

Biodiversity Bioinformatics – Case study

# Moulting genes across arthropod diversity



SIB course 31-08-2023



Giulia Campli, PhD student  
Prof. Robert Waterhouse group  
University of Lausanne, Department of Ecology and Evolution

# The question: why moulting?

Arthropods represent the richest and most diverse Metazoan phylum (reported close to 1 million species); it comprehends:

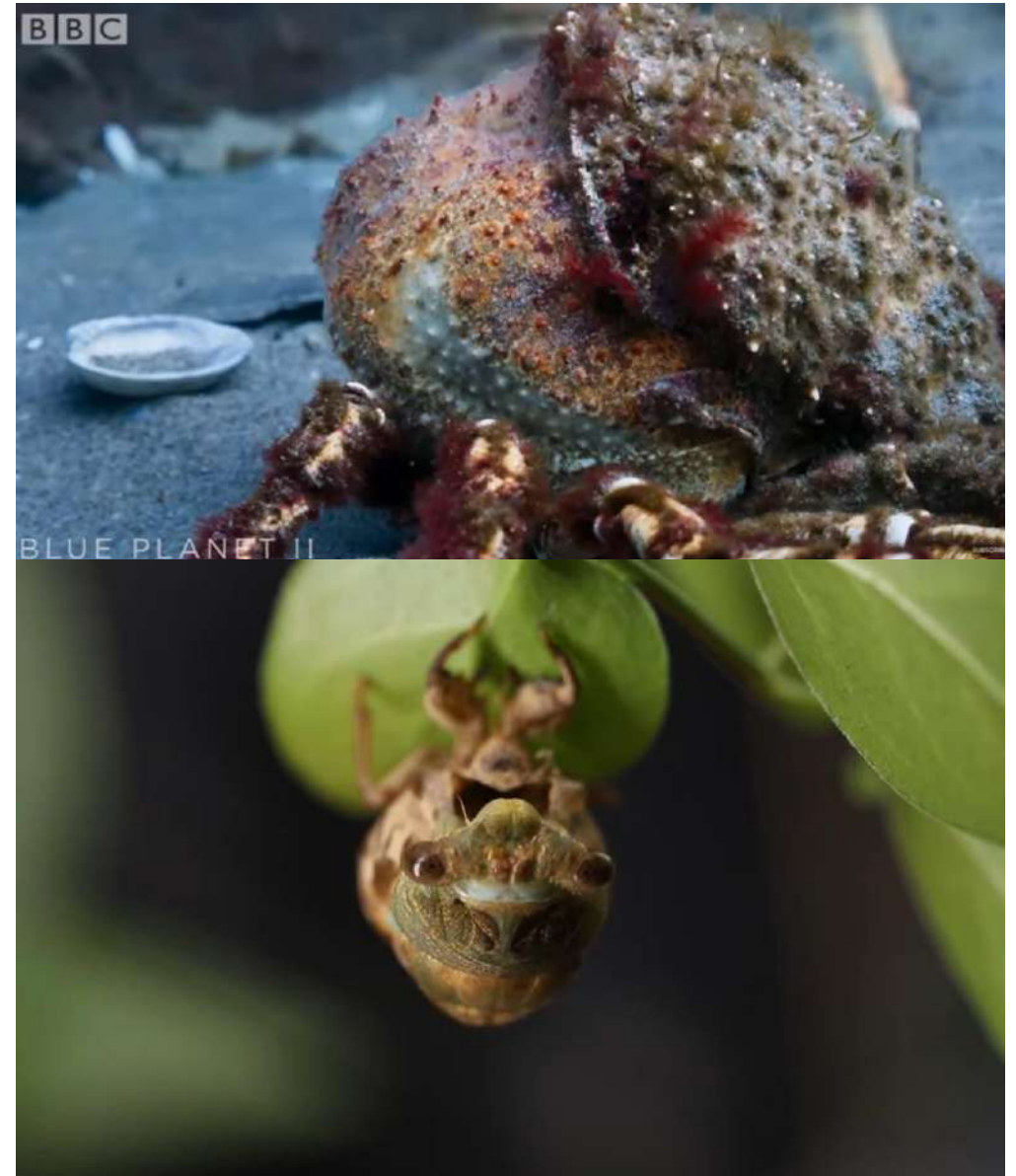
- Chelicerata (spiders, scorpion, mites...)
- Myriapoda (millipedes, centipedes)
- Crustacea (crabs, water fleas, barnacles ... )
- Hexapoda (insects and non-insects)

They share a segmented, modular body plan, jointed appendages and an outer exoskeleton.

