Supplementary Figures

Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments

Ning Wang, Email: ningwum@umich.edu

Stephen A. Smith, E-mail: eebsmith@umich.edu

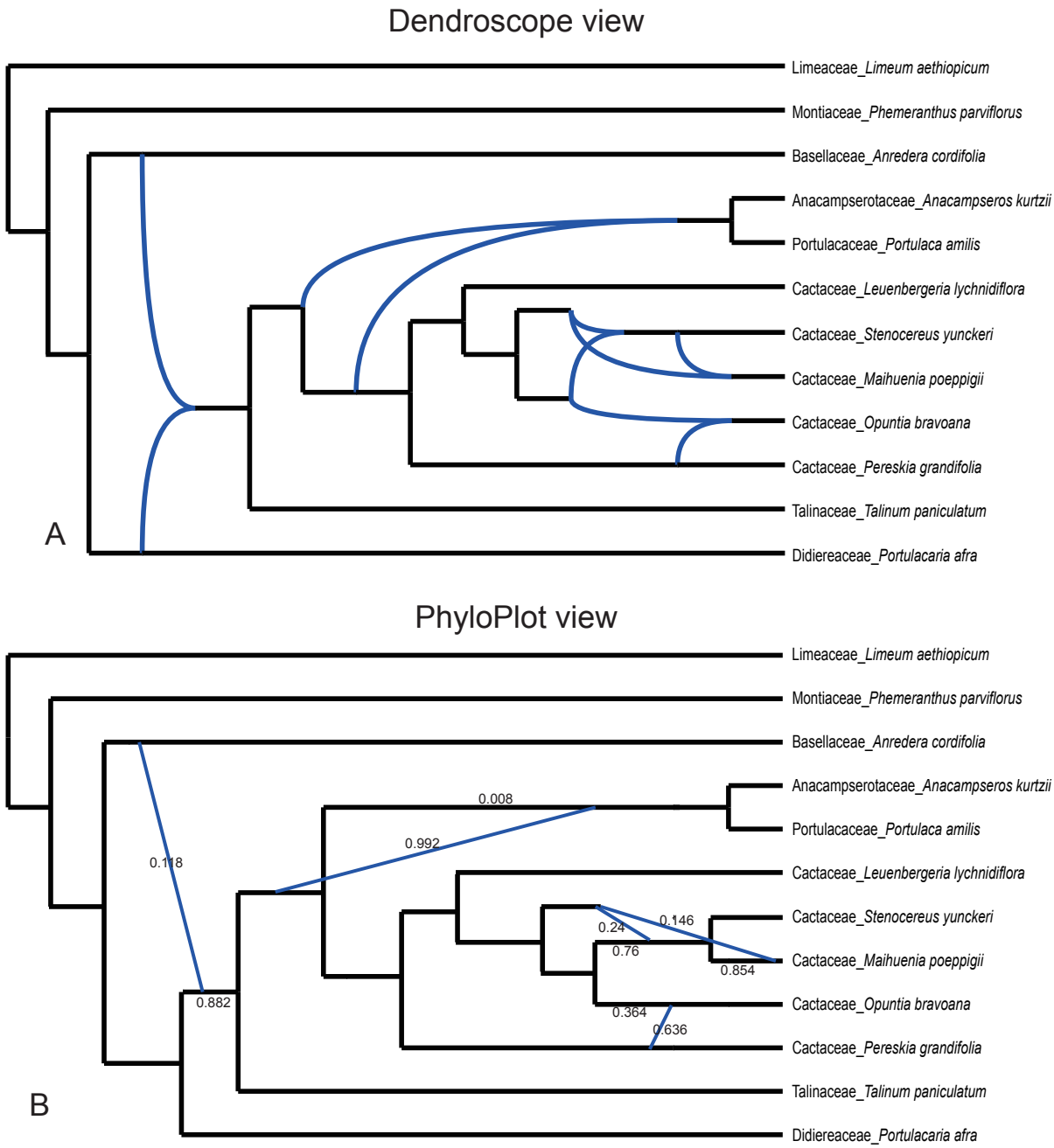

FIG. S1. The phylogenetic network inferred using MPL method in PhyloNet. Taxa were selected from each plant family based on their gene occupancy statistics. A: network visualized in Dendroscope, and B: the same network with inheritance probabilities between hybridization lineages visualized by PhyloPlot that implemented in PhyloNetworks (Solís-Lemus et al. 2017).

