

09:00-12:00 Genome annotation

- 09:00-09:30 Introduction to the study system and infrastructure
 - Making sure you have access to Fox
 - Submit the first set of jobs
 - i. Repeat mask
 - ii. Mapping protein sets
 - iii. *Ab initio* gene prediction
- 09:30-11:55 Introduction to genome annotation
 - Work through the rest of the programs
 - i. EvidenceModeler
 - ii. BUSCO
 - iii. Functional annotation
- 11:55-12:00 Summary

12:00-13:00 Lunch

13:00-14:00 Comparative genomics

- Introduction to comparative genomics
- Setting up OrthoFinder

14:00-14:15 Break

14:15-16:00

- Running OrthoFinder on proteins and CDS

Genome annotation and comparative genomics

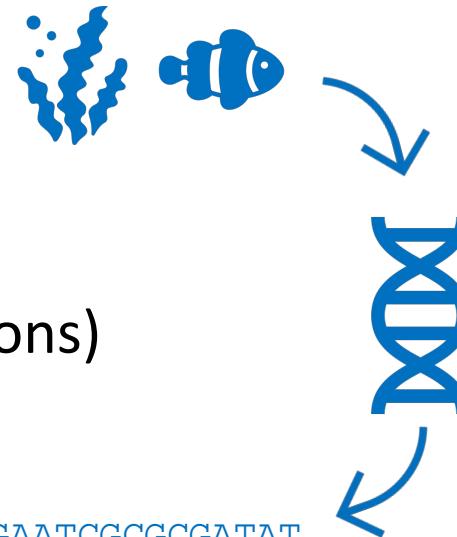
Part 1, afternoon

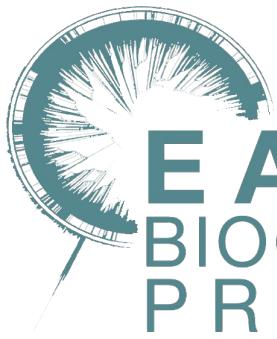
Teachers: Bram Danneels, Helle Tessand Baalsrud, José Cerca, Ole K. Tørresen
Oslo Bioinformatics Workshop Week 2023
11th December

New sequencing technology revolutionize genomic biodiversity research

- **fast** (days, not years)
- **cheap** (thousands \$, not millions)

```
ATCGCGCGATATGACTGACGAAATCGCTGATTAGTGGGTATGAATCGCGCGATAT  
GACTGACGAAATCGCTGATTAGTGGGTATGAATCGCGCGATATGACTGACGAAAT  
CGCTGATTAGTGGGTATGAATCGCGCGATATGACTGACGAAATCGCTGATTAGTG  
GGTATGAATCGCGCGATATGACTGACG
```

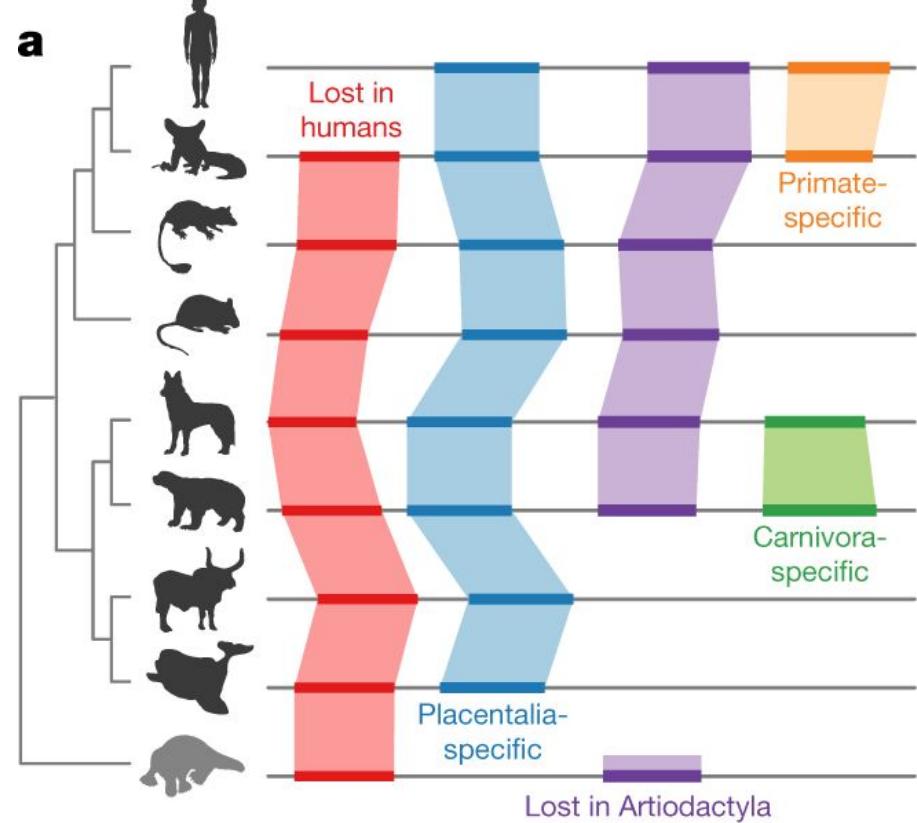
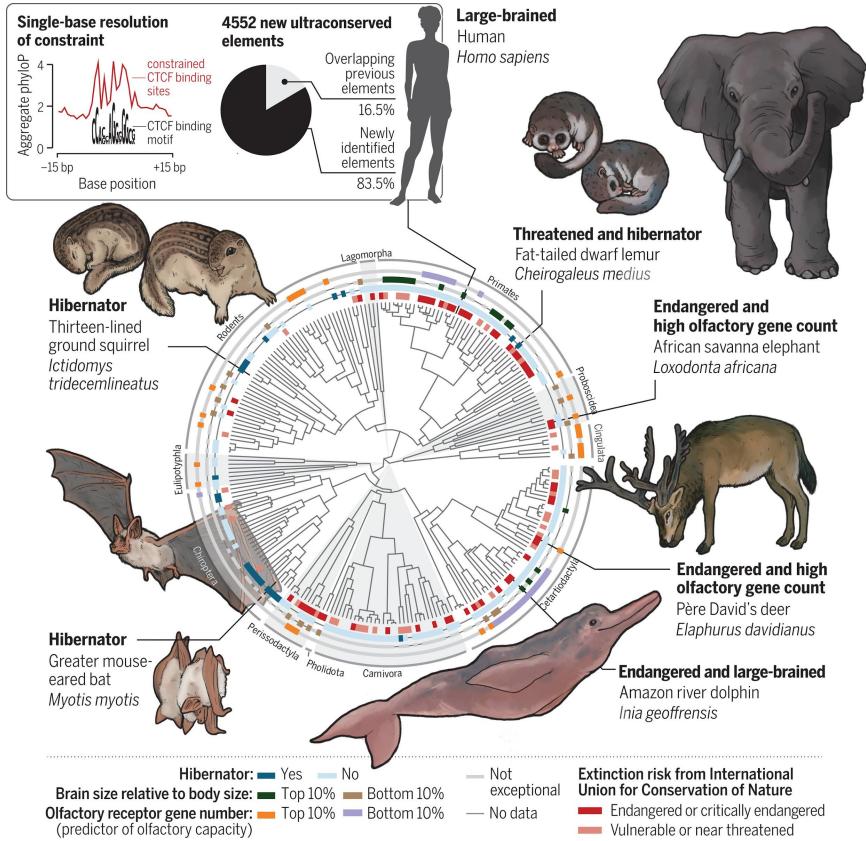




