

New Compositae specific probe set reduces paralogs in complex systems

Applications in Plant Sciences



Erika R. Moore-Pollard¹, Daniel S. Jones², Jennifer R. Mandel¹
¹University of Memphis, Memphis, TN, USA; ²Auburn University, Auburn, AL, USA

Asteraceae • Compositae-ParaLoss-1272 • Hyb-Seq • discordance • NGS • phylogenomics • polyploidy

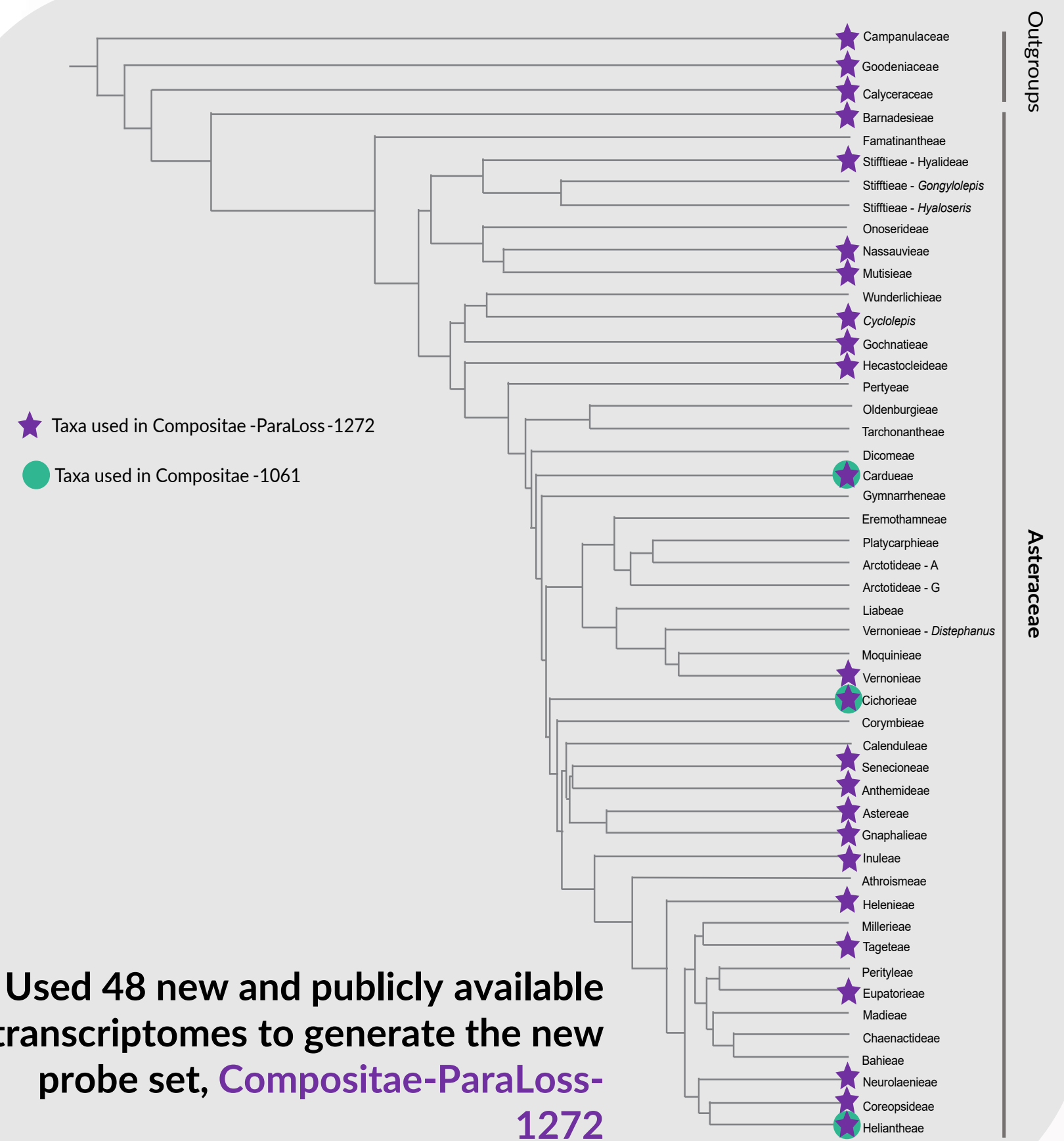
The Asteraceae specific probe set, **Compositae-1061**¹, has become popular among researchers studying members of Asteraceae, and has paved the way for investigations at lower taxonomic levels and non-model organisms. Though **Compositae-1061** has shown to be highly efficient at the higher- and some lower-taxonomic levels within the family, it generally lacks resolution at the genus to species level, especially in groups with complex evolutionary histories including polyploidy and hybridization.

