

1 TARGETED ENRICHMENT OF GENE FAMILIES IN PORTULLUGO

2

3 Targeted Enrichment of Large Gene Families for Phylogenetic Inference: Phylogeny and Molecular

4 Evolution of Photosynthesis Genes in the Portullugo (Caryophyllales)

5

6 Abigail J. Moore^{1,2,5}, Jurriaan M. de Vos^{1,3}, Lillian P. Hancock¹, Eric Goolsby^{1,4}, Erika J. Edwards^{1,4}

7

8 1 Dept. of Ecology and Evolutionary Biology, Brown University, Box G-W, Providence, RI 02912

9 2 current address: University of Oklahoma, Dept. of Microbiology and Plant Biology and Oklahoma

10 Biological Survey, 770 Van Vleet Oval, Norman, OK 73019

11 3 current address: Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, United Kingdom

12 4 current address: Dept. of Ecology and Evolutionary Biology, Yale University, PO Box 208105, New

13 Haven, CT 06520

14 5 corresponding author: (405) 325-5923, abigail.j.moore@ou.edu

15

16 ABSTRACT

17 Hybrid enrichment is an increasingly popular approach for obtaining hundreds of loci for
18 phylogenetic analysis across many taxa quickly and cheaply. The genes targeted for sequencing are
19 typically single-copy loci, which facilitate a more straightforward sequence assembly and homology
20 assignment process. However, single copy loci are relatively uncommon elements of most genomes, and
21 as such may provide a biased evolutionary history. Furthermore, this approach limits the inclusion of most
22 genes of functional interest, which often belong to multi-gene families. Here we demonstrate the
23 feasibility of including large gene families in hybrid enrichment protocols for phylogeny reconstruction
24 and subsequent analyses of molecular evolution, using a new set of bait sequences designed for the
25 “portullugo” (Caryophyllales), a moderately sized lineage of flowering plants (~2200 species) that
26 includes the cacti and harbors many evolutionary transitions to C₄ and CAM photosynthesis. Including

27multi-gene families allowed us to simultaneously infer a robust phylogeny and construct a dense sampling
28of sequences for a major enzyme of C₄ and CAM photosynthesis, which revealed the accumulation of
29adaptive amino acid substitutions associated with C₄ and CAM origins in particular paralogs. Our final set
30of matrices for phylogenetic analyses included 75–218 loci across 74 taxa, with ~50% matrix
31completeness across datasets. Phylogenetic resolution was greatly improved across the tree, at both
32shallow and deep levels. Concatenation and coalescent-based approaches both resolve with strong support
33the sister lineage of the cacti: Anacampserotaceae + Portulacaceae, two lineages of mostly diminutive
34succulent herbs of warm, arid regions. In spite of this congruence, BUCKy concordance analyses
35demonstrated strong and conflicting signals across gene trees for the resolution of the sister group of the
36cacti. Our results add to the growing number of examples illustrating the complexity of phylogenetic
37signals in genomic-scale data.

38

39KEY WORDS

40CAM photosynthesis, C₄ photosynthesis, bait sequencing, Cactaceae, protein sequence evolution, gene
41duplication.

42

43

44 Next-generation sequencing has revolutionized the field of phylogenetics, and there are now
45many approaches available to efficiently collect genome-scale data for a large number of taxa. In one way
46or another, they all involve downsampling the genome as a means to simultaneously sequence
47homologous genomic regions across multiple species. Transcriptome analysis was among the first
48approaches (Dunn et al. 2008; Jiao et al. 2011; Wickett et al. 2014), and this remains an effective method,
49but typically fresh or flash-frozen tissues must be used for RNA extraction. Many researchers have large
50and invaluable collections of stored genomic DNA collected over years of fieldwork that must remain
51relevant. More recently, approaches such as RAD-seq, genome skimming, and hybrid enrichment have
52been adopted as effective means of sub-sampling the genome to enable development of very large

53 datasets (1,000s of loci) across large numbers of individuals with multiplexed sequencing (McCormack et
54 al. 2013). For deeper phylogenetic problems spanning larger clades, hybrid enrichment is emerging as the
55 method of choice (Faircloth et al. 2012; Lemmon et al. 2012; de Sousa et al. 2014; Mandel et al. 2015;
56 Schmickl et al. 2015; 2016).

57 Hybrid enrichment studies tend to limit their scope to ‘single-copy loci’ (SCL), that is, genes that
58 do not appear to have maintained multiple paralogs within a genome after a gene duplication. Targeting
59 SCL has obvious appeal, as it facilitates straightforward contig assembly and reduces the risk of
60 constructing erroneous gene trees due to incorrect orthology assignment. However, the number of SCL in
61 a genome is relatively small, and especially in plants, they appear to be somewhat unusual. As all extant
62 flowering plants have undergone multiple rounds of whole genome duplication (De Bodt et al. 2005; Jiao
63 et al. 2011; Renny-Byfield and Wendel 2014; Soltis et al. 2015), SCL are likely under strong selection to
64 lose additional gene copies after undergoing duplication (Freeling 2009; de Smet et al. 2013). If gene loss
65 happens very quickly post duplication (i.e., prior to subsequent speciation events), these loci would be
66 especially useful for phylogenetics; if, on the other hand, gene loss is more protracted, these loci could
67 instead be especially problematic. Genome-wide estimates suggest that the rate of duplication is quite
68 high (0.01/gene/Ma) and subsequent loss is relatively slow, with the average half-life of a duplicate gene
69 estimated at ~4 Myr (Lynch and Conery 2000). It is at least worth considering that purported SCL may be
70 susceptible to “hidden” paralogy issues, due to differential loss of duplicates over longer periods of time
71 (Martin and Burg 2002; Álvarez and Wendel 2003).

72 An additional limitation of constraining analyses to SCL is the necessary omission of genes of
73 potential interest for other sorts of evolutionary or functional studies, independent of their utility in
74 phylogenetic inference. Beyond the primary goal of generating data for phylogenetic inference, hybrid
75 enrichment offers an unparalleled potential to affordably and efficiently build large comparative datasets
76 of important functional genes, enabling molecular evolution analyses of a scope not seen before. Because
77 the substrate of hybrid enrichment is whole genomic DNA, rather than transcriptomic data of expressed
78 genes, there is also the potential to isolate additional copies of genes that were not expressed at the time

79of tissue collection, providing a more complete picture on the evolutionary dynamics of gene duplication
80and function/loss of function. There are several methodological challenges to unlocking this potential,
81including: 1) designing baits that can target multiple members of large gene families across disparate
82groups of taxa, 2) accurately joining fragmented contigs that belong to the same paralog within
83individuals together into a single non-chimeric locus, and 3) confident assignment of loci to their correct
84orthologs across species. Each of these tasks is difficult but, we demonstrate, not insurmountable.

85 We present a first attempt to include multi-gene families in a hybrid enrichment study of the
86“Portullugo” (Caryophyllales) (sensu Edwards and Ogburn 2012), a diverse clade of ~2200 species of
87flowering plants with a worldwide distribution (Fig. 1). The clade includes nine major lineages, is most
88commonly found in warm and arid or semi-arid environments, and includes such charismatic succulents
89as the cacti of the New World and the Didiereaceae of Madagascar. The portullugo has received a fair
90amount of phylogenetic attention over the decades (e.g., Hershkovitz and Zimmer 2000; Applequist and
91Wallace 2001; Nyffeler and Egli 2010; Ocampo and Columbus 2010; Ogburn and Edwards 2015), yet
92relationships among many of its major lineages remain stubbornly unresolved; one particularly
93recalcitrant region is the relationship between the cacti, *Portulaca*, and Anacampserotaceae.

94 Portullugo also harbors multiple origins of two plant metabolic pathways: C₄ and CAM
95photosynthesis, both complex syndromes that employ a shared set of enzymes to increase internal plant
96CO₂ concentrations and improve photosynthetic efficiency (Edwards and Ogburn 2012). We are
97especially interested in the molecular evolution of genes coding for the major C₃, C₄, and CAM
98photosynthesis enzymes during evolutionary transitions between these metabolic pathways, and included
9919 major photosynthesis gene families in our hybrid enrichment design. Phylogenetic analyses of our data
100resolve many outstanding issues in portullugo phylogeny, and we also present the utility of our dataset for
101analyzing adaptive protein sequence evolution, with a preliminary analysis of the PEP Carboxylase gene
102family. In both C₄ and CAM photosynthesis, PEP Carboxylase is the enzyme recruited to first fix
103atmospheric CO₂ in leaves, where it is temporarily stored as a 4-carbon acid and later decarboxylated in
104the presence of the Calvin cycle. The enzyme is a critical component of both pathways, and previous

105work has demonstrated convergent evolution of multiple amino acid residues associated with both C₄ and
106CAM origins (e.g., Christin et al. 2007; 2014).

107

108MATERIALS AND METHODS

109

110Terminology

111 We use the term paralog to describe gene copies that diverged from one another in a duplication
112event, hence multiple paralogs can be present in a single individual. In contrast, ortholog is used when
113referring to a set of homologous genes that originated via speciation events. Depending on the context, a
114single gene can therefore be included and discussed in the context of a paralog group or an ortholog
115group. In the context of phylogenetic inference involving all sequenced genes, we refer to all orthologs
116from all of the various gene families as loci.

117

118Data Availability

119 All scripts are available in a public repository. One folder contains the analysis pipeline
120(<https://github.com/abigail-Moore/baits-analysis>) and a second folder contains scripts for bait design and
121gene tree/species tree analysis (https://github.com/abigail-Moore/baits-suppl_scripts).

