



Ancestral reconstruction and investigations of genomic recombination on *Campanulids* chloroplasts

Bashar Al-Nuaimi, Roxane Mallouhi, Bassam AlKindy, Christophe Guyeux, Michel Salomon, and Jean-François Couchot

FEMTO-ST Institute, UMR 6174 CNRS, DISC Computer Science Department, Université de Franche-Comté, France.

SeqBio 2015

Outline

- Introduction
- Ancestral reconstruction
- Case study
- The proposal
- Conclusion and future work

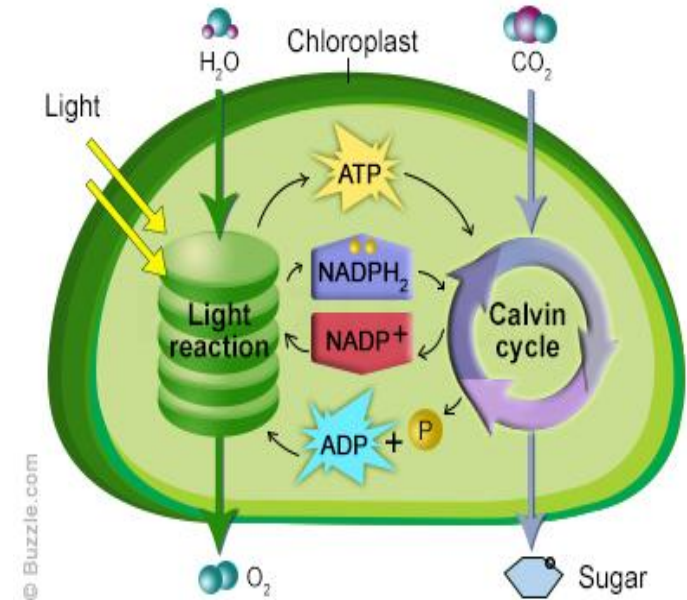




Introduction

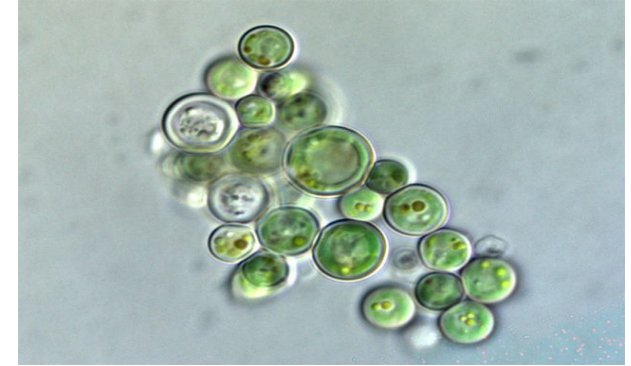
What are chloroplasts

- Key elements in living organisms history
 - Photosynthetic organisms at the base of most ecosystems trophic chains
 - Responsible for photosynthesis in Eukaryote
- Origin of the presence of oxygen in the atmosphere
- Primary resource of carbon storage
- First endosymbiosis lineages
 - Red algae
 - Green algae
 - Land plant



Data set

- 450 chloroplasts:
 - Complete sequences downloaded from NCBI
 - From algae to oak: representative distribution of plant species
 - Each genome has approximately 100,000 nucleotides, and 150 coding sequences (“genes”)
 - Gene content evolves with species
- **Core genes** of a set of genomes: genes present everywhere
- **Pan genes**: genes in at least one genome



Pedinomonas

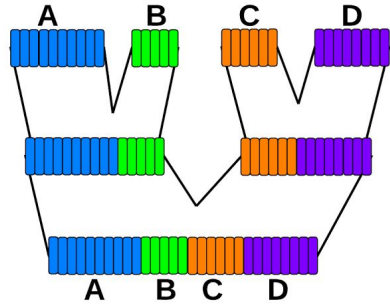


Apiales

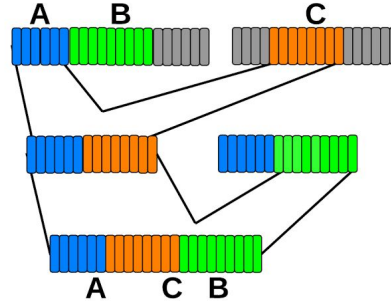
Recombination in chloroplasts



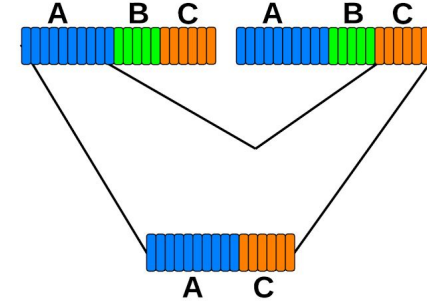
A. Composition
 $A+B+C+D \rightarrow ABCD$



