Case studies:
Combining orthology and expression data

Diving into the *Drosophila* immune response

Antonin THIÉBAUT - SIB Biodiversity bioinformatics, Day 2

Case studies:
Combining orthology and expression data

Diving into the *Drosophila* immune response

Drosophila orthology

- Get orthology data for 36 species of *Drosophila* + 4 mosquitoes (outgroup)
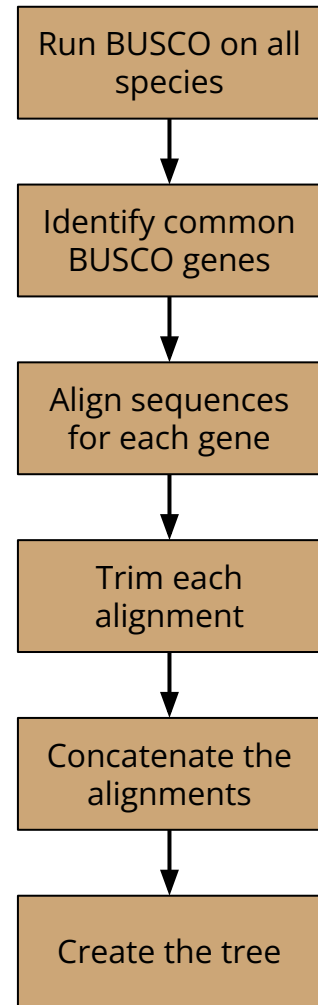
Drosophila orthology

- Get orthology data for 36 species of *Drosophila* + 4 mosquitoes (outgroup)

- Apply the OMA methodology

- Requires phylogenetic tree to guide the determination of orthology relationships
  ⇒ How do we get this tree?

Drosophila orthology

- Get orthology data for 36 species of *Drosophila* + 4 mosquitoes (outgroup)

- Apply the OMA methodology

- Requires phylogenetic tree to guide the determination of orthology relationships
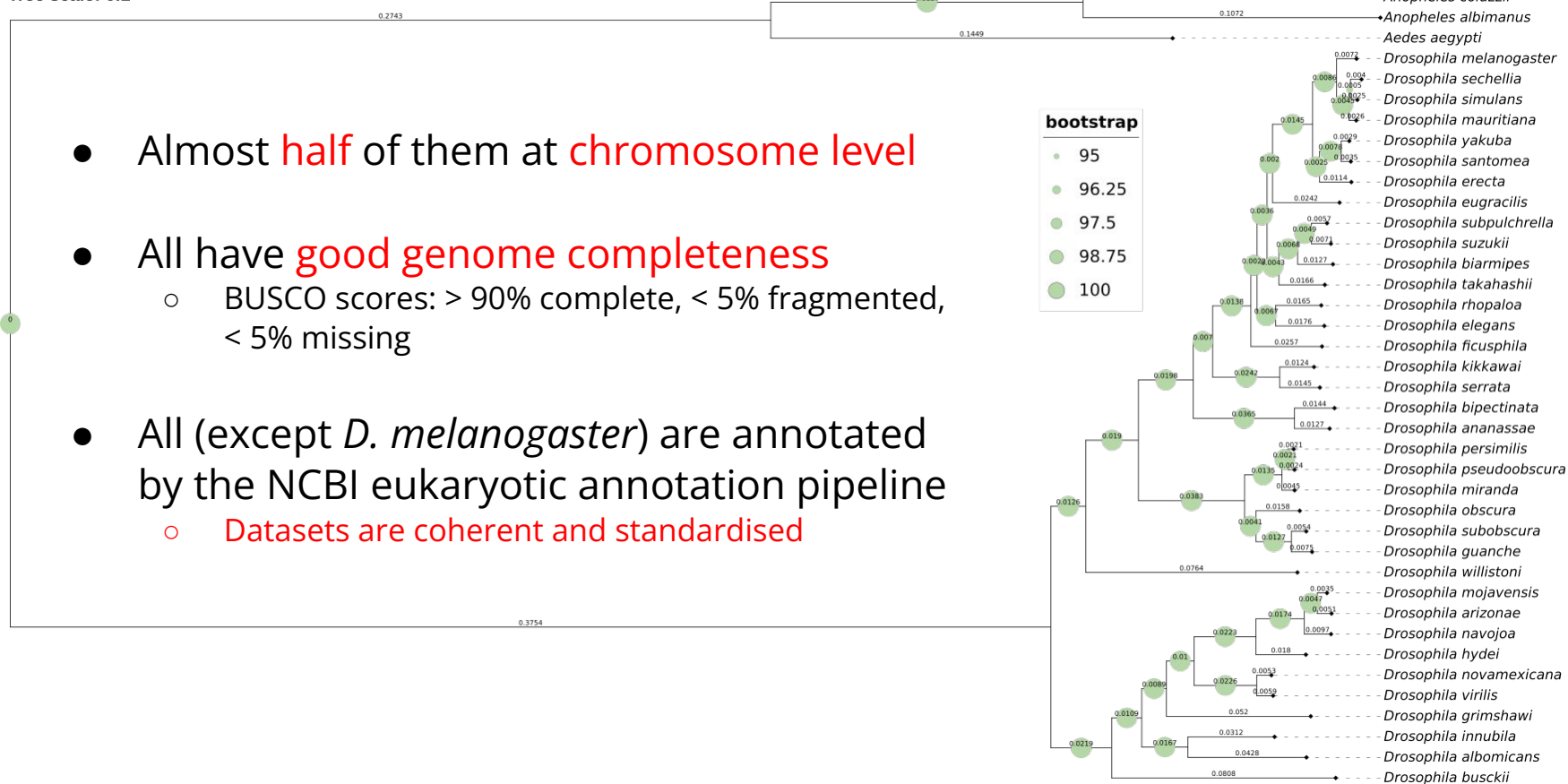 ⇒ How do we get this tree?