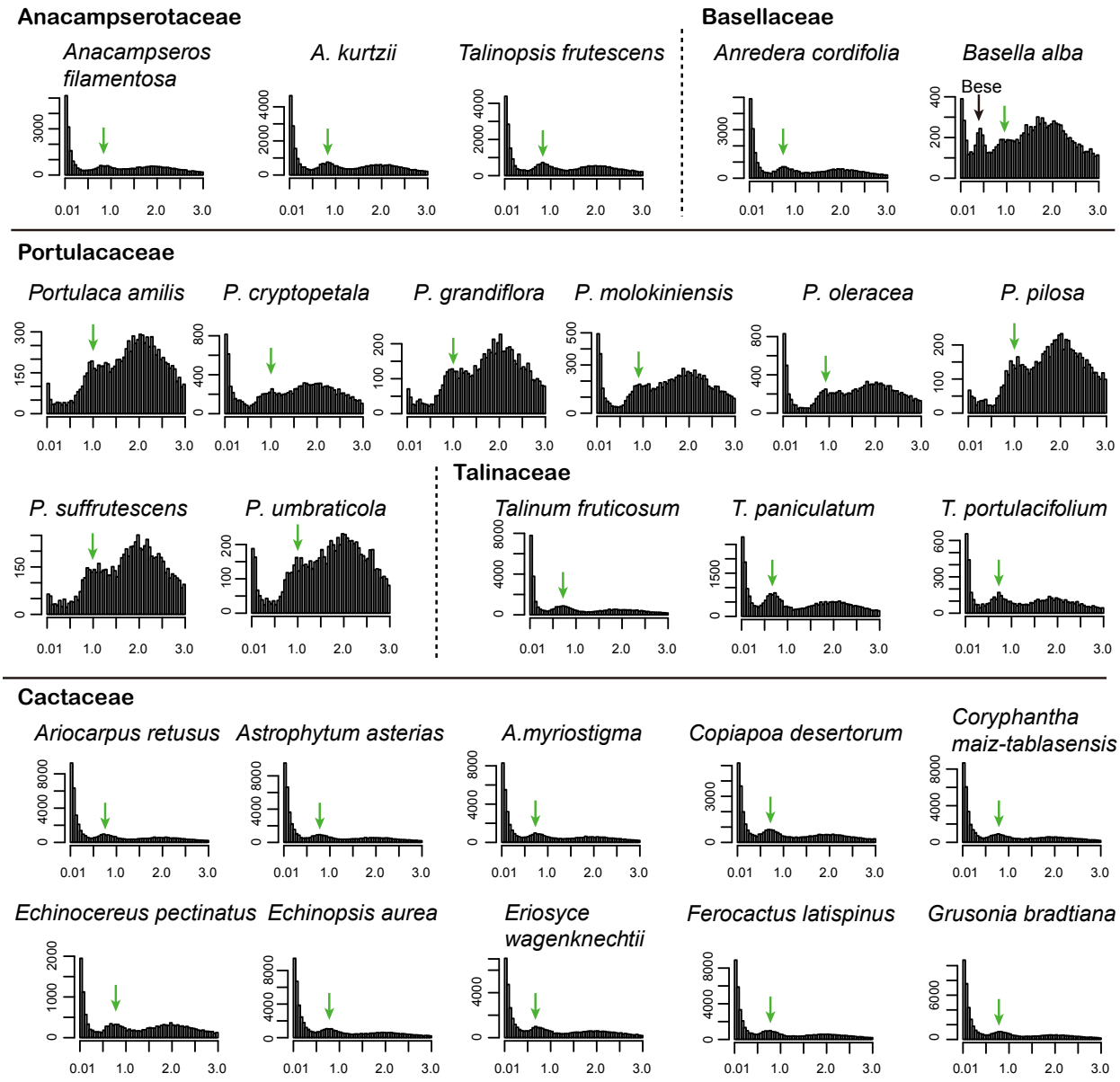

FIG. S2. Distribution of within-taxon synonymous distances (*Ks*) among paralogs gene pairs based on BLASTP hits. Zoomed-in *Ks* plots of certain species from Didiereaceae, Montiaceae, Molluginaceae and Outgroups were shown next to corresponding species. Potential WGD has been marked by arrows. Yang* represents Yang et al. (2018). Node 13: the origin of Portulacineae, Node 27: the origin of Didiereoideae, Node 17: the origin of *Calyptridium*, and Node 19: the MRCA of *Calandrinia* and *Claytonia* in Montiaceae.

