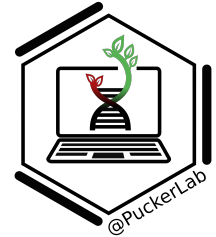
UNIVERSITÄT BONN



Prof. Dr. Boas Pucker

PBPM-BP-07

Availability of slides

- All materials are freely available (CC BY) - after the lectures:
  - eCampus: PBPM0 - Plant Biochemistry, Physiology and Molecular Biology (LEC)
  - GitHub: https://github.com/bpucker/teaching/PBPM

- Questions: Feel free to ask at any time

- Feedback, comments, or questions: pucker[a]uni-bonn.de

UNIVERSITÄT BONN

Repetition

- Transformation methods
- Transformation examples: purple tomato, golden rice
- Genome editing: CRISPR/Cas9
- Legal implications (EU)
- Plant SynBio (iGEM)

Motivation

"Nothing in biology makes sense except in the light of evolution"

Theodosius Dobzhansky

https://biologie-lernprogramme.de/daten/programme/js/homologer/daten/lit/Dobzhansky.pdf

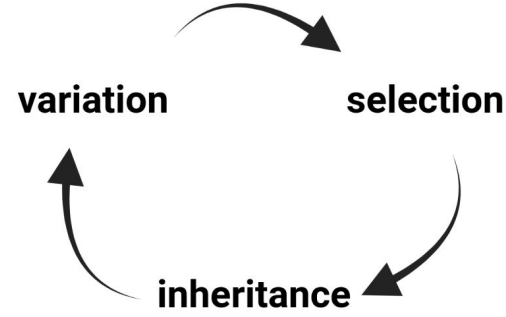
UNIVERSITÄT BONN

Evolutionary concepts

UNIVERSITÄT BONN

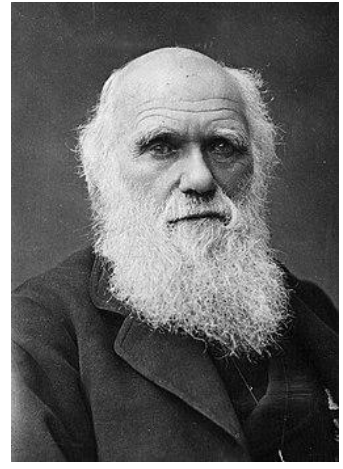
Basis of evolution

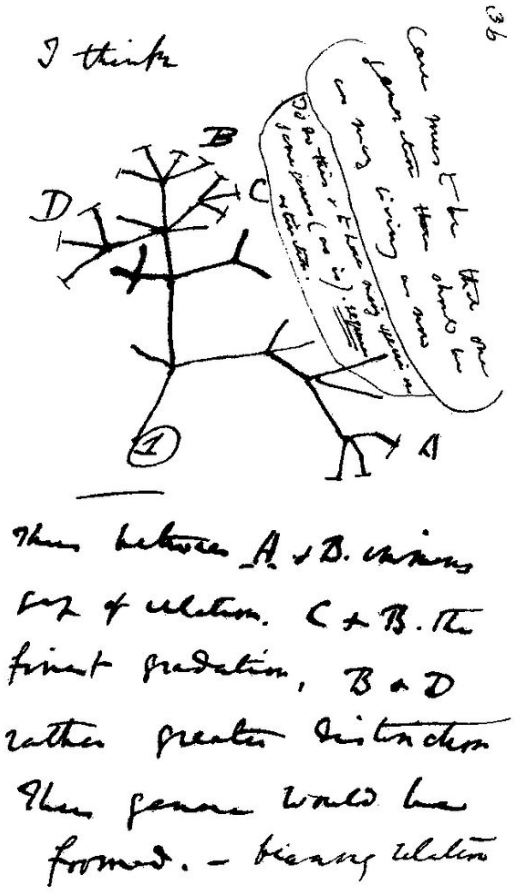
- More offspring than needed
- Genetic basis of traits responsible for fitness
- Selection based on fitness



Charles Darwin: I think

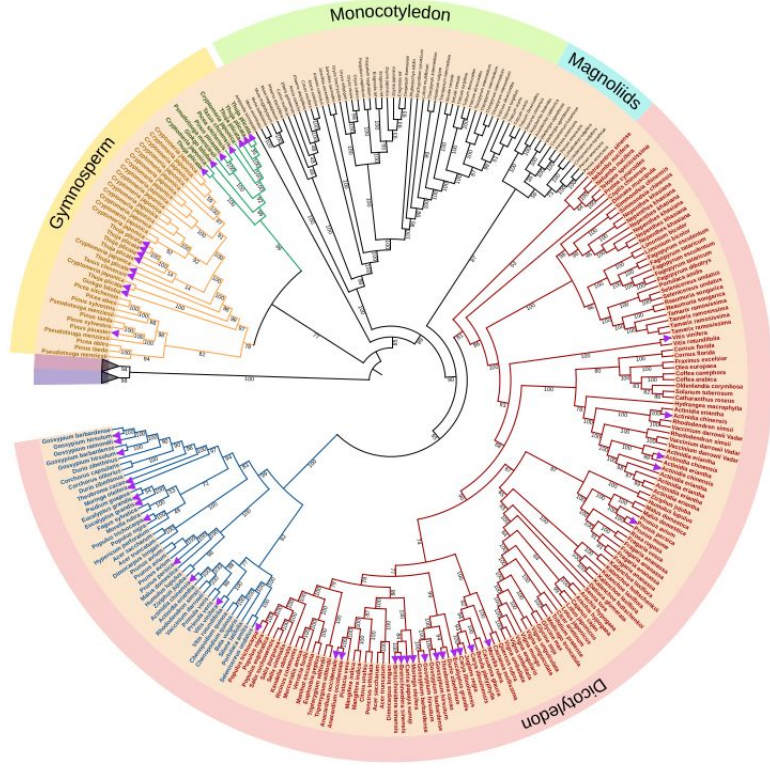
- Developed a theory to explain observations
- Most influential theory in biology?
- Theory vs. hypothesis
- Proving a theory?





UNIVERSITÄT BONN

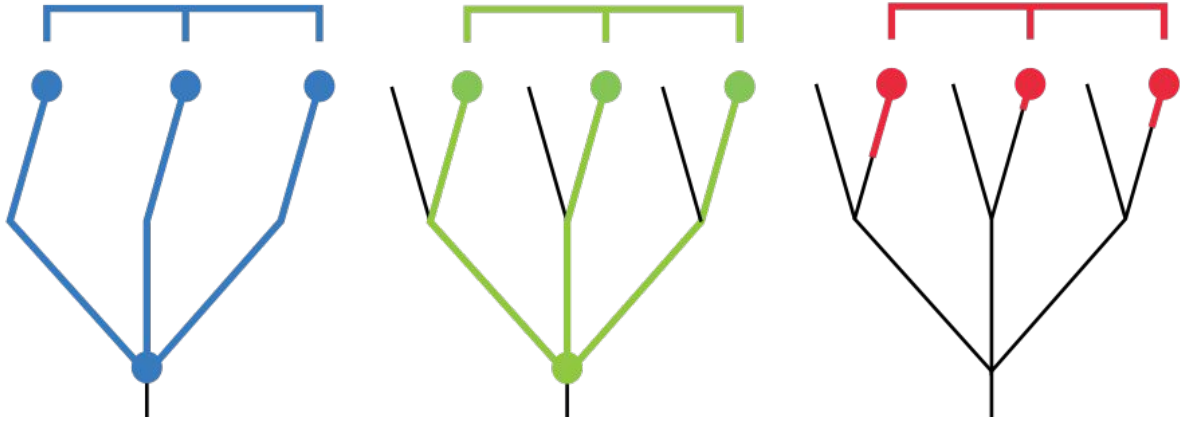
Concept of a tree



https://commons.wikimedia.org/wiki/File:Wild_tree_in_nature.jpg
Marin-Recinos & Pucker, 2025: 10.1101/2025.08.06.668978

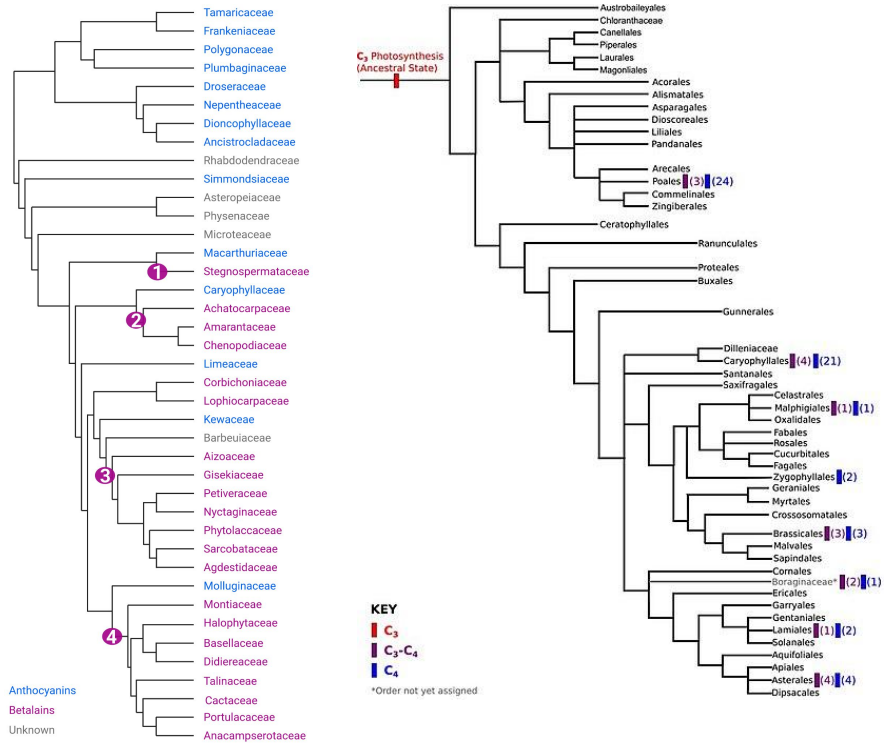
Monophyletic, paraphyletic, and polyphyletic

- Monophyletic = all members of group have common ancestor in the group and all descendants of that ancestor are in group
- Paraphyletic = Group members have common ancestor in the group, but not all descendants of that ancestor are included in group
- Polyphyletic = Group members do not have common ancestor in the group