The exoskeleton has to be periodically shed, in a process defined as Moulting (ecdysis), which represents a key event in all arthropods life cycle.

→ Understanding the mechanisms of moulting and the diversity of moulting patterns is a key to understand arthropod evolutionary history and success.

<https://www.youtube.com/watch?v=COD1cR5hPN8> above: *Spider crab* (BBC, Blue Planet II)  
<https://www.youtube.com/watch?v=yj8G17dessg&t=28s> bottom: *Cicada* (tdtangents)

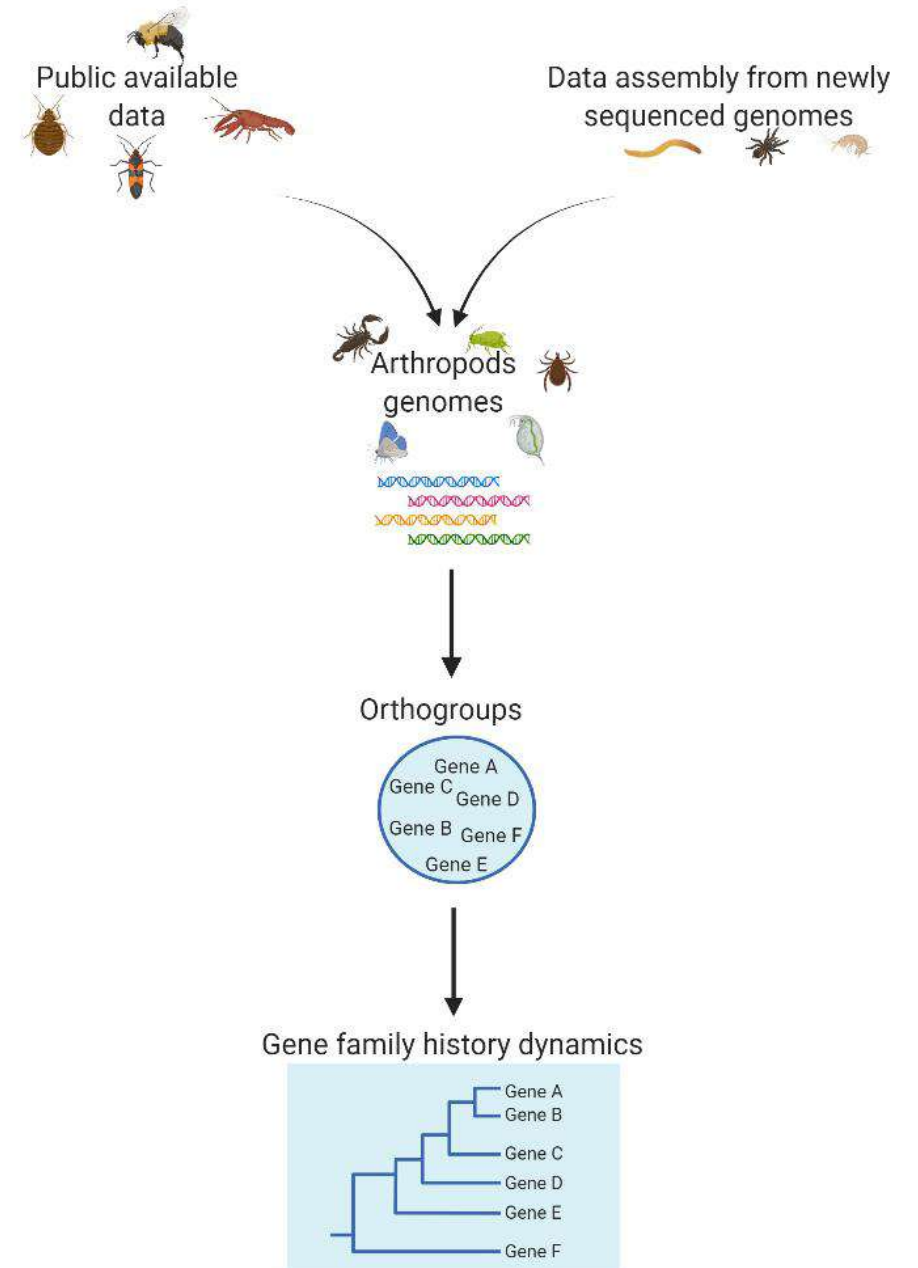


# Outline

Genomics to investigate changes in moulting gene repertoires and their functional role

