

# Leveraging Genome-Scale Data to Infer Patterns and Phylogenies

## from Discordant Signals in *Primula*

Rebecca L. Stubbs<sup>1</sup> and Elena Conti

Department of Systematic and Evolutionary Botany, University of Zürich



*P. veris* ssp. *veris*

*P. elatior* ssp. *meyeri*

*P. elatior* ssp. *elatior*

*P. vulgaris* ssp. *vulgaris*

*P. grandis*

*P. megaseifolia*

### BACKGROUND

- Utilizing genome-scale data, such as whole genome resequencing (WGR), provides unprecedented marker density and enables detecting and disentangling reticulate evolution across entire genomes.

- Analyses that draw upon different facets of data and rely on disparate methods to identify introgression aid in informing how confounding biological signals can be integrated in phylogenomics.

- Primula* sect. *Primula* (primroses, Primulaceae) represents an ideal model to test the power of WGR for investigating sources of phylogenetic discordance.

- Hemizygous S-locus supergene (Fig. 1)
- Chromosome-level genome of *P. veris*
- Seven species
- Documented gene flow between species

- We generated WGR data for 157 *Primula* sect. *Primula* accessions to examine the role that introgression, recombination, and incomplete lineage sorting (ILS) play in the evolutionary history of this clade.

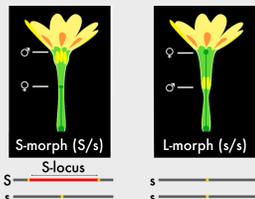


FIGURE 1. Short-styled (S) and long-styled (L) morphs of *P. veris*. The S-locus (red) is hemizygous in S-morphs (S haplotype), absent in L-morphs (s haplotype). Figure adapted from Potente et al. 2021.

### OBJECTIVES

- 1) What are the best approaches to disentangle gene tree discordance resulting from gene flow and ILS, when genome-scale data are available?
- 2) Do sources of gene tree discordance impact the phylogenetic resolution, relationships, and species monophyly in *Primula* sect. *Primula*?
- 3) Do areas of low recombination and smaller effective population size, such as the S-locus supergene in *Primula*, exhibit less ILS and gene tree discordance?

### OBJECTIVE 1 RESULTS

- Four different methods were used to detect reticulations at different timescales and signal strengths (Fig. 3).
- Results from multiple analyses were synthesized into an optimized topology that encompassed all highly supported hybridization edges across different analyses (Fig. 3).
  - Hybridization plays a more significant role than ILS in explaining gene tree discordance.

- Examining entire genomes in conjunction with targeting different facets of the data with complementary analyses produced a more complete picture of evolution.

### OBJECTIVE 2 RESULTS

- The ML tree and MSC tree supported the same relationships at shallow nodes (Figs. 2,5).
- Six species were strongly supported as monophyletic, *P. elatior* was non-monophyletic.
- P. elatior* had the most admixture (Fig. 5).
- High discordance was detected at most nodes, but was more pronounced at deeper nodes (Fig. 2).
- Our results, supported by multiple analyses, recovered six hybridization events during the evolution of *Primula* sect. *Primula* (Fig. 3).

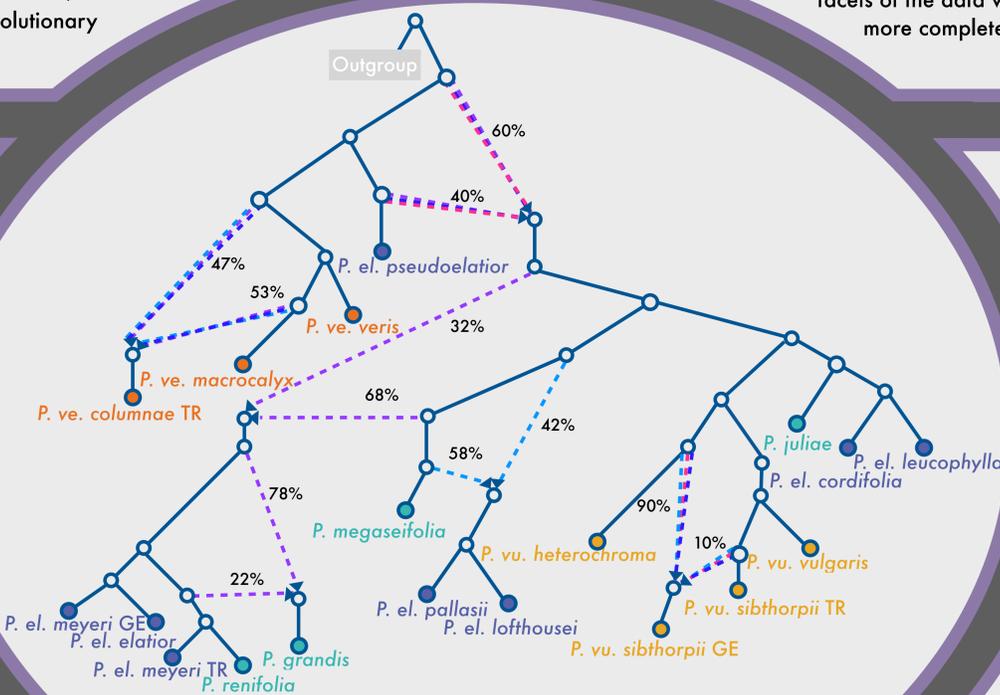


FIGURE 3. The optimized topology encompasses all highly supported hybridization edges across different analyses, and represents the most complete evolutionary hypothesis for *Primula* sect. *Primula*. Dashed lines represent reticulations at different phylogenetic depths. Hybridization edges: SNaQ (purple), QuBL (light blue), ADMIXTURE (pink), and/or Dsuite (dark blue). Percentages indicate the average percentage of alleles derived from a given hybridization edge.