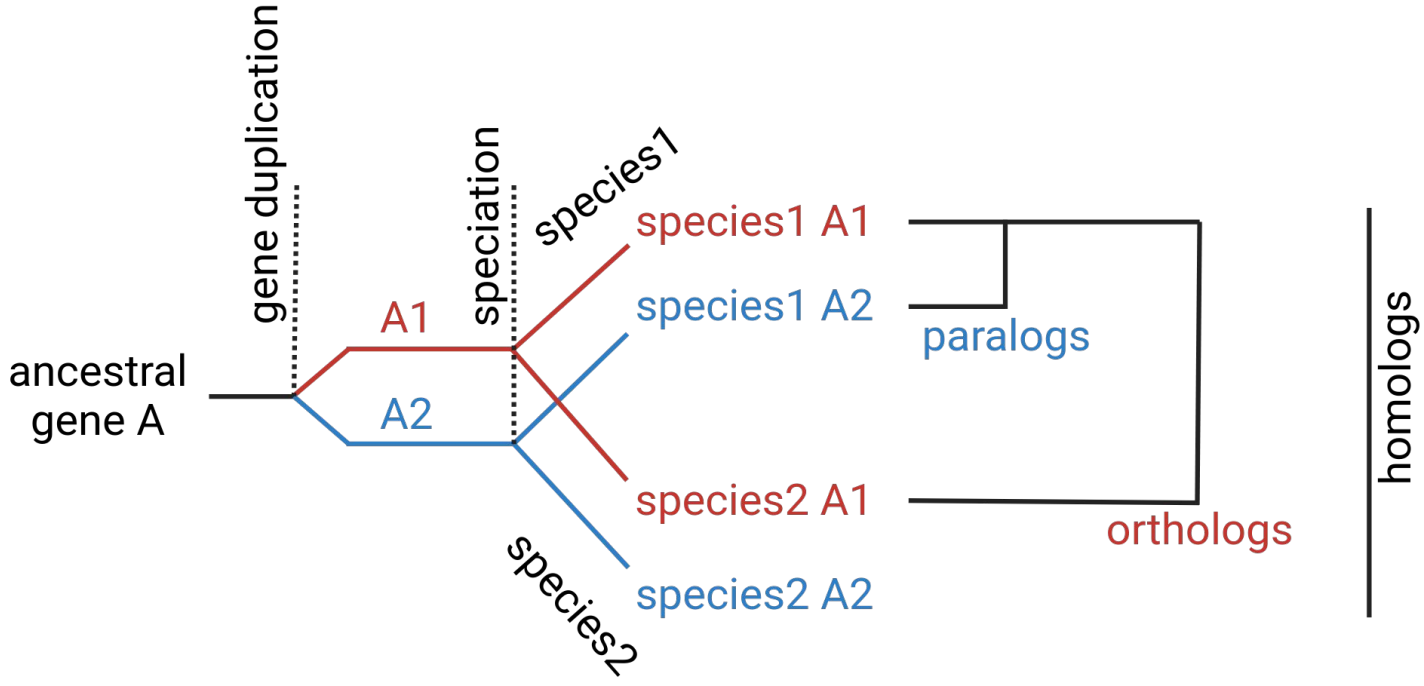
Convergent and independent evolution

- Trait appears multiple times independently
- Same genes can be involved and modified in similar ways
- Examples: pigment biosynthesis in Caryophyllales; photosynthesis type



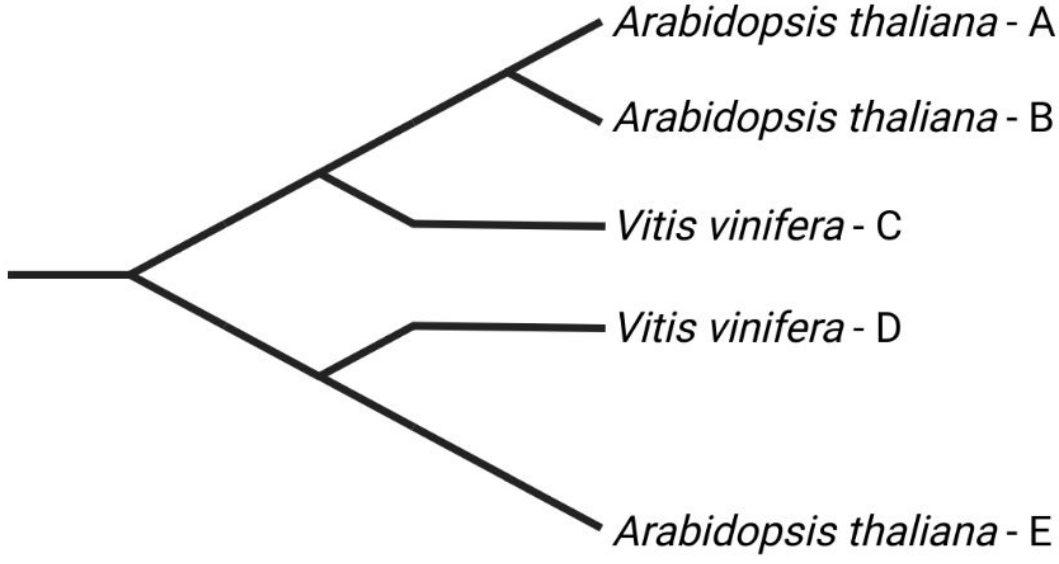
Timoneda *et al.*, 2019: 10.1111/nph.15980
Sheehan *et al.*, 2020: 10.1111/nph.16089
Williams et al., 2013: https://doi.org/10.7554/eLife.00961

Homologs, paralogs, orthologs

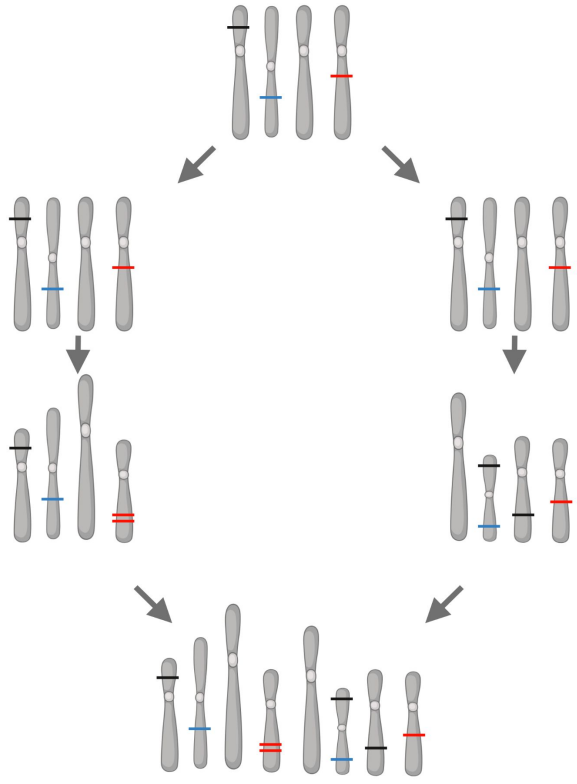


UNIVERSITÄT BONN

Which are orthologs/paralogs?



Homeologs



Speciation

Gene duplications, translocations, and rearrangements

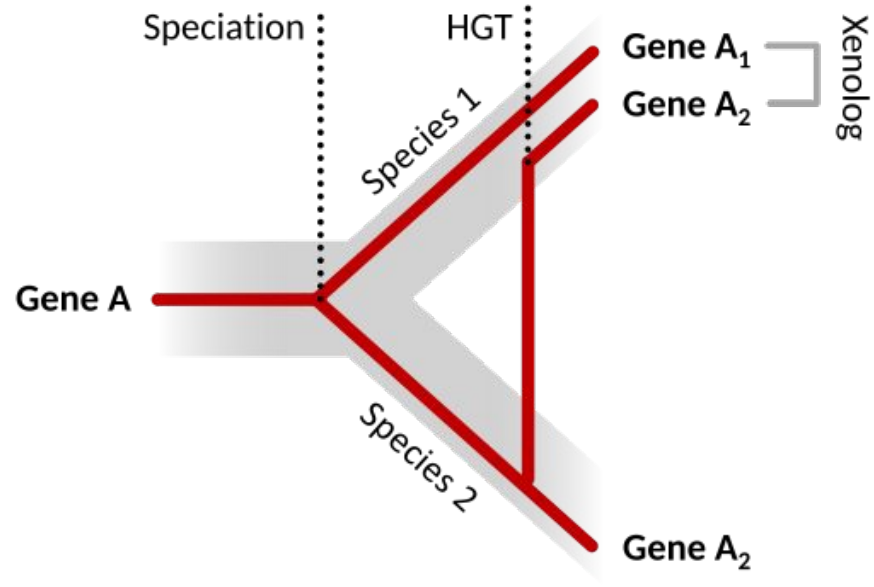
Hybridization leads to polyploidization

Glover et al., 2016: 10.1016/j.tplants.2016.02.005

UNIVERSITÄT BONN

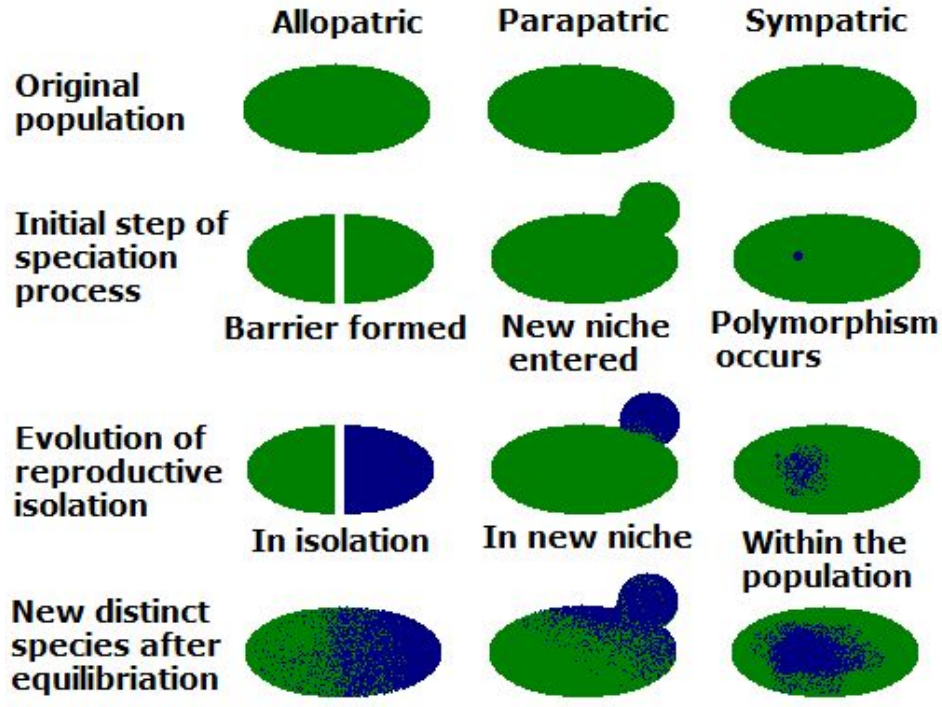
Xenologs

- Genes can be moved between species (horizontal gene transfer)
- Relationships turn from tree into networks
- Bacteria/viruses can transfer genes between plant species