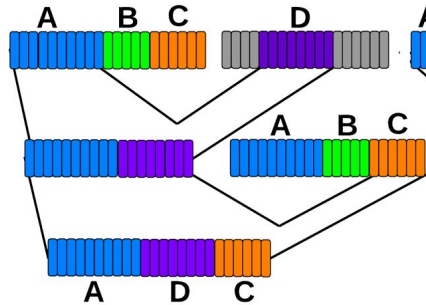
B. Insertion
 $AB+C \rightarrow ACB$



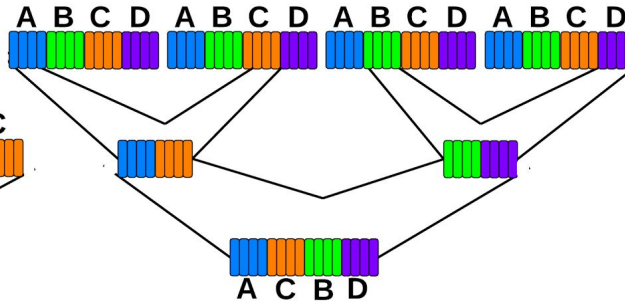
C. Deletion
 $ABC \rightarrow AC$



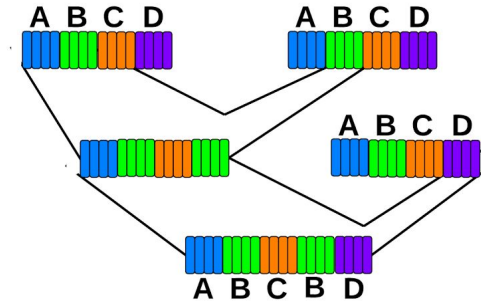
D. Replacement
 $ABC+D \rightarrow ADC$



E. Inversion
 $ABCD \rightarrow ACBD$



F. Duplication
 $ABCD \rightarrow ABCBD$



Our objective

➤ We work at gene level:

- INPUT: list of gene names provided by Dogma
- OUTPUT: ancestral genome reconstruction for chloroplasts

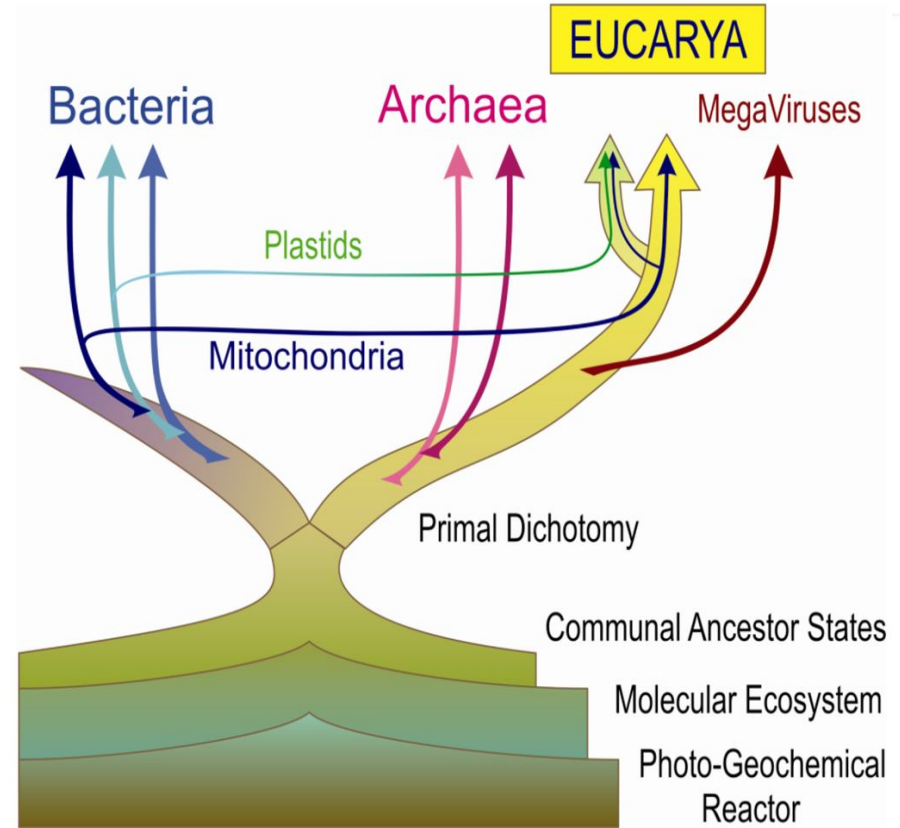
Organism name	Accession	Genome Id	Sequence length	Number of genes	Lineage
<i>Daucus carota</i>	NC_008325.1	114107112	155911 bp	138	Apiaceae
<i>Anthriscus cerefolium</i>	NC_015113.1	323149061	154719 bp	132	Apiaceae
<i>Panax ginseng</i>	NC_006290.1	52220789	156318 bp	132	Araliaceae
<i>Eleutherococcus senticosus</i>	NC_016430.1	359422122	156768 bp	134	Araliaceae
<i>Aralia undulata</i>	NC_022810.1	563940258	156333 bp	135	Araliaceae
<i>Brassaiopsis hainla</i>	NC_022811.1	558602891	156459 bp	134	Araliaceae
<i>Metapanax delavayi</i>	NC_022812.1	558602979	156343 bp	134	Araliaceae
<i>Schefflera delavayi</i>	NC_022813.1	558603067	156341 bp	134	Araliaceae
<i>Kalopanax septemlobus</i>	NC_022814.1	563940364	156413 bp	134	Araliaceae