- BUSCOphile

BUSCOphile

- Snakemake workflow relying on BUSCO

- 2892 common genes

- Sequence alignment: MUSCLE

- Alignment trimming: trimAl

- Tree creation: IQ-TREE

```
Run BUSCO on all
species
        ↓
Identify common
BUSCO genes
        ↓
Align sequences
for each gene
        ↓
Trim each
alignment
        ↓
Concatenate the
alignments
        ↓
Create the tree
```
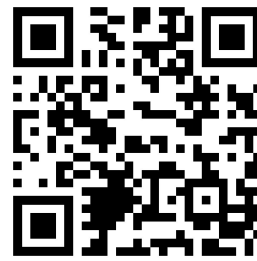
Drosophila phylogeny

- Almost half of them at chromosome level

- All have good genome completeness
  - BUSCO scores: > 90% complete, < 5% fragmented, < 5% missing

- All (except *D. melanogaster*) are annotated by the NCBI eukaryotic annotation pipeline
  - Datasets are coherent and standardised

Drosophila orthology

- Get orthology data for 36 species of *Drosophila* + 4 mosquitoes (outgroup)

- Apply the OMA methodology

- Requires phylogenetic tree to guide the determination of orthology relationships
  ⇒ How do we get this tree?

- BUSCOphile

- OMA

DrosOMA

- 40 diptera

- 24'896 HOGs

- 568'796 proteins

https://drosoma.dcsr.unil.ch/

Thiébaut A, Altenhoff AM, Campli G et al. DrosOMA: the Drosophila Orthologous Matrix browser. F1000Research 2023, 12:936

Case studies:
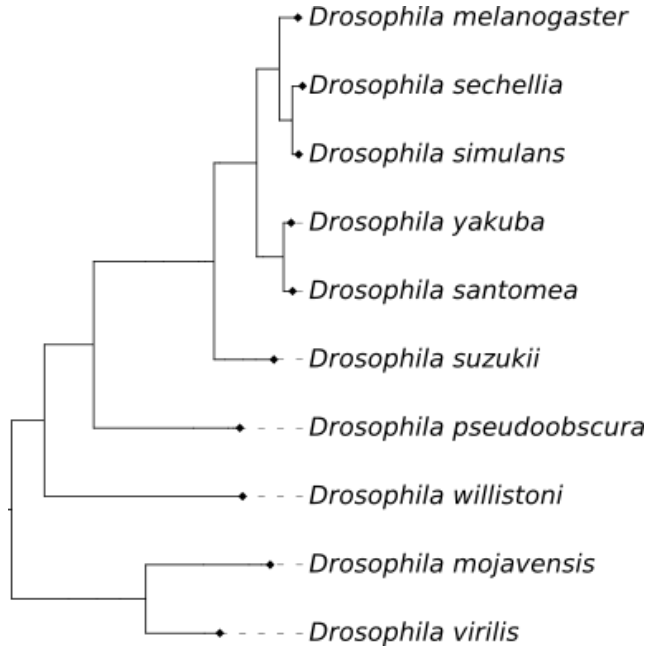# Combining orthology and expression data