EARTH BIOGENOME PROJECT

NORWAY

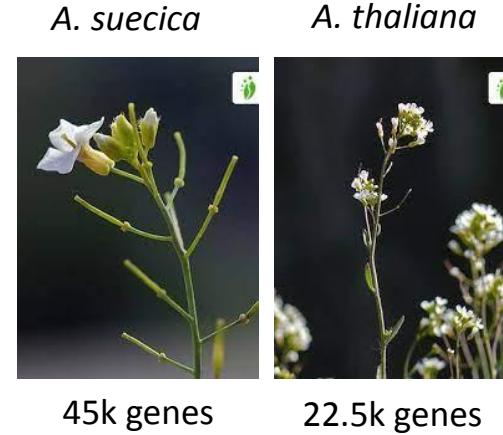
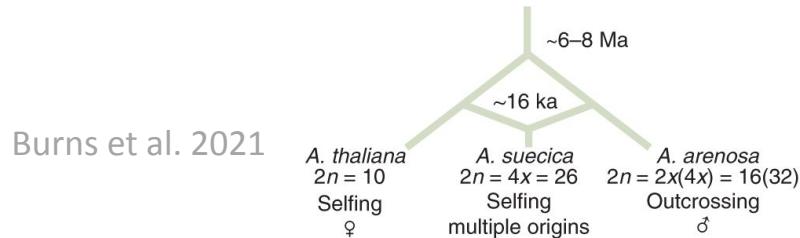




Comparing gene content

- Gene numbers can tell you something..