Ancestral Reconstruction

Ancestral reconstruction objectives

- Understand the evolutionary history of a set of organisms
- Investigate gene content evolution
- Improve a phylogenetic



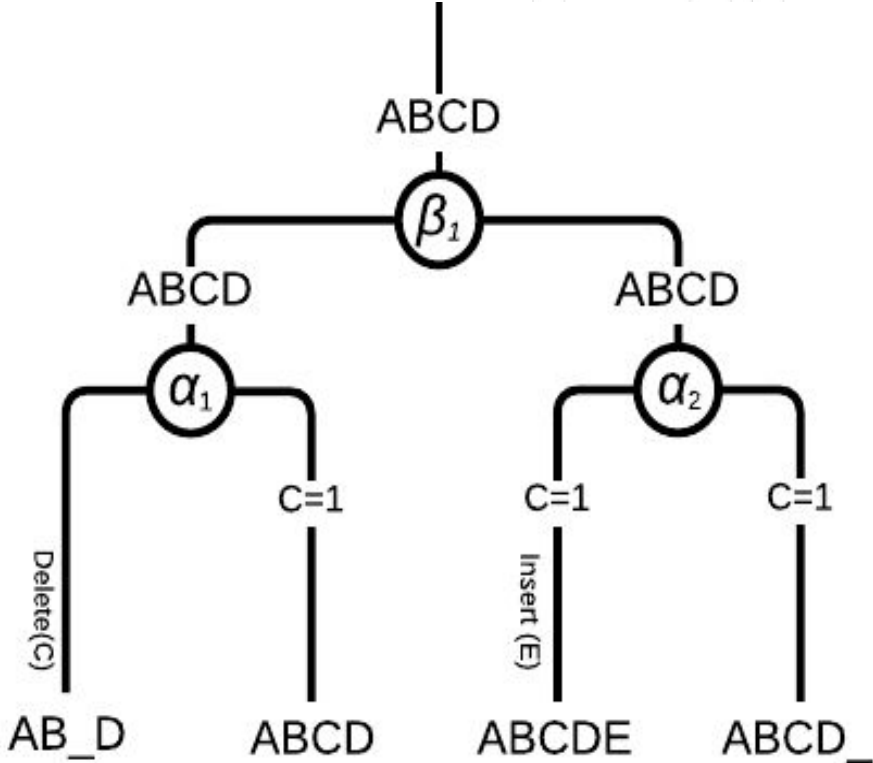
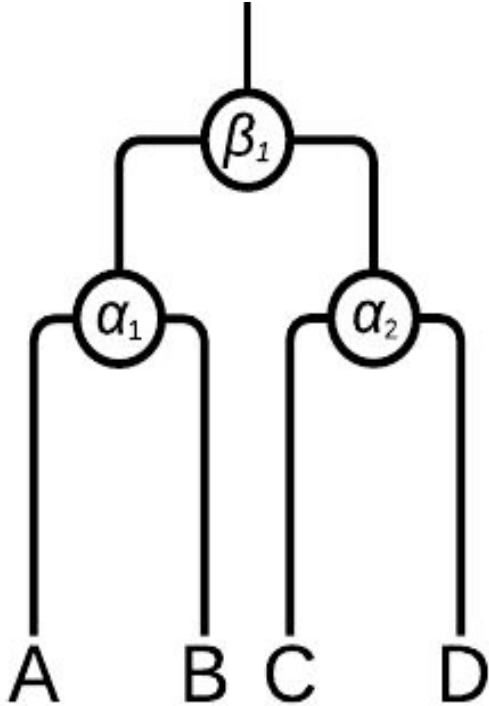
Ancestral reconstruction in chloroplasts case

- Genomes of various length and with repeated/missing genes
- Genomes contain multiple rearrangement events
- Gene order and content evolve so much compared to related other species
- Chloroplast genomes have lower genes than bacterial or nucleus ones