In this study, we developed a new probe set, **Compositae-ParaLoss-1272**², designed to target known single-copy orthologous loci across Asteraceae family members using MarkerMiner³. We tested its effectiveness at lower taxonomic levels using 19 taxa within the genus *Packera*, which has a complex evolutionary and taxonomic history, and was previously studied using **Compositae-1061**⁴. Sequences from both probe sets, along with a double-capture approach, were used to generate phylogenies, compare topologies, and assess node support.

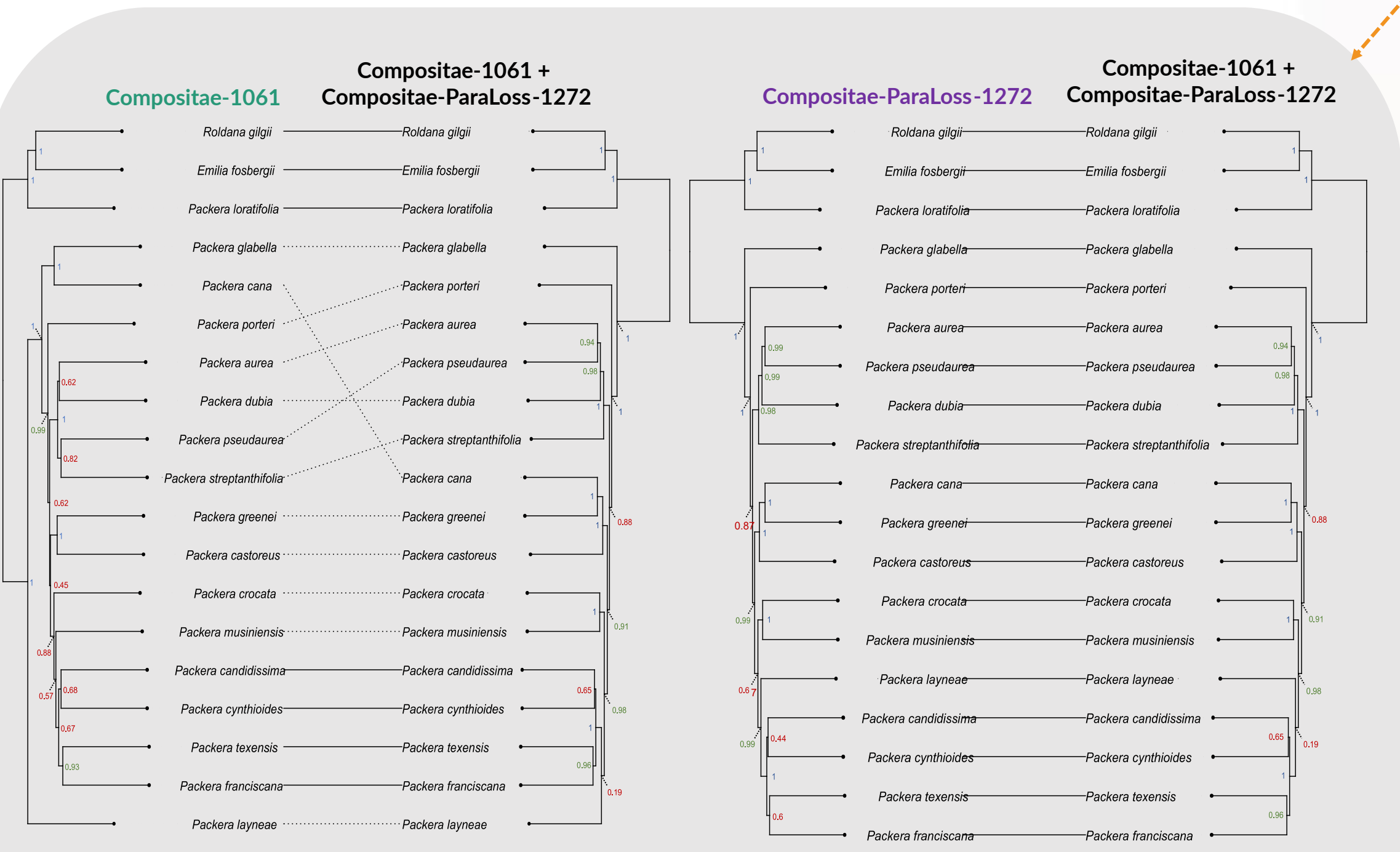
Differences between Compositae-ParaLoss-1272, Compositae-1061, and the angiosperm-wide probe set, Angiosperm-353⁵