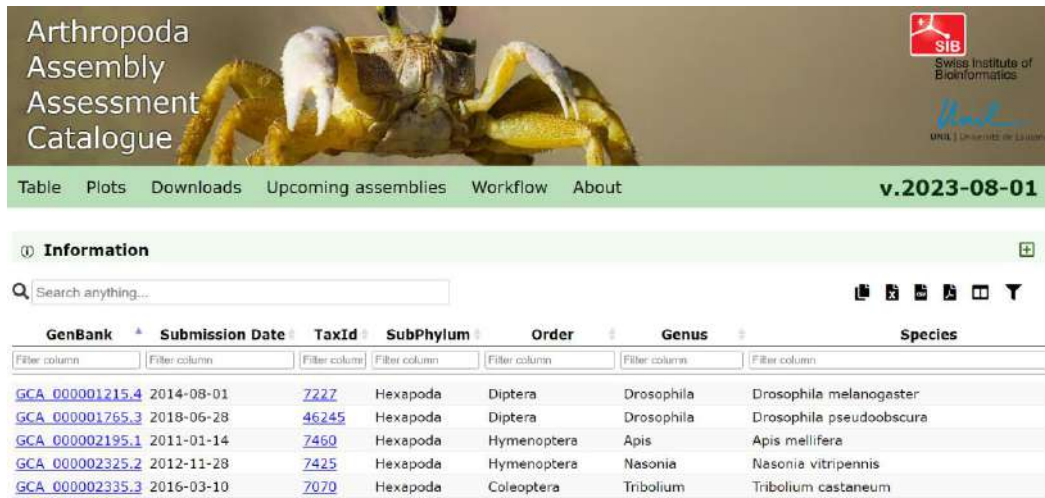
- 1) Which genomic resources can be used to study the moulting gene repertoires ?
- 2) Is it possible to identify genes coding for moulting pathway components across Arthropoda evolution ?
  - OrthoGroup\* inference
- 3) Which are the evolutionary events occurred in their gene family histories?
  - Ancestral state reconstruction and gene-tree-species-tree reconciliation

OrthoGroup\*: set of genes descended from a single gene in the last common ancestor of the compared genomes



# 1) Input data selection

## a) Arthropoda Assembly Assessment Catalogue (<https://evofunvm.dcsr.unil.ch/table.html>)



Arthropoda Assembly Assessment Catalogue

Table Plots Downloads Upcoming assemblies Workflow About v.2023-08-01

Information

Search anything...

GenBank	Submission Date	TaxId	SubPhylum	Order	Genus	Species
<a href="#">GCA_000001215.4</a>	2014-08-01	<a href="#">7227</a>	Hexapoda	Diptera	Drosophila	Drosophila melanogaster
<a href="#">GCA_000001765.3</a>	2018-06-28	<a href="#">46245</a>	Hexapoda	Diptera	Drosophila	Drosophila pseudoobscura
<a href="#">GCA_000002195.1</a>	2011-01-14	<a href="#">7460</a>	Hexapoda	Hymenoptera	Apis	Apis mellifera
<a href="#">GCA_000002325.2</a>	2012-11-28	<a href="#">7425</a>	Hexapoda	Hymenoptera	Nasonia	Nasonia vitripennis
<a href="#">GCA_000002335.3</a>	2016-03-10	<a href="#">7070</a>	Hexapoda	Coleoptera	Tribolium	Tribolium castaneum

b) Phylogeny from single-copy orthologs (BUSCOphile)