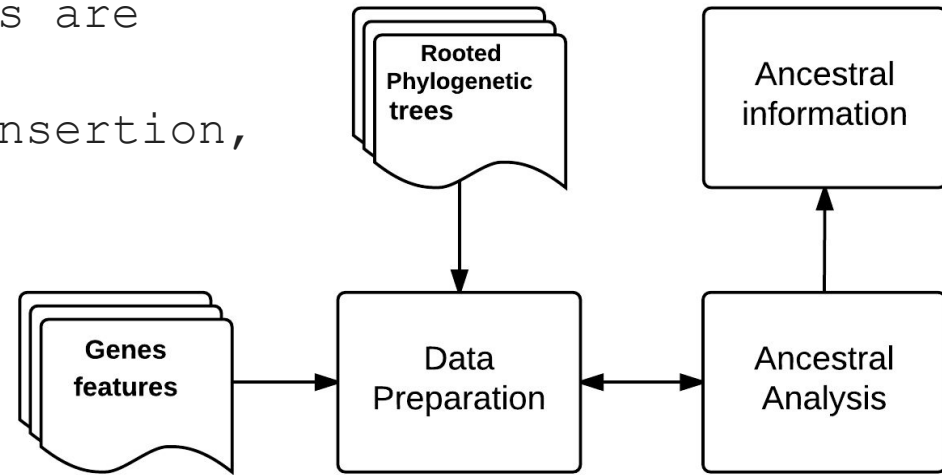
The proposal

General idea



General overview of the proposed method

- Working on previously obtained well-supported phylogenetic trees^{1,2}
- Gene duplication is taken into consideration
- Different chloroplast groups are applied
- Rearrangement operations (insertion, deletion, etc.)





Ancestral reconstruction algorithm

- INPUT: Family tree and genomes with list of genes
- OUTPUT: Ancestor genomes
- Select two brother genomes
- Find the list of cousin genomes
- Compute the gene duplication



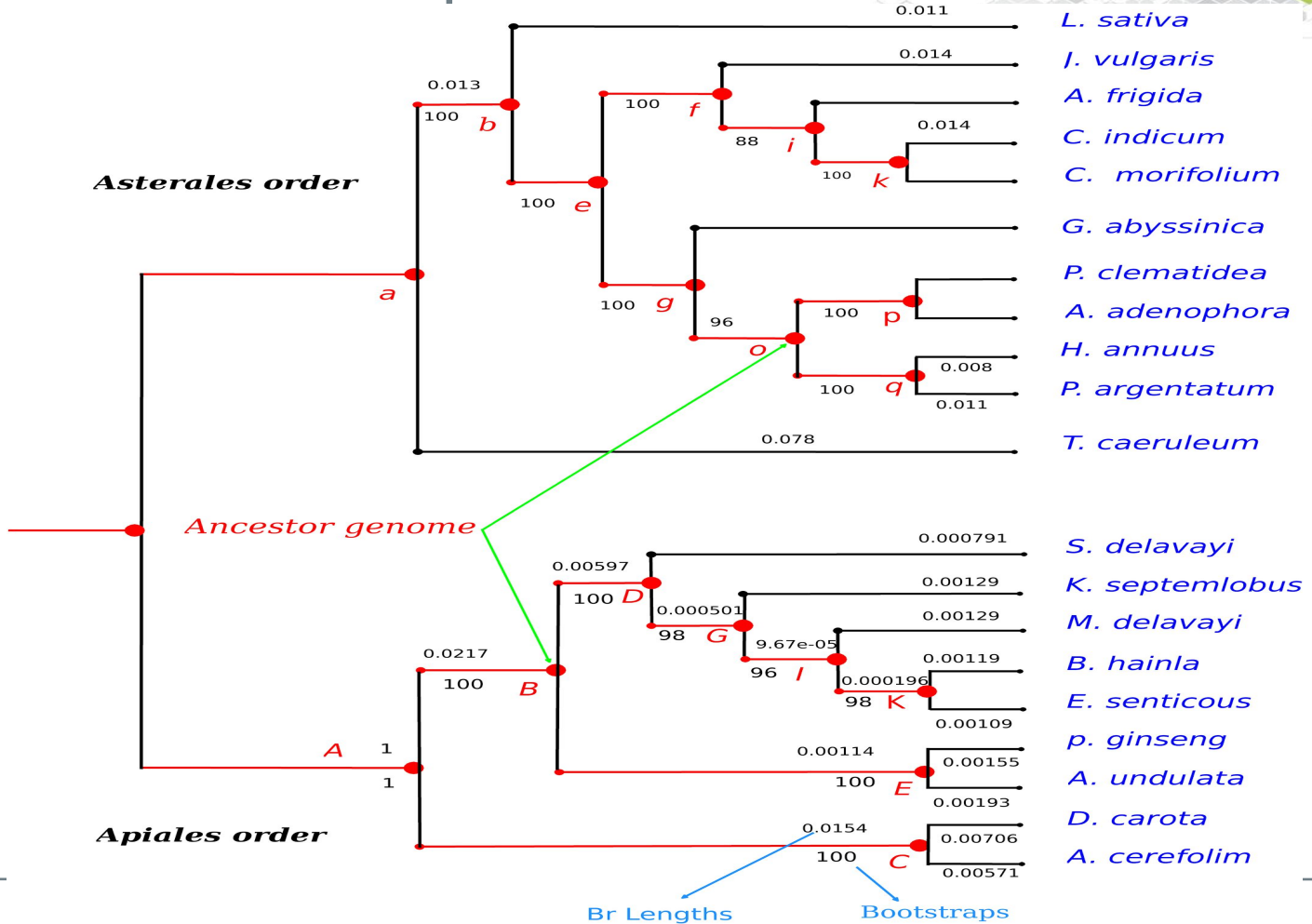
Ancestral reconstruction algorithm

- Matching operation:
 - Evaluate the matching score, deletion and insertion genes with SequenceMatcher
 - For each gene in brother genomes
 - In case of matching: add it to father genome
 - For non-matching gene:
 - Compare it with gene in cousin genome
 - In case of matching: add the gene to father genome
 - If non-matching:
 - Select cousin genome
 - If matching gene: add the gene to father
 - Repeat the matching operation