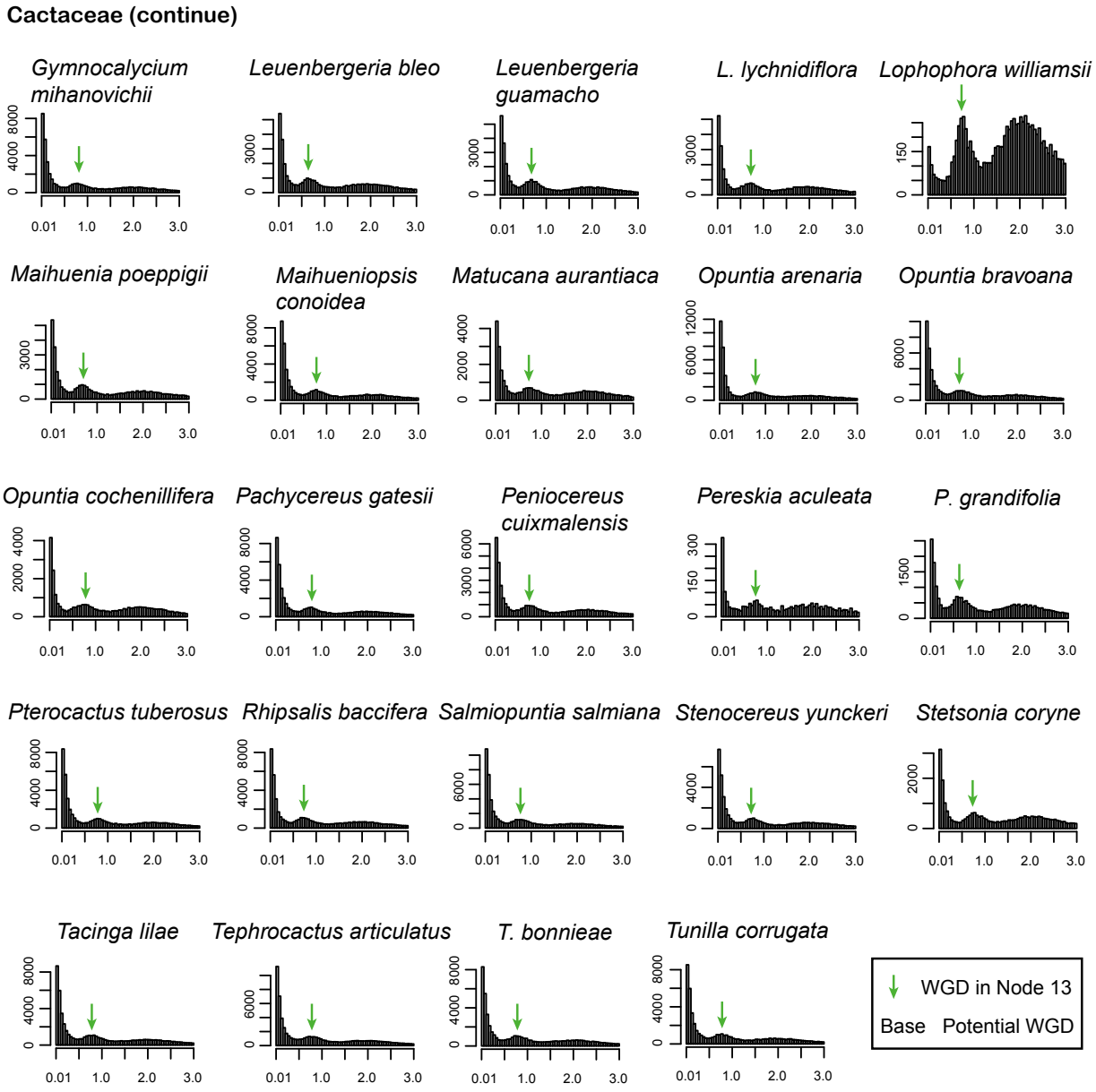

FIG. S2. Continued

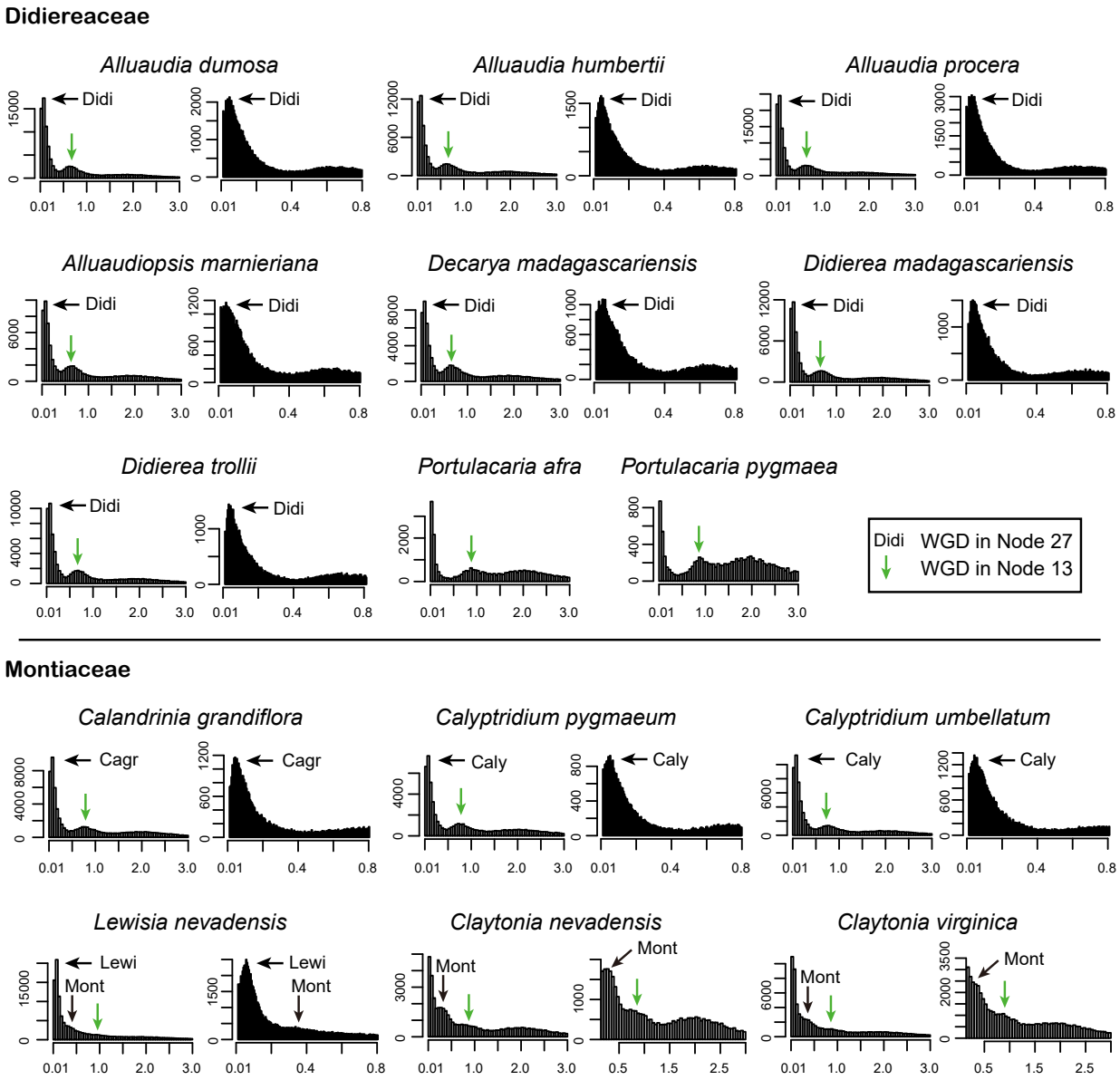

FIG. S2. Continued

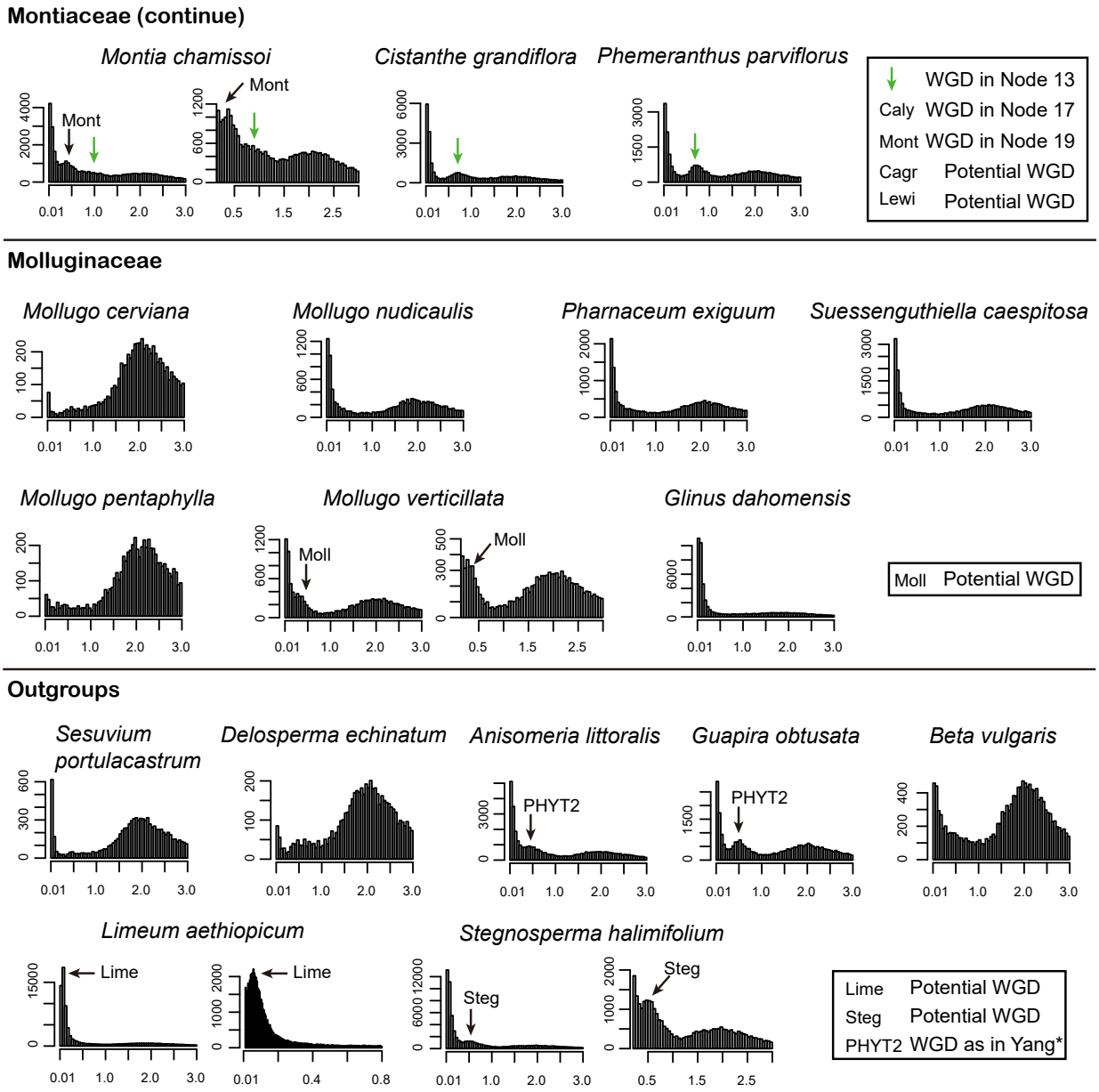

FIG. S2. Continued

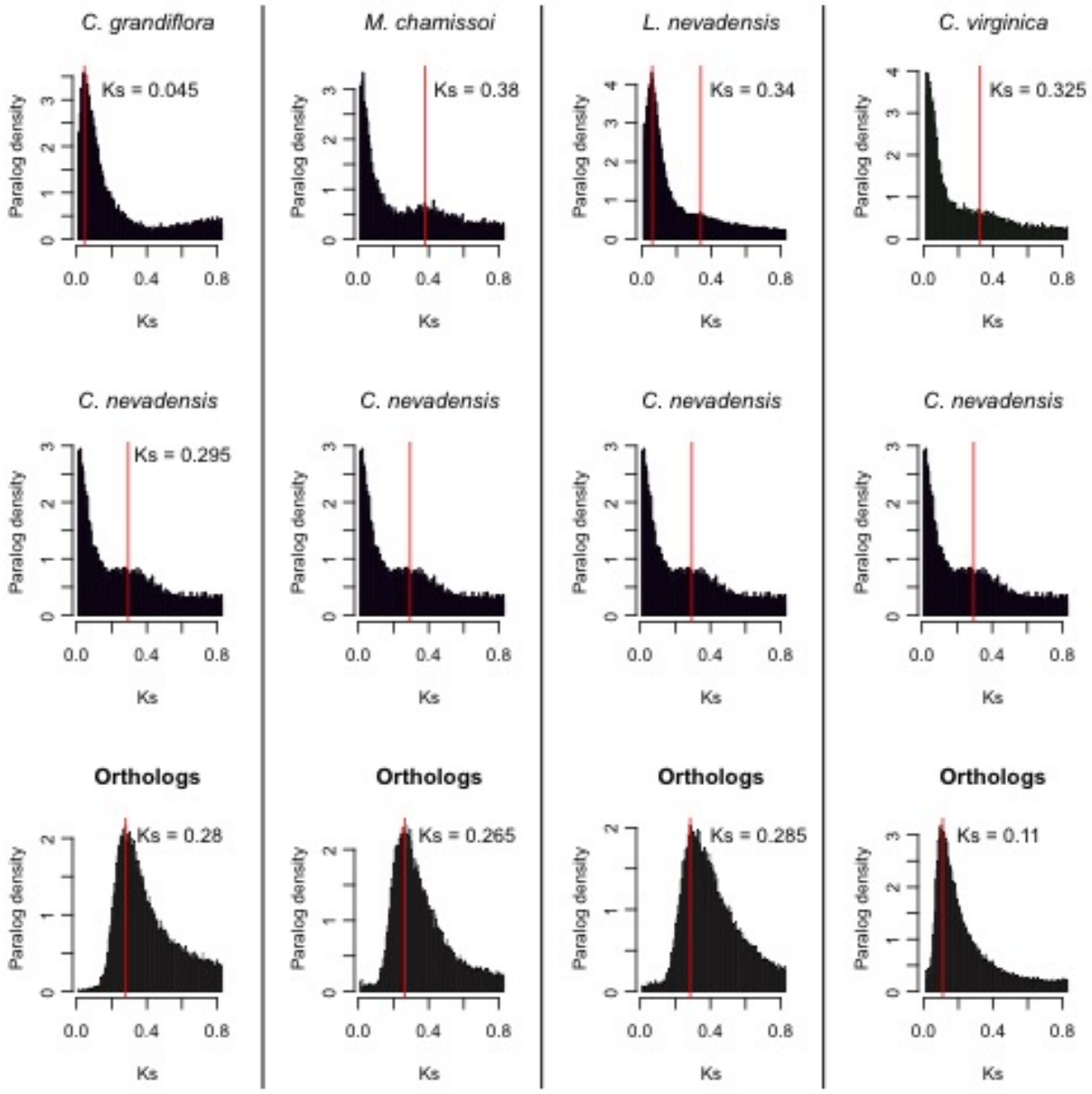

FIG. S3. The probability density *Ks* plots inferred form the within species paralogs and reciprocal orthologs (within the vertical panel). Considering the high number of gene duplication events mapping on the phylogenetic tree (fig. 1), the WGD probably occurred before the speciation of *Montia chamissoi* and *Claytonia*, and at approximately the same time