- E.g. genome duplications / polyploidizations

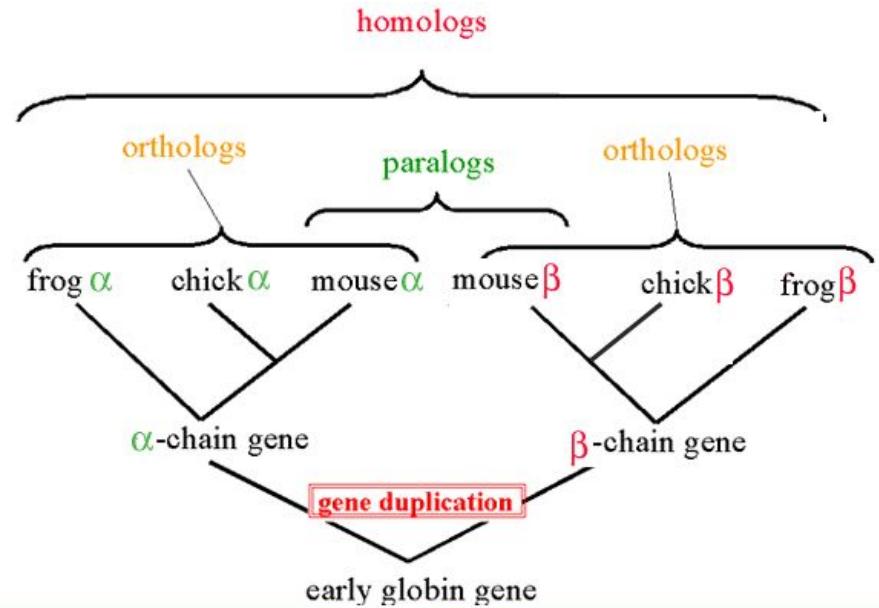


- Which genes species share (orthologs), often more important

- The concept of orthology and homology

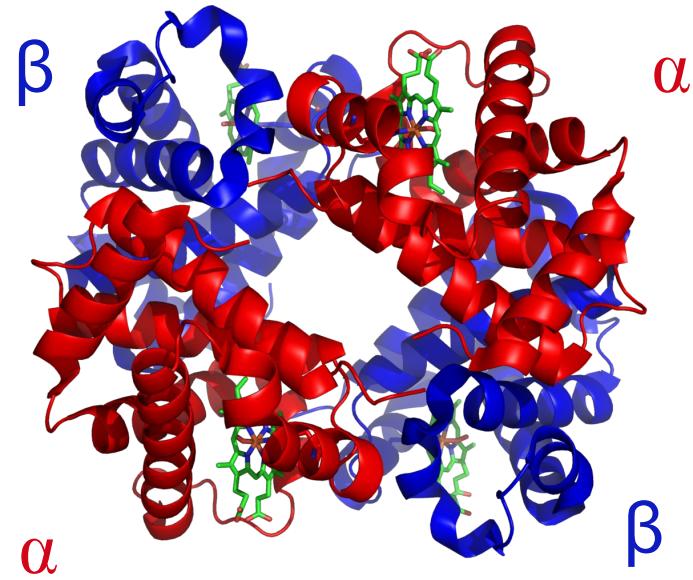
Detecting orthologs

- Identify gene duplications
- Reconstruct gene trees
- Infer natural selection on genes