	Compositae-ParaLoss-1272	Compositae-1061	Angiosperm-353
# loci	1,272	1,061	353
# baits	60,158	9,678	75,151
# loci overlap	0	30 (with Angiosperm-353)	30 (with Compositae-1061)
# species as input	48	3	42
Input data	Transcriptomes	Expressed sequence tags	Transcriptomes
Tool	MarkerMiner	BLAST	k-medoid clustering

Used 48 new and publicly available transcriptomes to generate the new probe set, **Compositae-ParaLoss-1272**

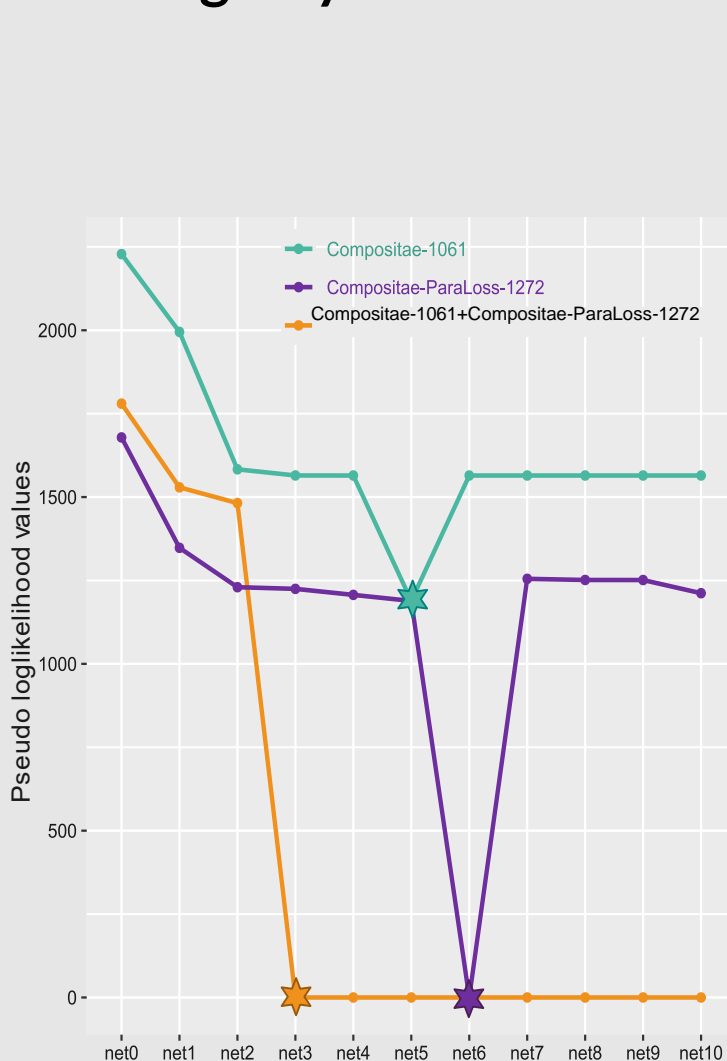


Compositae-ParaLoss-1272 provides improved resolution and utility for studies of complex groups and lower-taxonomic levels in the sunflower family.