c) Rooting

d) Calibration

→ 150 species spanning almost 600 mya evolution

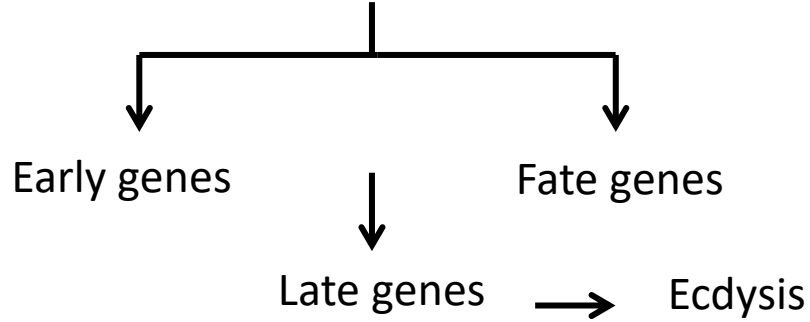




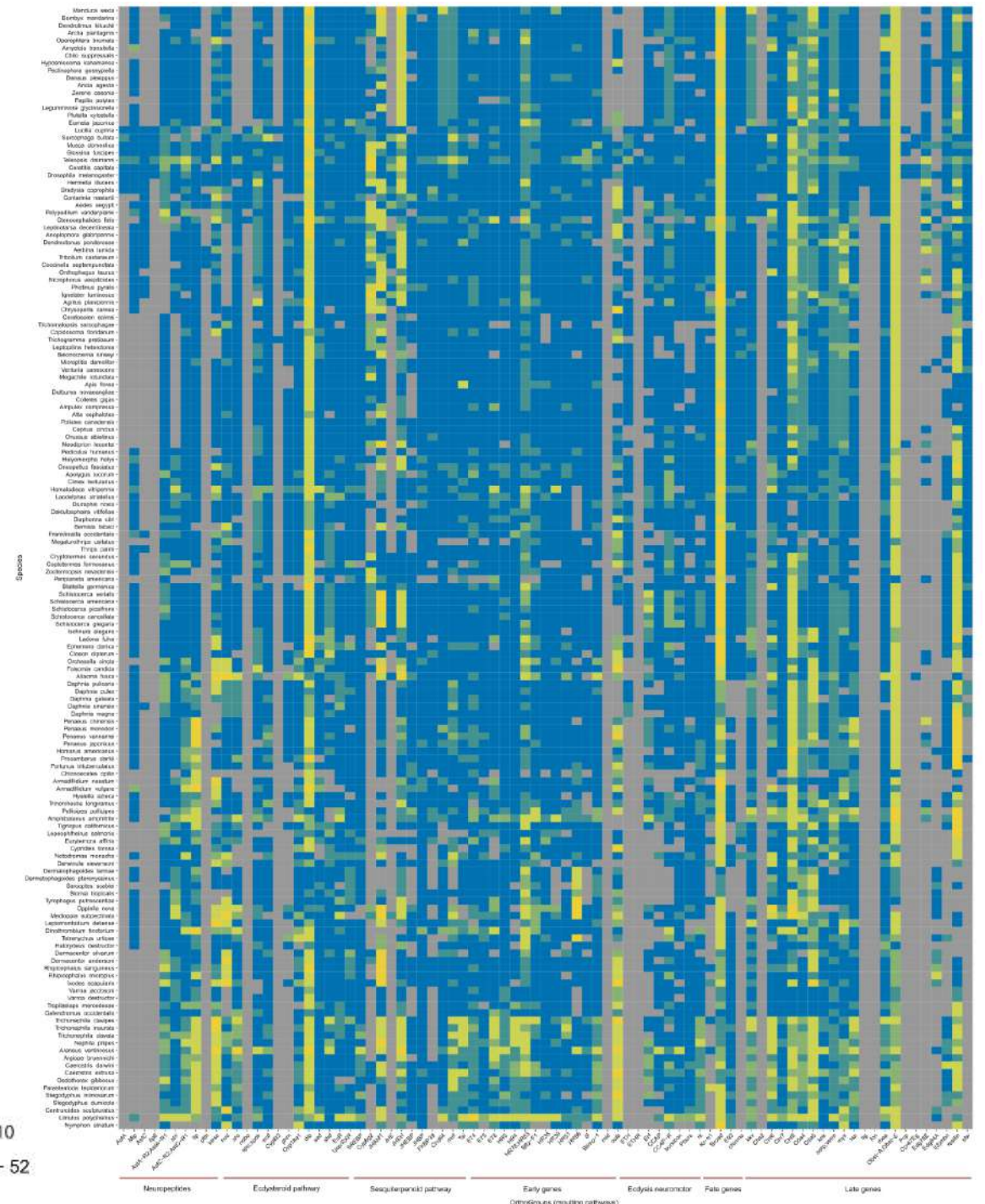
# 2) Identification of «moulting genes» across arthropods