122

123Probe Design

124 Probes for targeted enrichment were designed for use across the portullugo based on analyses of
125eight previously sequenced transcriptomes from the Portulacineae (Christin et al. 2014, 2015;
126Anacampserotaceae: *Anacampseros filamentosa*; Cactaceae: *Echinocereus pectinatus*, *Nopalea*
127*cochenillifera*, *Pereskia bleo*, *Pereskia grandifolia*, *Pereskia lychnidiflora*; Portulacaceae: *Portulaca*
128*oleracea*; and Talinaceae: *Talinum portulacifolium*) and four from its sister group Molluginaceae (Matasci
129et al. 2014; *Hypertelis cerviana* (called *M. cerviana* in 1KP), *Mollugo verticillata*, *Paramollugo*
130*nudicaulis* (called *M. nudicaulis* in 1KP), and *Trigastrotheca pentaphylla* (called *M. pentaphylla* in

1311KP)). MyBaits probes were designed from two sets of genes: 19 gene families that were known to be
132important in CAM and C₄ photosynthesis, and 52 other low- or single-copy nuclear genes (Table S1;
133MYcroarray, Ann Arbor, MI, USA).

134 Sequences for photosynthesis-related genes were taken from the alignments from Christin et al.
135(2014, 2015), which included the transcriptomic data, sequences from GenBank, and individual loci from
136other members of the portullugo clade. Gene family identities for the remaining genes in the portullugo
137transcriptomes were assigned by blasting (BLASTN 2.2.25, default settings; Altschul et al. 1990) them
138against sets of orthologous sequences of known identity from six model plants (Ensembl database; Kersey
139et al. 2016; <http://plants.ensembl.org/>, accessed 4 Dec. 2013). Similarly, we also assigned gene family
140identities to genes from five additional Caryophyllales transcriptomes, which had previously been
141sequenced (*Amaranthus hypochondriacus*, Amaranthaceae; *Boerhavia coccinea*, Nyctaginaceae;
142*Mesembryanthemum crystallinum*, Aizoaceae; *Trianthema portulacastrum*, Aizoaceae; Christin et al.
1432015; *Beta vulgaris*, Amaranthaceae, Dohm et al. 2014) to be able to include them in subsequent
144analyses. Further details of probe design are provided in the Supplementary Methods.

145

146*Taxon Sampling*

147 Sixty portullugo individuals were sequenced (Supplemental Table 2), including multiple
148representatives of all major lineages (with the exception of the monotypic Halophytaceae, which was
149represented by *Halophytum ameghinoi*), and relevant sequences from transcriptomes of two further
150species were added (*Pereskia bleo*, Cactaceae; *Portulaca oleracea*, Portulacaceae). Eleven outgroups
151were added by extracting the relevant sequences from the five non-portullugo, Caryophyllales
152transcriptomes and the six model plant genomes, for a total of 73 taxa.

153

154*Molecular Sequencing*

155 Leaf material was first extracted using the FastDNA Spin Kit (MP Biomedicals, Santa Ana, CA).
156After DNA extraction, samples were cleaned using a QIAquick PCR Cleanup Kit (Qiagen Inc., Valencia,

157CA), following the manufacturer's protocol. DNA was fragmented using sonication and libraries were
158prepared using the NEBNext Ultra or NEBNext Ultra II DNA Library Prep Kits for Illumina (New
159England Biolabs, Ipswich, MA), including addition of inline barcodes (see supplementary methods for
160details). For hybridization with MyBaits probes, samples were combined into groups of 8 or 9 with
161approximately equal amounts of DNA for each sample, resulting in a total of 100–500 ng of DNA in 5.9
162µl of buffer. A low stringency hybridization protocol was followed, because species used for bait design
163were sometimes distantly related to the species sequenced (Li et al. 2013). The remainder of the
164hybridization and cleanup protocol followed version 2 of the manufacturer's protocol, except that the
165cleanup steps took place at 50°C instead of 65°C. Final quantification, combination, and sequencing of
166most samples were performed at the Brown University Genomics Core Facility on an Illumina HiSeq
1672000 or 2500, to obtain 100-bp, paired end reads. Further details are given in the supplementary methods.
168Reads for each individual were submitted to the NCBI SRA (BioProject PRJNA387599, accession
169numbers in Supplemental Table 2).

170

171*Methods Summary for Data Processing and Orthology Assignment*

172 We designed a three-part bioinformatics pipeline to reconstruct gene sequences (Fig. 2). Part I
173aimed to extract all relevant reads for each gene family and then assemble them into contigs. Part II then
174constructed longer sequences from contigs and assigned them to particular paralogs within a gene family.
175Part III identified gene duplications within gene families, extracted phylogenetically useful sets of
176orthologs, and used them for phylogenetic analysis. The major steps in the pipeline are summarized
177below; details are given in the supplementary methods.

178 In part I, paired reads were classified into gene families using BLASTN version 2.2.29 (Altschul
179et al. 1990) and assembled into contigs. A read pair was assigned a gene family if either read matched (e-
180value < 10⁻¹⁶) the sequences used to design its baits. For each gene family, reads were then pooled among
181the individuals that belonged to each of the 9 major lineages, and SPAdes version 3.1.0 (Bankevich et al.
1822012) was used to assemble them into 9 preliminary contigs. By using reads from different individuals

183and different species in the same assembly, we maximized contig number and lengths by also assembling
184chimeric contigs containing reads from multiple individuals; this step allowed us to pull significantly
185more reads into the pool for analysis. In the next step, a new BLAST database was created from the
186chimeric contigs and the sequences from which the baits were designed. The reads were then blasted to
187this larger database, again extracting both reads of a pair if either matched. For each individual and gene
188family, reads were assembled using SPAdes. Finally, the resulting contigs were blasted against the bait
189sequences to identify exons, and only exons were used for all subsequent analyses.

190 Part II of the pipeline identified the paralog that each contig from Part I belonged to, in order to
191combine contigs and maximize the sequence length for each paralog. This iterative process began with
192initial backbone alignments and trees for each gene family; these consisted of the sequences used to
193design the baits as well as the model plant and non-portullugo Caryophyllales sequences. In each
194iteration, all contigs for a gene family were first added to the backbone alignment using MAFFT version
1957.017 (Kato and Standley 2013) and then placed in the backbone gene family tree using the short-read
196classification algorithm in RAxML version 8.0.22 (option “-f v”; Berger 2011, Stamatakis 2014). These
197two steps yielded gene-family trees that contained one or several clusters of contigs. Each cluster was
198treated as a putative paralog and extracted for further testing. For each cluster and individual, contigs
199were combined into a consensus sequence (based on their positions in the backbone alignment) if the
200number of conflicting bases in overlapping contigs (e.g., due to presence of multiple alleles) was
201acceptably low. If a consensus sequence was successfully produced from a cluster of contigs, it was added
202to the backbone tree. If not, those contigs were analyzed again in the next iteration. After six iterations,
203some contigs could still not be combined into acceptable consensus sequences (e.g. due to recent gene
204duplications that were absent from the backbone tree). Here, a single contig per individual and paralog
205was selected.

206 Part III of the pipeline extracted paralogs as separate phylogenetic loci from the gene-family
207trees, by identifying the positions of gene duplications in comparison with a preliminary species tree, and
208used these loci to reconstruct gene trees. Part III was performed twice, first with a preliminary species tree

209constructed from three chloroplast loci (*matK*, *ndhF*, and *rbcL*) and the nuclear internal transcribed spacer
210(ITS) region, all recovered as off-target reads, and then with an updated species tree, reconstructed from
211the loci recovered from the pipeline. For each round of analysis, NOTUNG version 2.8.1.6 (Chen et al.
2122000, Stolzer et al. 2012) was used to find gene duplications in the gene family trees, based on the given
213species tree. While the topology of the species tree was taken as given, poorly supported nodes (< 90%
214bootstrap) on the gene family trees were rearranged to correspond to the species tree, to minimize the
215impact of lack of support on paralog classification. Besides accounting for poorly-supported
216incongruences between the gene-family tree and species tree, we also employed a conservative strategy
217involving a variety of criteria to accept duplications. Most importantly, a duplication was accepted if the
218two sister groups subtended by a putative duplication contained at least one shared individual or two
219shared taxonomic families represented by different individuals. This strategy prevented us from accepting
220putative duplications that in fact represent incongruence between the topologies of the gene family tree
221and species tree. After inspection, at each node that subtended an accepted duplication, the smaller sister
222group was pruned off as a distinct locus, while the larger group was retained on the gene tree. (Note that
223after pruning, the larger group represents more than a single paralog, as it also contains the unduplicated
224sequences from the tree partition not affected by the focal duplication.) This strategy maximized the
225number of loci that contained all or most of the individuals, facilitating phylogenetic inference. We then
226calculated the number of individuals and number of major lineages present in each locus, and removed all
227sites with >90% missing data prior to analysis.

228

229*Reconstruction of Species Trees and Estimating Gene Tree Congruence*

230 To evaluate phylogenetic relations and branch support, we used concatenation and coalescent-
231based approaches on each of five data sets. We selected the following data sets, after dividing species into
232eleven taxonomic groups (the nine major clades recognized within the portullugo, and two additional
233paraphyletic groups for the outgroups, namely non-portullugo Caryophyllales and non-Caryophyllales;
234Table S3, Fig. S1): all loci that were present in two or more groups (g2 matrix, 218 loci, 42.7% missing

235loci, where "missing loci" are loci that were completely absent for certain individuals), all loci present in
236at least 5 taxonomic groups (g5 matrix, 163 loci, 27.7% missing loci), all loci present in at least 50% of
237individuals (i36 matrix, 136 loci, 20.6% missing loci), all loci present in at least 9 taxonomic groups (g9
238matrix, 115 loci, 18.1% missing loci), and all loci present in at least 80% of individuals (i57 matrix, 75
239loci, 10.1% missing loci). Concatenation analyses were performed in RAxML with 100 bootstrap
240replicates. Coalescent-based species trees were reconstructed using ASTRAL II version 4.10.2 (Mirarab
241and Warnow 2015, Erfan and Mirarab 2016) using gene trees from RaxML as input.