as the divergence of *Lewisia*, but possibly after the divergence of *Calandrina grandiflora*. *Ks* values > 0.8 and < 0.01 are not shown. The *Ks* peaks are identified by eye and marked with the red line with an estimated value.

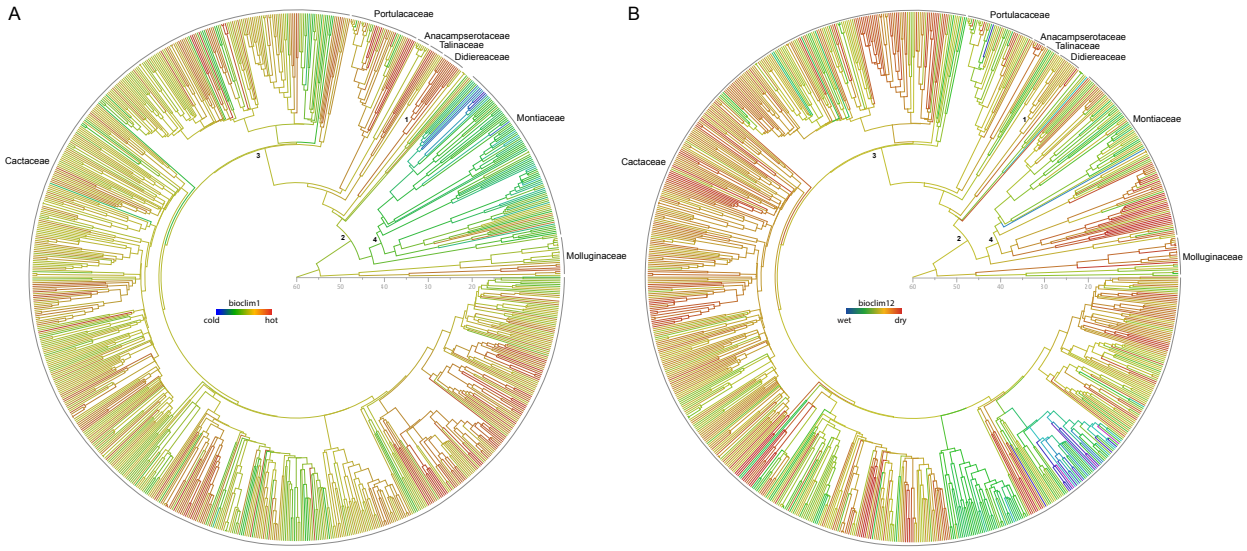

FIG. S4. Reconstruction of mean annual temperature (left) and mean annual precipitation (right) on the 'big time' tree of Portulacineae built with NCBI data. Node labels indicate clades with WGD/extensive gene duplication (1: Didiereoideae, 2: Portulacineae, 3: Cactaceae, 4: Montiaceae except for *Phemeranthus*).