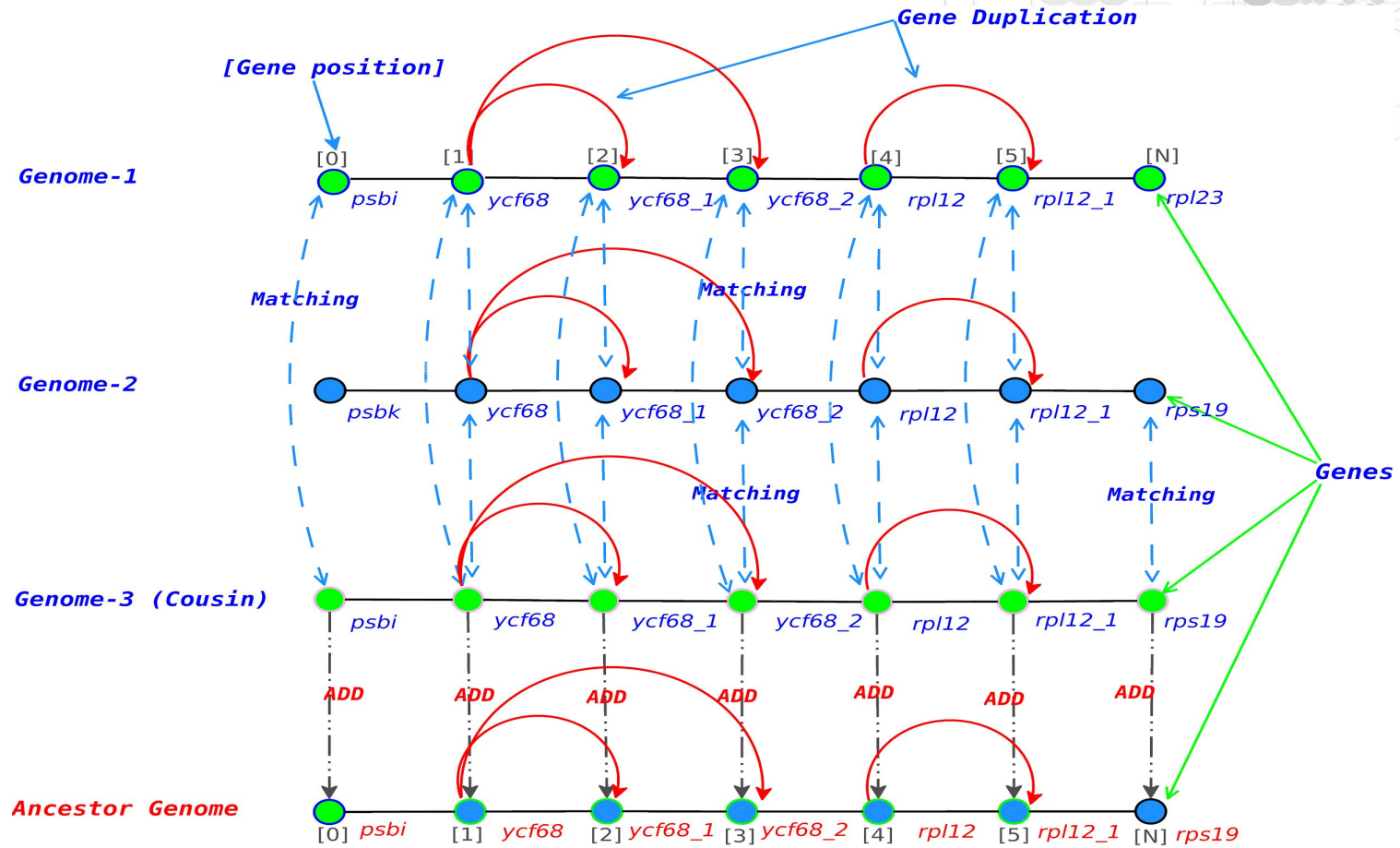


Case study

Apiales and Asterales species

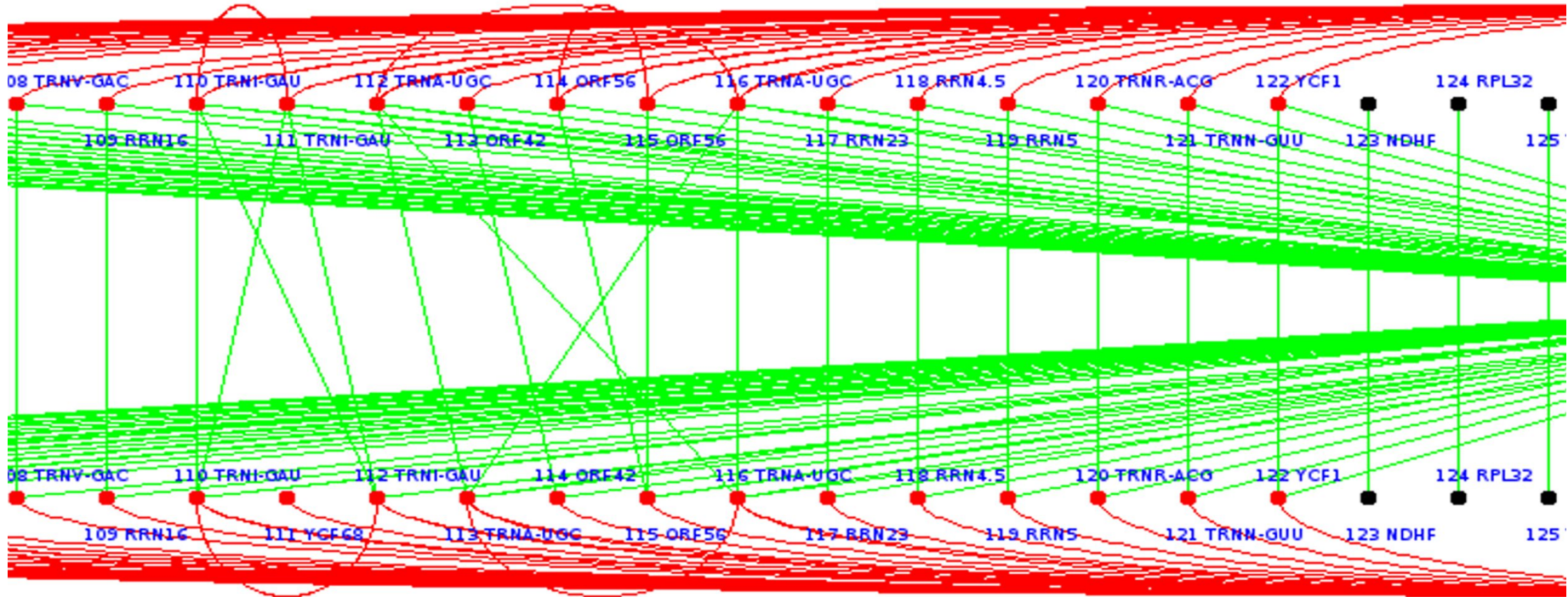


Simulation of ancestral reconstruction



Recombination in sister species (*Apiales*)