242 In addition, to evaluate genomic support for relations among major clades of portullugo, Bayesian
243concordance analysis was performed using BUCKy version 1.4.4 (Larget et al. 2010) based on the
244posterior distribution of gene trees from analyses in MrBayes 3.2 (Ronquist et al. 2012). BUCKy
245estimates the genomic support as a concordance factor for each relationship found across analyses of all
246individual loci (Ane et al. 2006; Baum 2007).. This way, groups of genes supporting the same topology
247are detected, while accounting for uncertainty in gene tree estimates. BUCKy thus alleviates the concern
248that methods to reconcile gene trees, such as ASTRAL, may underestimate uncertainty of species tree
249(Leache & Rannala 2011), by highlighting genomic conflict. The implementation of BUCKy analyses
250requires identically named tips to be present in trees for all loci. In order to maximize the number of loci
251that could be simultaneously analyzed, taxa were renamed to their major lineage (outlined above), and all
252but one random exemplar per family was pruned from each sample of the posterior distribution of
253MrBayes trees. Although there is strong support for the monophyly of the major portullugo lineages
254(Nyffeler and Eggli 2010), our renaming and pruning approach does not require it to be so for each
255individual gene tree. Rather, the phylogenetic position of a family is averaged out over all probable
256positions, because a large number of renamed, pruned trees from each posterior distribution are input.

257 We conducted two sets of BUCKy analyses: one focusing on the position of Cactaceae within the
258ACPT clade (Anacampserotaceae + Cactaceae + Portulacaceae + Talinaceae), and a broader
259Portulacineae-wide analysis, focusing on the remaining relationships after collapsing the ACPT clade to a
260single taxon. MrBayes analyses generated a posterior distribution of gene trees for each locus and

261 consisted of two runs of 4,000,000 generations with default MCMCMC settings, sampling every 4,000
262 generations, employing a GTR substitution model with gamma-distributed rate variation across sites.
263 After confirming the adequacy of these settings and excluding 25% of samples as burnin, runs were
264 combined to a full posterior distribution of 1500 samples and subsequently thinned to 200 samples. The
265 full posterior distribution was subjected to the renaming and thinning approach described above. For each
266 locus, posterior probabilities of the monophyly of each lineage and their relationships to one another were
267 scored from the thinned posterior distribution as the fraction of sampled trees that contained the node of
268 interest. Analyses were conducted on all loci in which all focal lineages were present (ACPT: n=143;
269 Portulacineae: n=132). For all analyses, we ran BUCKy using four runs of 100,000 generations each and
270 computed genome-wide concordance factors (in which loci are interpreted as a random sample from the
271 genome) of all possible relationships, as well as the posterior probability of each locus pair to support the
272 same tree. All processing of MrBayes and BUCKy files was performed using custom R scripts.

273

274 *Molecular Evolution of PEP Carboxylase*

275 Phylogenetic trees of the three major lineages of PEP Carboxylase in eudicots (*ppc-1E1*, *ppc-1E2*,
276 and *ppc2*) were inferred using RAxML. Coding sequences were translated into amino acid sequences and
277 numbered according to *Zea mays* sequence CAA33317 (Hudspeth and Guala 1989). Fourteen amino acid
278 residues (466, 517, 531, 572, 577, 579, 625, 637, 665, 733, 761, 780, 794, and 807) that were previously
279 determined to be under positive selection in C₄ grasses (Christin et al. 2007; Besnard et al. 2009), as well
280 as position 890, which is associated with malate sensitivity (Paulus et al. 2013), were examined across the
281 three major paralogs separately. Some residues could not be identified due to missing data or ambiguity.
282 For these residues, marginal ancestral state reconstruction was performed using the *rerootingMethod*
283 function in the R package phytools (Revell 2012) to determine the amino acid with the highest marginal
284 probability.

285

286 RESULTS

287

288 *Sequence Coverage and Dataset Structure*

289 We obtained between 682,702 and 13,008,046 reads per individual (mean $3,385,697 \pm 2,953,383$;
290 Table S2). Percent enrichment, expressed as the percentage of read pairs yielding a blast hit to the bait
291 design alignments, ranged from 0.26% to 12.32% across individuals, with a mean of $2.66 \pm 1.81\%$ after
292 two rounds of blasting (0.17% to 6.20%, with a mean of $1.72 \pm 0.98\%$ after one round of blasting; Table
293 S2) and did not differ between species closely related to individuals with transcriptomes (Cactaceae,
294 *Portulaca*, *Mollugo*) and more distantly related species ($2.73 \pm 1.01\%$ vs $2.63 \pm 2.56\%$, $N = 59$, $p = 0.40$
295 for 1-tailed t-test, samples having unequal variance).

296 Part I of the analysis pipeline yielded a widely varying number of contigs per individual and gene
297 family (mean of 15.6 ± 35.5 , range of 0 (in numerous cases) to 1816 (PEPC genes in *Ullucus tuberosus*)).
298 The total number of contigs per individual ranged from 500 (*Calandrinia lehmannii*) to 2576 (*Ullucus*
299 *tuberosus*), with a mean of 1123 ± 451 (Table S2). Part II of the pipeline consolidated these contigs into
300 longer sequences, and the total number of sequences per individual ranged from 62 (*Calandrinia*
301 *lehmannii*) to 221 (*Alluaudia procera*), with a mean of 149 ± 28 (Table S2). The number of loci per
302 individual per gene family was also variable (mean of 1.85 ± 1.61 , range of 0 (in numerous cases) to 13
303 (*nadmdh* in *Alluaudia procera*)).

304 The pipeline yielded a total of 665 phylogenetic loci, with the number of individuals per locus
305 ranging from 1 to 72 and the number of loci per gene family varying between 1 (i.e., putatively single-
306 copy; 5 loci) and 34 (*nadmdh*). Taxon sampling across these loci was quite variable, with some loci being
307 present in all major lineages, and others only being present in a single group (because they were paralogs
308 due to a gene duplication near the tips). The mean sequence length per locus varied between 152 and
309 4075 bp (812 ± 616 bp). There was considerable variation in the number of loci per gene family, both
310 between gene families and between individuals. Photosynthesis genes generally had many more paralogs
311 than the non-photosynthesis genes, although there was variation in both groups of gene families (Fig. S2

312for heatmaps showing the number of recovered per gene family for each individual; Fig. S3 for
313duplication numbers across all branches of the species tree).

314

315Phylogenetic Analyses

316 Most nodes were congruent and well supported across all analyses of all matrices (Fig. 3; Fig. S4,
317all remaining trees). The major differences are summarized in Table S5. Most of the conflict between
318analyses concerned relationships within the cacti, particularly the various species of *Pereskia*, and some
319of the relationships among closely related species of the Montiaceae. All nine of the major clades within
320the portullugo were well supported (>95% bootstrap) in both coalescent (ASTRAL) and concatenated
321(RAxML) species trees. The following larger clades were also always well supported: Anacampserotaceae
322and Portulacaceae as sister lineages; the clade comprised of Anacampserotaceae, Cactaceae, and
323Portulacaceae (ACP); ACP plus Talinaceae (ACPT); ACPT plus Didiereaceae; and the Portulacineae
324(portullugo without Molluginaceae). The analyses consistently recovered Montiaceae alone as sister to the
325seven remaining clades in the Portulacineae, and Basellaceae as sister to Halophytaceae, although with
326lower support in both cases.

327 The major conflict between analyses resided within Cactaceae. While all analyses recover the
328“core cacti” (sensu Edwards et al. 2005) as monophyletic, four of the five concatenated analyses show
329*Maihuenia* to be sister to Opuntioideae + Cactoideae with high support, while all ASTRAL analyses and
330the i57 concatenated analysis recover Opuntioideae as sister to *Maihuenia* plus Cactoideae. The
331relationships within *Pereskia* are quite variable and are generally poorly supported. Two analyses recover
332a monophyletic *Pereskia*, two recover *P. lychnidiflora* alone as sister to the core cacti, and the remaining
333six recover a clade composed of *P. grandifolia*, *P. sacharosa*, and *P. horrida* as sister to the core cacti, a
334relationship first proposed by Edwards et al. (2005; the “caulocacti”).

335 Even though the various species tree analyses demonstrated congruence in the resolution of major
336relationships, Bayesian Concordance Analysis highlighted significant underlying genome-wide conflict
337among loci. First, the primary concordance tree from the portullugo-wide analysis revealed similar

338topologies as the ASTRAL and concatenation analyses, but with low to medium genome-wide
339concordance factors (CF; from 0.63 for ACP to 0.28 for Portulacineae except Montiaceae; Table S6),
340indicating that significant portions of the genome support relationships that deviate from the dominant
341signal (Fig. 4). For instance, in the ACPT analysis, the sister group of Cactaceae as Portulacaceae +
342Anacampserotaceae was supported by half of our sampled loci (mean CF 0.52), while the other half
343supported either Anacampserotaceae (mean CF 0.25) or Portulacaceae (mean CF 0.23) alone as sister to
344Cactaceae (Fig. 4). In the Portulacineae-wide analysis, the ASTRAL and concatenation-inferred position
345of Halophytaceae as sister to Basellaceae received a mean CF of 0.35, somewhat higher than an
346alternative placement as sister to Montiaceae (mean CF 0.19; Fig. S4).

347

348Molecular Evolution of PEP Carboxylase

349 One of the major *ppc* paralogs, *ppc-1E1*, has undergone multiple rounds of duplication in
350ancestral Portulacineae, with sequences clustering into five main paralogs, denoted *ppc1E1a–e* (following
351Christin et al. 2014; Fig. 5). In addition, *ppc-1E1a* underwent a further duplication (*ppc-1E1a'*) in
352ancestral *Portulaca*. Members of Didiereaceae and *Lewisia* appear to possess additional copies of *ppc-*
353*1E1* distinct from the *ppc-1E1a–e* duplications, though their placement is poorly supported. Non-
354Portulacineae Caryophyllales possess a single copy of *ppc-1E1*.