Diving into the *Drosophila* immune response
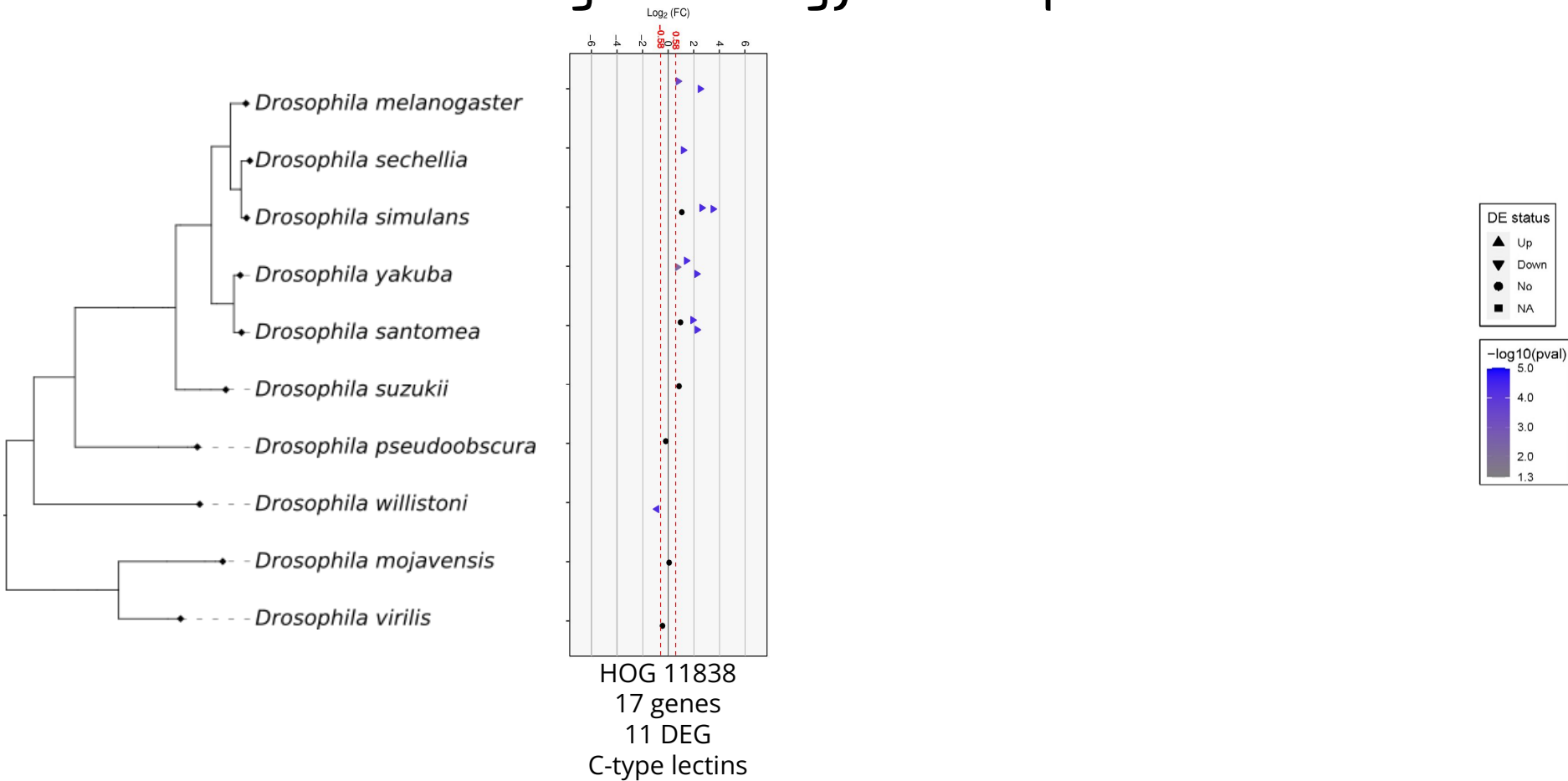
Expression data of *Drosophila* immune response