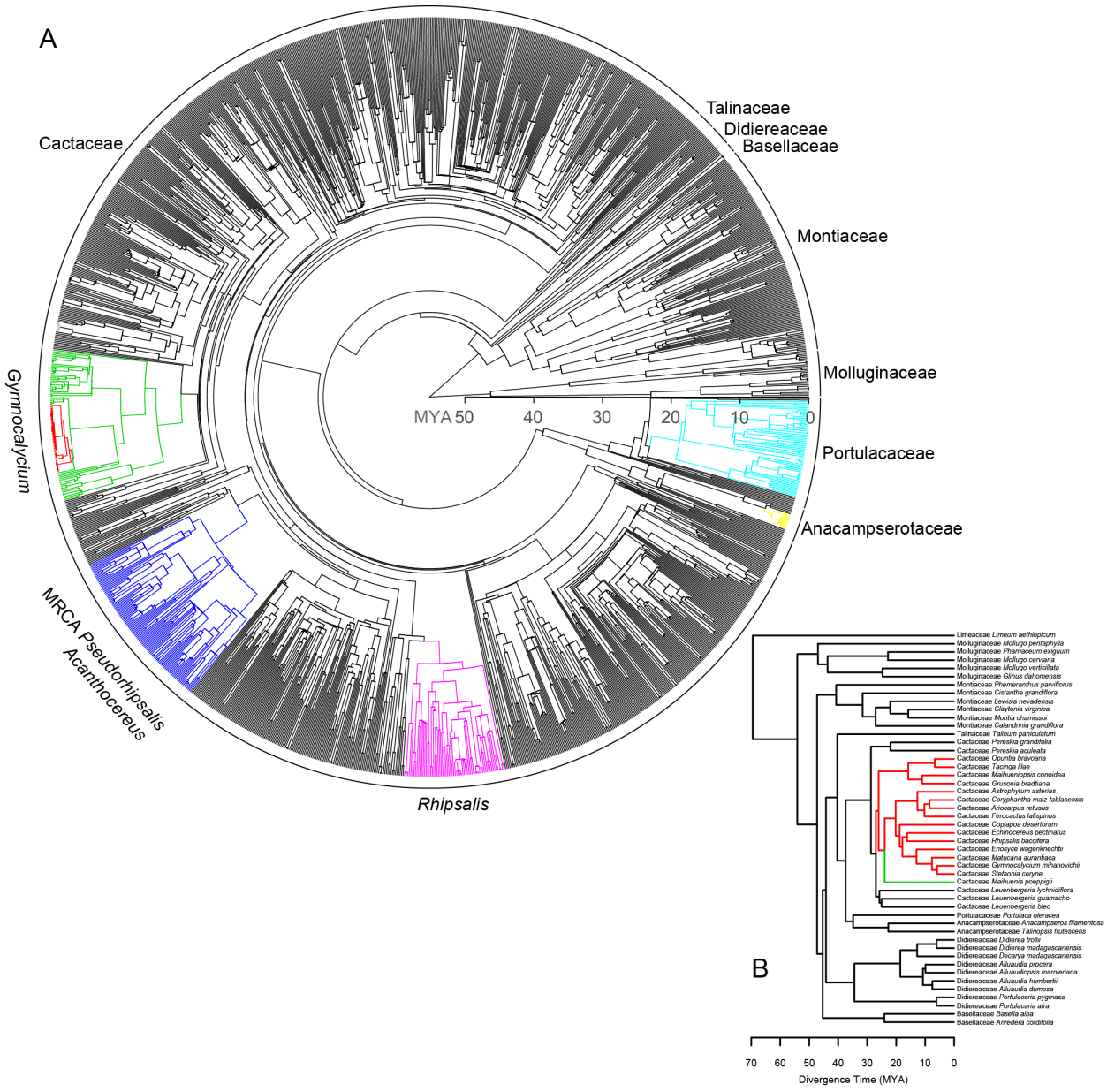

FIG. S5. Diversification rate shifts on a species level phylogeny (A) and a reduced phylogeny of major lineages (B).

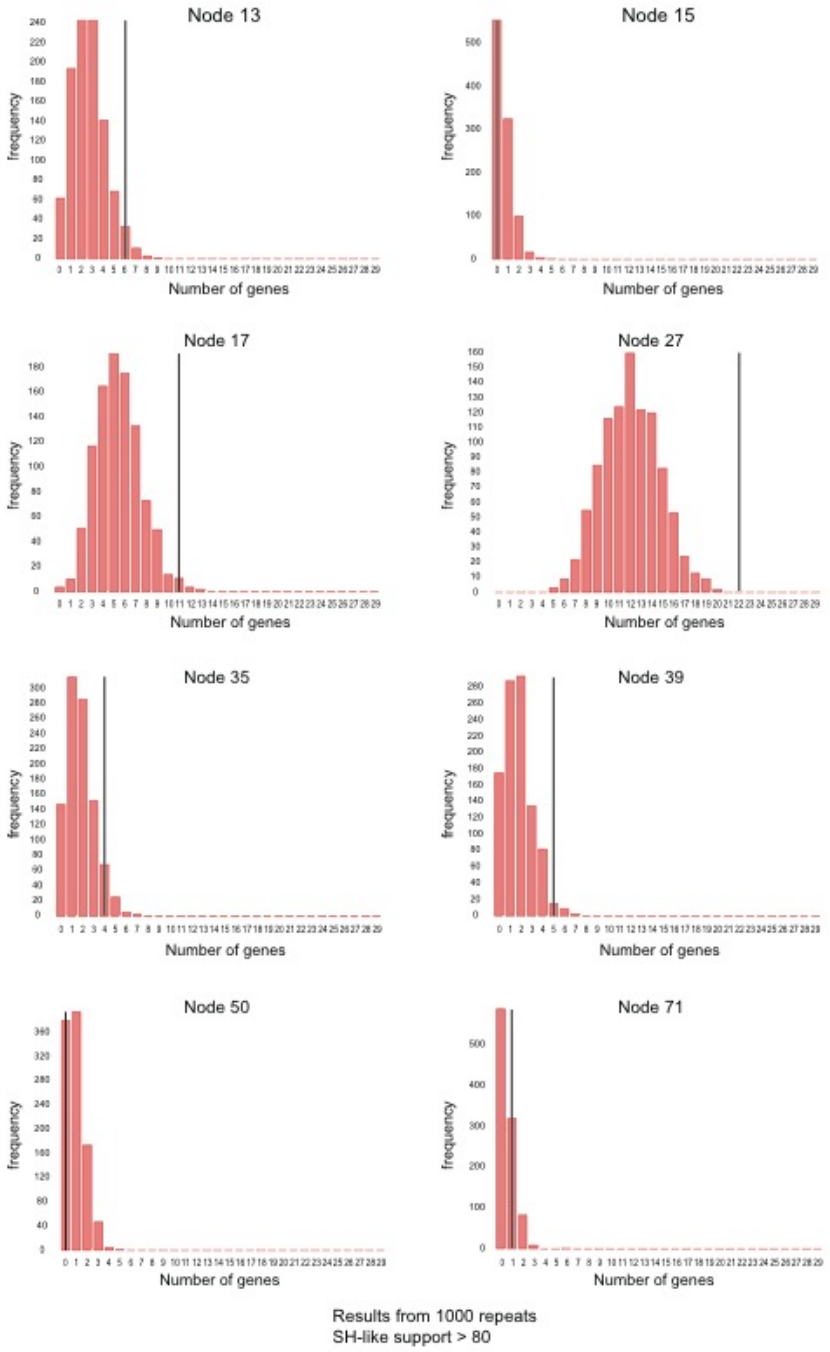

FIG. S6. The frequency distribution of the number of duplication genes at the focal nodes (under SH-like support > 80) generated from 1000 replicates of randomly selecting 29 homologs. The node number can be referred to in fig. 1. Nodes 13, 17 and 19 were identified with WGD events. The solid black line indicates the number of duplication genes from the 29 stress responding genes that targeted on in the main text.