D. carota

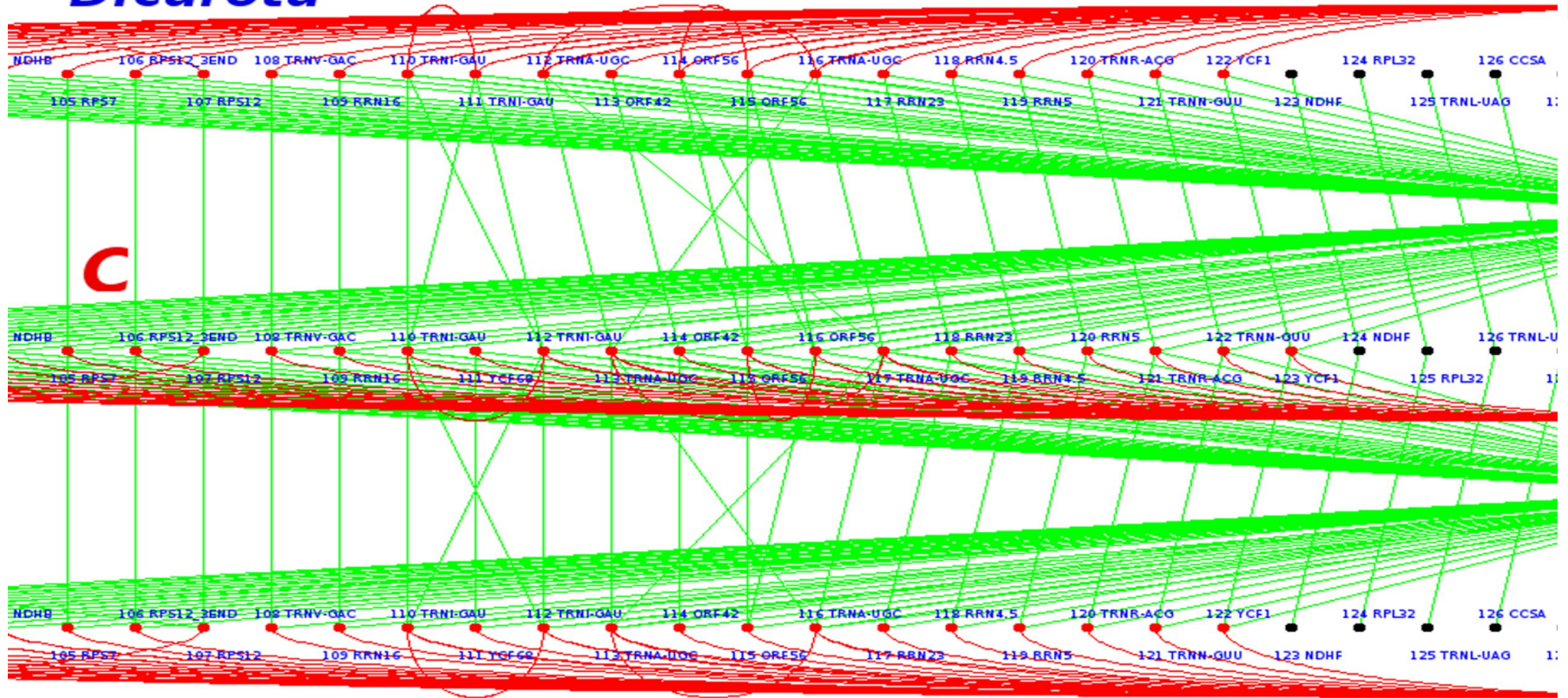


A. cerefolim

Ancestor (*Apiales*)



D. carota