Speciation

- sympatric = speciation due to genetic changes
- parapatric = speciation in a niche
- allopatric = physical separation of population leads to speciation



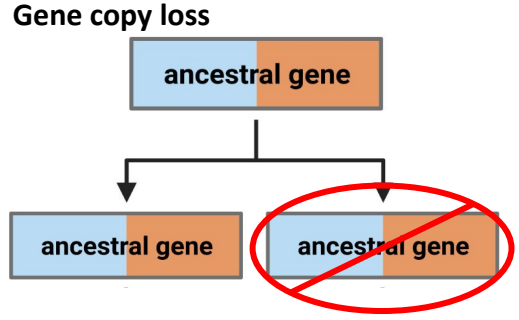
https://en.wikipedia.org/wiki/File:Speciation_modes.png

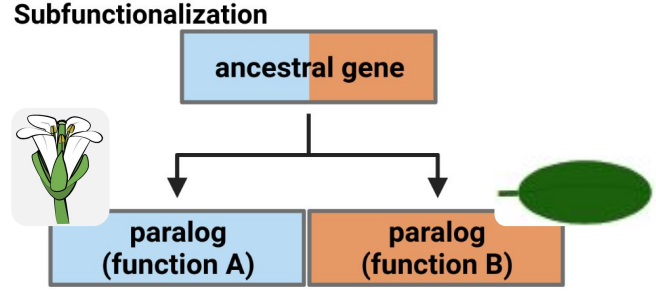
UNIVERSITÄT BONN

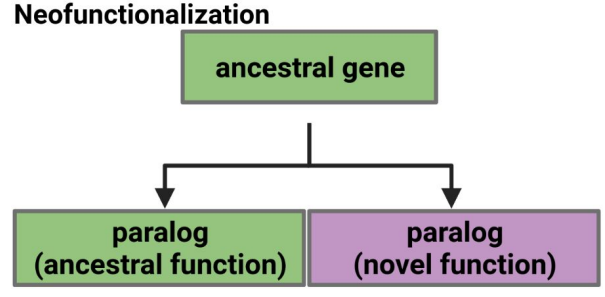
Gene duplication

Importance of gene duplications in plants

- 65% of annotated plant genes have a copy
- Most copies do not persist during evolution
- Gene duplications are crucial for metabolic novelty in specialized plant metabolism

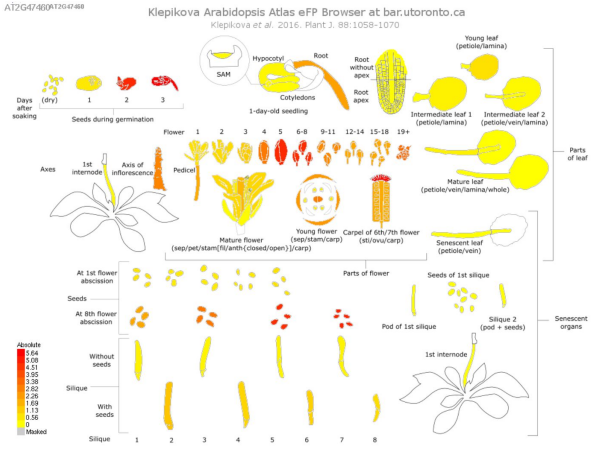


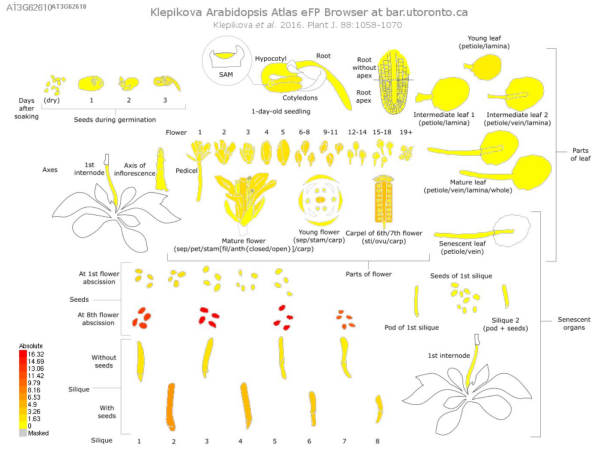


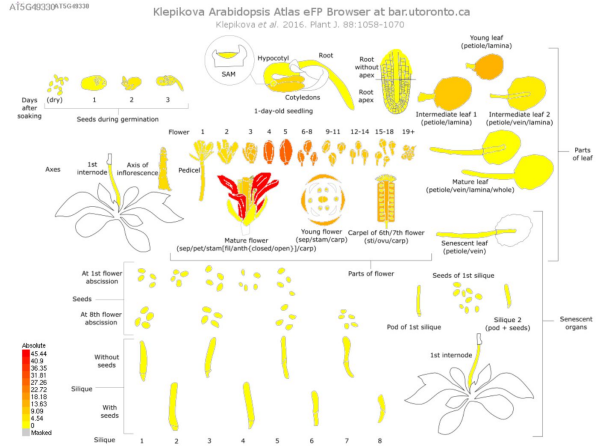


Panchy et al., 2016: 10.1104/pp.16.00523

Example: SG7 MYB expression

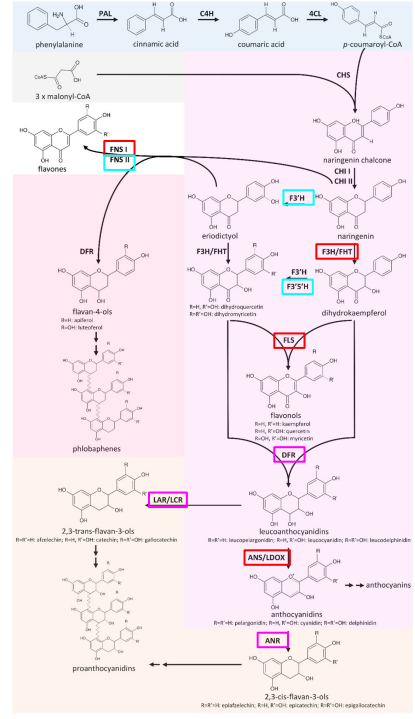






Importance of gene duplications in plants

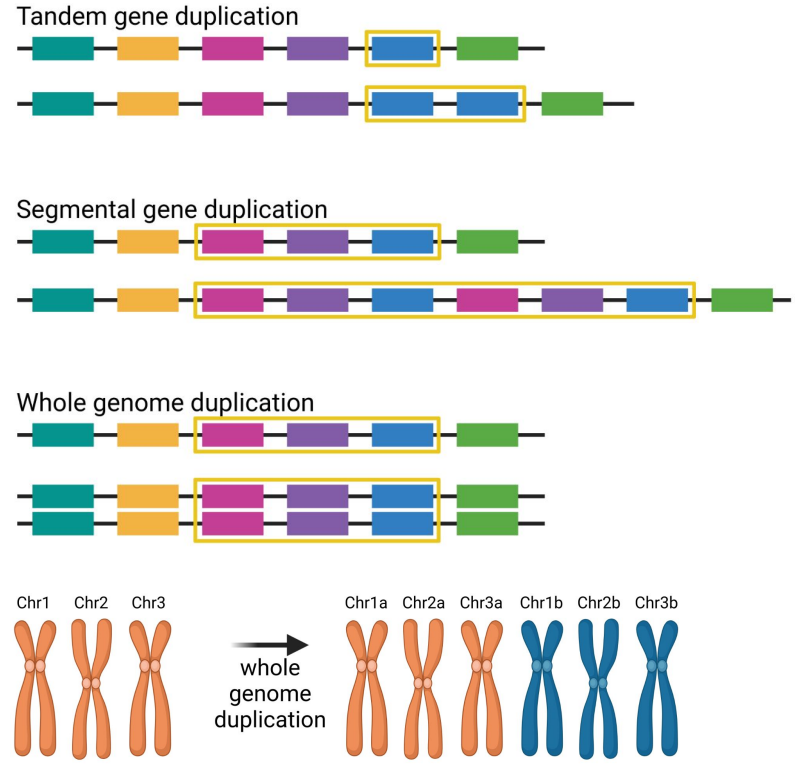
- 65% of annotated plant genes have a copy
- Most copies do not persist durin evolution
- Gene duplications are highly relevance in specialized metabolism:
  - FNSI / F3H / FLS / ANS (2-ODDs)
  - FNS II / F3’H / F3’5’H (CYPs)
  - DFR / LAR / ANR (SDRs)



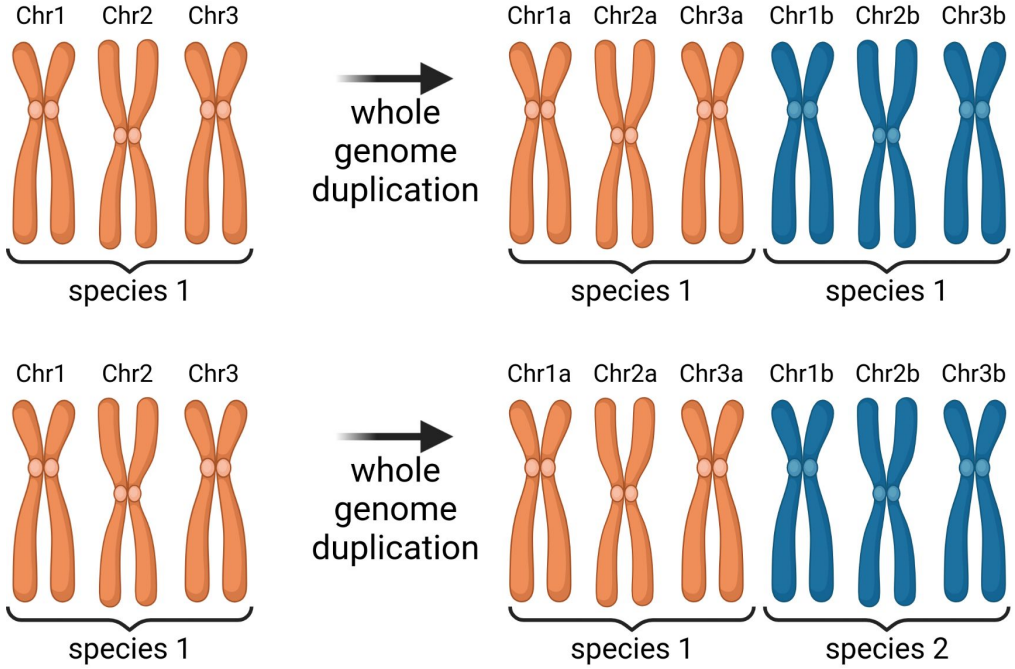
Panchy et al., 2016: 10.1104/pp.16.00523
Pucker et al., 2020: 10.3390/plants9091103