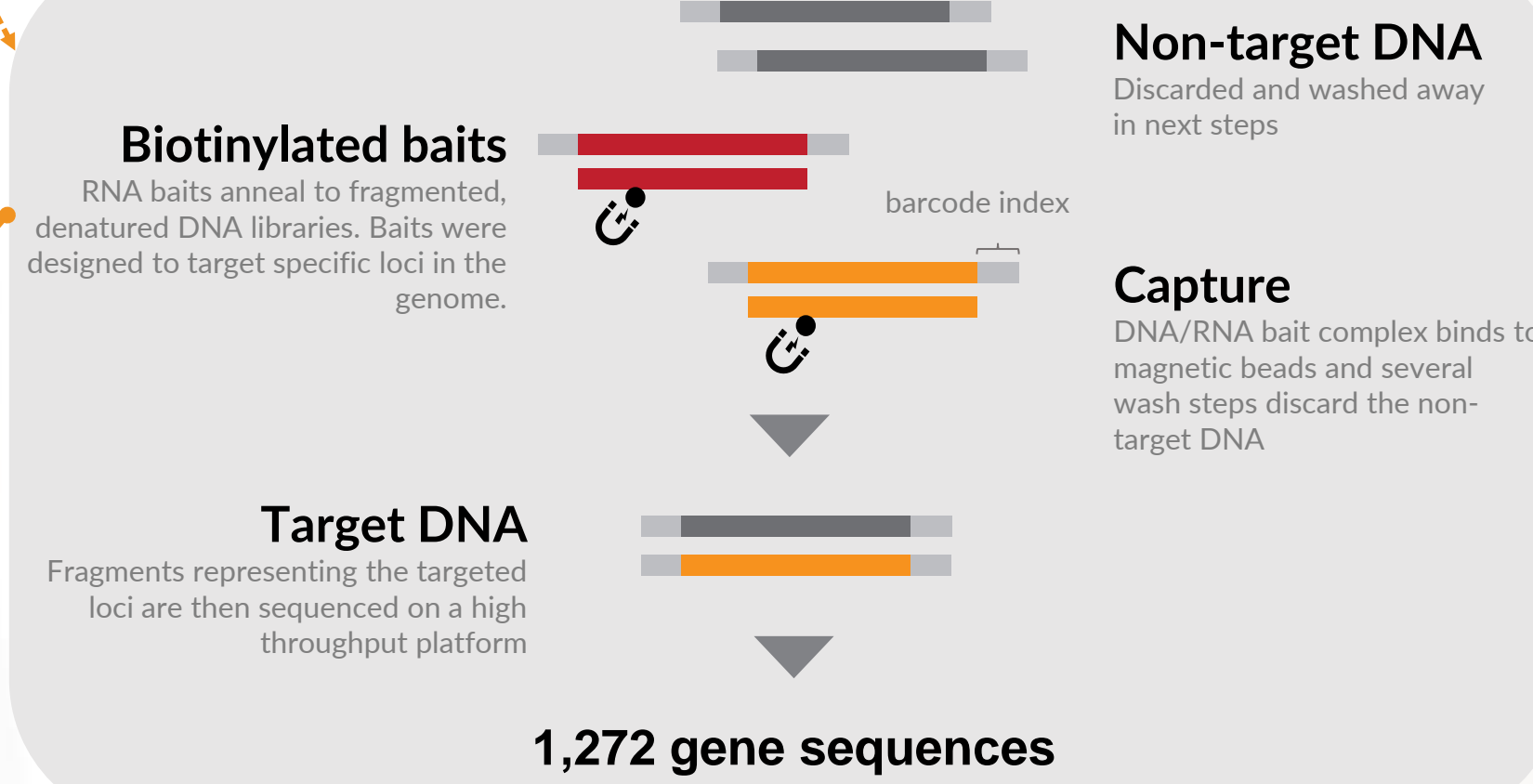