355 We inferred multiple C₄-associated amino acid substitutions in *ppc-1E1*, both inside and outside
356of Portulacineae. In particular, *ppc-1E1a'* within the *Portulaca* lineage, which has been previously shown
357to be associated with C₄ activity (Christin et al. 2014), contains an elevated number of C₄-associated
358amino acids relative to *ppc-1E1a* across Portulacineae. We also looked for C₄-specific AA substitutions in
359CAM species, with the hypothesis that there may be convergence in coding sequences between these two
360syndromes due to their shared function of PEP Carboxylase. We discovered a number of C₄-adaptive
361osubstitutions in *ppc-1E1b–e* in different CAM lineages, with most lineages showing the greatest
362accumulation (five) in *ppc-1E1c*. However, some species in particular (*Ceraria* + *Portulacaria*) show a
363very broad distribution of putative C₄-like substitutions across *ppc-1E1b–e*. The most ubiquitous and

364 consistent C₄-adaptive AA in other plant groups, Ser780, has appeared only in *ppc-1E1c* and *ppc-1E1e*
365 (and the C₄ *ppc-1E1a*). Sequences of *ppc-1E2* and *ppc2* were also examined, and both paralogs exhibit
366 very low rates of evolution in general, and a low number of C₄-associated amino acid substitutions (0.59
367 and 0.29 C₄-associated substitutions per site per unit branch length, respectively) relative to *ppc-1E1a-e*
368 (1.44, 1.12, 1.47, 2.37, and 2.45, respectively).

369

370 DISCUSSION

371

372 *Targeted Gene Enrichment with Multi-Gene Families*

373 Hybrid bait enrichment is becoming increasingly common in phylogenetics, with researchers
374 developing specialized probe sets designed specifically for unique lineages, much in the way we have
375 done with portulugo (e.g., Lemmon et al. 2012; Nicholls et al. 2015; Heyduk et al. 2016). The critical
376 difference is that most studies focus on capturing ‘single-copy’ genes in plant genomes (De Smet et al.
377 2013; Chamala et al. 2015), which facilitate contig assembly and homology assignment. Despite the
378 obvious practical advantages of single-copy loci for phylogenetic inference, many questions require
379 sequencing other parts of the genome. By broadening sampling to include multi-copy gene families,
380 targeted enrichment provides an effective means to collect data for phylogenetic reconstruction while
381 simultaneously accumulating comparative datasets on particular gene families of interest. In this study,
382 we attempted to target a wide array of genes in our probe design, and included gene families of major
383 photosynthesis proteins that are relevant to our broader research program.

384 While all phylogenomic datasets undoubtedly contain some errors in orthology assignment,
385 overall, we feel confident that we are accurately sorting paralogs into their correct ortholog groups,
386 resulting in accurate phylogenetic inference. First, our phylogenetic findings are congruent with those
387 recovered in phylogenetic studies that used a non-controversial, sanger-sequencing-based approach (e.g.,
388 the ACPT clade, Portulacineae ex Montiaceae; Nyffeler & Egli 2010). Secondly, for classic recalcitrant
389 nodes, our BUCKy analyses reveal significant underlying genomic conflict, despite an emerging

390 phylogenetic resolution (discussed below). Thus, rather than conflicting previous studies, our results are
391 largely consistent with them, and provide greater insight into the real genomic conflict underlying
392 historically recalcitrant nodes.

393 Overall, we have identified two key challenges to working with multi-gene families in hybrid
394 enrichment approaches. First, if baits are designed only from exons (e.g., RNA-seq data, which are the
395 most likely genomic resource for most groups), we can only sequence across short introns. Thus, long
396 introns in the genomic data prevent the entire gene from assembling into one contig, which increases the
397 chance of assembling chimeric sequences derived from multiple paralogs in later steps. Second, the
398 sampling density of the gene family ‘backbone tree’ used to assign orthology has an enormous effect on
399 the ability to accurately classify contigs. Both of these limitations were mostly (though not entirely)
400 overcome by our iteration of the short-read classification step, as confidently placed contigs were
401 maintained in the backbone tree for further rounds of homology assignment. We admit that our approach
402 is largely one of “brute force” at this point, with massive iteration and refinement of key steps. However,
403 as researchers continue to sequence additional taxa with their designed baits, the backbone trees in this
404 step will become more densely sampled, and confidence in contig classification should increase.

405

406 *Strong Consensus for Major Relationships within the Portullugo, and the Sister Lineage of the Cacti*

407 The primary goal of our targeted enrichment study was phylogenetic inference, and our analyses
408 provide robust support for most major relationships within the portullugo. In nearly all cases,
409 concatenation and coalescent-based inference methods are congruent and show similar levels of support.
410 Although previous analyses presented conflicting support for the branch uniting Portulacineae with
411 Molluginaceae (Arakaki et al. 2011; Soltis et al. 2011; Yang et al. 2015; Brockington et al. 2015; Thulin et
412 al. 2016), our analyses across all datasets confidently support this node, though our sampling outside of
413 the portullugo is sparse and not designed to directly address this question. Montiaceae consistently
414 appears as sister to the remaining Portulacineae, though with lower support in both ASTRAL and
415 concatenated analyses than we would have predicted. Long recognized clades, like ACPT

416(Anacampserotaceae, Cactaceae, Talinaceae, Portulacaceae) and ACP (ACPT without Talinaceae) remain
417strongly supported. More importantly, relationships among other difficult taxa are beginning to
418crystallize. Our analyses confirm the monophyly of the Didiereaceae *s.l.* (Bruyans et al, 2014) and its
419placement as sister to the ACPT clade. In addition, Halophytaceae, a monotypic subshrub endemic to the
420arid interior of Argentina, is now placed with moderate support as sister to Basellaceae in all of our
421analyses, which is a new finding.

422 One of the more complex phylogenetic problems in the Portulacineae has been identifying the
423sister lineage of Cactaceae. The cacti are among the most spectacular desert plant radiations, with ~ 1500
424species of mostly stem succulents that diversified recently, during the late Miocene-Pliocene time period
425(Arakaki et al. 2011). They are closely related to *Portulaca*, a globally widespread, herbaceous and
426succulent C₄ lineage, and the Anacampserotaceae, another unusual succulent lineage with most species
427endemic to South Africa. The relationship among these three clades has remained uncertain, despite
428numerous phylogenetic studies aimed at resolving it (Hershkovitz and Zimmer 2000; Applequist and
429Wallace 2001; Nyffeler and Eggli 2010; Ocampo and Columbus 2010; Ogburn and Edwards 2015). We
430present strong support for *Portulaca* + Anacampserotaceae together as the sister lineage of the cacti, in
431both concatenation (100% BS) and coalescent (98%) analyses.

432 In spite of this congruence, our BUCKy analyses revealed strong and significant discord among
433loci for these relationships, with roughly half of our sampled genome (mean CF 0.52) supporting
434((A,P),C) and roughly 25% supporting either (A(P,C)) or (P(A,C)) (Fig. 4; abbreviating lineages by their
435first letter). It is important to note that this discord among individual gene trees is not derived from poorly
436supported topologies of individual loci; on the contrary, posterior probabilities for the alternative
437topologies are routinely very high, mostly with 100% support (Fig 4, panels E–H). Due to the congruence
438and overall strong support for ((A,P)C) by multiple inference methods and alternative matrices, we
439tentatively accept this topology and present it as the best working hypothesis for ACP relationship.
440Nevertheless, we find the amount and strength of conflicting signal throughout the genome quite
441remarkable. The reconstruction of a single, bifurcating species tree has generally been seen as the ultimate

442goal of phylogenetics and lack of resolution is typically regarded as a problem that will be solved with the
443addition of more or better data. However, in a growing number of cases, additional data have only shown
444the problem to be more complicated, and strong conflict in genome-scale data appears to be the rule,
445rather than the exception (Sally et al. 2012; Suh et al. 2015; Pease et al. 2016; Brown and Thomson
4462016; Shen et al. 2017).

447*Genomic Conflict in Deep Time Phylogenetics*

448 Commonly proposed reasons for the existence of recalcitrant nodes in phylogeny reconstruction,
449beyond lack of phylogenetic information, include homoplasy, incomplete lineage sorting, incorrect
450homology assignment due to gene duplication and loss, protracted gene flow, and hybridization.
451Homoplasy was long the preferred explanation for lack of resolution due to conflict (as exemplified by
452long-branch attraction and the Felsenstein zone), when it was assumed that, in general, gene trees would
453be congruent with the species tree. Newer data are showing that, while some degree of incomplete lineage
454sorting (ILS) would always be expected, ILS may be a reasonable explanation for recalcitrant nodes in
455some instances (Oliver 2013; Suh et al. 2015; Hahn and Nakhleh 2015). Strongly supported incongruence
456of our various gene trees could be evidence for widespread incomplete lineage sorting at several nodes in
457our phylogeny, including the split between Anacampserotaceae, Cactaceae, and Portulacaceae, and the
458relationship of the various species of *Pereskia* to the remainder of the Cactaceae.

459 The adoption of coalescent theory to resolve ancient nodes has been quickly accepted (e.g.,
460Edwards et al. 2007; Mirarab et al. 2014), though not without some skepticism (Gatesy and Springer
4612013; Springer and Gatesy 2014; Gatesy and Springer 2014). Clearly, there is obvious value in evaluating
462gene trees independently of one another, as they may represent distinct evolutionary histories.
463Concatenation of very large matrices has also been shown to cause inflated support values, masking
464significant phylogenetic conflict in the underlying data (Salichos and Rokas 2013). Under the coalescent,
465the expected degree of deviation of gene trees from the species tree depends on effective ancestral
466population sizes and generation times (Degnan & Rosenberg 2009). In a series of simulations, Oliver

467(2013) provided some estimate of ancestral population sizes and generation times needed for the signal of
468ILS to be recovered in practise; while not impossible, our intuition is that these conditions are not often
469met in plants, at least in our study system.

470 We cannot help but consider the diffuse and significant numbers of inferred gene duplications in
471our dataset (Fig. S3), including many potential losses of paralogs in certain groups. It is true that
472inference of both paralog presence *and* absence is compromised in any genome sub-sampling approach
473(hybrid enrichment, RNA-seq, etc.), because the absence of a particular paralog could simply be because
474the paralog was not captured in the sub-sampling or, in the case of transcriptomes, expressed in the
475collected tissue. Nevertheless, the ubiquitous and phylogenetically dispersed pattern of our inferred
476duplications across the portullugo (Fig. S3) implies that, regardless of where precisely these duplications
477are located, isolated duplications are common along the vast majority of reconstructed branches, and not
478confined to occasional WGD events. Considering estimated genome-wide rates of gene duplication and
479loss in other groups (Lynch and Conery 2000; Liu et al. 2014), we wonder if the ILS signal in some of
480these deep-time phylogenetic studies may be better considered as the “incomplete sorting” of paralogs
481due to differential paralog fixation following gene duplication and subsequent speciation, rather than a
482persistent signal of incomplete sorting of alleles alone. In datasets like ours, which span deep nodes and
483typically include no measure of intraspecific sequence variation, we find it difficult to distinguish
484between these two scenarios when accounting for gene tree-species tree incongruence.