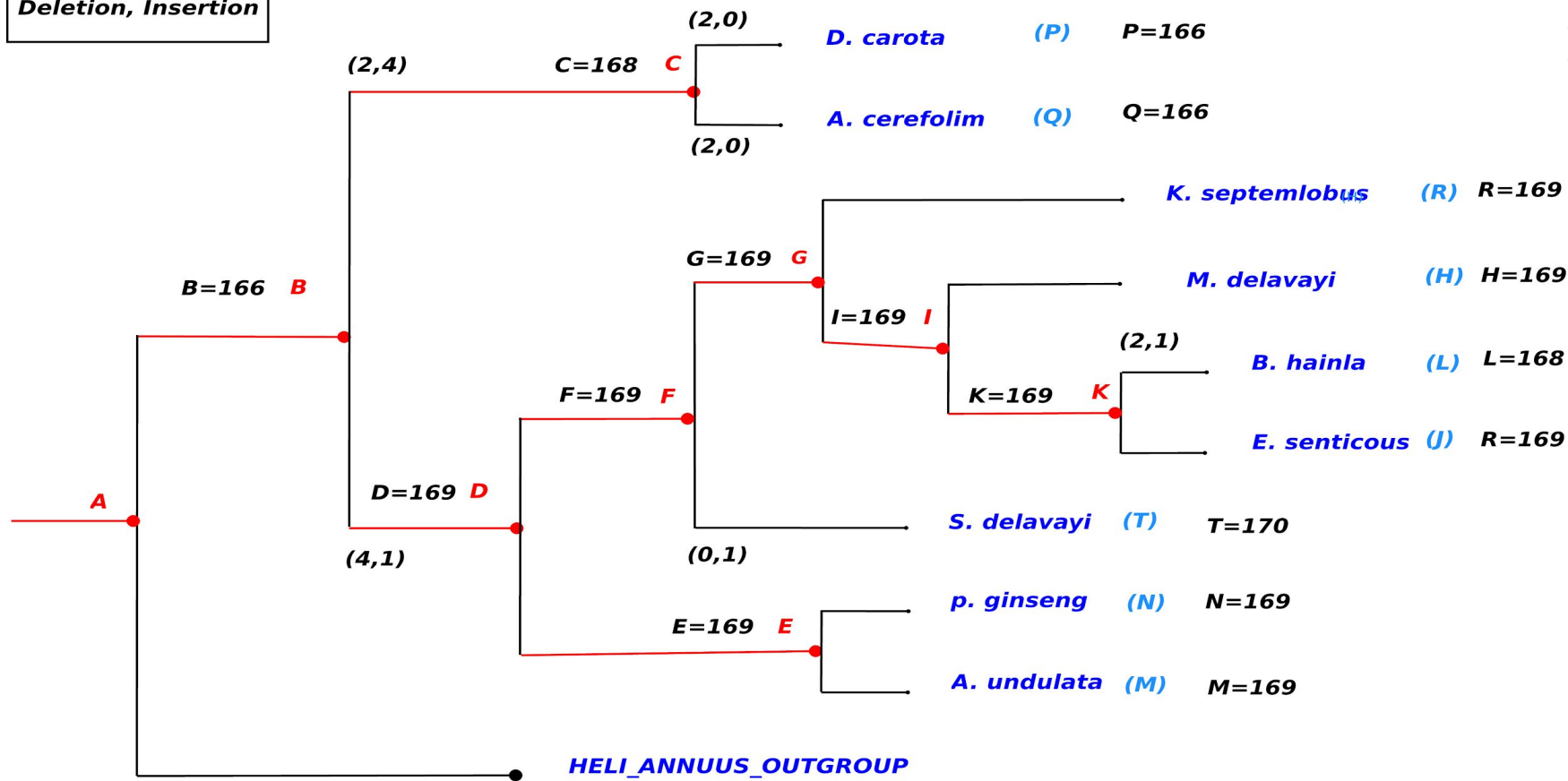


A. cerefolim

Ancestral reconstruction results (*Apiales*)