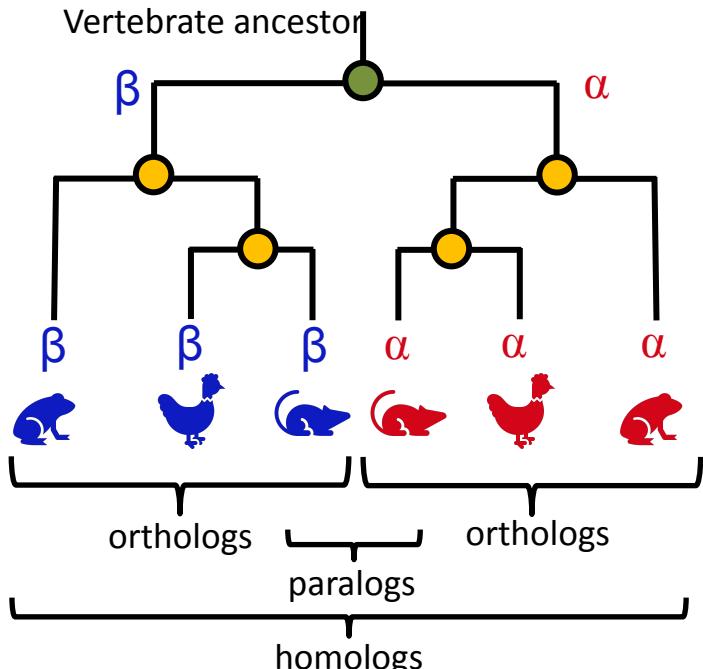
Speciation
events

Homologs, parologs and orthologs



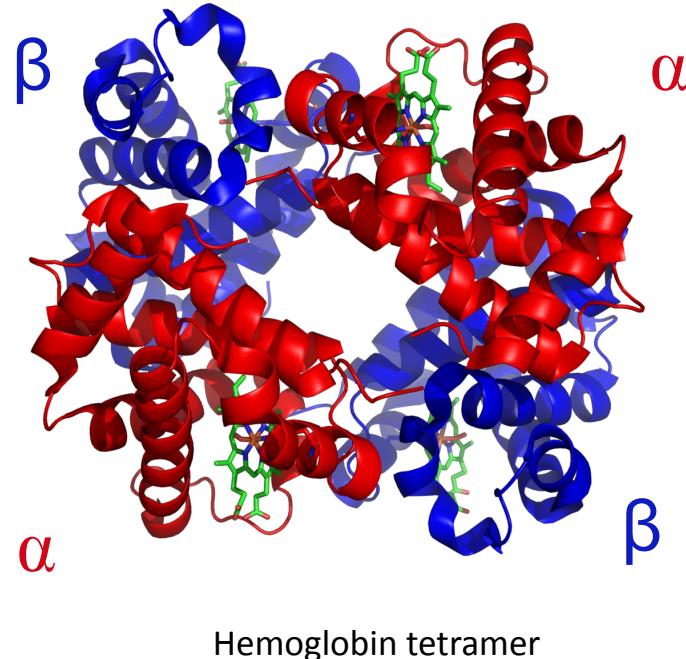
Hemoglobin tetramer

Homologs, parologs and orthologs



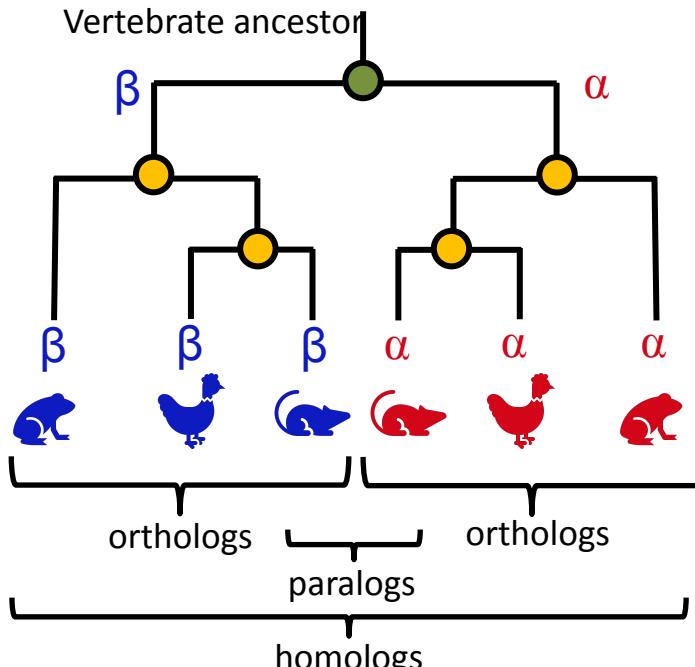
Gene duplication ●

Species split ●



Hemoglobin tetramer

Homologs, parologs and orthologs



Homologs:

- Similar sequences
- Descent from a common ancestor
- May or may not have a similar function

Paralogs:

- Homologous sequences within a species
- Arose by gene duplication
- May or may not have a similar function

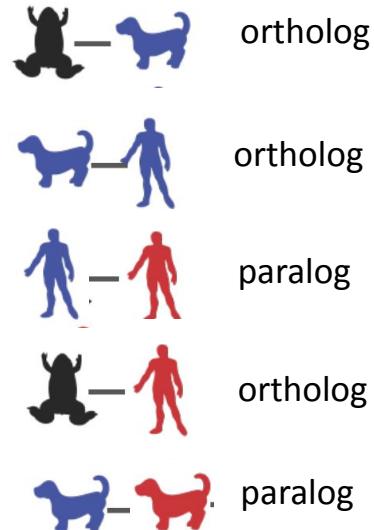
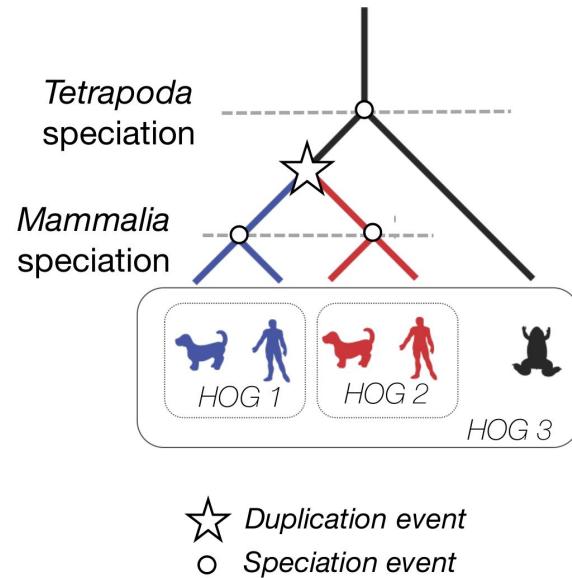
Orthologs:

- Homologous sequences in different species
- Arose from a speciation event in a common ancestor
- May or may not have a similar function

Gene duplication ●

Species split ●

Homology: ortholog or paralog?