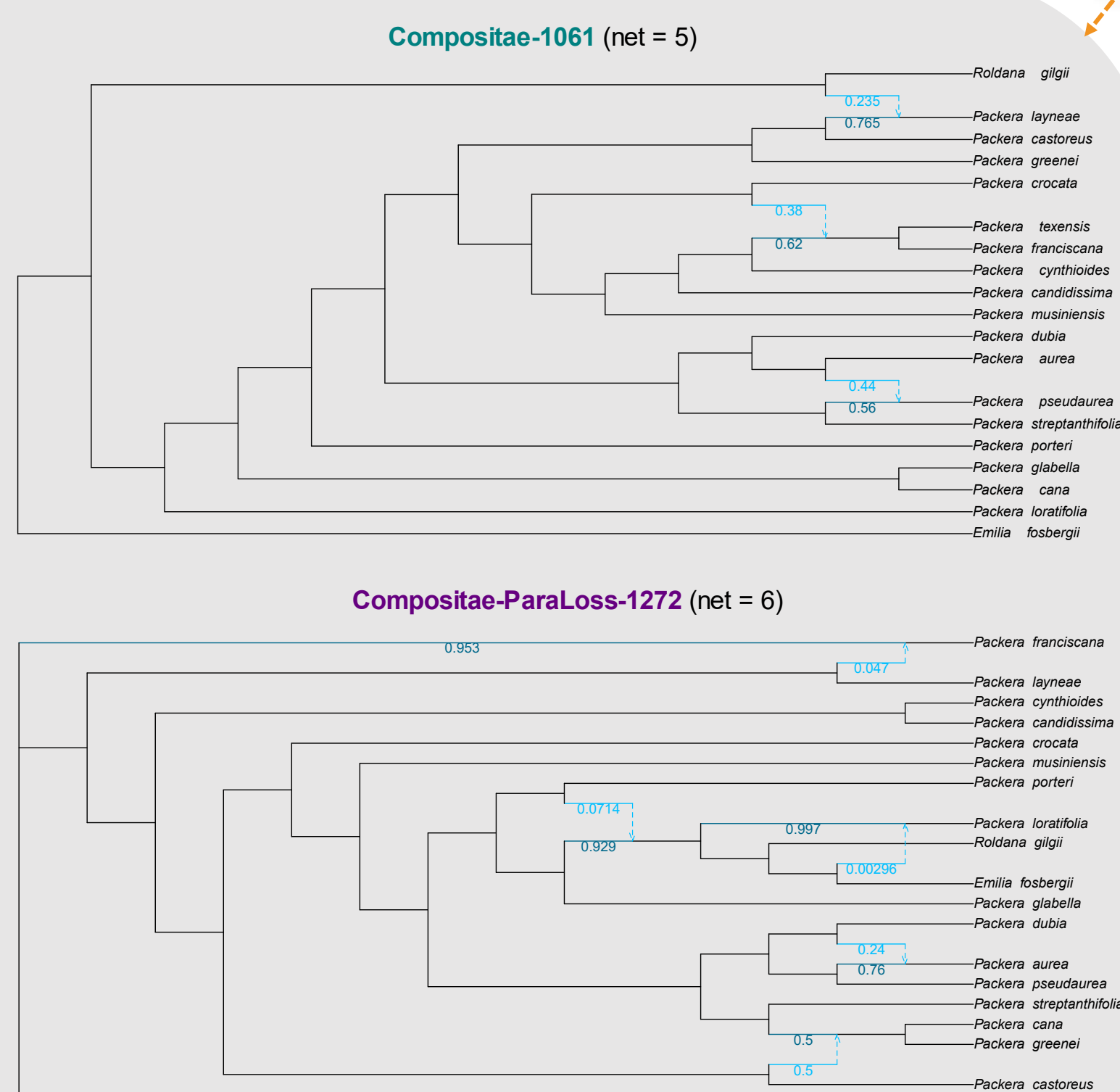


