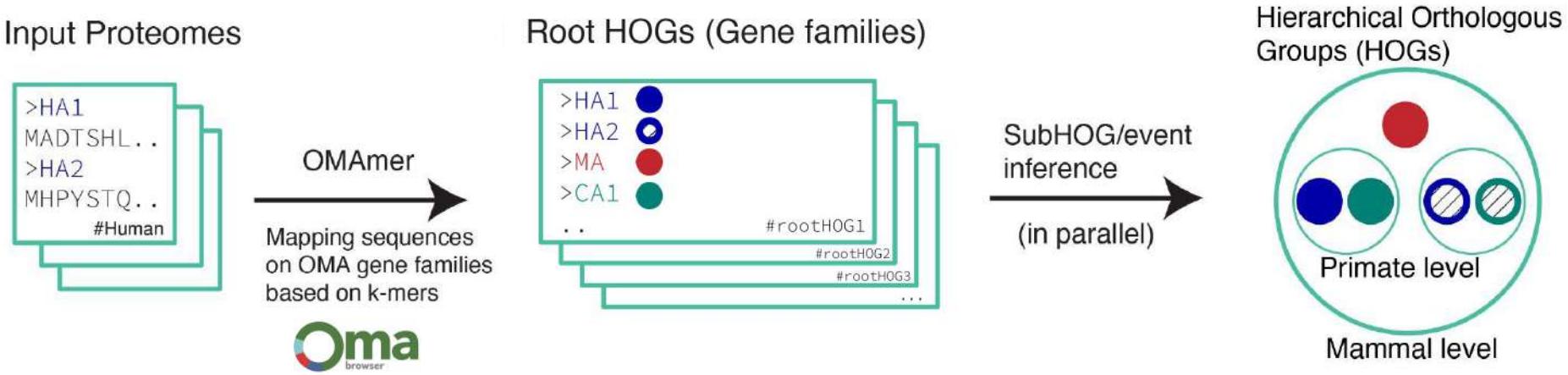
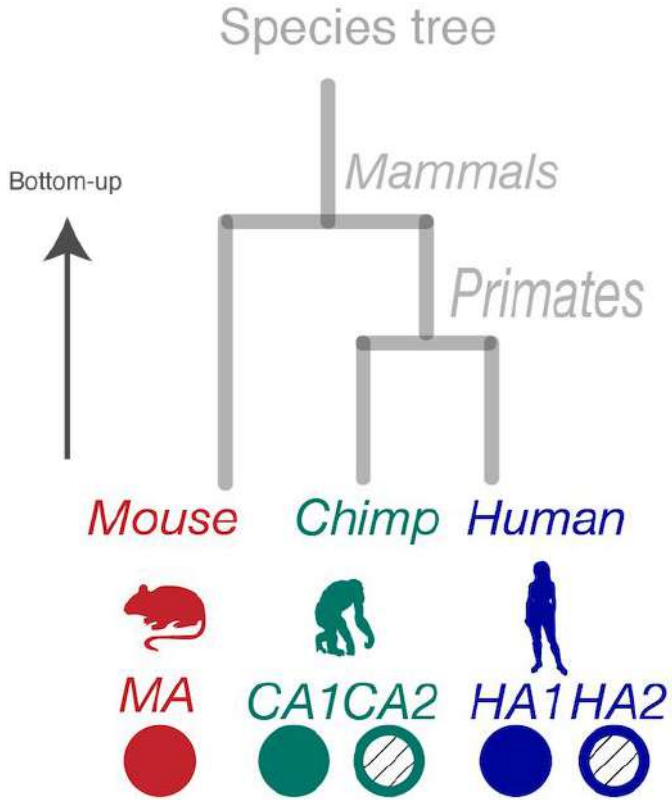
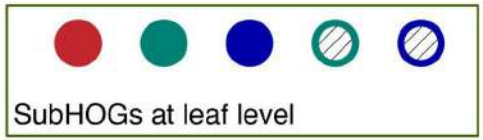
Module 3: FastOMA