Gene duplication mechanisms

- Tandem duplication: copies are located next to each other on same chromosome
- Segment duplication: like tandem duplications, but multiple adjacent genes are also duplicated
- Whole genome duplication (WGD): all genes are duplicated

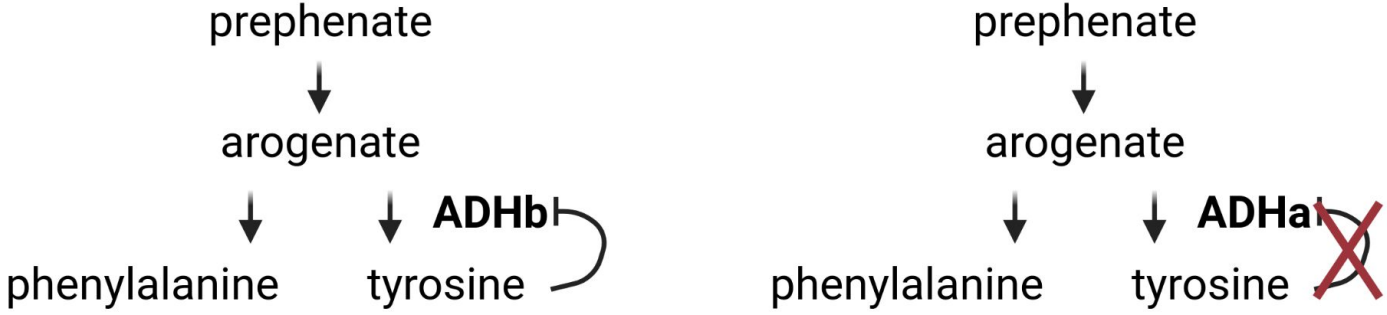


WGD: autopolyploidy vs. allopolyploidy



Example: ADH in Caryophyllales

- ADH was duplicated in Caryophyllales
  - ADHa = feedback-resistance enzyme
  - ADHb = feedback-inhibited enzyme

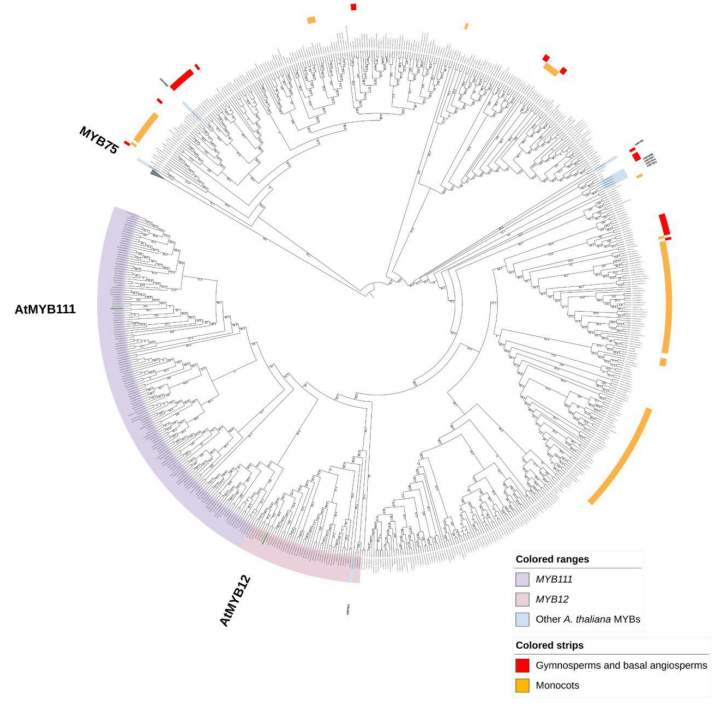


Lopez-Nieves et al., 2017: 10.1111/nph.14822

Example: Flavonol regulating MYBs

- AtMYB111 and AtMYB12 are regulators of the flavonol biosynthesis

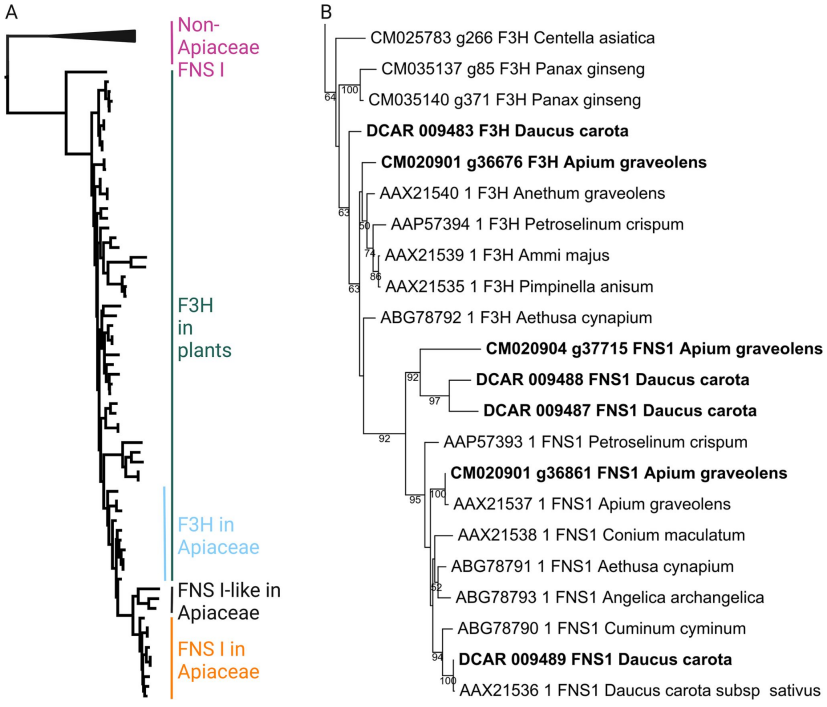
- Deep duplication of a flavonol-regulating MYB in angiosperms

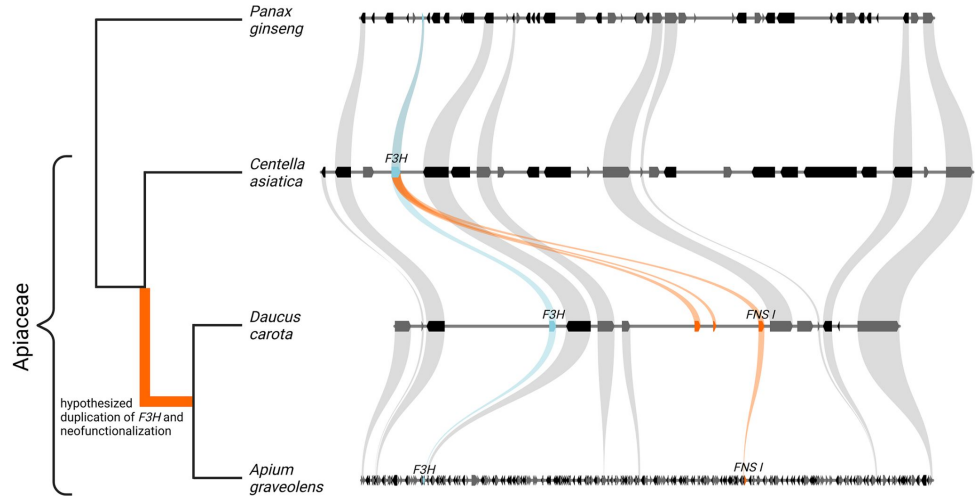


Natarajan & Pucker, 2025: 10.1101/2025.10.10.681656

UNIVERSITÄT BONN

Example: *F3H/FNS I* in Apiaceae





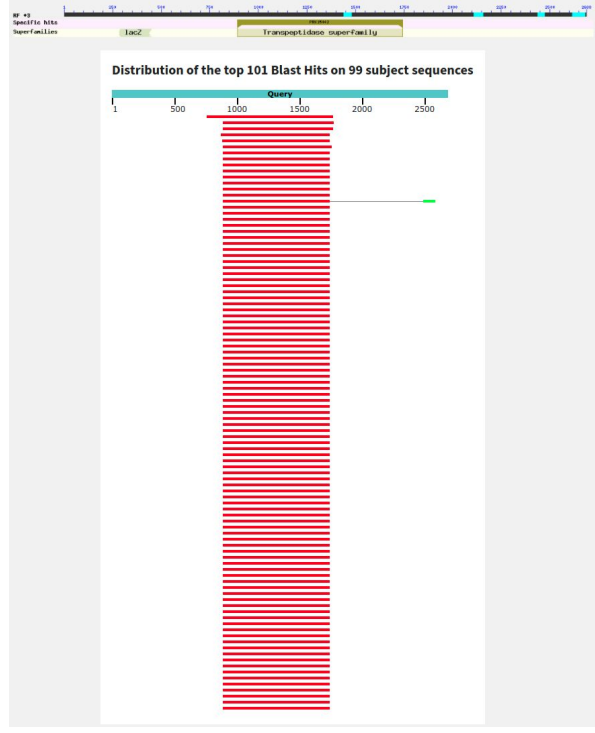
https://doi.org/10.1371/journal.pone.0280155

UNIVERSITÄT BONN

Tools in phylgenetics

Collecting sequences: BLAST

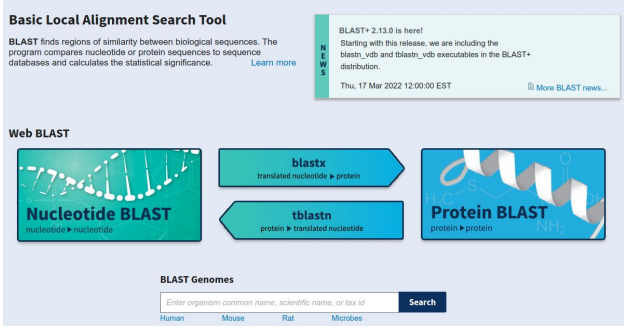
- BLAST = Basic Local Alignment Search Tool
- Identification of short stretches of sequence similarity
- Quick search against a large database
- Sequence similarity suggest homology i.e. common ancestors and similar function

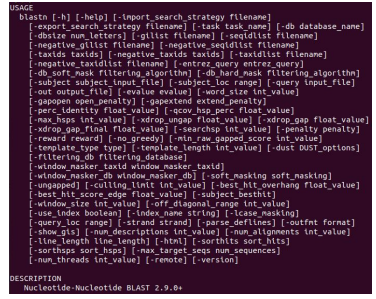


UNIVERSITÄT BONN

Online vs. local BLAST

NCBI website	Local BLAST
Convenient to use	Requires some command line knowledge
Requires no local computational resources	Computational resources needed
Transfer of query sequence	Query kept secret
Large databases available	Download of databases required