- RNA-seq experiments done by Jean-Luc Imler's team @IBMC, Strasbourg

- 10 *Drosophila* species:
  - *D. melanogaster*
  - *D. mojavensis*
  - *D. pseudoobscura*
  - *D. santomea*
  - *D. sechellia*
  - *D. simulans*
  - *D. suzukii*
  - *D. virilis*
  - *D. willistoni*
  - *D. yakuba*

- Injection of 2,3-cGAMP vs Tris
  - 2,3-cGAMP = compound mimicking viral infection

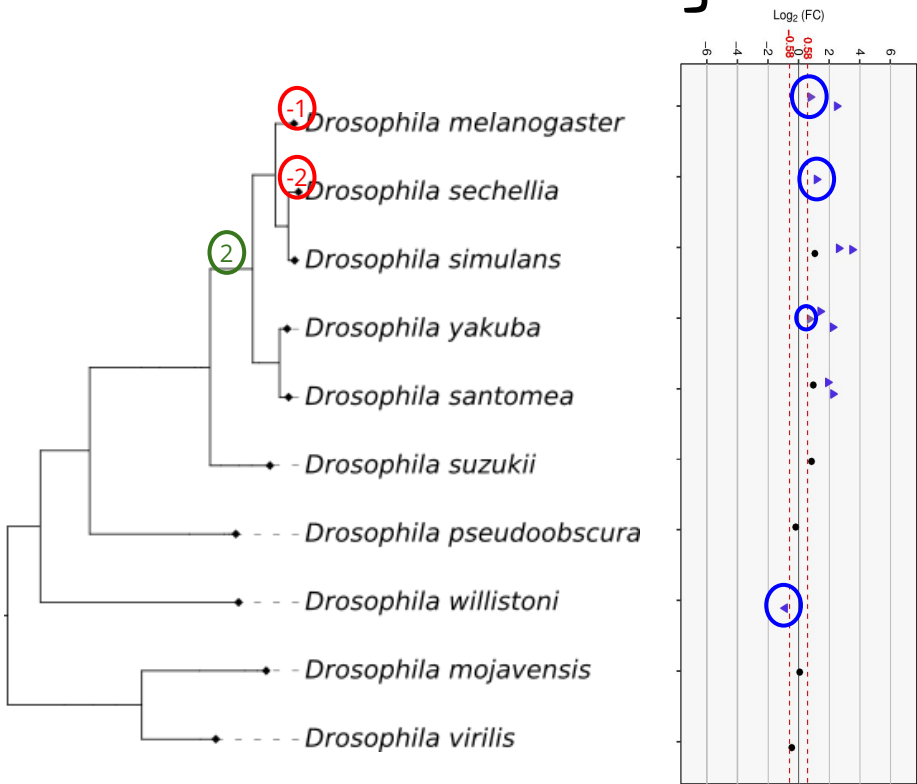- Sequencing, trimming, mapping… ⇒ Differential expression analyses
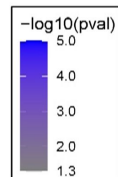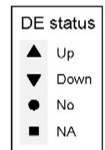 - https://gitlab.com/aathbt/rnaseq-analysis-workflow