Node support is higher when probe sets are combined and relationships remain the same to **Compositae-ParaLoss-1272**

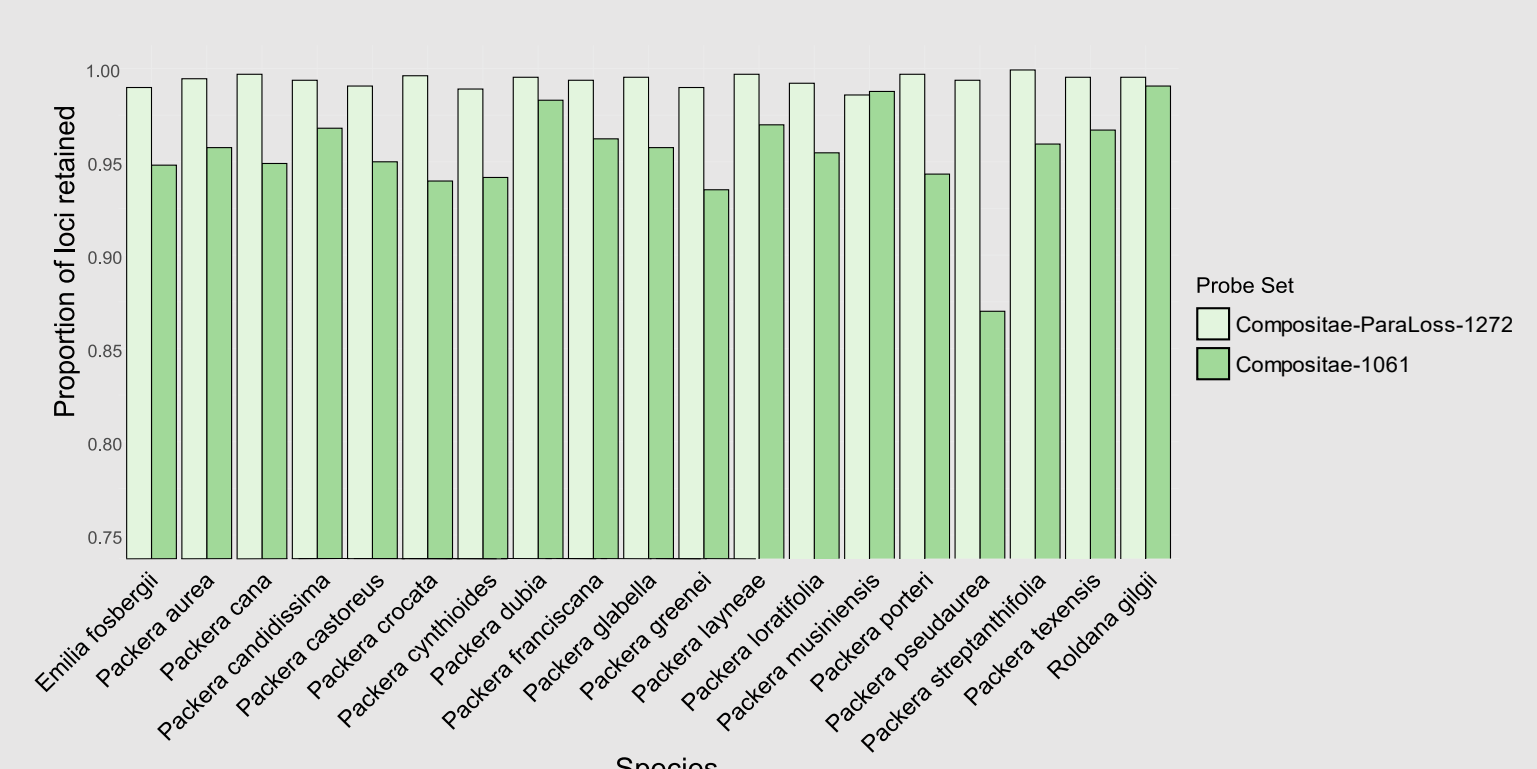