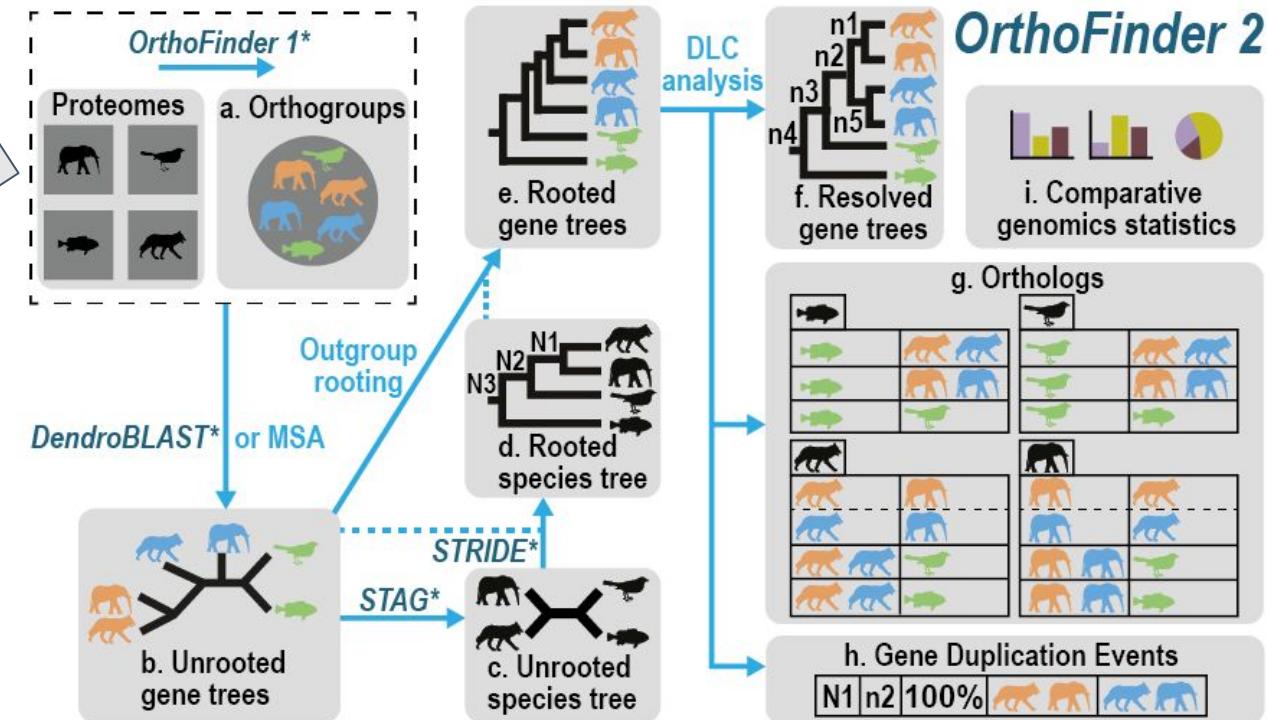


Orthofinder

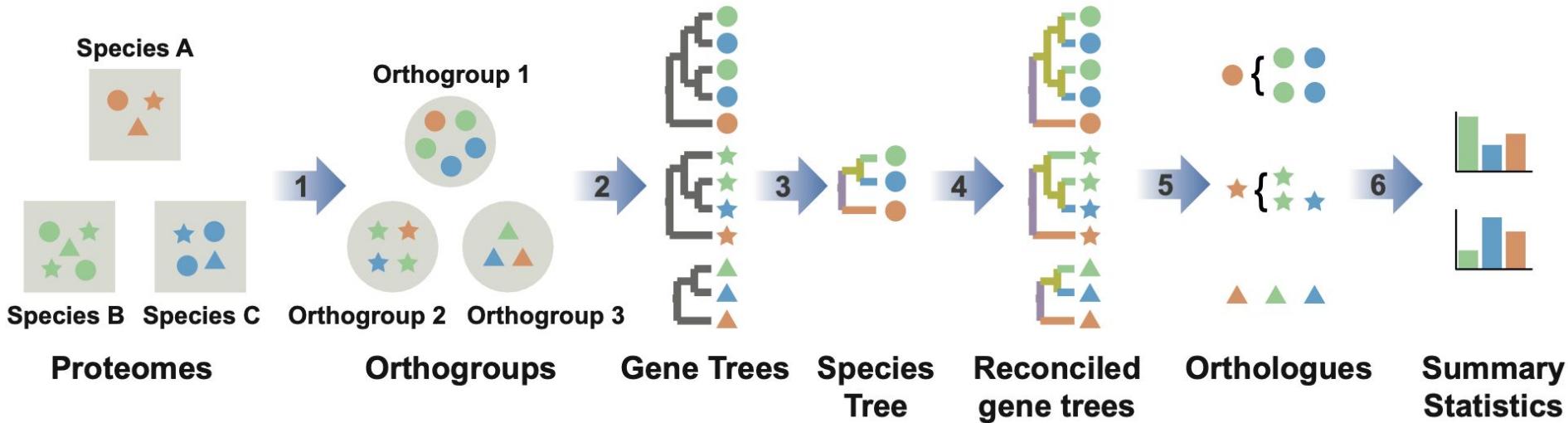
Proteomes:

Fasta files with protein sequences for all genes, for each species

```
GaiGall.fa — Edited
>NSGALG000100111107
MVLEDLGFLRLPWPFTRVWIPPEEEWIKSAEITQDWDGDRSGLLULLEDEGTIVDOLPGRGGGGPLFLNPPILVGNDLT
LGLGDLGVYGMGIPVKLCLYTACAGIHPHCKLPCVCDGTONTSSLKOPPYGLYKRDRSPAYDLDLDEFMEATDTRY
GONTLQEDFGNNNAFRFLKYREKTYCTNDTGTQTAVALAGLLAAQKATGPKTAEHVKVFLGAGRLGEALGIANLIVMA
MVENGVPVEDAARRRIWMDKIGLVRGDRREEKVQNOQEPFAHPAPDSDIPADFEAVATTLKPSAIIVGAGARLFTPSVLOK
SDPFLCAAKASNLPTAAKEAATAYTTLTEGRCLFASGSPDVVTLPDGRSFKAGQNAIYFPGVAVLVISSSVRH
PDVYQHPRNAAPS
```