BLAST alternatives

- HMMER: http://hmmer.org/
  - Dedicated to the identification of specific domains

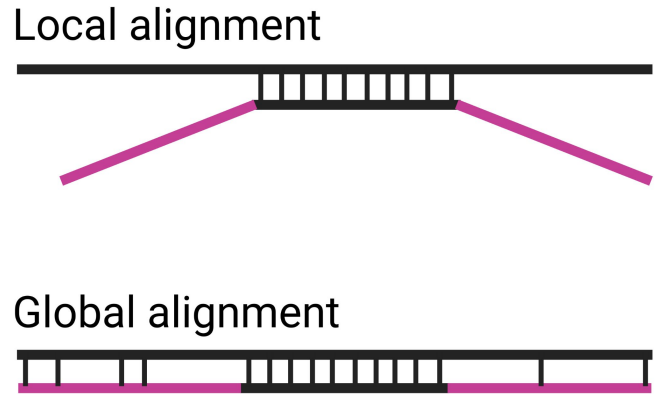
- DIAMOND: https://github.com/bbuchfink/diamond
  - Much faster, but higher memory requirements

https://github.com/bbuchfink/diamond

Local vs. global alignment

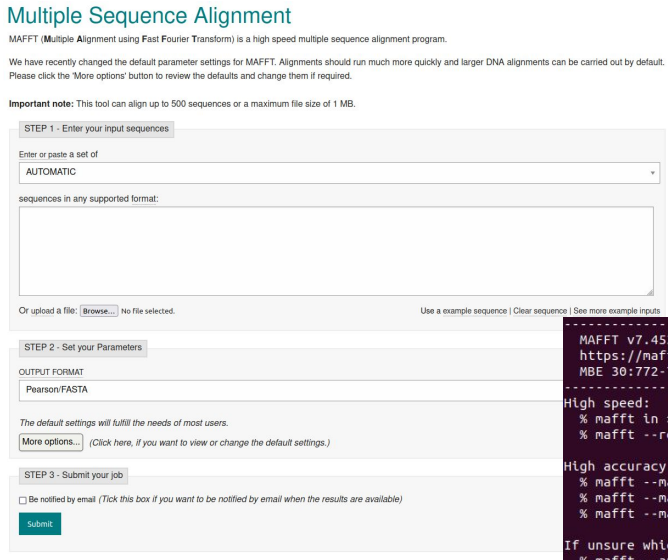
- Local alignment:
  - highlights small stretches with high similarity
  - random hit possible
  - better for strong sequence length differences

- Global alignment:
  - only works by overall high similarity
  - similar sequence length required
  - full length similarity supports similar function



Alignment construction

- MAFFT
- MUSCLE
- CLUSTAL (omega, W)




```
--------------------------------------------------------------------------------
  MAFFT v7.453 (2019/Nov/8)
  https://mafft.cbrc.jp/alignment/software/
  MBE 30:772-780 (2013), NAR 30:3059-3066 (2002)
--------------------------------------------------------------------------------
High speed:
  % mafft in > out
  % mafft --retree 1 in > out (fast)

High accuracy (for <~200 sequences x <~2,000 aa/nt):
  % mafft --maxiterate 1000 --localpair  in > out (% linsi in > out is also ok)
  % mafft --maxiterate 1000 --genafpair  in > out (% einsi in > out)
  % mafft --maxiterate 1000 --globalpair in > out (% ginsi in > out)

If unsure which option to use:
  % mafft --auto in > out

--op # :         Gap opening penalty, default: 1.53
--ep # :         Offset (works like gap extension penalty), default: 0.0
--maxiterate # : Maximum number of iterative refinement, default: 0
--clustalout :   Output: clustal format, default: fasta
--reorder :      Outorder: aligned, default: input order
--quiet :        Do not report progress
--thread # :     Number of threads (if unsure, --thread -1)
--dash :         Add structural information (Rozewicki et al, submitted)
```


https://www.ebi.ac.uk/Tools/msa/mafft/

UNIVERSITÄT BONN

Alignment construction

- Letters representing amino acid residues are arranged to highlight similarities
- Gaps (-) are inserted to show differences

Input FASTA file with sequences

```
>seq1
MDTEKYMEKWIDQGHALFPEEDQ
>seq2
MDTDIKYSEKWIQGHSLFPDDQ
>seq3
METEKYMEKWIQGHSIFPEDQ
>seq4
MDTEIKYMEKWIQGHALFPDDQP
```

Alignment of sequences

```
seq1   MDTE-KYMEKWIDQGHALFPEEDQ-
seq2   MDTDIKYSEKWI-QGHSLFPD-DQ-
seq3   METE-KYMEKWI-QGHSIFPE-DQ-
seq4   MDTEIKYMEKWI-QGHALFPD-DQP
```

UNIVERSITÄT BONN

Alignment formats: CLUSTAL vs. FASTA

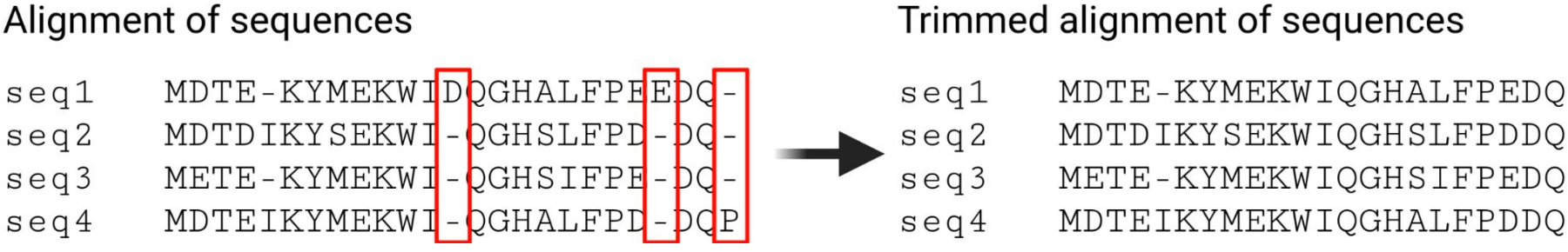
```
AthCHS          MVMAGASSLDEIRQAQRADGPAGILAIGTANPENHVLQAEYPDYYFRITNSEHMTDLKEK
BvuCHS          M---ATPSVQEIRDAQRSNGPATILAIGTANPANEMYQAEYPDFYFRVTKSEHMSELKQK
                *   .:.*::***:***::*** ********* *.: ******:***:*:****::**:*

AthCHS          FKRMCDKSTIRKRHMHLTEEFLKENPHMCAYMAPSLDTRQDIVVVEVPKLGKEAAVKAIK
BvuCHS          FKRMCDKSMIKKRYMHVTQELLEENPHMCDYNASSLNTRQDILATEVPKLGKEAAVKAIK
                ******** *:**:**:*:*:*:****** * *.**:*****:..***************

AthCHS          EWGQPKSKITHVVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRIAK
BvuCHS          EWGQPRSKITHVIFCTTSGVDMPGADYQLTKLLGLRPSVKRFMLYQQGCYAGGTVLRLAK
                *****:******:*****************************:*:*****:*******:**

AthCHS          DLAENNRGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFSDGAAALIVGSDPDTSVGEK
BvuCHS          DIAENNRGARVLVVCAEITVICFRGPTETHLDSMIGQALFGDGAGAVIVGADLDESI-ER
                *:**************:***.: ****::*****::******.***.*:***:* * *: *:

AthCHS          PIFEMVSAAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNIVKSLDEAFKPLGISD
BvuCHS          PIFQLVWAAQTILPDSEGAIDGHLREVGLAFHLLKDVPGLISKNIEKALVEAFKPIGIDD
                ***::* *********:************:*************** *:* *****:**.*

AthCHS          WNSLFWIAHPGGPAILDQVEIKLGLKEEKMRATRHVLSEYGNMSSACVLFILDEMRRKSA
BvuCHS          WNSIFWAAHPGGPAILDQVESKLGLKQDKLSTTRHVLSEFGNMSSACVLFILDEMRKRSM
                ***:** ************* *****::*: :*******:****************::*

AthCHS          KDGVATTGEGLEWGVLFGFGPGLTVETVVLHSVPL--
BvuCHS          KEGMATTGEGLEWGVLFGFGPGLTVETVMLHSVPIAN
                *:*:*************************:*****:
```

```
>AthCHS
MVMAGASSLDEIRQAQRADGPAGILAIGTANPENHVLQAEYPDYYFRITNSEHMTDLKEK
FKRMCDKSTIRKRHMHLTEEFLKENPHMCAYMAPSLDTRQDIVVVEVPKLGKEAAVKAIK
EWGQPKSKITHVVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMYQQGCFAGGTVLRIAK
DLAENNRGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFSDGAAALIVGSDPDTSVGEK
PIFEMVSAAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNIVKSLDEAFKPLGISD
WNSLFWIAHPGGPAILDQVEIKLGLKEEKMRATRHVLSEYGNMSSACVLFILDEMRRKSA
KDGVATTGEGLEWGVLFGFGPGLTVETVVLHSVPL--
>BvuCHS
M---ATPSVQEIRDAQRSNGPATILAIGTANPANEMYQAEYPDFYFRVTKSEHMSELKQK
FKRMCDKSMIKKRYMHVTQELLEENPHMCDYNASSLNTRQDILATEVPKLGKEAAVKAIK
EWGQPRSKITHVIFCTTSGVDMPGADYQLTKLLGLRPSVKRFMLYQQGCYAGGTVLRLAK
DIAENNRGARVLVVCAEITVICFRGPTETHLDSMIGQALFGDGAGAVIVGADLDESI-ER
PIFQLVWAAQTILPDSEGAIDGHLREVGLAFHLLKDVPGLISKNIEKALVEAFKPIGIDD
WNSIFWAAHPGGPAILDQVESKLGLKQDKLSTTRHVLSEFGNMSSACVLFILDEMRKRSM
KEGMATTGEGLEWGVLFGFGPGLTVETVMLHSVPIAN
```

UNIVERSITÄT BONN

Alignment trimming

- Removal of alignment columns with low abundance
- Columns with low abundance are not informative
- Reduced alignment lengths makes tree construction easier



