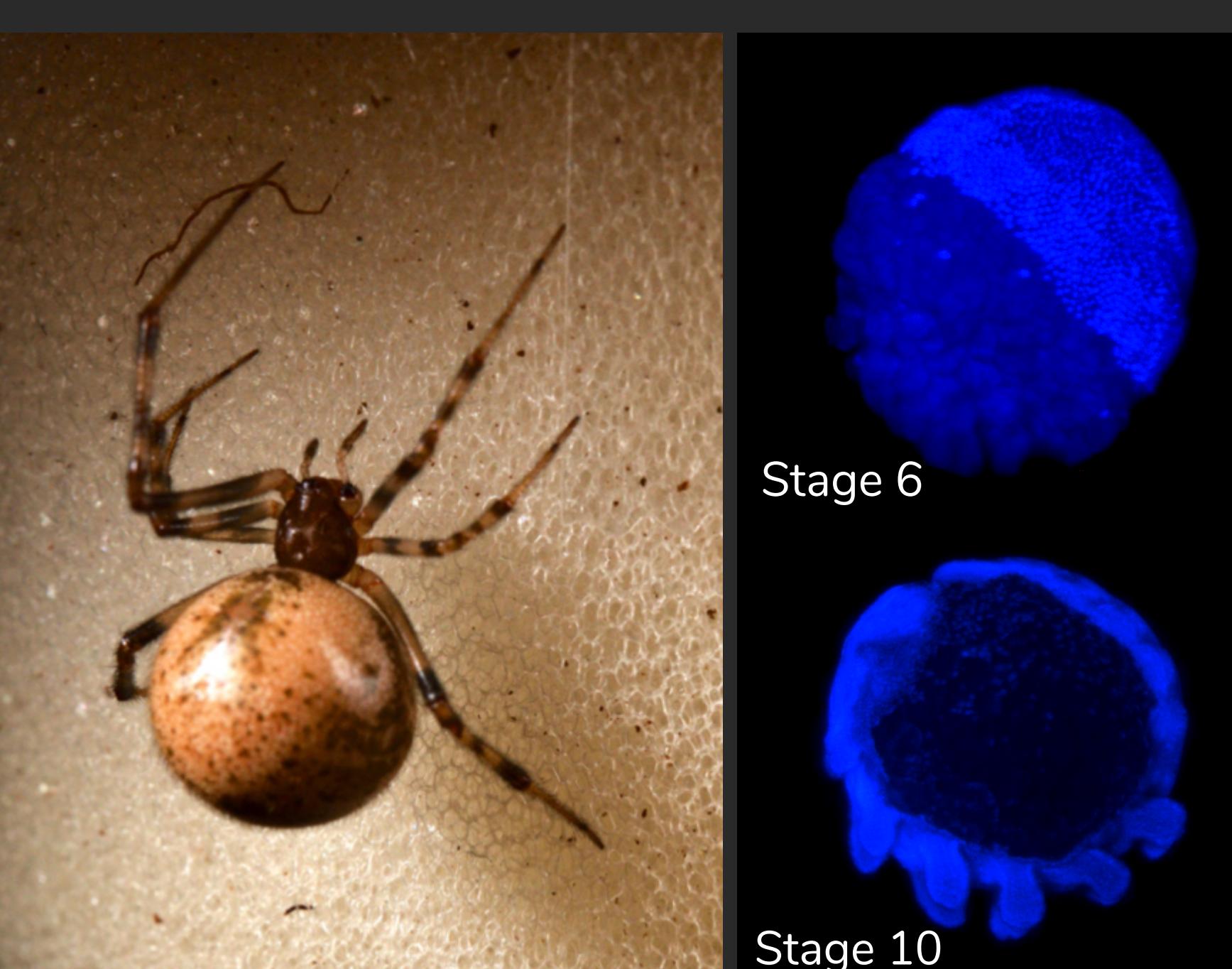


We investigated the evolutionary origin of the genes expressed in the limb bud stage of the Arizona bark scorpion. Individual appendage primordia were dissected and pooled for RNA extraction and sequencing. After quantification [4], we identified uniquely expressed genes to each appendage as well as the set of genes universally expressed in all tissues. The relative frequency of phylostrata for the genes expressed in each appendage was variable but revealed a combination genes from older and younger strata. Accounting for orthology revealed a tendency of BLAST-based searches to yield older phylostrata with respect to UPhO-based searches.