Orthology delineation (OrthoLogger) and extrapolation of 65 gene families for downstream analysis:

Ecdysteroids pathway      Sesquiterpenoids pathway



- Mostly single-copy genes across all the species
- Multicopy genes in nearly all the species
- Mostly restricted to few species

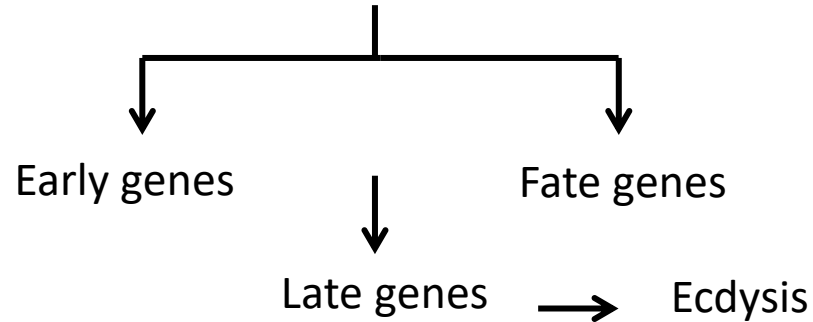


Gene copy number

## 2) Identification of «moulting genes» across arthropods: are some lineages more dynamic than others?

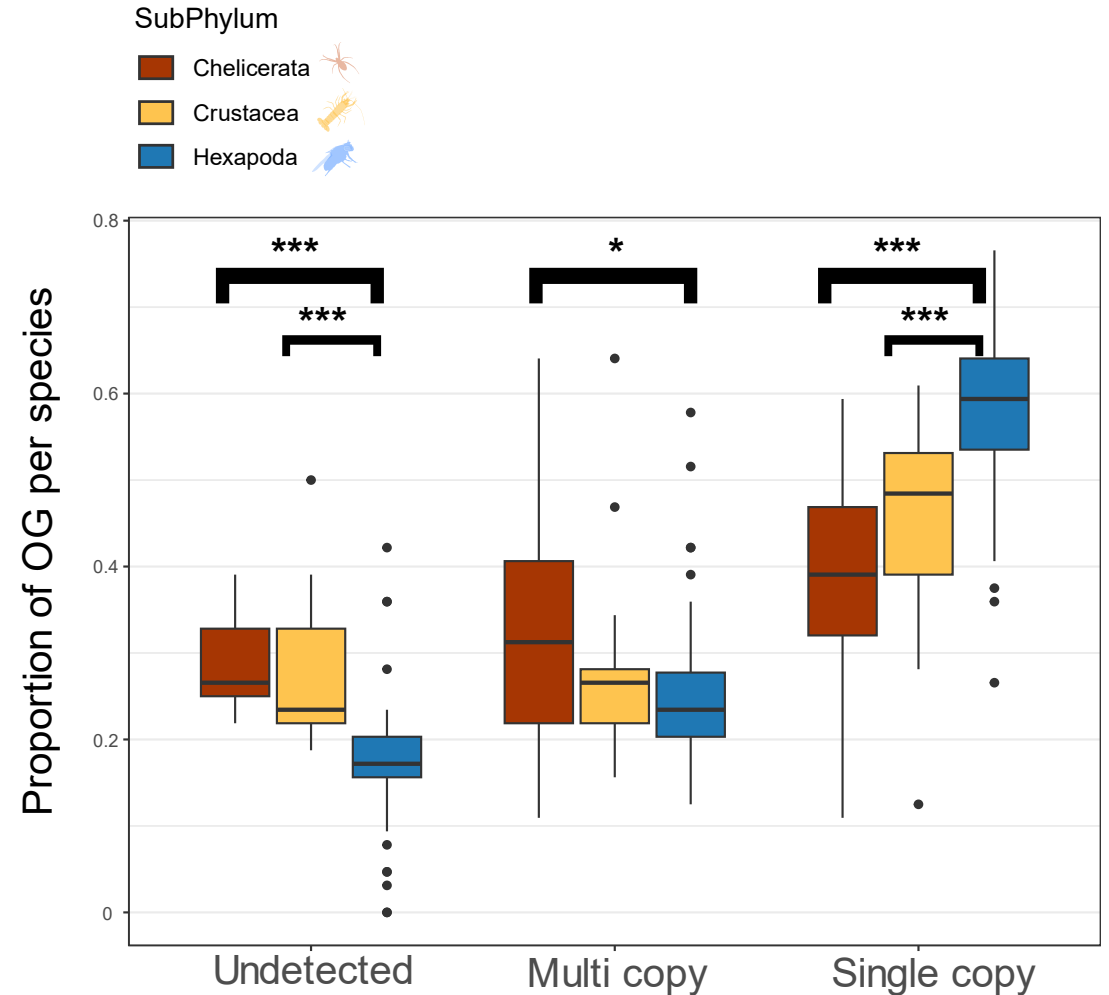
Orthology delineation (OrthoLogger) and extrapolation of 65 gene families for downstream analysis:

Ecdysteroids pathway      Sesquiterpenoids pathway



- Mostly single-copy genes across all the species
- Multicopy genes in nearly all the species
- Mostly restricted to few species

→ Orthology delineation suggest gene history in Chelicerata are highly dynamic





# 3) Characterization of family history

## Ancestral count reconstruction of gene copy number

Given orthology data and a species phylogeny, the ancestral gene copy number in ancestral species is estimated from gene copy number in extant species

→ It allows quantification of potential gene gains and gene losses in the framework of the species evolution




1449 total expansions and 1401 total contractions occurred across all the families

Point big and red = many families expanded and contracted, Point small and blue = no events

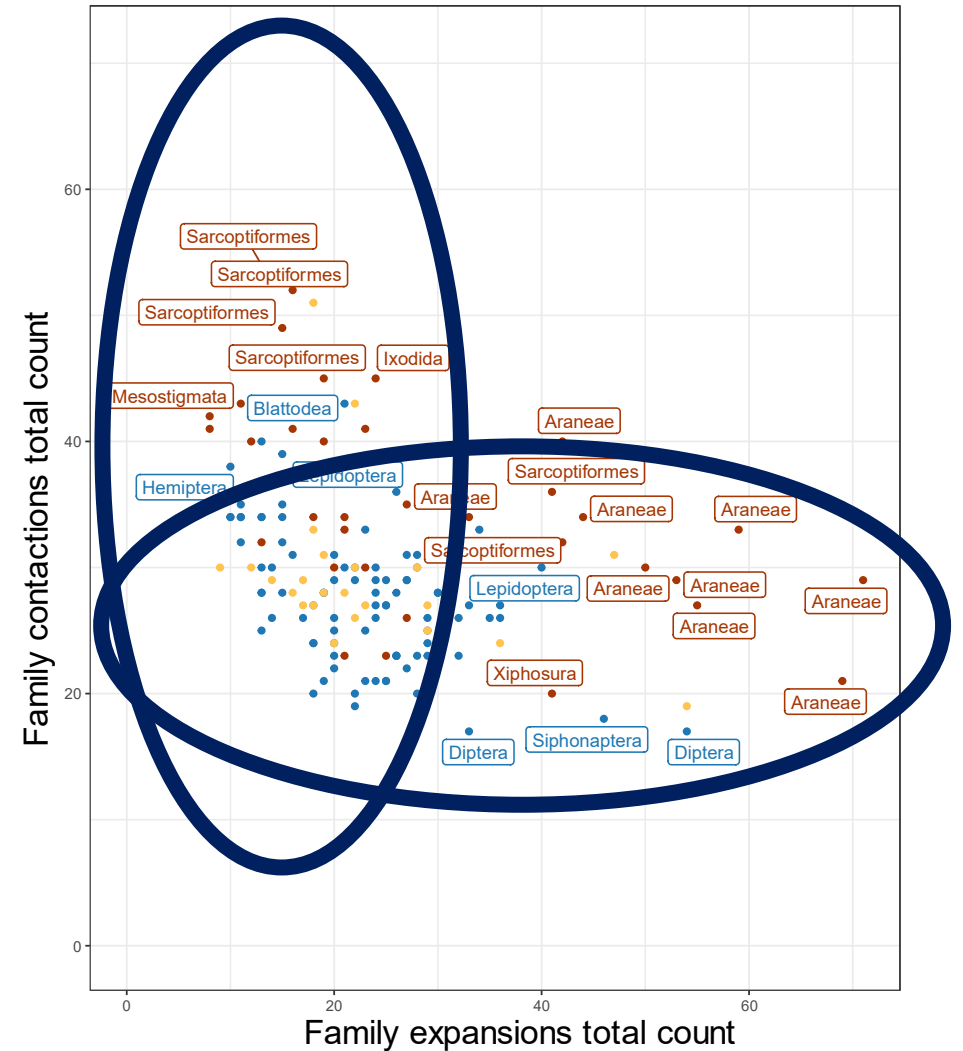
### D. How many expanding/contracting families are found along the species evolutionary path?