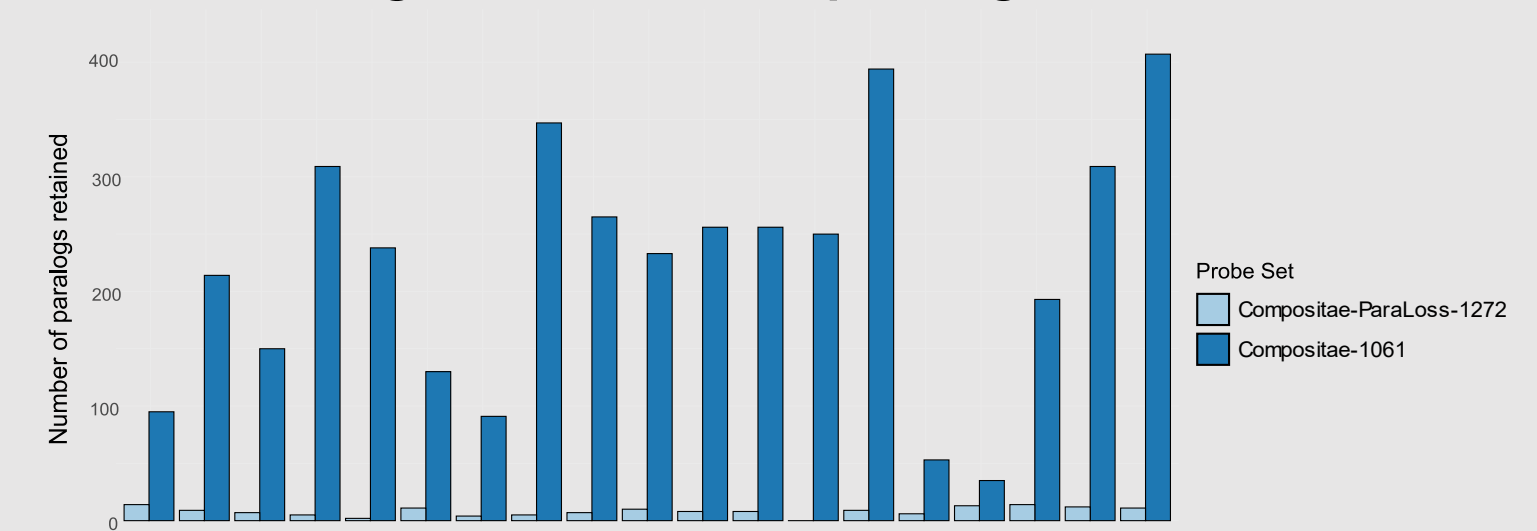
Inferred reticulation events differ between Compositae-1061 and Compositae-ParaLoss-1272 using PhyloNetworks^{6,7}