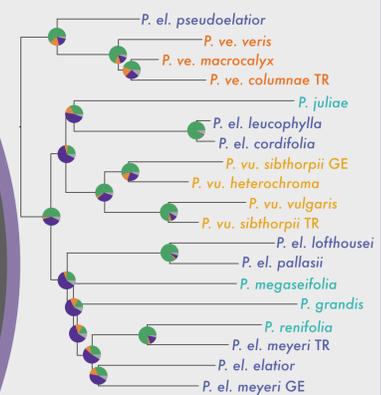


FIGURE 2. MSC phylogeny with pie charts showing the proportion of 5,674 genome fragment trees that are in concordance (green), discordance (purple), support the second most common topology (orange), or have less than 70% BS support (grey) at that node.

### CONCLUSIONS

- We have presented the first completely resolved species phylogeny for *Primula* sect. *Primula*.
- Our results demonstrate that gene flow and introgression can be persistent features during evolution, yet phylogenetic resolution is possible through full genomic data and a combination of population and species-level analyses.
- As genome-scale data becomes increasingly accessible, focus should shift towards a more complete understanding of the biological processes operating in the evolution of different lineages.
- Selecting analyses that encompass multiple aspects of data and understanding how these analyses both differ from and complement each other are paramount to this endeavor.

This work was supported by the Swiss National Science Foundation (grant 10031A\_175556 to EC). We thank Emiliano Mora-Carrera, Barbara Keller, Narjes Yousefi, Giacomo Potente, Étienne Léveillé-Bourret, Ferhat Celep, Judita Kochjarová, Giorgi Tedoradze, Spyros Theodoridis, and Deren A.R. Eaton for input on this project. We are grateful to Hüseyin Baykal, Bryan Drew, Monika Janišová, Dimitri Khushivadze, D. Blaine Marchant, Parvin Yousefi, and Stina Weststrand for field work assistance and sample collection.

### OBJECTIVE 3 RESULTS

- The S-locus trees were more similar to each other than trees obtained from other regions in the genome (Fig. 4).
- The S-locus has been less impacted by discordance due to suppressed recombination and smaller effective population size.

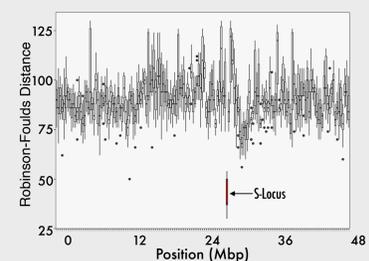


FIGURE 4. Comparison of Robinson-Foulds (RF) distances for genome fragment (GF) trees in the S-locus (red) and all other regions (white) of the first chromosome. The GF trees in S-locus are significantly ( $p < 0.001$ ) more similar to each other than those in other regions of the first chromosome.

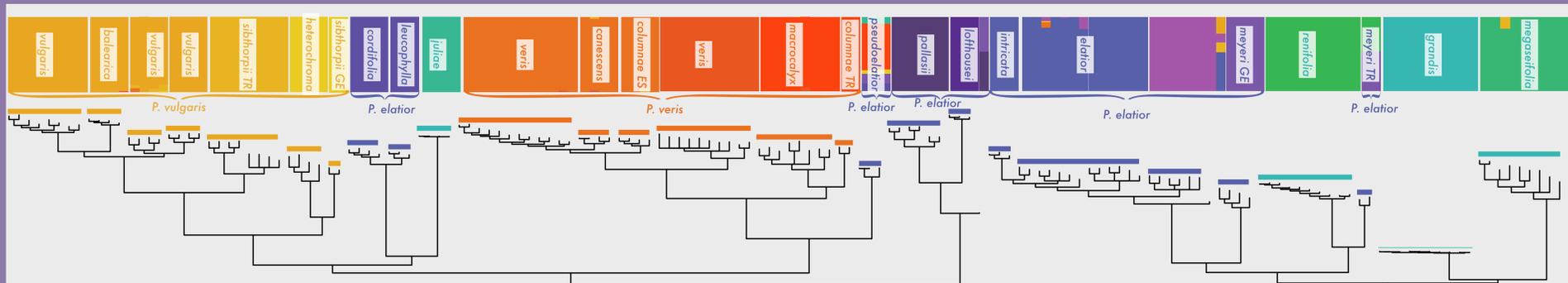


FIGURE 5. Overview of *Primula* sect. *Primula* with 157 ingroup specimens. UPPER: ADMIXTURE plot showing genetic clusters. For *P. veris* (orange), *P. vulgaris* (yellow), and *P. elatior* (purple) subspecies name shown in white boxes and species name shown below. For other specimens (in teal) species names are shown in white boxes. LOWER: Maximum likelihood phylogeny inferred from IQ-TREE with 1,375 concatenated low-copy nuclear loci. Tips correspond to name in ADMIXTURE plot and are colored by species. Branches <95% bootstrap are collapsed, and all other branches recovered with >95% support.