→ Two main evolutionary trajectories

SubPhylum

- Chelicerata 
- Crustacea 
- Hexapoda 

Ancestral state reconstruction analysis



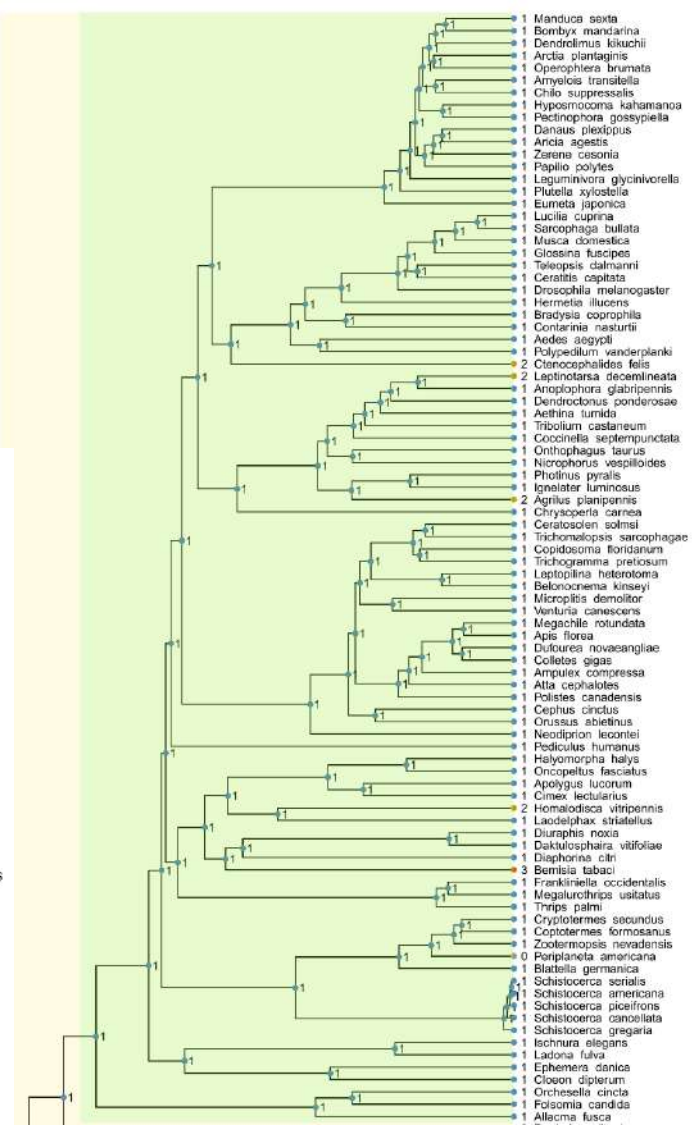
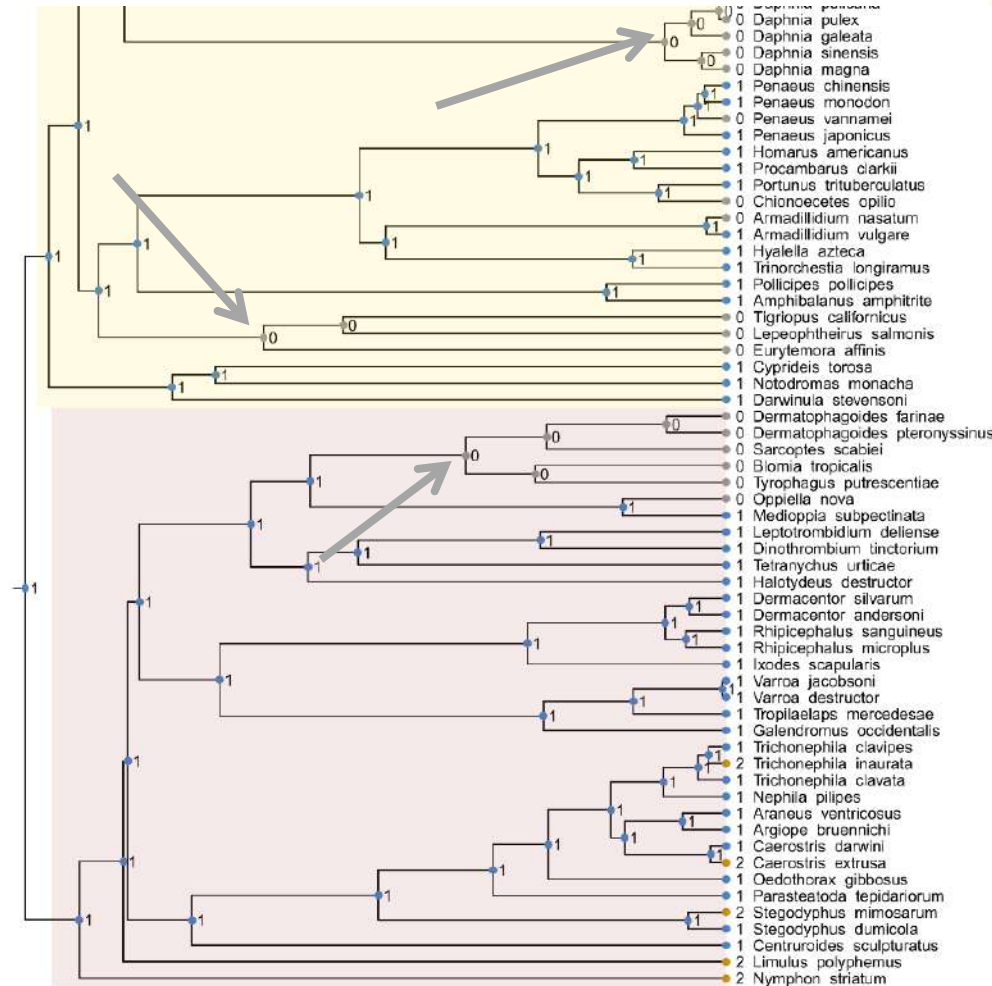


# 3) Characterization of family history

## Ancestral count reconstruction of gene copy number

A closer look to individual families:  
E93 gene family

- 11 total gains and 8 total losses
- Single copy in Hexapoda
- Present beyond Hexapoda
- Lost in sublineages: Sarcoptiformes, Daphnids, Copepoda
- Duplicated in 5 chelicerate and 5 insect species



Gene copy number ● 0 ● 1 ● 2 ● 3

# 3) Characterization of family history

## Gene-tree-species-tree reconciliation and results comparison

- A. Reconciliation (NOTUNG): 3 total duplications and 20 total losses
- B. Ancestral reconstruction: 11 total gains and 8 total losses

- candidate gained copies are fragmented sequences, which are removed from the alignment and counted as losses (MAFFT-TrimAl-RAXML)
- Duplications in *L. polyphemus*, *N. striatum* and *A. planipennis* confirmed
- Ancestral losses Sarcoptiformes, Copepoda and Daphnida confirmed

THANKS FOR YOUR ATTENTION !