UNIVERSITÄT BONN

Tree inference

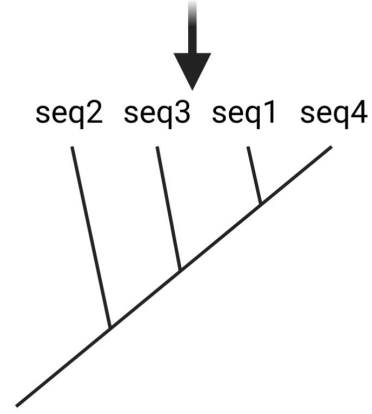
- Construction of phylogenetic trees is based on sequence alignments
- Similar sequences (recent shared ancestors) are grouped
- Tools: RAxML, FastTree, IQ-TREE, MEGA

Trimmed alignment of sequences

```
seq1    MDTE-KYMEKWIQGHALFPEDQ
seq2    MDTDIKYSEKWIQGHSLFPDDQ
seq3    METE-KYMEKWIQGHSIFPEDQ
seq4    MDTEIKYMEKWIQGHALFPDDQ
```

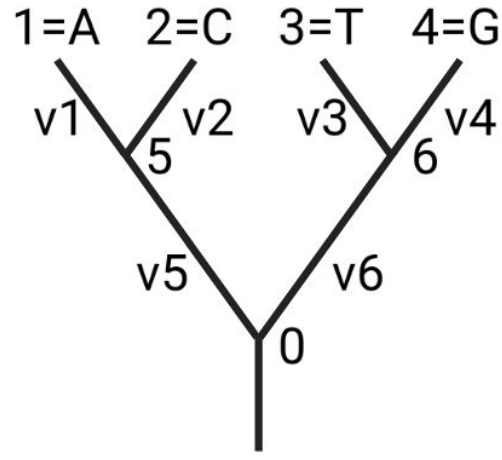
Distance calculation

	seq1	seq2	seq3	seq4
seq1	.	5	3	2
seq2	.	.	6	3
seq3	.	.	.	5
seq4	.	.	.	.



Maximum likelihood (ML) trees

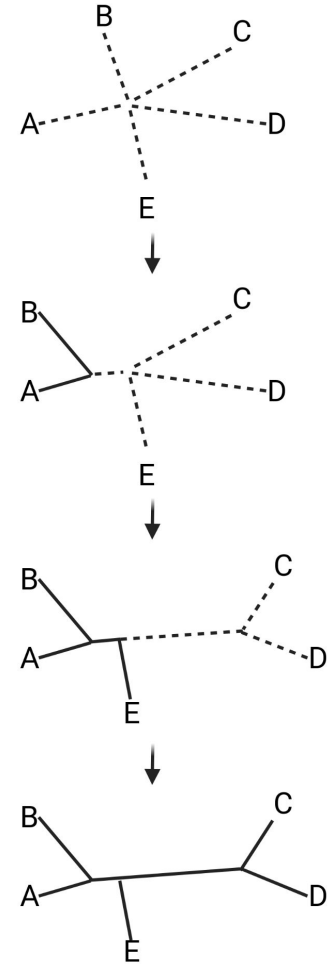
- Finding the tree with highest likelihood to explain the observed sequences
- Number of possible trees depends on number of sequences
- Often impossible to explore all possible trees (very computationally demanding)
- Tool: RAxML-NG, IQ-TREE



Cho, 2012: https://lnk.tu-bs.de/Ym3RKP
https://github.com/amkozlov/raxml-ng

Neighbour Joining (NJ) tree

- Combination of sequences (groups of sequences) with smallest distance
- Quick and computationally cheap approach
- Less reliable than maximum likelihood (ML) trees
- Tool: MEGA

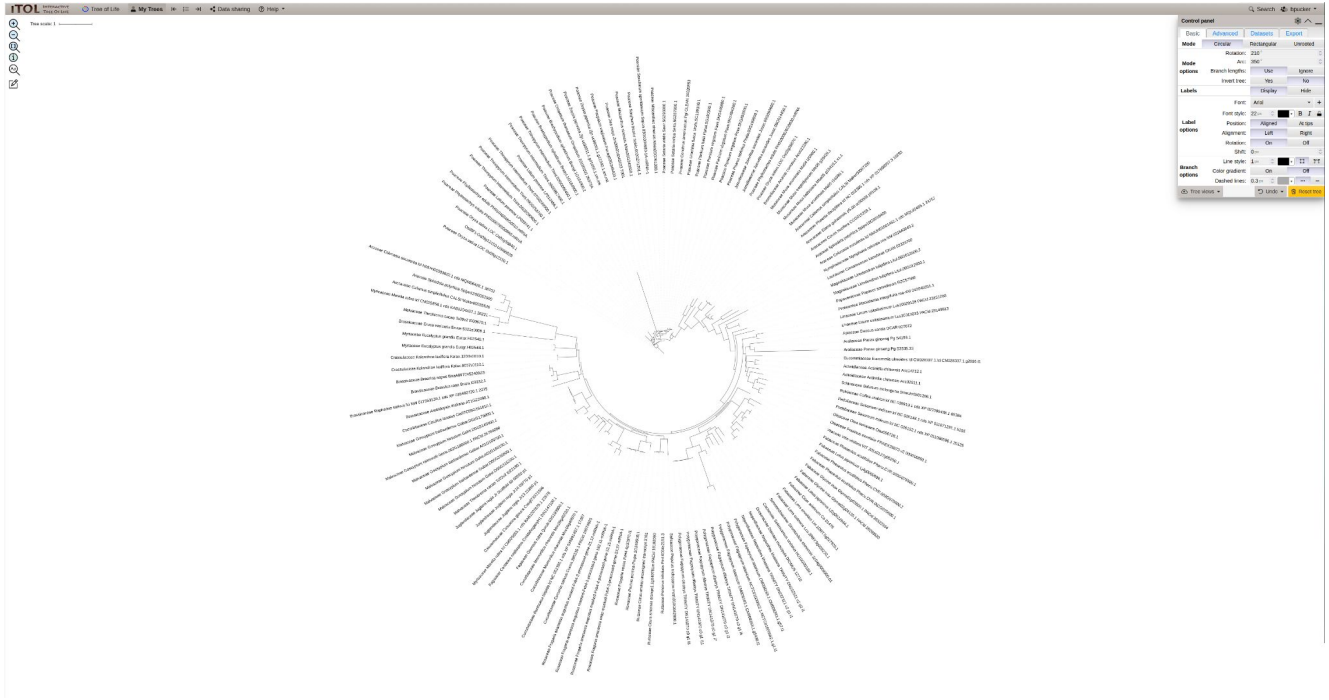


Kumar et al., 2016: 10.1093/molbev/msw054

UNIVERSITÄT BONN

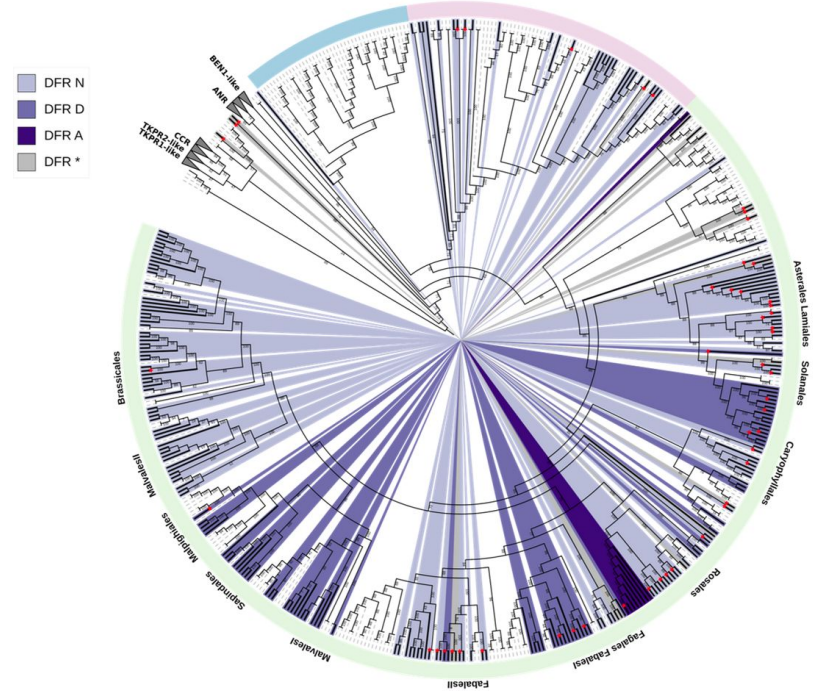
Tree visualization

- iTOL: https://itol.embl.de
- FigTree: https://tree.bio.ed.ac.uk/software/figtree/



Example: DFR evolution

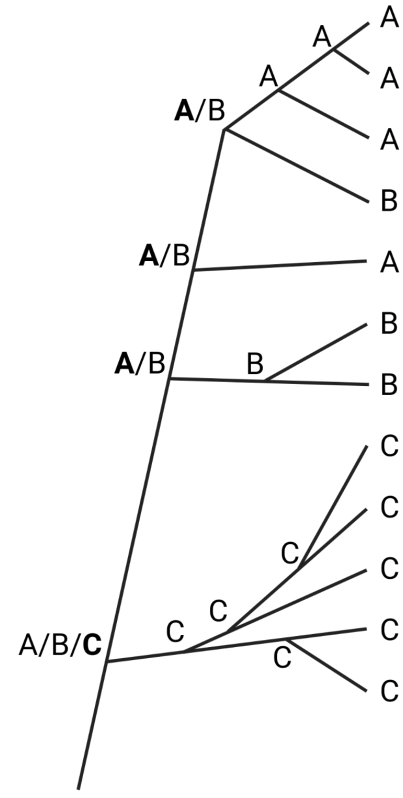
- Evolution of *DFR* in land plants
- Multiple clades visible
- Sequences of closely related species cluster in lineages



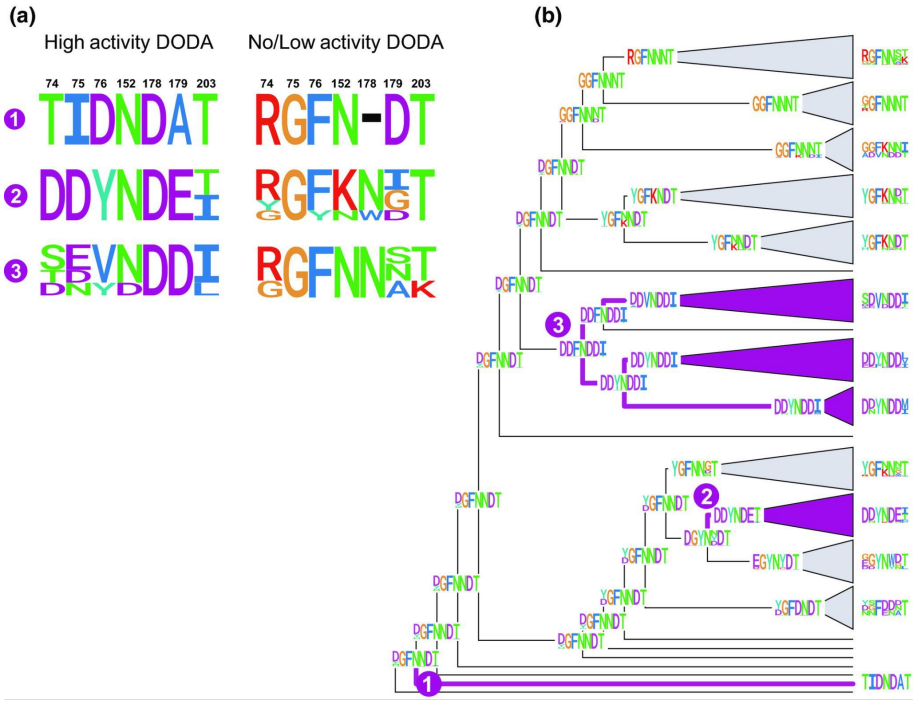
Choudhary & Pucker, 2023: 10.1371/journal.pone.0305837