485

486*Molecular Evolution of PEP Carboxylase*

487 A secondary goal of our study was to design an enrichment scheme that would allow us to
488simultaneously build a large database of genes relevant to the evolution of C₄ and CAM photosynthesis.
489Our previous work on PEPC evolution in this lineage identified five Portulacineae-specific gene
490duplications within *ppc-1E1*, the major *ppc* paralog that is most often recruited into C₄ function across
491eudicots (These duplications all appeared to take place after the separation of the Molluginaceae; Christin
492et al. 2014; 2015). We also previously identified specific amino acid substitutions in C₄ and CAM *ppc* loci

493 consistent with changes seen in C₄ origins in grasses, suggesting that there may be shared adaptive AA
494 residues associated with both C₄ and CAM function, likely due to the enzyme's similar function in both
495 syndromes (Christin et al. 2014; 2015). Our small analysis presented here (Fig. 5) is preliminary, and only
496 meant to illustrate the feasibility of performing large-scale comparative molecular evolution studies with
497 bait sequence data by focusing on an already well known gene family as a proof of concept.

498 Our expanded baits sampling and analysis is consistent with our previous findings. We confirmed
499 the additional duplication of *ppc-1E1a* within the *Portulaca* lineage (*ppc-1E1a'*) that was associated with
500 the evolution of C₄ photosynthesis in this group, and the use of this specific paralog in C₄ function has
501 already been documented (Christin et al. 2014). Multiple residues of *ppc-1E1a'* overlap with amino acids
502 associated with C₄ photosynthesis in grasses, whereas *ppc-1E1a* possesses 0–1 of the C₄-associated AA
503 residues in all taxa examined. Strikingly, *ppc-1E1a'* sequences also exhibit substantial variation within the
504 major clades of *Portulaca*, suggesting that differing C₄ origins in *Portulaca* were associated with the
505 fixation of distinct C₄-adaptive AA residues within *ppc-1E1a'* (Christin et al. 2014).

506 Less is known about the relationship between CAM function and the molecular evolution of
507 PEPC-coding genes. We have discovered the Ser780 residue, which is ubiquitous in C₄ PEPC in multiple
508 CAM species (Christin et al. 2014); however, in orchids, CAM-expressed PEPC does not seem to require
509 Ser780 (Silvera et al. 2014). Furthermore, expression studies have found nighttime up-regulation of
510 primarily *ppc-1E1c* (with Ser780; Christin et al. 2014; Brillhaus et al. 2015) and in one case each *ppc-*
511 *1E1a* (Brillhaus et al. 2015), *ppc-1E1d* (Christin et al. 2014), and *ppc-1E1e* (Brillhaus et al. 2015), all
512 without Ser780, in multiple Portulacineae engaged in a CAM cycle. In this first broader look at amino
513 acid substitutions across the entire *ppc* gene family, we can observe a few patterns. First, many CAM
514 species appear to have accumulated multiple AA residues that have been identified as important to C₄
515 function, suggesting that there may be a shared selection pressure for both syndromes at the molecular
516 sequence level. Second, only two of the five *ppc1* copies (with the exception of the *Portulaca*-specific
517 *1E1a'*) in Portulacineae have acquired a Ser780: *ppc-1E1c* and *ppc-1E1e*. In general, *ppc-1E1c* and *ppc-*
518 *1E1e* are the paralogs that have acquired the most C₄-adaptive AA residues.

519 A peculiar case is presented by *Ceraria fruticulosa* and *Portulacaria afra* in the Didiereaceae (the
520sister group to the ACPT clade). These species demonstrate a relatively high number of C₄-associated
521amino acids in *ppc-1E1b*, *ppc-1E1c*, *ppc-1E1d*, and *ppc-1E1e*; furthermore, the specific residues that
522overlap with C₄-associated amino acids largely differ across paralogs, and the only copy in these species
523with a Ser780 is *ppc-1E1e*. Considering that a *ppc* copy with a Ser780 has, to our knowledge, never been
524found in non-C₄ or non-CAM plants, we predict that *ppc-1E1e* in these taxa is primarily used for CAM
525function. However, the three additional paralogs that also exhibit putatively adaptive AA residues may
526also contribute to CAM function. This type of scenario has never been demonstrated for C₄
527photosynthesis, though we have previously documented significant upregulation of both *ppc-1E1c* and
528-*1E1d* in *Nopalea* (CAM, Cactaceae) at night (Christin et al. 2015) and Brillhaus et al. (2015) documented
529significant upregulation of *ppc-1E1c*, -*1E1e*, and likely -*1E1a* (*Talinum triangulare*, facultative CAM,
530Talinaceae).

531 In light of the broad distribution of putative adaptive residues across the multiple copies of *ppc*-
532-*1E1* in the Portulacineae, it seems that this lineage and gene family might be an especially powerful
533system for examining the dynamics of gene duplication and subsequent sub-functionalization (e.g. Ohno
5341970). Perhaps in some Portulacineae lineages, functional specialization of particular *ppc-1E1* paralogs
535took a considerable amount of time post-duplication, with many of them co-contributing to CAM function
536for millions of years, while accumulating adaptive AA changes independently. Transcriptome profiling of
537a broader array of Portulacineae could provide critical insight here; for instance, all members of the ACPT
538clade so far investigated show strong upregulation of *ppc-1E1c* during CAM, potentially because of the
539presence of the Ser780 residue. Perhaps this mutation occurred earlier in the ACPT clade than it did in the
540*Ceraria/Portulacaria* clade, which facilitated a more rapid functional specialization of *ppc-1E1c* to CAM
541function. If the Ser780 mutation is of large effect, then the timing of its appearance may have significant
542consequences for subsequent specialization of duplicated genes.

543 In conclusion, we show that it is possible to use targeted sequence capture to sequence gene
544families across a broad taxonomic range of plants. Phylogenetic studies need not be confined to single

545copy genes that may be of limited interest outside of their phylogenetic utility; rather, our sampling can be
546expanded to include large, multi-gene families. Not only does this allow the use of a greater proportion of
547the genome in targeted sequence capture studies, it also enables exhaustive sampling and analysis of any
548gene with relevance to a very broad range of evolutionary questions. This creates exciting opportunities
549for phylogenetic biology in general, opening the potential for systematics-centered research to fully grow
550into integrative and comprehensive analyses of whole-organism evolution.

551

552ACKNOWLEDGEMENTS

553

554 The authors would like to thank P.-A. Christin, M. Howison, M. Moeglein, C. Munro, and F.
555Zapata for helpful discussion; E. Johnson for the figures; B. Dewenter, J.A.M. Holtum, E. van Jaarsveld,
556F. Obbens, D. Tribble, and R. de Vos for field assistance; B. Dewenter and C. Schorl for lab assistance;
557the 1KP project for Molluginaceae transcriptome sequences; CapeNature (0028-AAA008-00140),
558SANParks, and the Government of South Australia (E26345-1) for permission to collect. L.P.H. was
559supported in part by NSF IGERT grant DGE-0966060.

560

561FUNDING

562This work was supported by the National Science Foundation (DEB-1252901 to E.J.E.).

563

564REFERENCES

565

566Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. 1990. Basic local alignment search tool. *J.*
567*Mol. Biol.* 215:403–410.

568Álvarez I., Wendel J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular*

569 *Phylogenetics and Evolution.* 29:417–434.

570Ané C., Larget B., Baum D.A., Smith S.D., Rokas A. 2006. Bayesian Estimation of Concordance among

- 571 Gene Trees. *Mol. Biol. Evol.* 24:412–426.
- 572 Applequist W.L., Wallace R.S. 2001. Phylogeny of the portulacaceous cohort based on *ndhF* sequence
573 data. *Syst. Bot.* 206: 406-419.
- 574 Arakaki M., Christin P.A., Nyffeler R., Lendel A., Eggli U., Ogburn R.M., Spriggs E., Moore M.J.,
575 Edwards E.J. 2011. Contemporaneous and recent radiations of the world's major succulent plant
576 lineages. *Proc. Natl. Acad. Sci.* 108:8379–8384.
- 577 Baum D.A. 2007. Concordance Trees, Concordance Factors, and the Exploration of Reticulate
578 Genealogy. *Taxon.* 56:417–426.
- 579 Berger S. A., Krompass D., Stamatakis A. 2011. Performance, accuracy, and web server for
580 evolutionary placement of short sequence reads under maximum likelihood. *Syst. Biol.*
581 60:291–302.
- 582 Besnard G., Muasya A.M., Russier F., Roalson E.H., Salamin N., Christin P.-A. 2009. Phylogenomics of
583 C4 photosynthesis in sedges (Cyperaceae): multiple appearances and genetic convergence. *Mol. Biol.*
584 *Evol.* 26:1909–1919.
- 585 Brilhaus D., Bräutigam A., Mettler-Altmann T., Winter K., Weber A.P.M. 2015. Reversible Burst of
586 Transcriptional Changes during Induction of Crassulacean Acid Metabolism in *Talinum triangulare*.
587 *Plant Physiology.* 170:102–122.
- 588 Brockington S.F., Yang Y., Gandia Herrero F., Covshoff S., Hibberd J.M., Sage R.F., Wong G.K.S., Moore
589 M.J., Smith S.A. 2015. Lineage-specific gene radiations underlie the evolution of novel betalain
590 pigmentation in Caryophyllales. *New Phytol.* 207:1170–1180.
- 591 Brown J.M., Thomson R.C. 2016. Bayes factors unmask highly variable information content, bias, and
592 extreme influence in phylogenomic analyses. *Syst. Biol.*:101–14.
- 593 Bruyns P.V., Oliveira-Neto M., Melo-de-Pinna G.F., Klak C. 2014. Phylogenetic relationships in the
594 Didiereaceae with special reference to subfamily Portulacarioideae. *Taxon.* 63:1053–1064.
- 595 Chamala S., García N., Godden G.T., Krishnakumar V., Jordon-Thaden I.E., De Smet R., Barbazuk W.B.,
596 Soltis D.E., Soltis P.S. 2015. MarkerMiner 1.0: A new application for phylogenetic marker