Orthofinder



Orthofinder

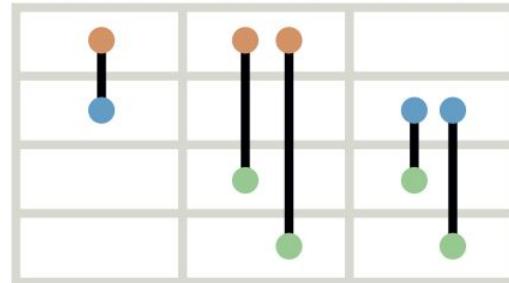
A. Orthogroup



Group of genes descended
from single gene in LCA
of group of species

B. Orthologues

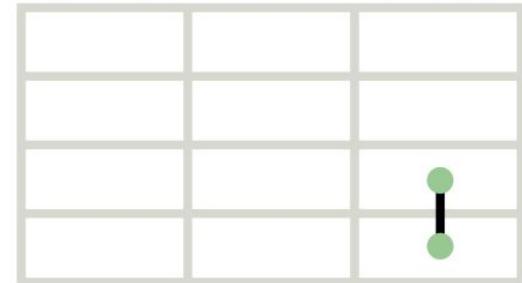
Hu-Mo **Hu-Ch** **Mo-Ch**



Pairs of genes descended
from single gene in LCA
of pair of species

C. Paralogues

Hu-Hu **Mo-Mo** **Ch-Ch**



Pairs of genes descended
from gene duplication
event

Orthofinder

Papers:

<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1832-y>

<https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0721-2?ref=https://githubhelp.com>

Github:

<https://github.com/davidemms/OrthoFinder> Includes a talk you can watch to learn more!

Tutorial:

<https://davidemms.github.io/>

Orthofinder exercise

Objective:

Identifying orthogroups and determine the species tree for Mucoromycota fungi

Fungi

- Absorb nutrients outside their body
 - Release enzymes that break down complex molecules
- Essential in terrestrial ecosystems – recycle nutrients



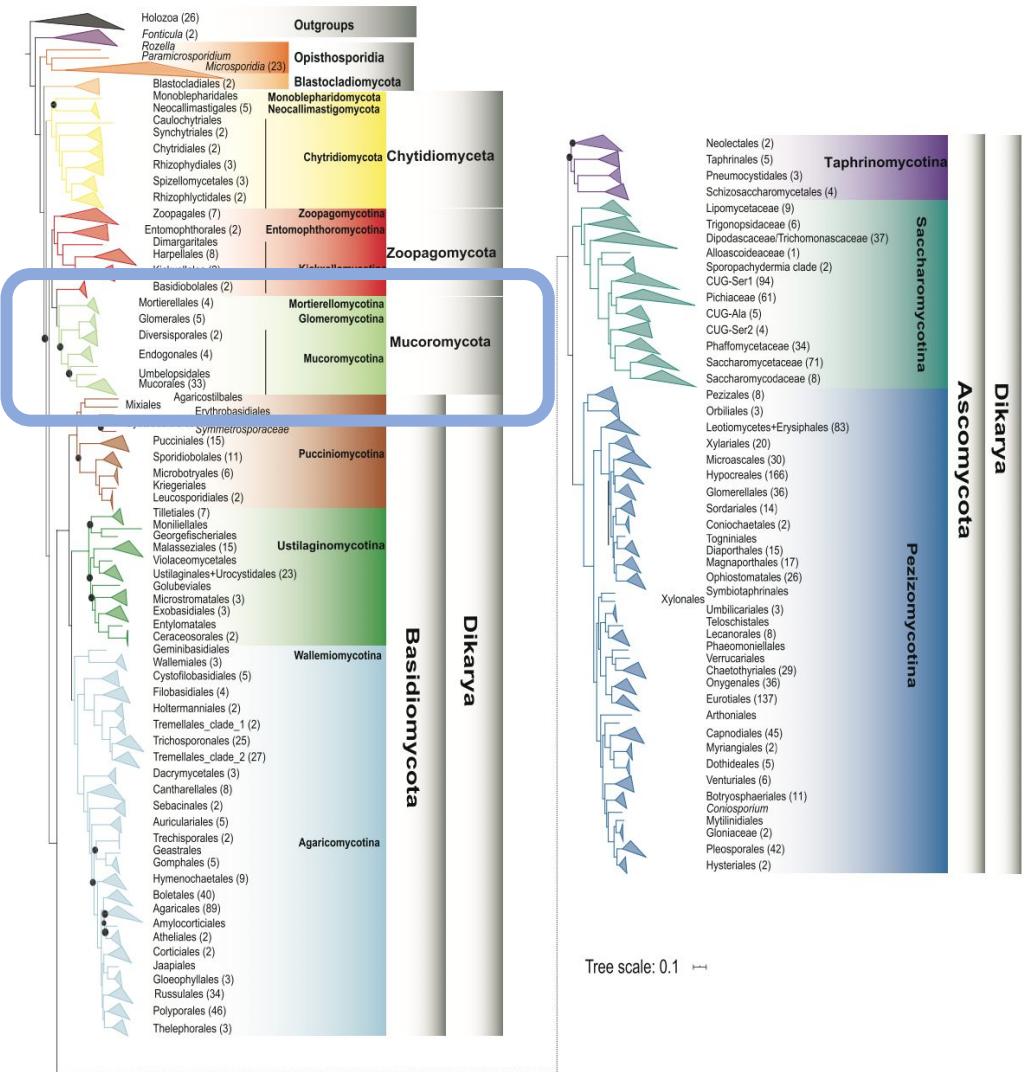
Fungi – biotechnological applications

- Drugs (antibiotics!)
- Beer, wine, bread etc
- Chemical production



Fungi

- 100 000 described species
- Estimated 2-5 million species ++
- Phylogenetic relationships still unresolved



