





# Solving conflicts when gene and genome evolution disagree in paleopolyploid plants

Lada Isakova, Alexandra Louis, Elise Parey and Hugues Roest Crollius

Genome Organization and Dynamics group, Institut de Biologie de l'École Normale Supérieure (IBENS), École Normale Supérieure



## Introduction

- Whole-genome duplications (WGDs) are evolutionary events that result in the doubling of a genome. Post-WGD diploidization (gene loss) leads to complex gene families where orthologs and paralogs are difficult to identify reliably.
- SCORPiOs was developed to correct phylogenetic gene trees at duplication nodes corresponding to a WGD using local syntenic context and tested on Teleost fish WGD.



- WGDs are especially common in plants **Aim:** to apply SCORPiOs approach to correct for a WGD event in plants

Fig. 2 Whole-genome duplications in plants. Red arrow: Papilionoideae WG duplication under study (see Methods).



## **Methods**

Whole-genome duplication selection. We selected an appropriate WGD (Papilionoideae) based on the availability of the duplicated species and outgroups, and their genome assembly quality.



## Results

#### **Trees correction by SCORPiOs.**

- Utilization of the sliding window of 20 genes and 2 outgroup species proved to correct the largest number of gene trees (1130 trees).
- Among 14 048 families tested, SCORPiOs identified 3371 (24%) synteny-inconsistent families and was able to correct 1322 (39%).

**Evolutionary categorization of** duplicated families and GO terms enrichments

- We classified 66% of all families as singletons, 23% as facultative and 11% as systematic ohnologs.
- After correction we found a considerable change in the counts of different categories - the number of systematic and facultative ohnologs increased by 32% and 89% respectively, and the number of singletons decreased by 32% compared to the counts in original trees.



Fig. 4 Improvement of the duplication confidence score on the Papilionoideae duplication nodes after the correction (probability density, KDE).



Rosales (outgroup)

Cannabis sativa female | Genes: 27236| Scaffolds: 136| N50: 2661| L70: 7 • Prunus persica | Genes: 26873| Scaffolds: 43| N50: 3161| L70: 5

**WGD** - Whole Genome Duplication

**WGT - Whole Genome Triplication** 

Fig. 3. Phylogenetic tree of the species used in the study. SCORPiOs was used to correct for Papilionoideae WGD event, which happened ~55 Ma years ago (red arrow). Table in grey (right): total gene number and genome assembly quality metrics (number of scaffolds, N50 and L70 statistics) calculated from Ensembl data (see next section) for each species and used for WGD selection. In the high-quality genome assembly L70 is usually low (>= chromosome number) and N50 is high (<= chromosome length).

#### Data preparation and SCORPiOs run.

24 524 gene trees, 10 879 of which can be Data from Ensembl Trees correction with SCORPiOs corrected (>2 genes, at least 1 outgroup) Plants v.49 (96 Parameters: default + tested: Gene trees reconciliation with species): Sliding window of 15, 20, 25 TreeBeST (duplication nodes - Gene trees neighboring genes (for synteny labeling) - Gene alignments analysis) - Gene - Outgroup(s): *Prunus persica* Filtering to select for duplicated and coordinates only, *Prunus persica* + *Cannabis* outgroup species genes - Species tree sativa.

**Evolutionary categorization of duplicated families.** In each gene tree we looked for gene families, defined as subtrees in a given gene tree with the root node in Papilionoideae. We then classified the gene families into three categories with respect to their fate after the duplication:

| <b>•••</b> | 1 | $\mathbf{\circ}$ | Λ |  |
|------------|---|------------------|---|--|

Singletons

genes found on a single

1 & 2 - scenarios of gene loss

branch in all species;

- Correction with SCORPiOs allowed to spot new GO terms enrichments in ohnologs and singletons.

Fig. 5 (A) Number of families by temation tologs category in the whole dataset after correction. (B) Fraction of Syst families corrected by SCORPiOs in each category <sup>-</sup>acultative ohnologs before and after correction. (C) GO terms enrichments in ohnologs before and after correction by SCORPiOs.

## **Conclusions and perspectives**

- The application of SCORPiOs on the Papilionoideae WGD lead to the correction of 39% gene families and an increase in ohnologs' counts,



**X** - gene loss, A and B - different species

Systematic ohnologs WGD-originated paralogs retained on two branches after the duplication in all descendant species

Facultative ohnologs WGD-originated paralogs retained on two branches in at least one species but not in all

Gene Ontology terms enrichments search in families with different evolutionary trajectories. AgriGO was used to search for GO enrichments in the sets of *Medicago truncatula* genes from families from different evolutionary categories. The threshold for significant enrichments: P value <0.01, FDR<0.01. Redundancy of GO terms was reduced with REVIGO.

consistent with the results reported for fish.

- Reliable identification of the orthology and paralogy relationships between genes can provide a new insight into functional (GO) enrichments of the genes with different evolutionary fates.
- The correct positioning of the duplication node in the gene trees will allow a more precise reconstruction of ancestral genomes.
- Corrected gene trees allow us to study the gene fate after the duplication (the probability of co-occurrence of gene families of the same evolutionary category in one tree, etc.)

### References

- 1. Parey E, Louis A, Cabau C, Guiguen Y, Roest Crollius H, Berthelot C. Synteny-guided resolution of gene trees clarifies the functional impact of whole genome duplications. Mol. Biol. Evol., 2020.
- 2. Ren R., Wang H., Guo C., Zhang N., Zeng L., Chen Y., Ma H., and Qi J. Widespread Whole Genome Duplications Contribute to Genome Complexity and Species Diversity in Angiosperms. Mol. Plant, 2018.