- 597development using angiosperm transcriptomes. *Appl. Plant Sci.* 3: 1400115.
- 598Chen K., Durand D., Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and
599 optimizing gene family trees. *J. Comp. Biol.* 7:429–447.
- 600Christin P.-A., Salamin N., Savolainen V., Duvall M.R., Besnard G. 2007. C4 Photosynthesis Evolved in
601 Grasses via Parallel Adaptive Genetic Changes. *Curr. Biol.* 17:1241–1247.
- 602Christin P.-A., Arakaki M., Osborne C.P., Bräutigam A., Sage R.F., Hibberd J.M., Kelly S.,
603 Covshoff S., Wong G.K.-S., Hancock L., Edwards E.J. 2014. Shared origins of a key enzyme
604 during the evolution of C4 and CAM metabolism. *J. Exp. Bot.* 65:3609–3621.
- 605Christin P.-A., Arakaki M., Osborne C.P., Edwards E.J. 2015. Genetic enablers underlying the clustered
606 evolutionary origins of C4 photosynthesis in angiosperms. *Mol. Biol. Evol.* 32:846–858.
- 607Crawford D. J. 2010. Progenitor-derivative species pairs and plant speciation. *Taxon* 59:1413–1423.
- 608De Bodt S., Maere S., Van de Peer Y. 2005. Genome duplication and origin of Angiosperms. *Trends*
609 *Ecol. Evol.* 20:591–597.
- 610De Smet R., Adams K.L., Vandepoele K., Van Montagu M.C.E., Maere S., Van de Peer Y. 2013.
611 Convergent gene loss following gene and genome duplications creates single-copy families in
612 flowering plants. *Proc. Natl. Acad. Sci. USA.* 110:2898–2903.
- 613de Sousa F., Bertrand Y.J.K., Nylinder S., Oxelman B., Eriksson J.S., Pfeil B.E. 2014. Phylogenetic
614 Properties of 50 Nuclear Loci in Medicago (Leguminosae) Generated Using Multiplexed Sequence
615 Capture and Next-Generation Sequencing. *PLoS ONE.* 9:e109704–10.
- 616Degnan J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies
617 coalescent. *Trends in Ecol. Evol.* 24:332–340.
- 618Dohm J. C., Minoche A. E., Holtgräwe D., Capella-Gutiérrez S., Zakrzewski F., Tafer H., Rupp O.,
619 Sörensen T. R., Stracke R., Reinhardt R., Goesmann A., Kraft T., Schulz B., Stadler P. F.,
620 Schmidt T., Gabaldón T., Lehrach H., Weisshaar B., Himmelbauer H. 2014. The genome of the
621 recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature.* 505:546–549.
- 622Dunn C.W., Hejnlol A., Matus D.Q., Pang K., Browne W.E., Smith S.A., Seaver E., Rouse G.W., Obst M.,

- 623 Edgecombe G.D., Sørensen M.V., Haddock S.H.D., Schmidt-Rhaesa A., Okusu A., Kristensen R.M.,
624 Wheeler W.C., Martindale M.Q., Giribet G. 2008. Broad phylogenomic sampling improves
625 resolution of the animal tree of life. *Nature*. 452:745–749.
- 626Edgar R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
627 *Nucl. Aci. Res.* 32:1792–1797.
- 628Edwards E.J., Nyffeler R., Donoghue M.J. 2005. Basal cactus phylogeny: Implications of *Pereskia*
629 (*Cactaceae*) paraphyly for the transition to the cactus life form. *Am. J. Bot.* 92:1177–1188.
- 630Edwards E.J., Ogburn R.M. 2012. Angiosperm Responses to a Low-CO₂ World: CAM and C₄
631 Photosynthesis as Parallel Evolutionary Trajectories. *Int. J. Plant Sci.* 173:724–733.
- 632Edwards S.V., Liu L., Pearl D.K. 2007. High-resolution species trees without concatenation. *Proc. Natl.*
633 *Acad. Sci. USA.* 104:5936–5941.
- 634Faircloth B.C., McCormack J.E., Crawford N.G., Harvey M.G., Brumfield R.T., Glenn T.C. 2012.
635 Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary
636 timescales. *Syst. Biol.* 61:717–726.
- 637Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-
638 genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* 60:433–453.
- 639Gatesy J., Springer M.S. 2013. Concatenation versus coalescence versus “concatalescence.” *Proc. Natl.*
640 *Acad. Sci. USA.* 110:1179–1179.
- 641Gatesy J., Springer M.S. 2014. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed
642 hidden support, and the coalescence/concatalescence conundrum. *Mol. Biol. Evol.* 80:231–266.
- 643Hahn M.W., Nakhleh L. 2015. Irrational exuberance for resolved species trees. *Evolution.* 70:7–17.
- 644Hershkovitz M.A., Zimmer E.A. 2000. Ribosomal DNA evidence and disjunctions of western American
645 *Portulacaceae*. *Mol. Phyl. Evol.* 15:419–439.
- 646Heyduk K., McKain M.R., Lalani F., Leebens-Mack J. 2016. Evolution of a CAM anatomy predates the
647 origins of Crassulacean acid metabolism in the Agavoideae (*Asparagaceae*). *Mol. Phyl. Evol.*
648 105:102–113.

- 649Hudspeth, R. L., & Guala, J. W. 1989. Structure and expression of the maize gene encoding the
650 phosphoenolpyruvate carboxylase isozyme involved in C4 photosynthesis. *Plant Mol. Biol.* 12:
651 579-589.
- 652Jiao Y., Wickett N.J., Ayyampalayam S., Chanderbali A.S., Landherr L., Ralph P.E., Tomsho L.P., Hu
653 Y., Liang H., Soltis P.S., Soltis D.E., Clifton S.W., Schlarbaum S.E., Schuster S.C., Ma H.,
654 Leebens-Mack J., dePamphilis C.W. 2011. Ancestral polyploidy in seed plants and angiosperms.
655 *Nature.* 473:97–100.
- 656Katoh K., Standley D. M. 2013. MAFFT multiple sequence alignment software version 7:
657 improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780.
- 658Kersey P.J., Allen J.E., Armean I., Boddu S., Bolt B.J., Carvalho-Silva D., Christensen M., Davis P.,
659 Falin L.J., Grabmueller C., Humphrey J., Kerhornou A., Khobova J., Aranganathan N.K.,
660 Langridge N., Lowy E., McDowall M.D., Maheswari U., Nuhn M., Ong C.K., Overduin B.,
661 Paulini M., Pedro H., Perry E., Spudich G., Tapanari E., Walts B., Williams G., Tello Ruiz M.,
662 Stein J., Wei S., Ware D., Bolser D.M., Howe K.L., Kulesha E., Lawson D., Maslen G., Staines
663 D.M. 2016. Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res.*
664 44:D574–80.
- 665Kircher M., Sawyer S., Meyer, M. 2012. Double indexing overcomes inaccuracies in multiplex
666 sequencing on the Illumina platform. *Nucl. Aci. Res.* 40:e3.
- 667Larget B.R., Kotha S.K., Dewey C.N., Ane C. 2010. BUCKy: Gene tree/species tree reconciliation with
668 Bayesian concordance analysis. *Bioinformatics.* 26:2910–2911.
- 669Leaché A.D., Rannala B. 2011. The accuracy of species tree estimation under simulation: a
670 comparison of methods. *Syst. Biol.* 60:126–137.
- 671Lemmon A.R., Emme S.A., Lemmon E.M. 2012. Anchored hybrid enrichment for massively high-
672 throughput phylogenomics. *Syst. Biol.* 61:727–744.

673Li C., Hofreiter M., Straube N., Corrigan S., Naylor G.J.P. 2013. Capturing protein-coding genes across
674 highly divergent species. *BioTechniques* 54:321–326.

675Liu S., Liu Y., Yang X., Tong C., Edwards D., Parkin I.A.P., Zhao M., Ma J., Yu J., Huang S., Wang X.,
676 Wang J., Lu K., Fang Z., Bancroft I., Yang T.-J., Hu Q., Wang X., Yue Z., Li H., Yang L., Wu J., Zhou
677 Q., Wang W., King G.J., Pires J.C., Lu C., Wu Z., Sampath P., Wang Z., Guo H., Pan S., Yang L., Min
678 J., Zhang D., Jin D., Li W., Belcram H., Tu J., Guan M., Qi C., Du D., Li J., Jiang L., Batley J.,
679 Sharpe A.G., Park B.-S., Ruperao P., Cheng F., Waminal N.E., Huang Y., Dong C., Wang L., Li J., Hu
680 Z., Zhuang M., Huang Y., Huang J., Shi J., Mei D., Liu J., Lee T.-H., Wang J., Jin H., Li Z., Li X.,
681 Zhang J., Xiao L., Zhou Y., Liu Z., Liu X., Qin R., Tang X., Liu W., Wang Y., Zhang Y., Lee J., Kim
682 H.H., Denoeud F., Xu X., Liang X., Hua W., Wang X., Wang J., Chalhoub B., Paterson A.H. 1AD.
683 The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. *Nat*
684 *Commun.* 5:1–11.

685Lynch M., Conery J.S. 2000. The evolutionary fate and consequences of duplicate genes. *Science*.
686 290:1151–1155.

687Mandel J.R., Dikow R.B., Funk V.A. 2015. Using phylogenomics to resolve mega-families: An example
688 from Compositae. *J. Syst. Evol.* 53:391–402.

689Martin A.P., Burg T.M. 2002. Perils of Paralogy: Using HSP70 Genes for Inferring Organismal
690 Phylogenies. *Syst. Biol.* 51:570–587.