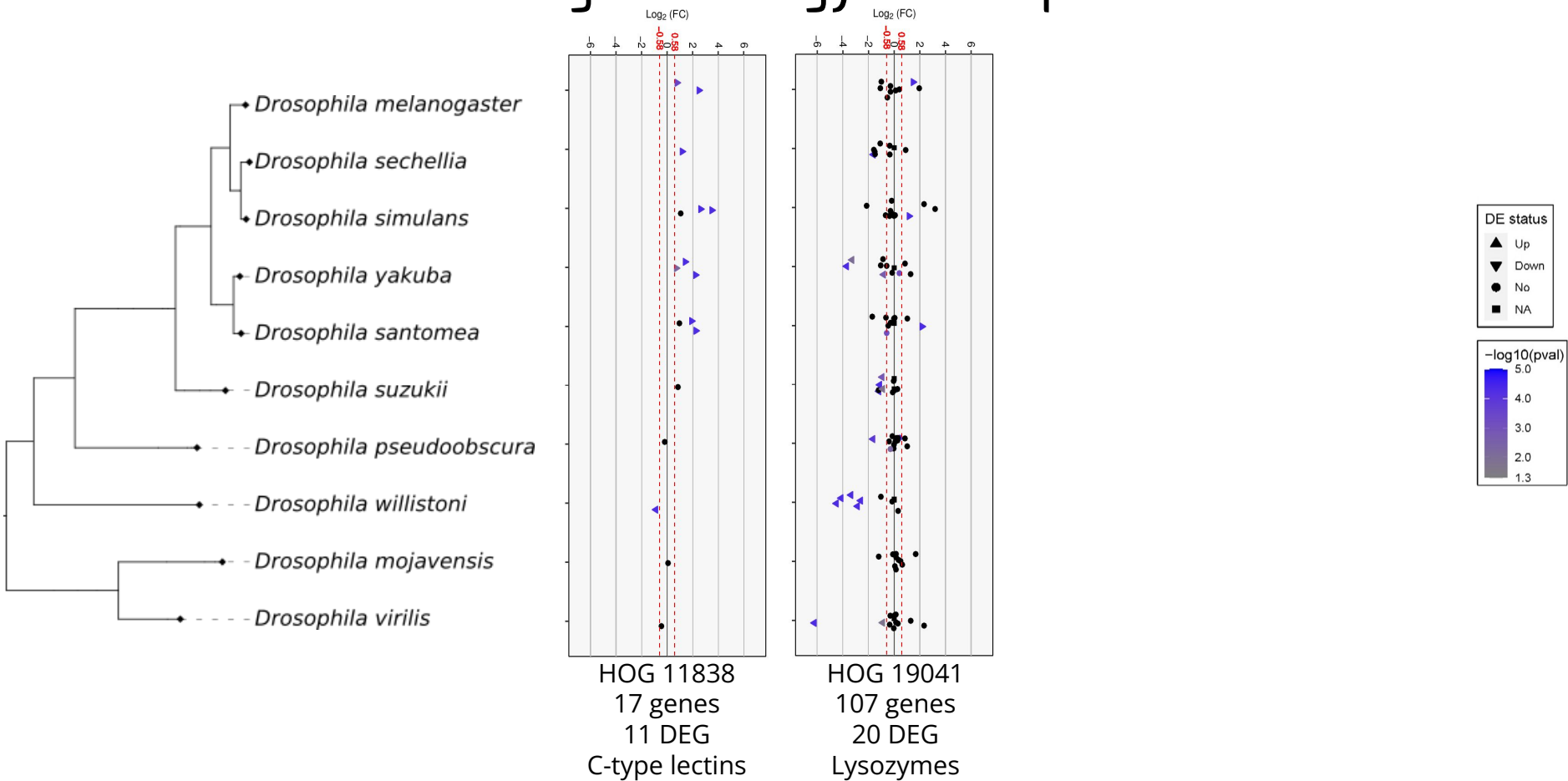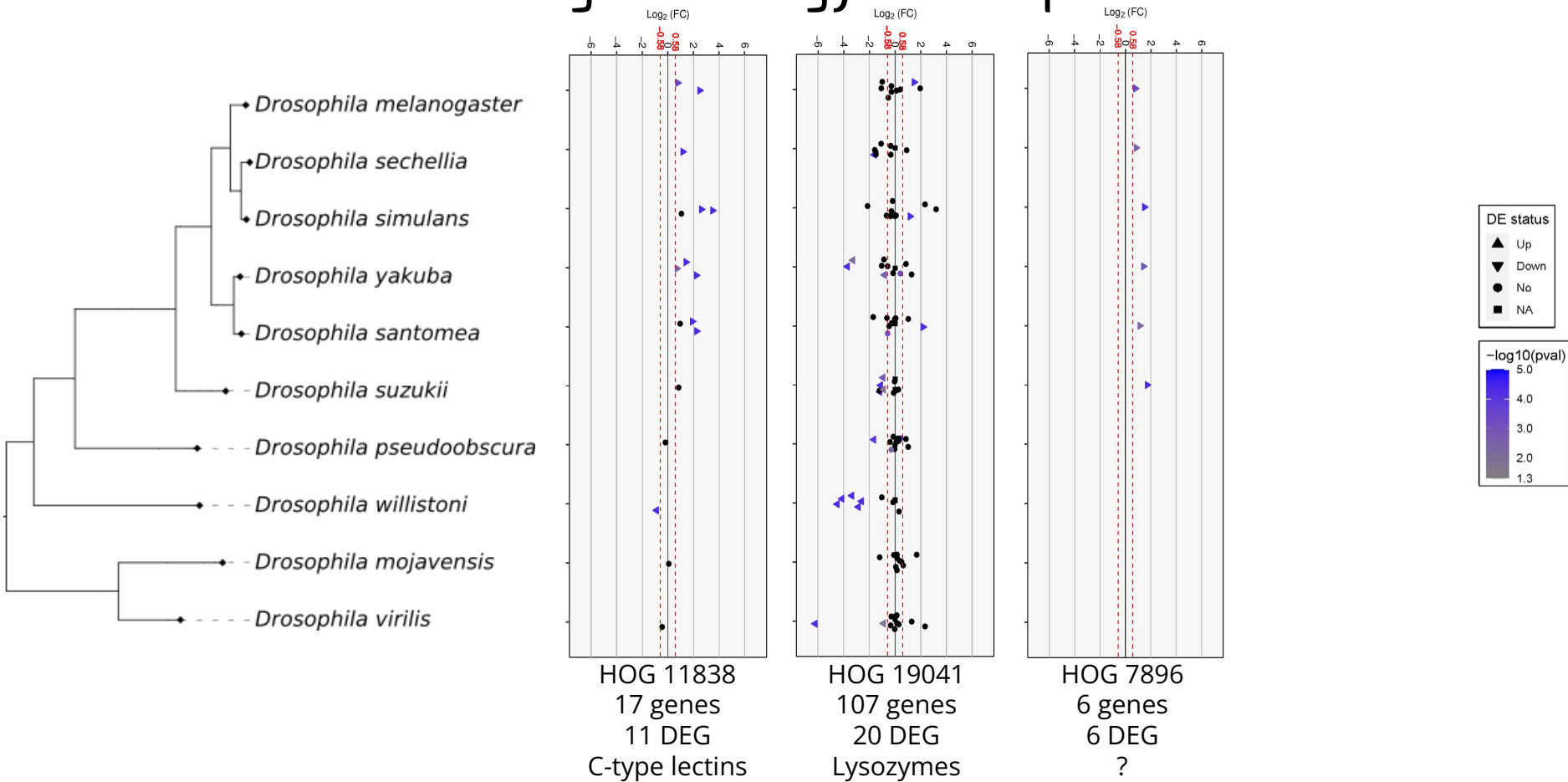# Combining orthology and expression

Combining orthology and expression

HOG 11838
17 genes
11 DEG
C-type lectins

Combining orthology and expression

HOG 11838
17 genes
11 DEG
C-type lectins

Combining orthology and expression

HOG 11838
17 genes
11 DEG
C-type lectins

HOG 19041
107 genes
20 DEG
Lysozymes

Combining orthology and expression

HOG 11838
17 genes
11 DEG
C-type lectins

HOG 19041
107 genes
20 DEG
Lysozymes

HOG 7896
6 genes
6 DEG
?

Combining orthology and expression

HOG 11838
17 genes
11 DEG
C-type lectins

HOG 19041
107 genes
20 DEG
Lysozymes

HOG 7896
6 genes
6 DEG
?

HOG 9797
6 genes
6 DEG
Asparaginases

Questions?