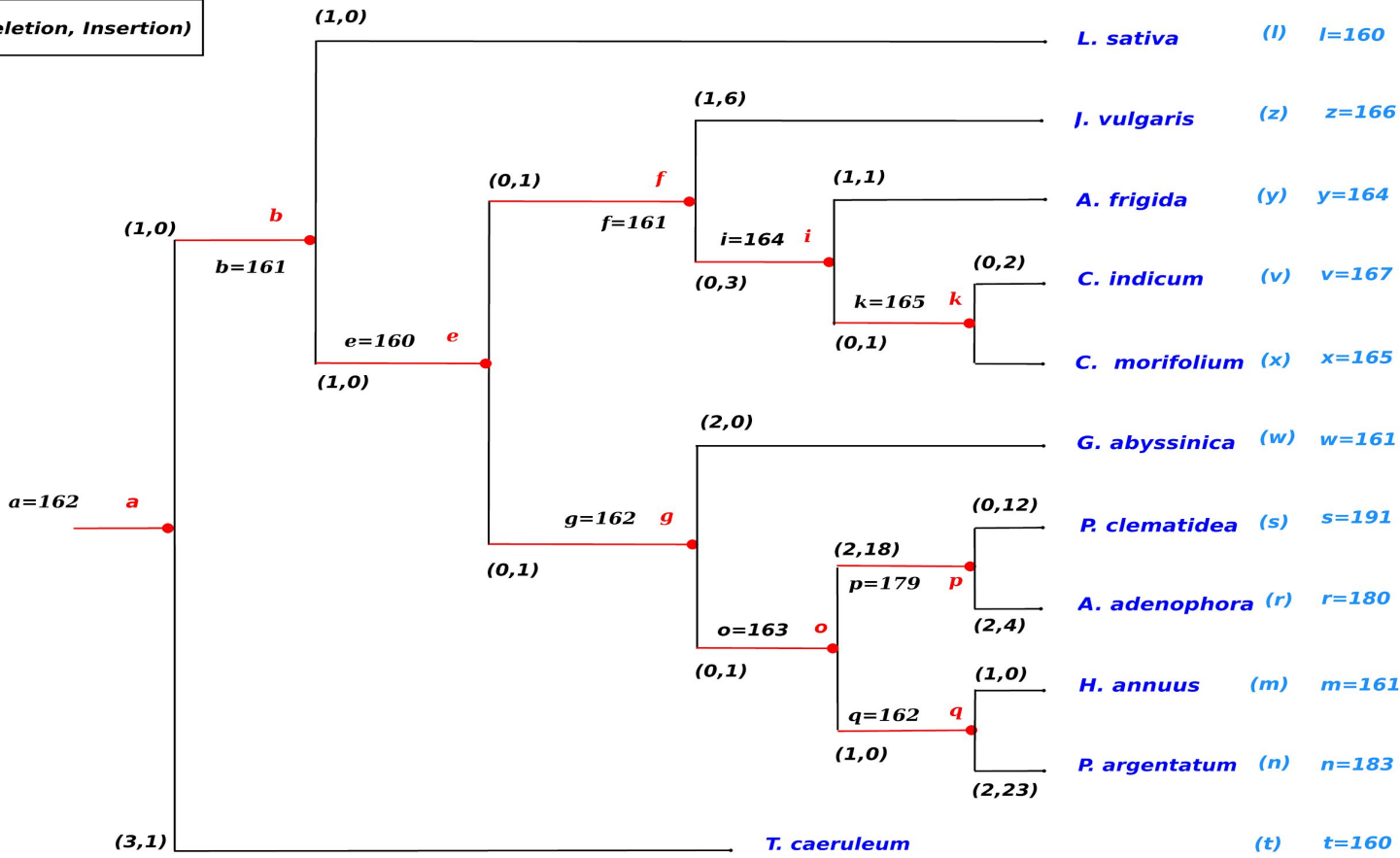
Apiales

Deletion, Insertion



Ancestral reconstruction results (Asterales)

Asterales
 (Deletion, Insertion)





Conclusion and future work

Conclusion

- The two studied families have not faced the same kind of genomic recombination
- Insertions and deletions: in *Asterales* and not in *Apiales*
- Different of genes duplication between the families
- Ancestor genome provides relation between members of phylogenetic tree

Future work



The proposal work is ongoing regarding the design of ancestral reconstruction of chloroplastic genomes:

- Investigating best phylogenetic tree and reconstruct the ancestral genomes
- Continue both the theoretical investigations and its applications to the whole set of 450 chloroplasts
- Investigating the ancestral DNA sequences



Thank you !
Any question ?