Mucoromycota

- Diverse phylum including various molds
 - (e.g. the common bread molds *Mucor* and *Rhizopus*)
- Mycorrhizal fungi, root endophytes, and plant decomposers

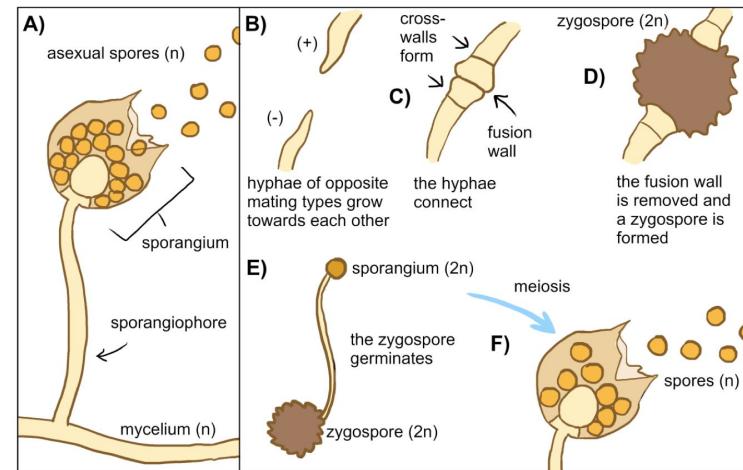
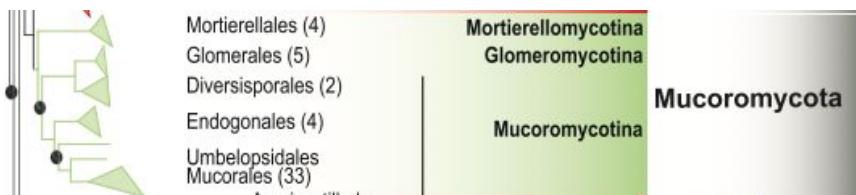
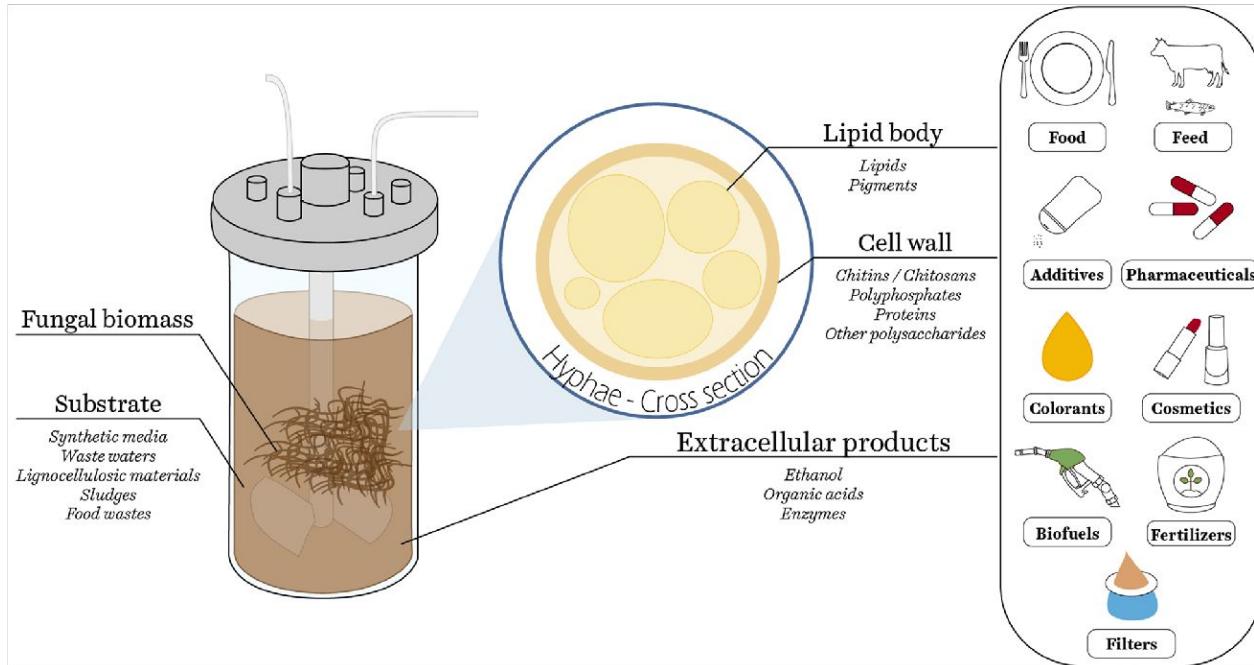


Figure 1.3. Reproduction of Mucoromycota fungi. A) Asexual reproduction by the production of sporangiospores (mitospores). B-F) Sexual reproduction by the production of zygospores. A-D are adapted from Blakeslee (1904), Hocking (1967) and O'donnell et al. (1976). E is adapted from Gauger (1961). Ploidy is denoted by n (haploid) and $2n$ (diploid).