FastOMA, our new tool

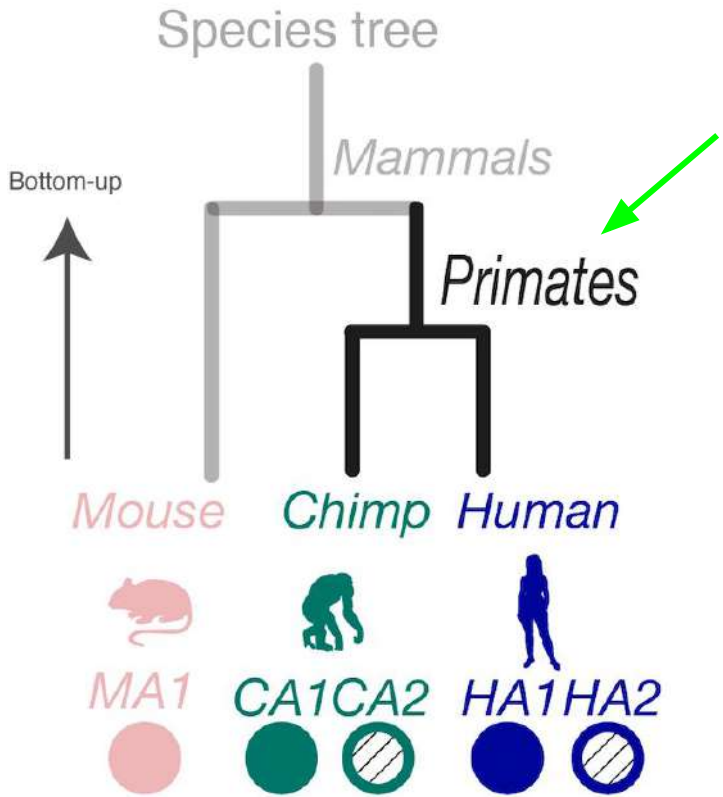


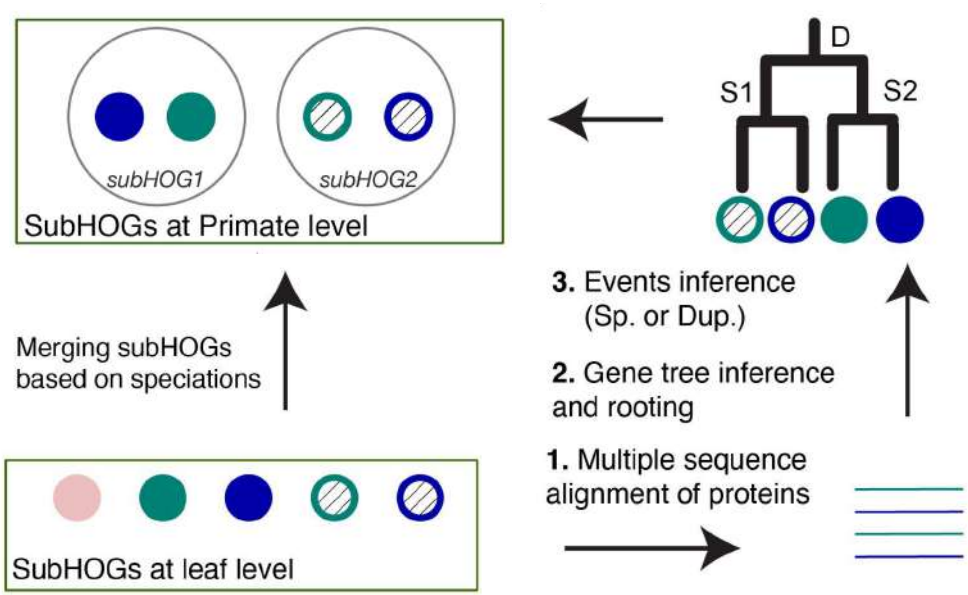
HOG inference



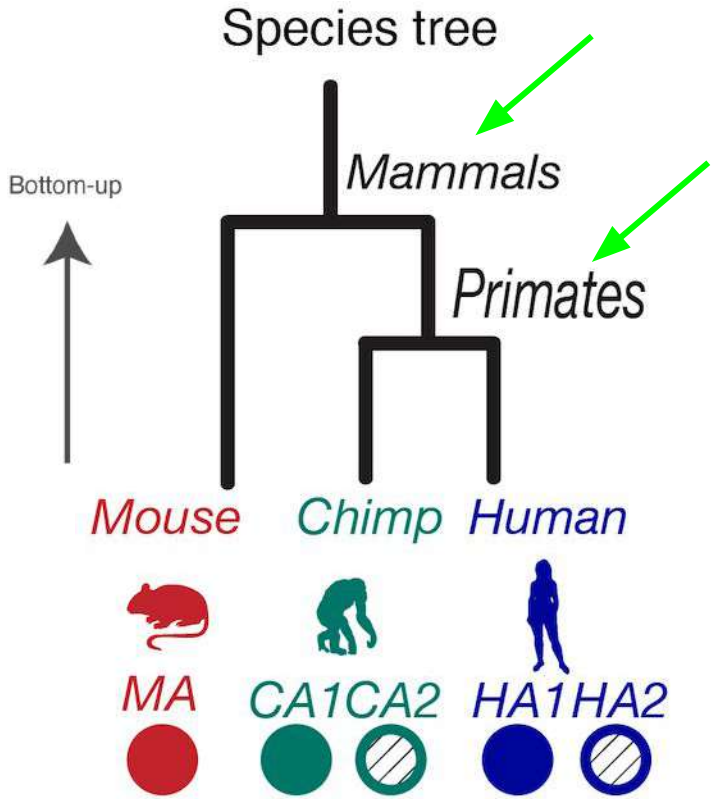


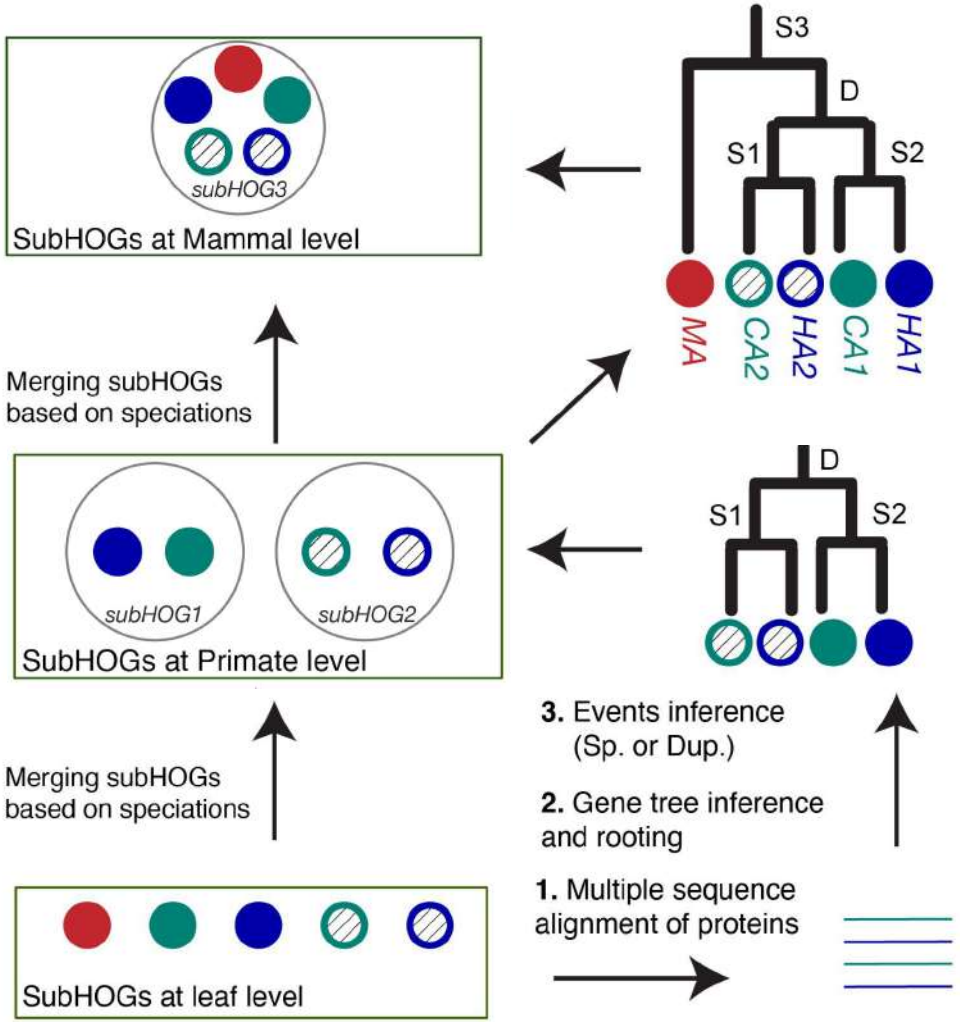
HOG inference



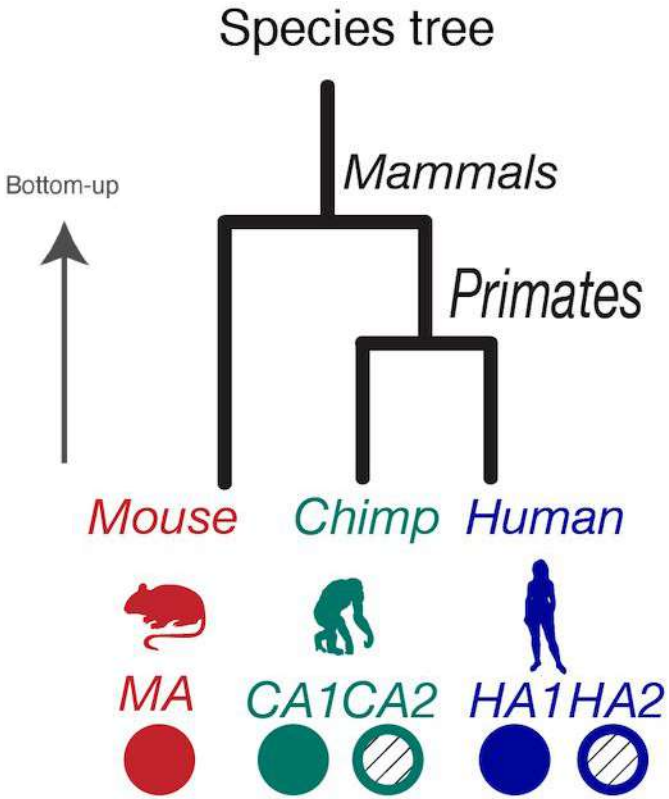


HOG inference



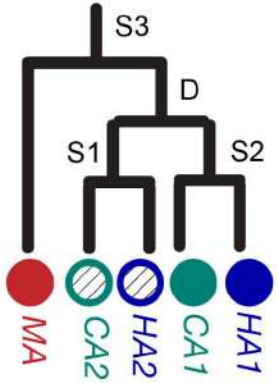


HOG inference

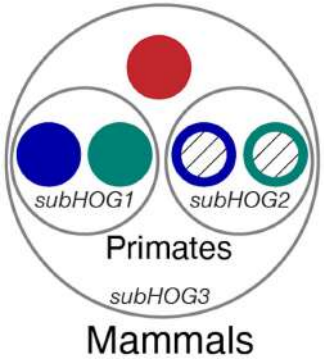


HOG in orthoXML

```
<orthoGroup3        Mammals>
  MA