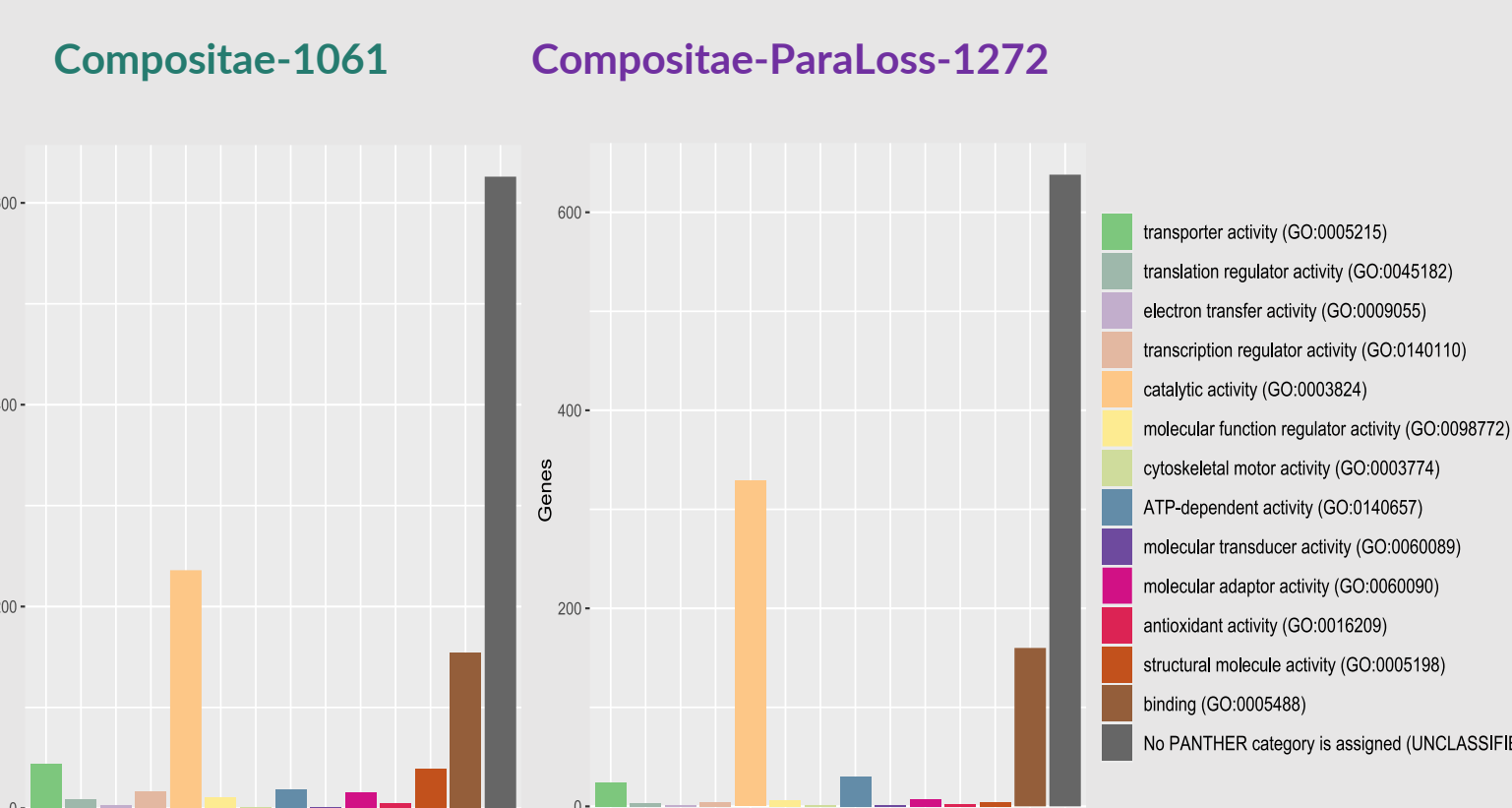
Compositae-1061 and Compositae-1061+Compositae-ParaLoss-1272 are likely best in network analyses



Compositae-ParaLoss-1272 retains longer genes and fewer paralogs



Compositae-ParaLoss-1272 has a narrower range of targeted and conserved gene functions



Questions? Contact me!
moore.erika.r@gmail.com
ermore-pollard.com
 @erikarmore11



CITATIONS:
¹Mandel et al. (2014), *APPS*, 2(2), 1300085; ²Moore-Pollard et al. (2023), *APPS*, in press; ³Chamala et al. (2015), *APPS*, 3(4), 1400115; ⁴Moore-Pollard & Mandel (2023), *bioRxiv*; ⁵Johnson et al. (2016), *APPS*, 4(7), 1600016; ⁶Solis-Lemus et al. (2017), *MBE*, 34(12); ⁷Solis-Lemus & Ané (2016), *PLoS Genet.*, 12(3): e1005896.