Mucoromycota metabolites



Our data

- Annotated assemblies for 7 species
- Haploid genomes
- Small genomes: 22-50 Mb

Umbelopsis ramanniana

Umbelopsis vinacea

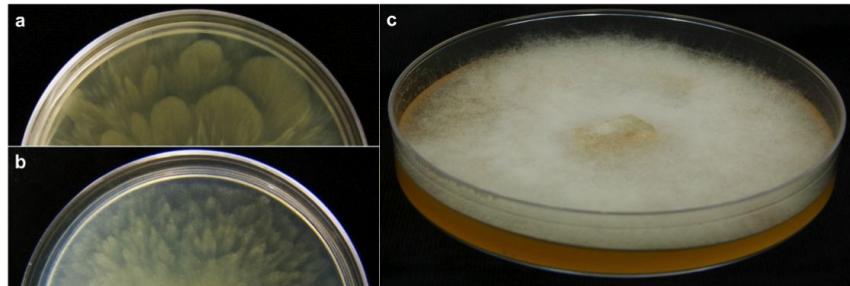
Umbelopsis isabellina

Mortierella zonata

Mortierella alpina

Podila humilis

Linnemannia hyalina

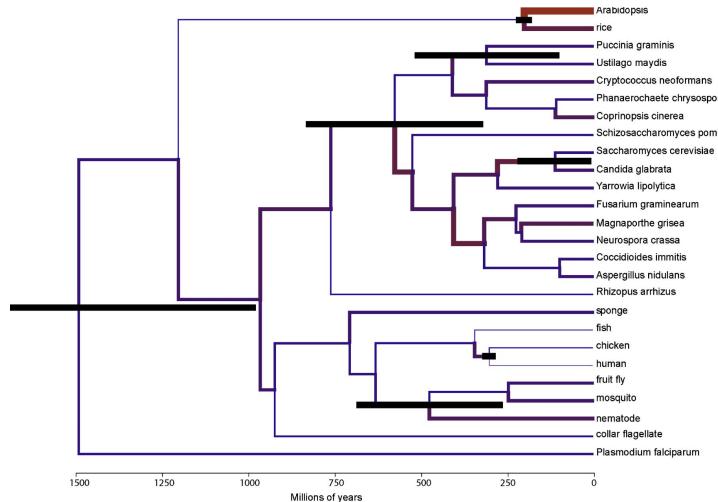


Podila humilis

Outgroup species?

Aspergillus nidulans

- filamentous fungi in the phylum Ascomycota
- Extensive functional annotation, model species
 - beneficial to interpret putative function of genes in orthogroups
- Divergence time from Mucoromycota is 750 MY
 - should add more outgroup species to ensure proper grouping of orthogroups



Outgroup species?

```
wget  
http://ftp.ensemblgenomes.org/pub/fungi/release-54/fa  
sta/aspergillus_nidulans/pep  
/Aspergillus_nidulans.ASM1  
142v1.pep.all.fa.gz -O  
AspNid.pep-transcripts.fa.gz
```

← → ⌂ Not Secure | http://ftp.ensemblgenomes.org/pub/fungi/release-54/fasta/

Index of /pub/fungi/release-54/fasta

	Name	Last modified	Size	Description
	Parent Directory			-
	aphanomyces_astaci/	2022-05-06 17:17	-	
	aphanomyces_invadans/	2022-05-06 14:55	-	
	ashbya_gossypii/	2022-05-06 09:46	-	
	aspergillus_clavatus/	2022-05-06 18:05	-	
	aspergillus_flavus/	2022-05-06 14:49	-	
	aspergillus_fumigatus/	2022-05-06 17:22	-	
	aspergillus_fumigatusa1163/	2022-05-06 19:58	-	
	aspergillus_nidulans/	2022-05-06 16:08	-	
	aspergillus_niger/	2022-05-06 15:23	-	
	aspergillus_oryzae/	2022-05-06 10:05	-	
	aspergillus_terreus/	2022-05-06 15:01	-	
	beauveria_bassiana/	2022-05-06 10:02	-	
	blumeria_graminis/	2022-05-06 15:12	-	
	botrytis_cinerea/	2022-05-06 15:13	-	
	candida_albicans/	2022-05-06 09:49	-	
	candida_auris/	2022-05-06 15:29	-	
	candida_duobushaemulonis/	2022-05-06 15:29	-	
	candida_glabrata/	2022-05-06 16:42	-	

Outgroup species?

Advice:

- pick species as close to your ingroup as possible
 - from public databases, or sequence it yourself
- include a species with functional annotation, a “model species” close to your ingroup (if possible)



GENEONTOLOGY
Unifying Biology

Our data

<https://id.tol.sanger.ac.uk/>



AspNid.pep-transcripts
ggMorZona1.proteins
gzLinHyal1.proteins
gzMorAlpi1.proteins
gzPodHuml1.proteins
gzUmbIlsab1.proteins
gzUmbRama1.proteins
gzUmbVina1.proteins

Orthofinder options

- proteins or DNA?
- different alignment / tree building options

Let's start running Orthofinder!

Please work together in groups of 5-6 when doing this, because OrthoFinder can create lots of small files that might overflow our allocation. So one of 5-6 starts OrthoFinder after discussions.

Evaluation form



Please fill out and give feedback!