  <paraGroup>
    <orthoGroup1    Primates>
      HA1
      CA1
    </orthoGroup1>
    <orthoGroup2    Primates>
      HA2
      CA2
    </orthoGroup2>
  </paraGroup>
</orthoGroup3>
```

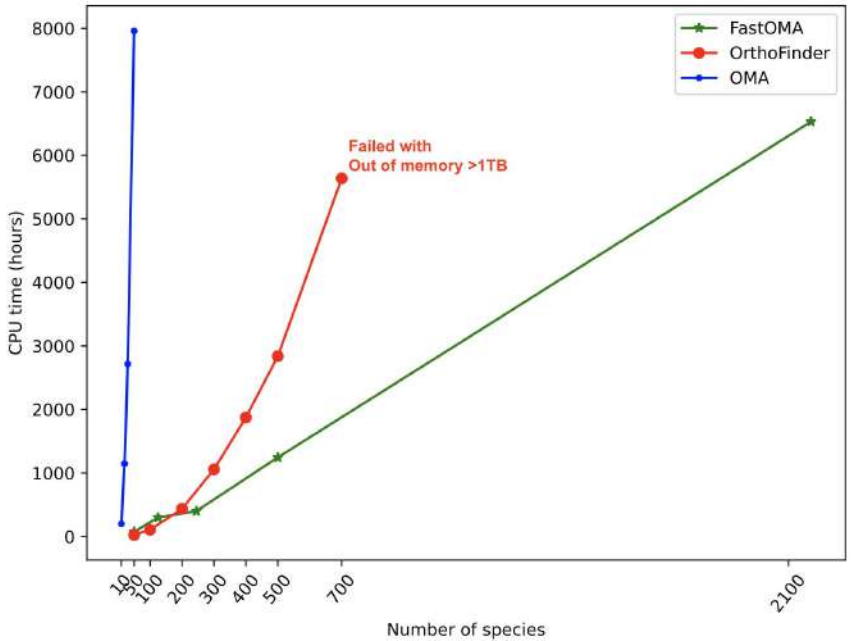


Nested structure of HOG



Orthology inference for Eukaryote dataset

- 2180 eukaryotic species
- Uniprot reference proteomes
- in a single day using 300 CPUs



github.com/DessimozLab/FastOMA

Module 3.3

- **4. How many Root HOGs are in the HOG file?**

✕ Hint ✕ Answer

Each line in the output file denotes a gene family. After running, check the end of the file rootHOGs.tsv. Note that the indexing starts from 0.

There are 6793 rootHOG (gene families) in this file.

- **5. Consider the gene "60S ribosomal protein L15-A" in *Schizosaccharomyces pombe* with protein ID: RL15A_SCHPO. How many proteins are in the gene family (for these 5 species of interest)?**