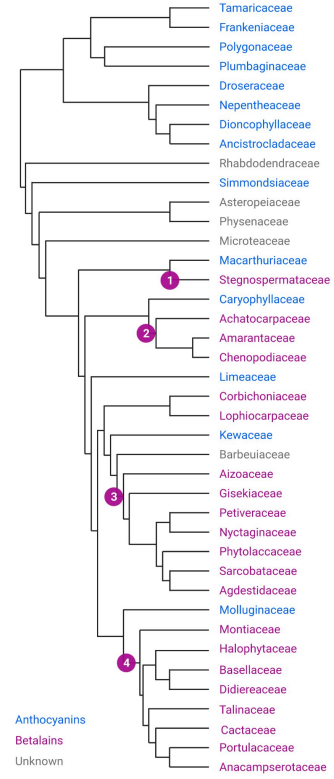
Ancestral state reconstruction

- Identify the state of a certain sequence/trait in a last common ancestor
- Multiple states are possible in some cases
- Ancestral states are often predicted with a probability



Example: ancestral state reconstruction





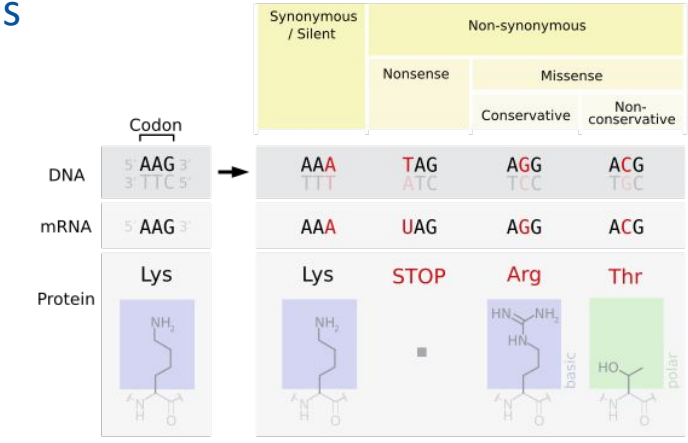
Sheehan et al., 2020: 10.1111/nph.16089

Selection pressure analysis

- Nonsynonymous (Ka) are compared against synonymous (Ks) substitutions in protein-coding genes

- Interpretation:
  - Ka/Ks < 1 ⟶ purifying selection
  - Ka/Ks = 1 ⟶ neutral evolution
  - Ka/Ks > 1 ⟶ positive selection

- Requires orthologous protein-coding sequences, correct codon alignment, and sufficient evolutionary divergence for reliable estimates

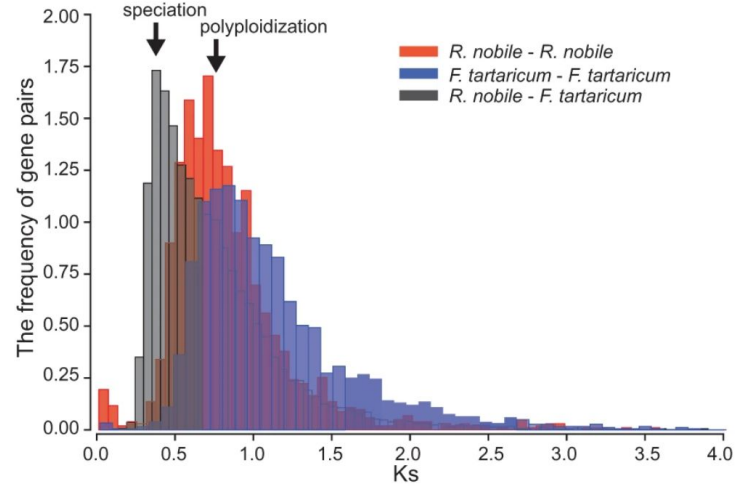


https://commons.wikimedia.org/wiki/File:Point_mutations-en.svg

UNIVERSITÄT BONN

Dating evolutionary events

- Ks distribution peak reveal timing of duplication events (e.g., WGDs)
- Low vs. high Ks values indicate recent vs. ancient evolutionary events
- Fossil calibrations provide absolute time anchors
- Limitations: rate heterogeneity, misidentified orthologs/paralogs, and incomplete fossil records



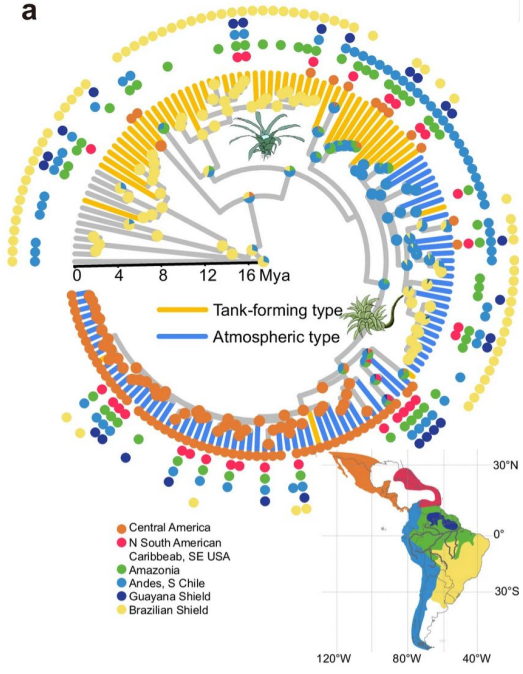
Feng et al., 2022: 10.1038/s42003-023-05044-1

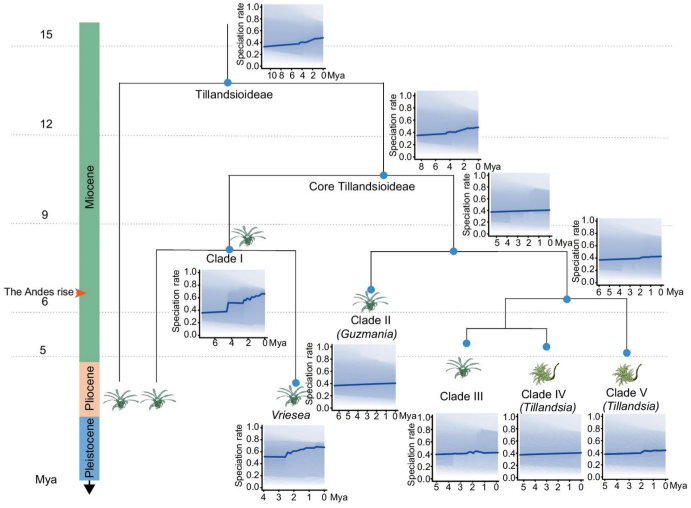
UNIVERSITÄT BONN

Evolution of plant traits

Example: Tillandsioideae

- Tank formation to capture water and nutrients
- Absorptive trichomes on tank forming leaves
- Development of air roots

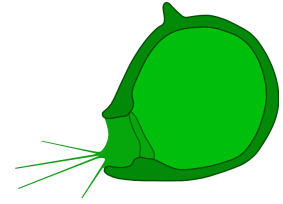




Lyu et al., 2024: 10.1038/s41467-024-53756-7

Example: Carnivory in Lentibulariaceae

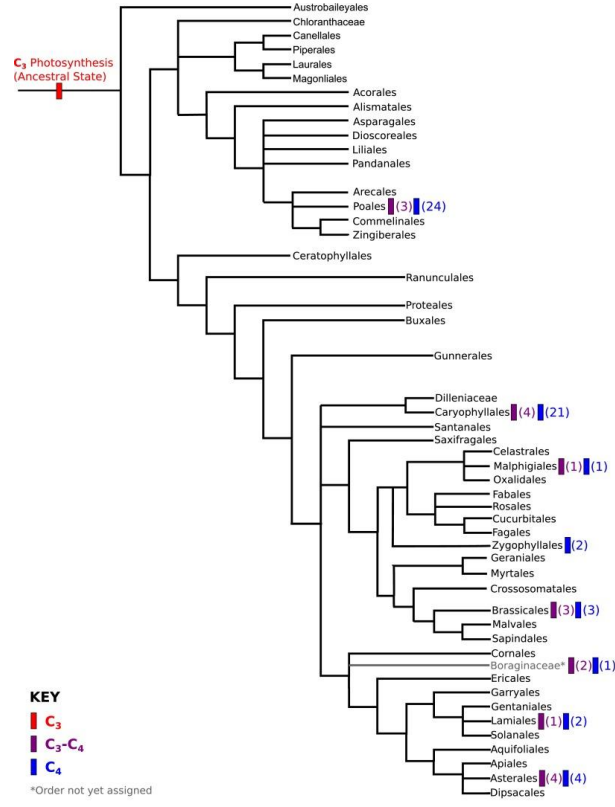
- Large and diverse carnivorous plant family (400+ species)
- Carnivory evolved to compensate for nutrient-poor habitats (low N+P)
- *Utricularia* (bladderworts): suction traps
- *Pinguicula* (butterworts): sticky leaf traps
- *Genlisea* (corkscrew plants): lobster-pot traps



Müller et al., 2004: 10.1055/s-2004-817909
https://commons.wikimedia.org/wiki/File:Pinguicola_Antarctica.JPG, CC BY 4.0
https://commons.wikimedia.org/wiki/File:Genlisea_violacea_giant.jpg, CC BY-SA 4.0