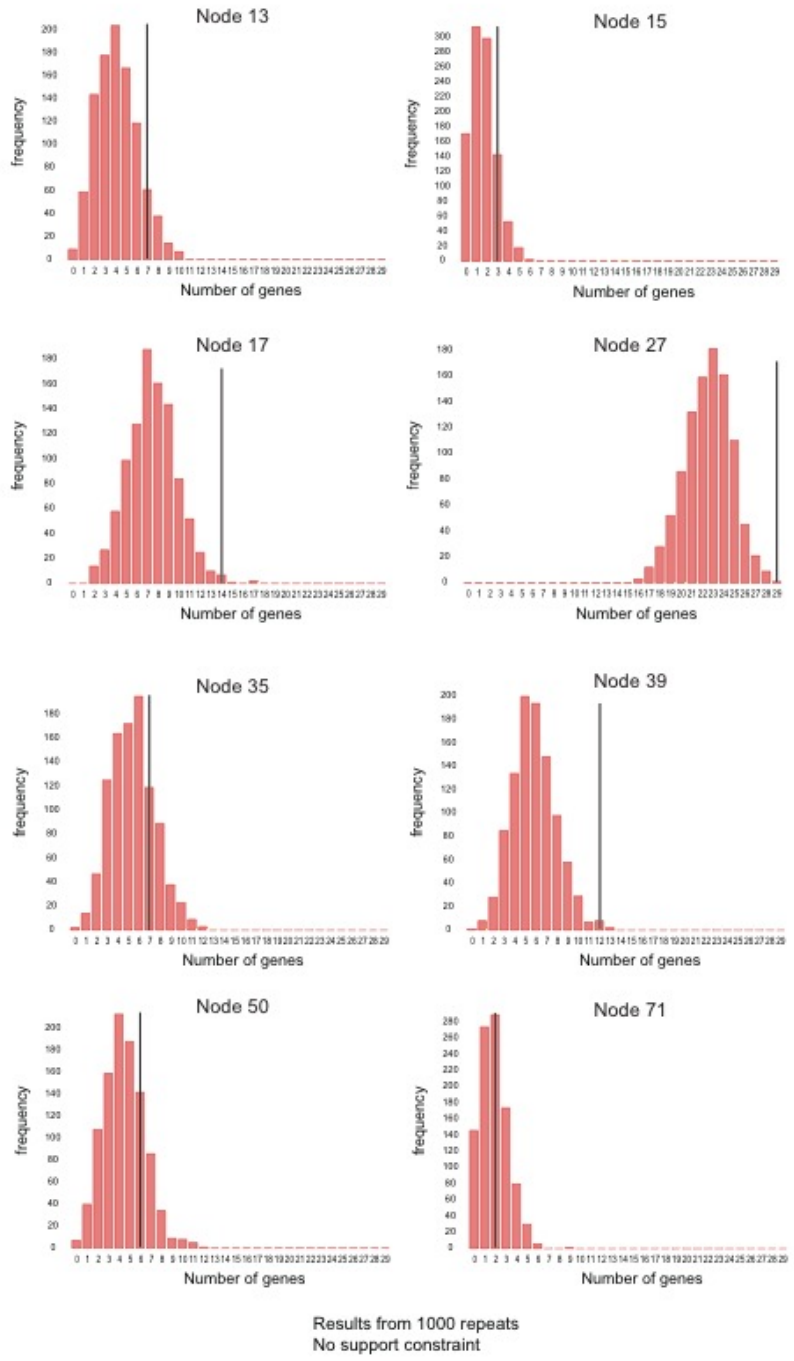

FIG. S6. Continued. Calculation of gene duplication without support constraint.