691Matasci N., Hung L.-H., Yan Z., Carpenter E.J., Wickett N.J., Mirarab S., Nguyen N., Warnow T.,
692 Ayyampalayam S., Barker M., Burleigh J.G., Gitzendanner M.A., Wafula E., Der J.P., dePamphilis
693 C.W., Roure B., Philippe H., Ruhfel B.R., Miles N.W., Graham S.W., Mathews S., Surek B.,
694 Melkonian M., Soltis D.E., Soltis P.S., Rothfels C., Pokorny L., Shaw J.A., DeGironimo L.,
695 Stevenson D.W., Villarreal J.C., Chen T., Kutchan T.M., Rolf M., Baucom R.S., Deyholos M.K.,
696 Samudrala R., Tian Z., Wu X., Sun X., Zhang Y., Wang J., Leebens-Mack J., Wong G.K.-S. 2014.
697 Data access for the 1,000 Plants (1KP) project. *GigaSci.* 3:17–10.

698McCormack J.E., Hird S.M., Zellmer A.J., Carstens B.C., Brumfield R.T. 2013. Applications of next-

- 699 generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and*
700 *Evolution*. 66:526–538.
- 701 Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: genome-
702 scale coalescent-based species tree estimation. *Bioinformatics*. 30:i541–8.
- 703 Nicholls J.A., Pennington R.T., Koenen E.J.M., Hughes C.E., Hearn J., Bunnefeld L., Dexter K.G., Stone
704 G.N., Kidner C.A. 2015. Using targeted enrichment of nuclear genes to increase phylogenetic
705 resolution in the neotropical rain forest genus *Inga* (Leguminosae: Mimosoideae). *Front. Plant Sci.*
706 6:1382–1420.
- 707 Novikova P. Y., Hohmann N., Nizhynska V., Tsuchimatsu T., Ali J., Muir G., Guggisberg A., Paape
708 T., Schmid K., Fedorenko O. M., Holm S., Säll T., Schlötterer C., Marhold K., Widmer A.,
709 Sese J., Shimizu K. K., Weigel D., Krämer U., Koch M. A., Nordborg M. 2016. Sequencing of
710 the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant
711 trans-specific polymorphism. *Nature Genetics* 48:1077–1082.
- 712 Nyffeler R., Egli U. 2010. Disintegrating Portulacaceae: A new familial classification of the suborder
713 Portulacineae (Caryophyllales) based on molecular and morphological data. *Taxon*. 59:227–240.
- 714 Ocampo G., Columbus J.T. 2010. Molecular phylogenetics of suborder Cactineae (Caryophyllales),
715 including insights into photosynthetic diversification and historical biogeography. *Am. J. Bot.*
716 97:1827–1847.
- 717 Ogburn M.R., Edwards E.J. 2015. Life history lability underlies rapid climate niche evolution in the
718 angiosperm clade Montiaceae. *Mol. Phyl. Evol.* 92:181–192.
- 719 Ohno S. 1970. *Evolution by gene duplication*. London: George Allen & Unwin Ltd. Berlin, Heidelberg
720 and New York: Springer-Verlag
- 721 Oliver J.C. 2013. Microevolutionary processes generate phylogenomic discordance at ancient
722 divergences. *Evolution*. 67:1823–1830.
- 723 Paulus J.K., Schlieper D., Groth G. 2013. Greater efficiency of photosynthetic carbon fixation due to
724 single amino-acid substitution. *Nat. Commun.* 4:1518.

- 725 Pease J.B., Haak D.C., Hahn M.W., Moyle L.C. 2016. Phylogenomics Reveals Three Sources of Adaptive
726 Variation during a Rapid Radiation. *PLoS Biol.* 14:e1002379–24.
- 727 R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for
728 Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- 729 Renny-Byfield S., Wendel J.F. 2014. Doubling down on genomes: Polyploidy and crop plants. *Am. J. Bot.*
730 101:1711–1725.
- 731 Revell L.J. 2012. phytools: an R package for phylogenetic comparative biology (and other things).
732 *Methods Ecol. Evol.* 3:217–223.
- 733 Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard
734 M.A., Huelsenbeck J.P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model
735 choice across a large model space. *Syst. Biol.* 61:539–542.
- 736 Salichos L., Rokas A. 2013. Inferring ancient divergences requires genes with strong phylogenetic
737 signals. *Nature.* 497:327–331.
- 738 Scally A., Dutheil J.Y., Hillier L.W., Jordan G.E., Goodhead I., Herrero J., Hobolth A., Lappalainen T.,
739 Mailund T., Marques-Bonet T., McCarthy S., Montgomery S.H., Schwalie P.C., Tang Y.A., Ward
740 M.C., Xue Y., Yngvadottir B., Alkan C., Andersen L.N., Ayub Q., Ball E.V., Beal K., Bradley B.J.,
741 Chen Y., Clee C.M., Fitzgerald S., Graves T.A., Gu Y., Heath P., Heger A., Karakoc E., Kolb-
742 Kokocinski A., Laird G.K., Lunter G., Meader S., Mort M., Mullikin J.C., Munch K., O'Connor T.D.,
743 Phillips A.D., Prado-Martinez J., Rogers A.S., Sajjadian S., Schmidt D., Shaw K., Simpson J.T.,
744 Stenson P.D., Turner D.J., Vigilant L., Vilella A.J., Whitener W., Zhu B., Cooper D.N., de Jong P.,
745 Dermitzakis E.T., Eichler E.E., Flicek P., Goldman N., Mundy N.I., Ning Z., Odom D.T., Ponting
746 C.P., Quail M.A., Ryder O.A., Searle S.M., Warren W.C., Wilson R.K., Schierup M.H., Rogers J.,
747 Tyler-Smith C., Durbin R. 2012. Insights into hominid evolution from the gorilla genome sequence.
748 *Nature.* 483:169–175.
- 749 Schmickl R., Liston A., Zeisek V., Oberlander K., Weitemier K., Straub S.C.K., Cronn R.C., Dreyer L.L.,
750 Suda J. 2016. Phylogenetic marker development for target enrichment from transcriptome and

- 751 genome skim data: the pipeline and its application in southern African Oxalis (Oxalidaceae). *Mol.*
752 *Ecol. Resour.* 16:1124–1135.
- 753 Shen X.-X., Hittinger C.T., Rokas A. Contentious relationships in phylogenomic studies can be driven by
754 a handful of genes. *Nature Ecology & Evolution.* 1:0126 EP –.
- 755 Silvera K., Winter K., Rodriguez B.L., Albion R.L., Cushman J.C. 2014. Multiple isoforms of
756 phosphoenolpyruvate carboxylase in the Orchidaceae (subtribe Oncidiinae): implications for the
757 evolution of crassulacean acid metabolism. *J. Exp. Bot.* 65:3623–3636.
- 758 Soltis P.S., Marchant D.B., Van de Peer Y., Soltis D.E. 2015. Polyploidy and genome evolution in plants.
759 *Curr. Opin. Genet. Dev.* 35:119–125.
- 760 Springer M.S., Gatesy J. 2014. Land plant origins and coalescence confusion. *Trends Plant Sci.* 19:267–
761 269.
- 762 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
763 phylogenies. *Bioinformatics* 30:1312–1313.
- 764 Stolzer M., Lai H., Xu M., Sathaye D., Vernot B., Durand D. 2012. Inferring duplications, losses,
765 transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics*
766 28:409–415.
- 767 Suh A., Smeds L., Ellegren H. 2015. The Dynamics of Incomplete Lineage Sorting across the Ancient
768 Adaptive Radiation of Neoavian Birds. *PLoS Biol.* 13:e1002224.
- 769 Thulin M., Moore A.J., El-Seedi H., Larsson A., Christin P.-A., Edwards E.J. 2016. Phylogeny and
770 generic delimitation in Molluginaceae, new pigment data in Caryophyllales, and the new family
771 Corbichoniaceae. *Taxon.* 65:775–793.
- 772 Wickett N.J., Mirarab S., Nguyen N., Warnow T., Carpenter E., Matasci N., Ayyampalayam S., Barker
773 M.S., Burleigh J.G., Gitzendanner M.A., Ruhfel B.R., Wafula E., Der J.P., Graham S.W., Mathews
774 S., Melkonian M., Soltis D.E., Soltis P.S., Miles N.W., Rothfels C.J., Pokorny L., Shaw A.J.,
775 DeGironimo L., Stevenson D.W., Surek B., Villarreal J.C., Roure B., Philippe H., dePamphilis C.W.,
776 Chen T., Deyholos M.K., Baucom R.S., Kutchan T.M., Augustin M.M., Wang J., Zhang Y., Tian Z.,

777 Yan Z., Wu X., Sun X., Wong G.K.-S., Leebens-Mack J. 2014. Phylotranscriptomic analysis of the
778 origin and early diversification of land plants. *Proc. Natl. Acad. Sci. USA* 111:E4859–E4868.
779 Yang Y., Moore M.J., Brockington S.F., Soltis D.E., Wong G.K.-S., Carpenter E.J., Zhang Y., Chen L.,
780 Yan Z., Xie Y., Sage R.F., Covshoff S., Hibberd J.M., Nelson M.N., Smith S.A. 2015. Dissecting
781 molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing.
782 *Mol. Biol. Evol.* 32: 2001-2014
783

784FIGURE LEGENDS

785

786Figure 1

787 Species representatives across Portulacineae families. A) *Portulaca aff. filsonii* B)

788*Calandrinia hortiorum* C) *Anacampseros papyracea* D) *Halophytum ameghinoi* E) *Anredera*

789*diffusa*, F) *Pereskia portulacifolia*, G) *Neoraimondia arequipensis*.

790

791Figure 2

792 Assembly pipeline schematic. Part I extracts all relevant reads for each gene family and then

793assemble them into contigs. Part II constructs longer sequences from contigs and assigns them to

794particular paralogs within a gene family. Part III identifies gene duplications within gene families, isolates

795phylogenetically useful sets of orthologs, and uses them for phylogenetic analysis.

796

797Figure 3

798 Astral topology from the g5 locus sampling (including only the loci that are present in

799five or more groups). ASTRAL bootstrap values are above the branches, while RAxML bootstrap

800values are below the branches. Star indicates greater than 95% bootstrap support.