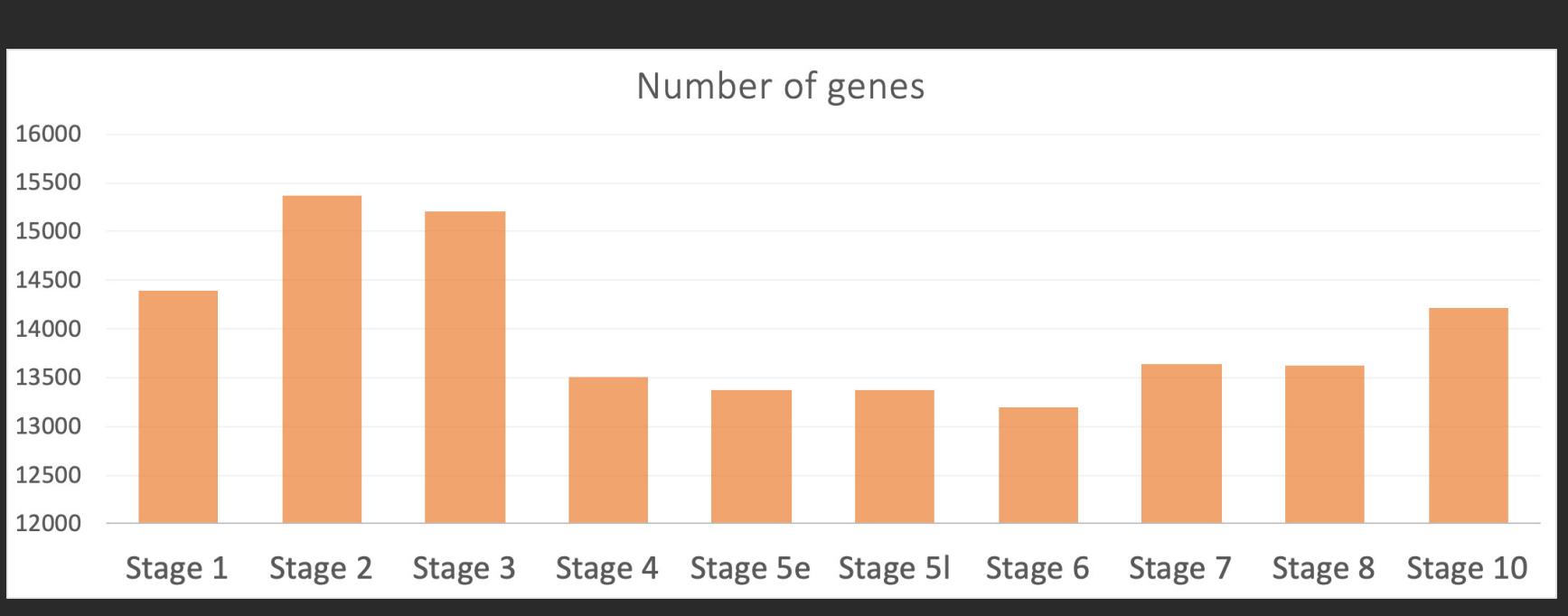


Case 2: Developmental stages in spiders

House spider (*Parasteatoda tepidariorum*)



We inspected the gene expression profile during the development of the common house spider *Parasteatoda tepidariorum* [5]. This dataset consists of RNA sequencing experiments from whole embryos in the first 10 developmental stages, spanning early cleavages to brain and limb differentiation. We estimated the phylostratum for every gene in the genome of this spider; these were later combined with measurement of transcript abundance. The expression levels of each gene combined with their relative ages, reveals that many genes expressed later during development are relatively old (contrary to von Baer's law), with gene origins as deep as the cellular organisms or Eukaryota strata.



References:

- [1] Domazet-Lošo, T., Brajković, J. and Tautz, D., 2007. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends in Genetics*, 23(11), pp.533-539.
- [2] Schwager, E.E., Sharma, P.P., Clarke, T., Leite, D.J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T. and Buffry, A.D., 2017. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC biology*, 15(1), p.62.
- [3] Ballesteros, J.A. and Hormiga, G., 2016. A new orthology assessment method for phylogenomic data: unrooted phylogenetic orthology. *Molecular biology and evolution*, 33(8), pp.2117-2134.
- [4] Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods*, 14(4), p.417.
- [5] Iwasaki-Yokozawa, S., Akiyama-Oda, Y. and Oda, H., 2018. Genome-scale embryonic developmental profile of gene expression in the common house spider *Parasteatoda tepidariorum*. *Data in Brief*, 19, pp.865-867.



HT CENTER FOR
HIGH THROUGHPUT
COMPUTING