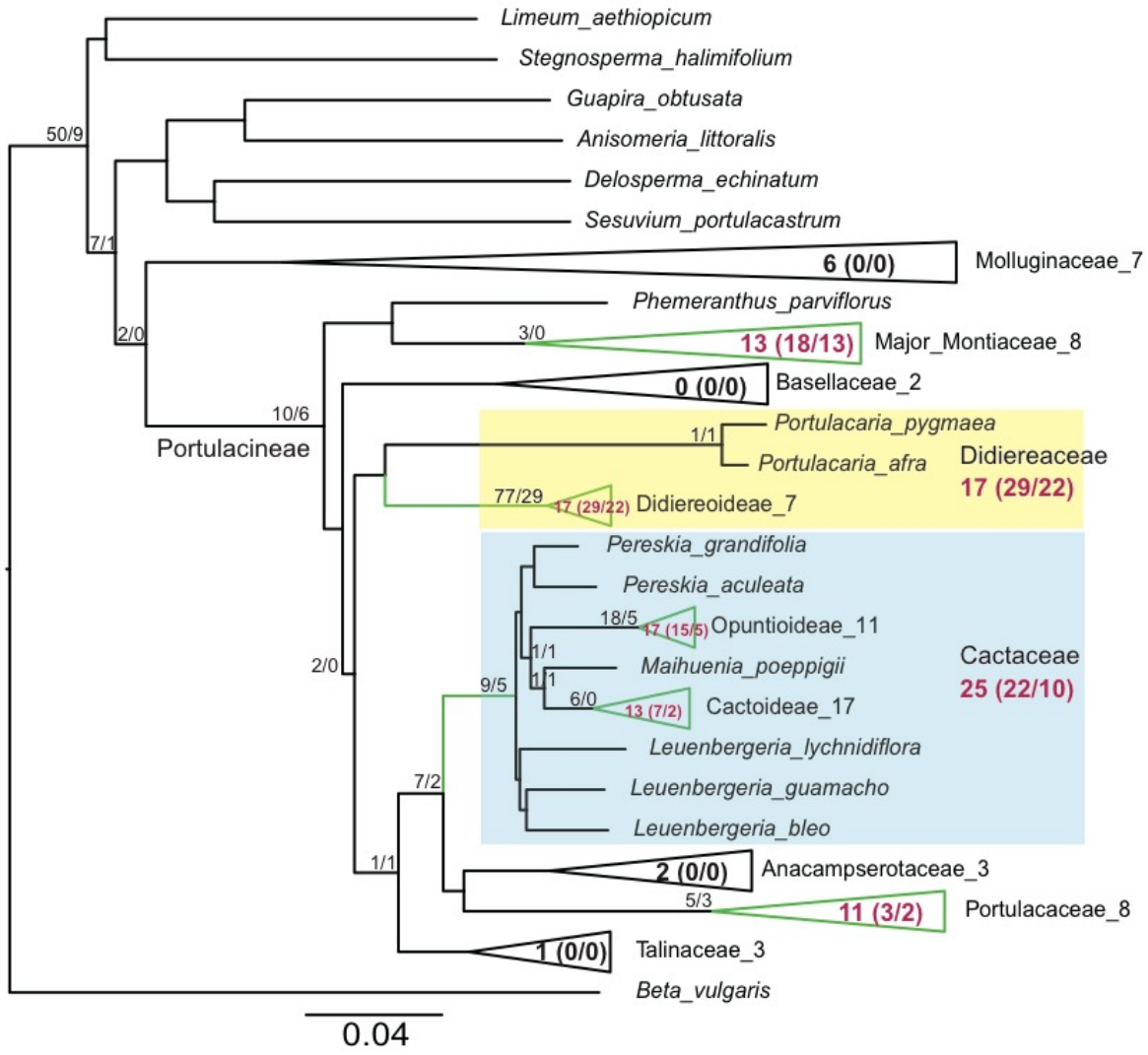

FIG. S7. Gene duplication and lineage positive selection patterns on a collapsed species tree from the 29 targeted genes (tables 2 and S7). Green labeled branches indicated the focal clades with more gene duplication and positive selected lineages. The number of sampling taxa in each major clade are indicated after clade name. The numbers above a certain internal edge represent the total number of duplication events (if any) for that edge from the 29 genes (generated directly from Phyparts analyses), with number before slash being generated under no support constraint, while number after slash being generated under SH-like support > 80. The bolded numbers within each clade are calculated based on table 2: the first number before parentheses represents the number of genes exhibiting positive selection within this clade. The two numbers within the parentheses are the number of genes showing duplication (i.e., duplicated homologs) within the whole clade (number before slash calculated based on "no support constraint" column in table S7 and number after slash calculated based on "SH80" column in table 2).

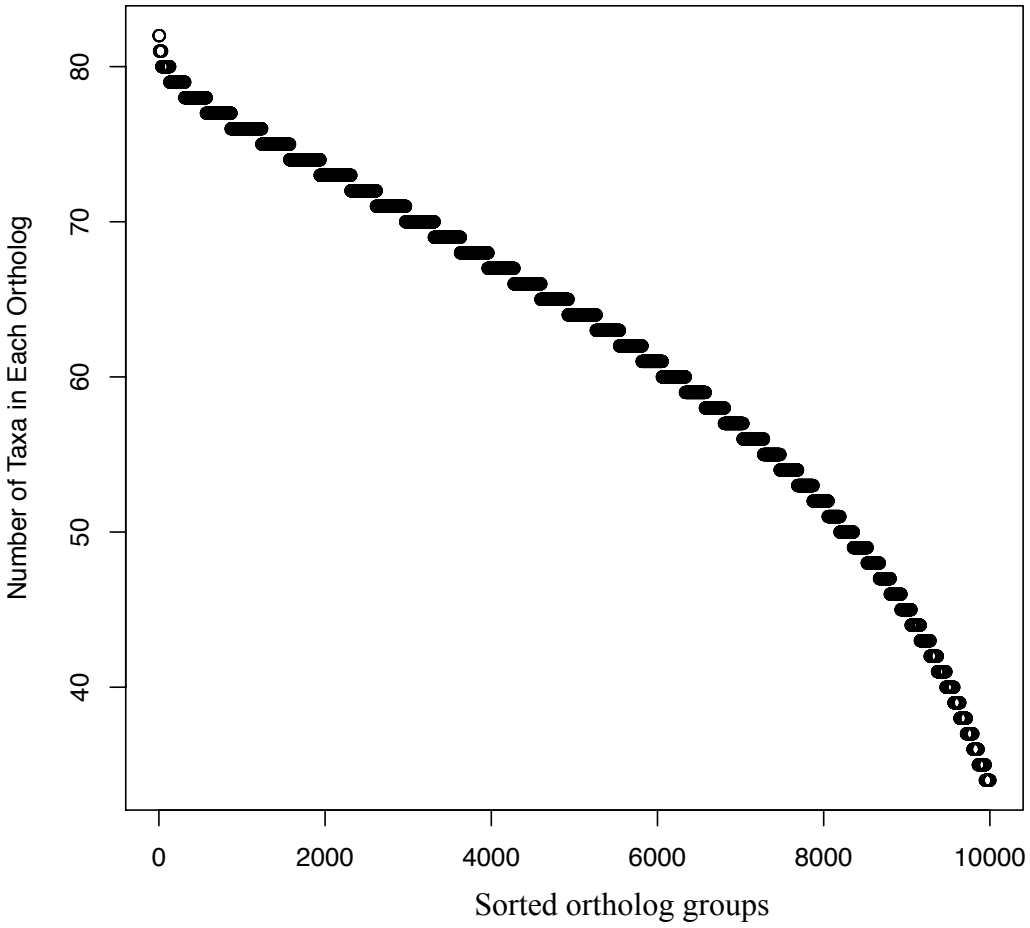

FIG. S8. Ortholog groups ranked from high to low by number of taxa represented. Only ortholog groups with at least 30 taxa were shown. Full taxon occupancy includes 82 taxa.