Example: photosynthesis

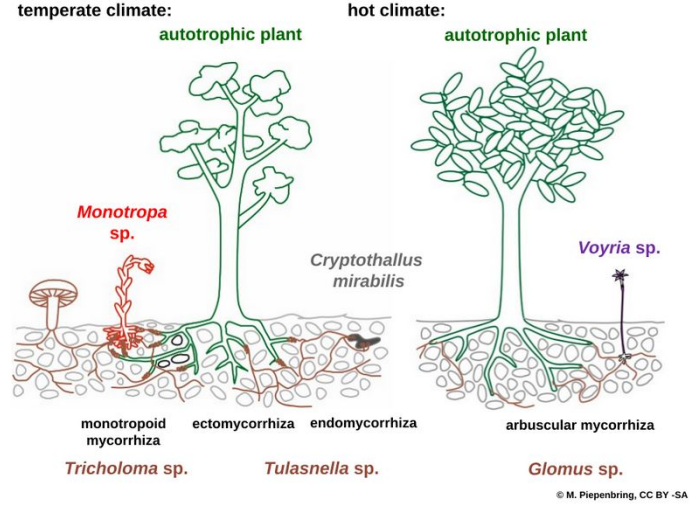
- Different types of photosynthesis evolved multiple times independently
- C3 was ancestral state
- C4 could be more efficient under certain conditions



William et al., 2013: 10.7554/eLife.00961

Myco-heterotrophic plants

- Not all plants do photosynthesis (*Monotropha*)
- Myco-heterotrophic plants take nutrients through fungus from photoautotrophic plants

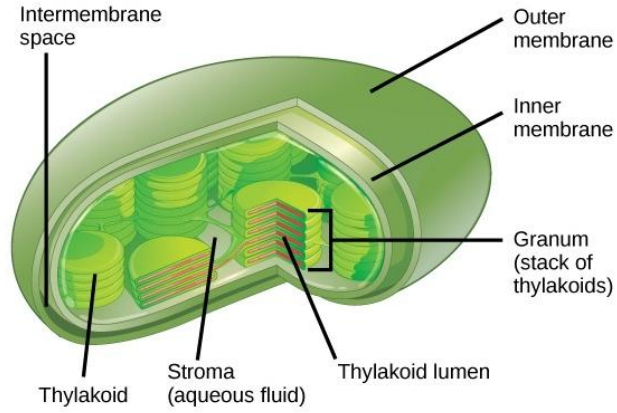


UNIVERSITÄT BONN

Plastids and mitochondria

Chloroplast

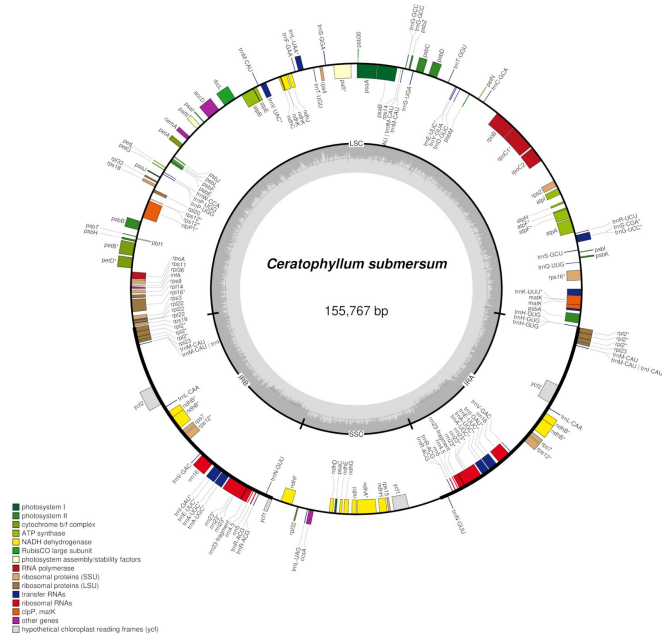
- Convert light energy into chemical energy through
- Stacks of thylakoids (grana) contain chlorophyll and other pigments essential for capturing light
- Contain circular DNA and 70S ribosomes; outer & inner membrane; prokaryotic gene structures
- Synthesize ATP, NADPH, and organic molecules like glucose, fueling plant metabolism



UNIVERSITÄT BONN

Plastome

- Circular DNA found in chloroplasts and other plastids
- Typically encodes ~100–120 genes (photosynthesis, transcription/translation machinery, and metabolic pathways)
- Structure and gene order are relatively conserved across land plants
- Inherited maternally in most plants (exceptions exist)
- Target for transplastomic engineering, offering high gene expression and reduced risk of gene flow

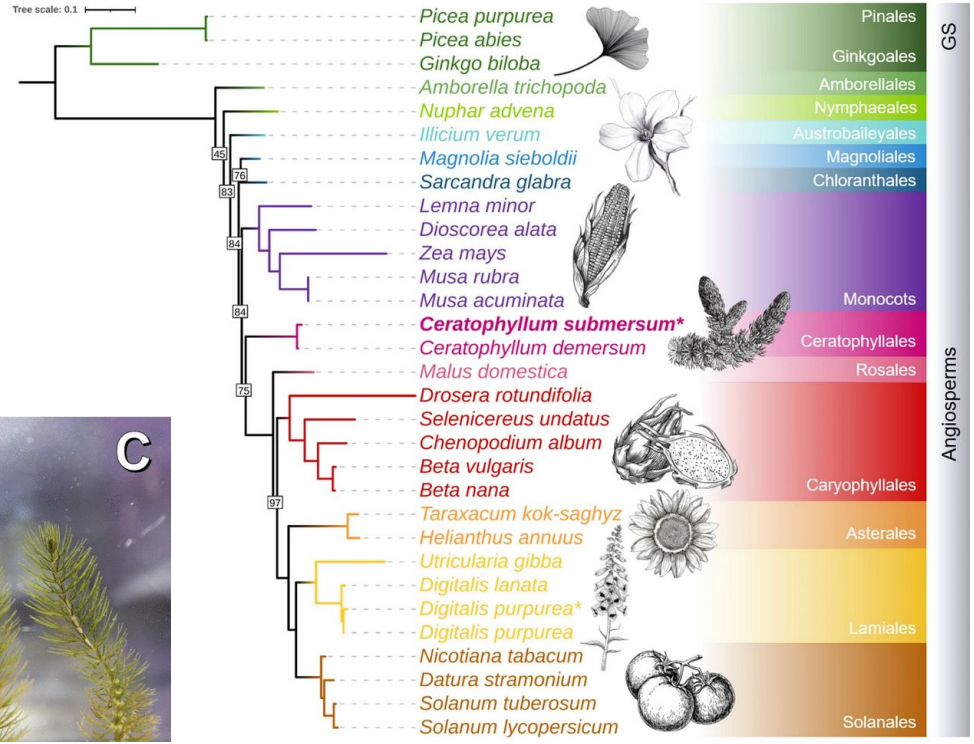


Meckoni et al., 2023: 10.1186/s13104-023-06459-z

UNIVERSITÄT BONN

Plastome as basis for phylogenetics

- *Ceratophyllum submersum* samples taken from local pond
- Long read sequencing resolved plastome
- Phylogenetic placement based on plastid marker genes

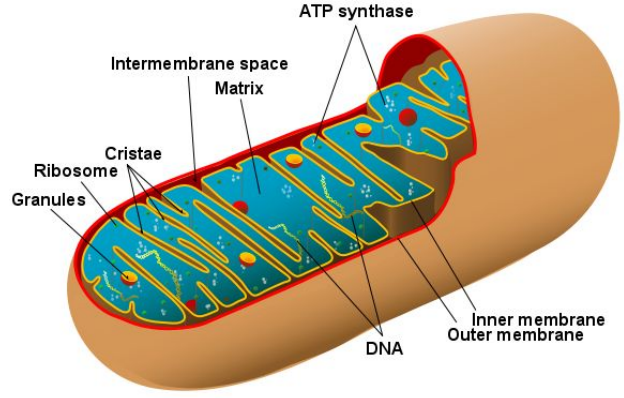


Meckoni et al., 2023: 10.1186/s13104-023-06459-z

UNIVERSITÄT BONN

Mitochondrium

- Generate ATP through oxidative phosphorylation and electron transport chain
- Outer membrane and highly folded inner membrane (cristae) enabling compartmentalized metabolism
- Circular DNA (mtDNA) and 70S-like ribosomes
- Citric acid cycle, β-oxidation of fatty acids, and amino acid metabolism

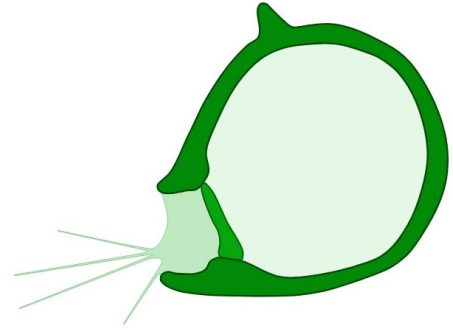


Chondrome

- Mitochondrial genome: large, multipartite, and highly recombinogenic DNA molecule
- Plant mitochondrial genomes are largest among eukaryotes
- Encodes core components of oxidative phosphorylation, ribosomal RNAs, tRNAs, and factors essential for mitochondrial function
- Sequence evolution is slow, yet genome structure changes rapidly
- Usually maternally inherited and valuable for studying plant evolution

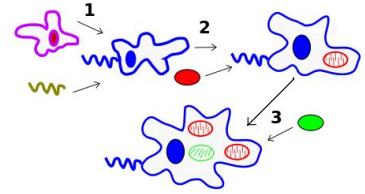
Endosymbionts

- Endophytes are microorganisms (fungi/bacteria) living inside plant tissues without causing disease
- Can enhance plant growth, nutrient uptake, defence, and stress tolerance
- Important in sustainable agriculture, biocontrol, and biofertilization strategies



https://www.izmb.uni-bonn.de/en/pbb/samuel_nestor_meckoni

Evidence for endosymbiont theory



- Mitochondria and chloroplasts retain their own circular DNA
- Genes lack introns (in many cases) and show operon-like arrangements
- Organelles contain 70S ribosomes and initiate prokaryotic translation
- Inner membranes contain bacterial-type lipids, distinct from eukaryotic membranes
- Dual-membrane structure reflects engulfment of a prokaryote by a host cell

UNIVERSITÄT BONN

Summary

- Evolutionary concepts: homologs, paralogs, orthologs
- Gene duplications: loss, subfunctionalization, neofunctionalization
- Phylogenetic tools: alignment, tree inference, tree visualization
- Evolution of traits in plants: carnivory, photosynthesis, pigmentation
- Plastids and mitochondria: endosymbionts

UNIVERSITÄT BONN

Time for questions!

UNIVERSITÄT BONN

Questions

1. What is a paralog and ortholog?
2. What is a monophyletic group?
3. What can happen to gene copies after a duplication event?
4. Provide an example for a gene duplication event in plants and explain its importance!
5. Which tool can quickly compare a sequence against a database?
6. Which tool can infer a phylogenetic tree?
7. What is a plastome?
8. Which theory explains the presence of chloroplasts in plants?