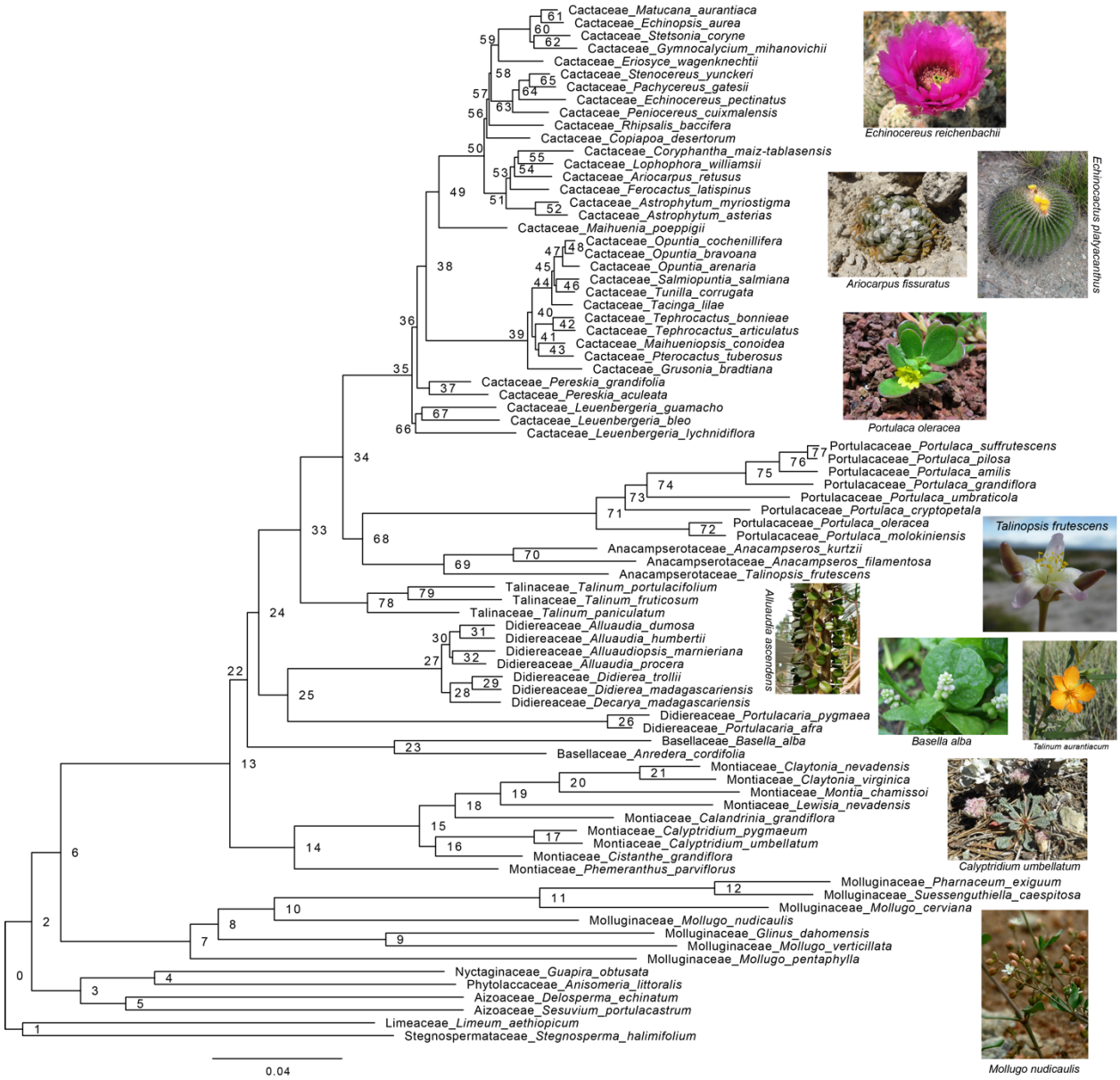

FIG. S9. Species tree from RAxML analysis of the 841-gene supermatrix, with number labeled for each node.

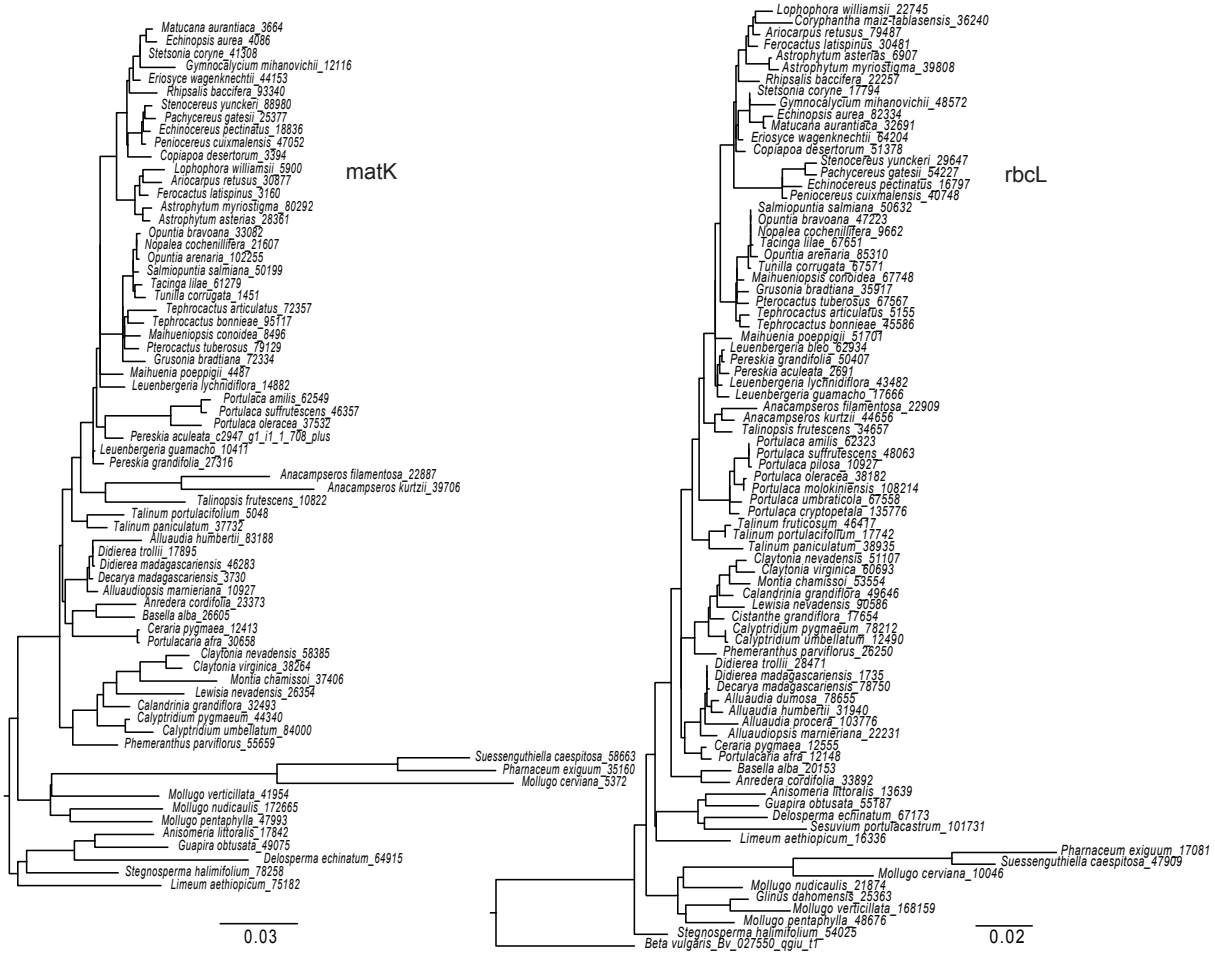

FIG. S10. ML trees built in RAxML with GTRGAMMA model based on matK (left) and rbcL (right) genes that extracted from the transcriptome data.

Refereces:

Solís-Lemus C, Bastide P, Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. *Mol Bio Evol.* 34(12): 3292–3298.

Yang Y, Moore MJ, Brockington SF, Mikenas J, Olivieri J, Walker JF, Smith SA. 2018. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytol.* 217(2):855–870