801

802Figure 4

803 Genomic conflict for phylogenetic relationships between Cactaceae and its putative sister groups,

804Portulacaceae and Anacampserotaceae. (A) Heatmap of the “calculate-pairs” analysis in BUCKy (i.e.,

805based on a posterior distribution of trees randomly pruned to one exemplar for each of Cactaceae,

806Portulacaceae, Anacampserotaceae, and Talinaceae), indicating the posterior probability (pp) that a pair of

807loci support the same topology (red: pp=1; white: pp=0). Thus, each row and column represents a locus;

808locus names and numbers are given to the left, posterior probability from the MrBayes analysis of

809 individual loci (i.e., based on unpruned trees) for three alternative sister group relations are given to the
810 right (light grey: $pp=0$; black: $pp=1$), and a dendrogram based on Euclidean distance between pp values is
811 drawn above. (B-D) Topologies for the three putative sister relations, with genome-wide concordance
812 factor (bold) and 95% credibility interval (in brackets) indicated. (E-H) Histograms of posterior
813 probability from MrBayes analyses of individual loci for four putative clades (E: Anacampserotaceae +
814 Cactaceae + Portulacaceae + Talinaceae; F: Portulacaceae + Anacampserotaceae; G: Cactaceae +
815 Anacampserotaceae; H: Cactaceae + Portulacaceae). Contribution to total frequency by loci in which
816 Anacampserotaceae + Cactaceae + Portulacaceae + Talinaceae is not supported ($pp=0$) is drawn in white,
817 contribution by other loci is drawn in gray.

818 [Note that strongly supported conflict across genes, rather than lack of information in individual genes, is
819 indicated through presence of multiple major red blocks in panel A, considerable concordance factors for
820 conflicting resolutions in panels B-D; and bimodal distributions in histograms of panels E-H.]

821

822 *Figure 5*

823 Molecular evolution of *ppc-1E1* in portullugo. Phylogenetic tree of *ppc-1E1* obtained via
824 RAxML. Where possible, named lineages are compressed. Amino acids residues for 14 positions shown
825 to be under positive selection in C_4 grasses, as well as position 890, which is associated with malate
826 sensitivity, are shown (numbering corresponds to *Zea mays* CAA33317). For compressed lineages, the
827 most frequent amino acid is shown. Amino acids are color-coded based on chemical properties. For
828 specific residues that could not be identified due to missing data or ambiguity, amino acids with highest
829 marginal probabilities are shown, and corresponding color codes are partially transparent
830 (*rerootingMethod* function in phytools, Revell 2012). Amino acids specifically associated with C_4 in
831 grasses are in boldface, and the number of boldface amino acids for a lineage are indicated with red, blue,
832 gray, and purple horizontal bars. Red bars indicate a C_4 lineage, blue bars indicate a lineage with CAM
833 activity, light blue bars indicate suspected CAM activity, purple indicates both C_4 and CAM, and gray
834 indicates a C_3 lineage. For *Portulaca*, a C_4 and CAM lineage, *ppc-1E1a'* bars are coded red because of the

835 known association with C₄ photosynthesis, and *ppc-1E1c* bars are coded blue because of the documented
836 association with CAM activity. Asterisks indicate lineages with drought-induced night time up-regulation
837 of transcript copy number, suggesting relevance to CAM activity.

838

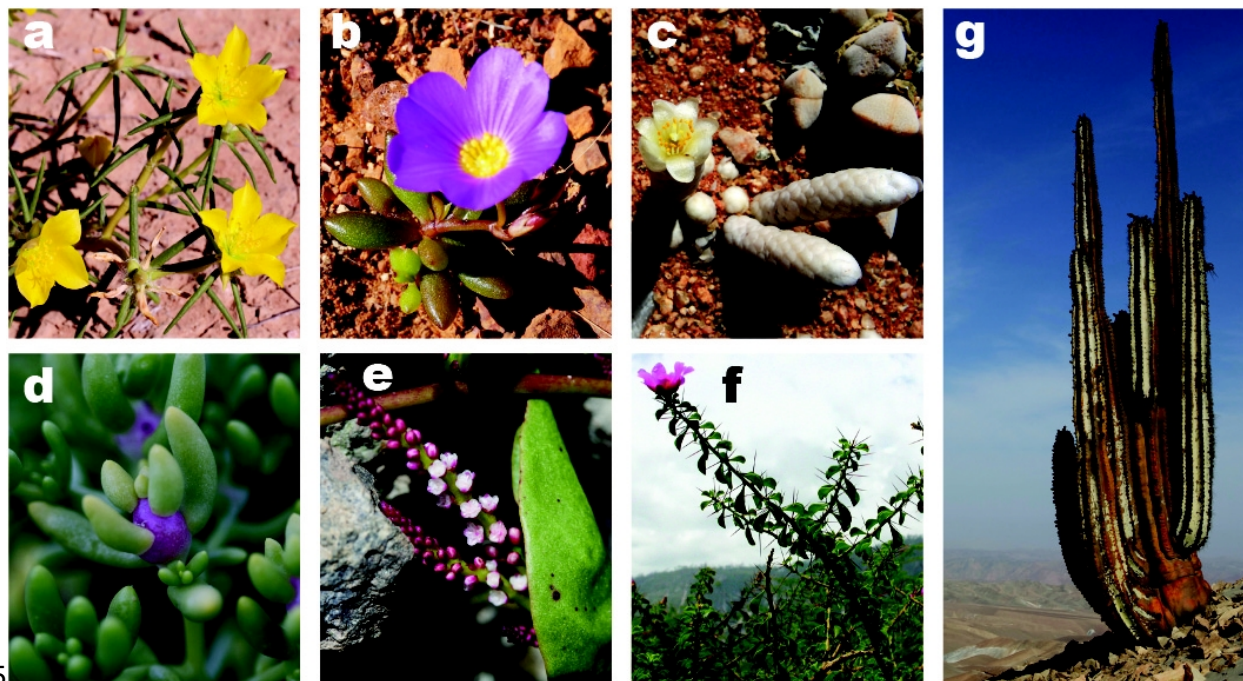
839 SUPPLEMENTARY MATERIAL

840

841 Tree files, concatenated alignments, and separate alignments for each locus are available from the Dryad
842 Digital Repository: [http://dx.doi.org/10.5061/dryad.\[NNNN\]](http://dx.doi.org/10.5061/dryad.[NNNN]) (supplementary tables)

843

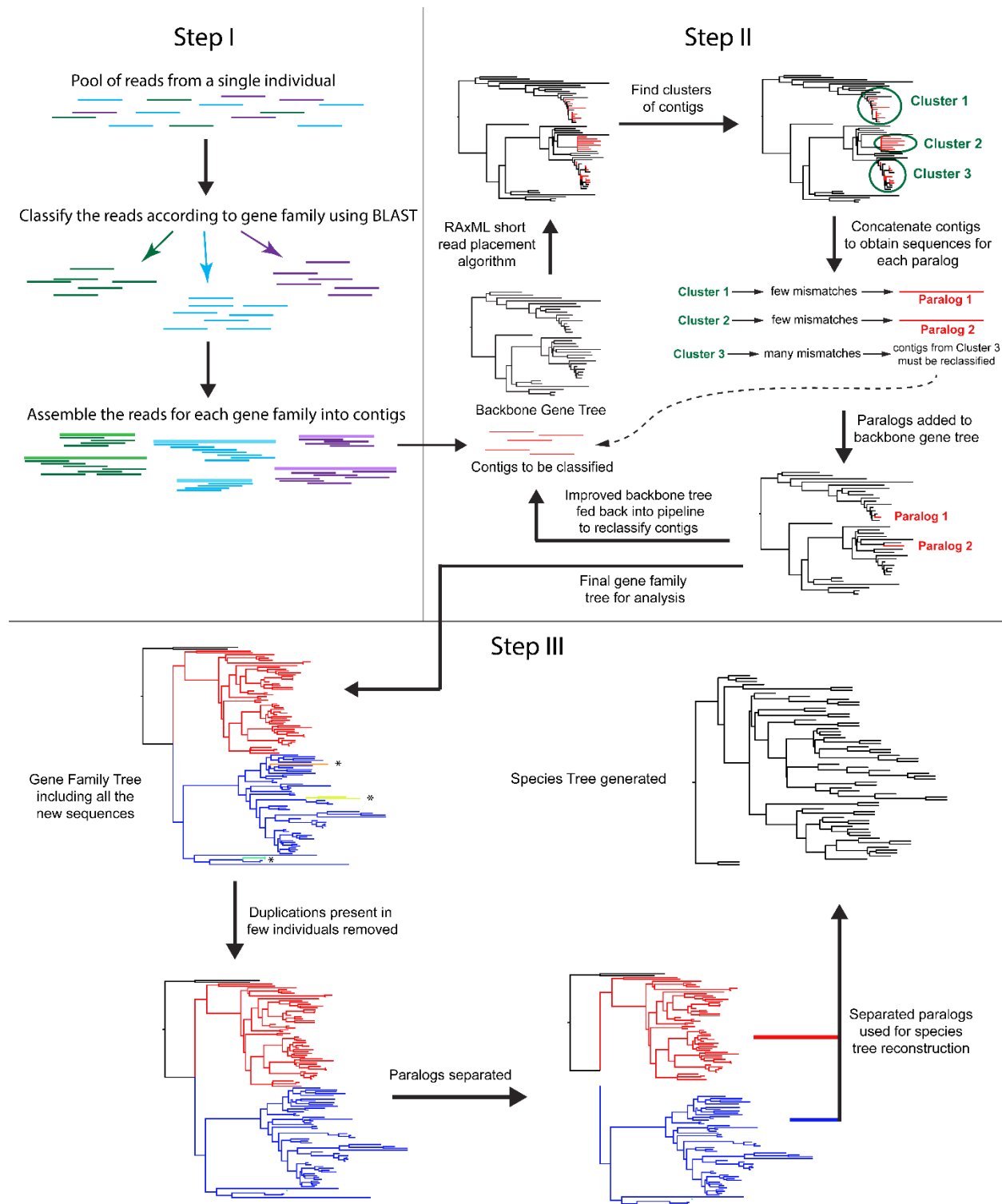
844Figure 1.



845

846

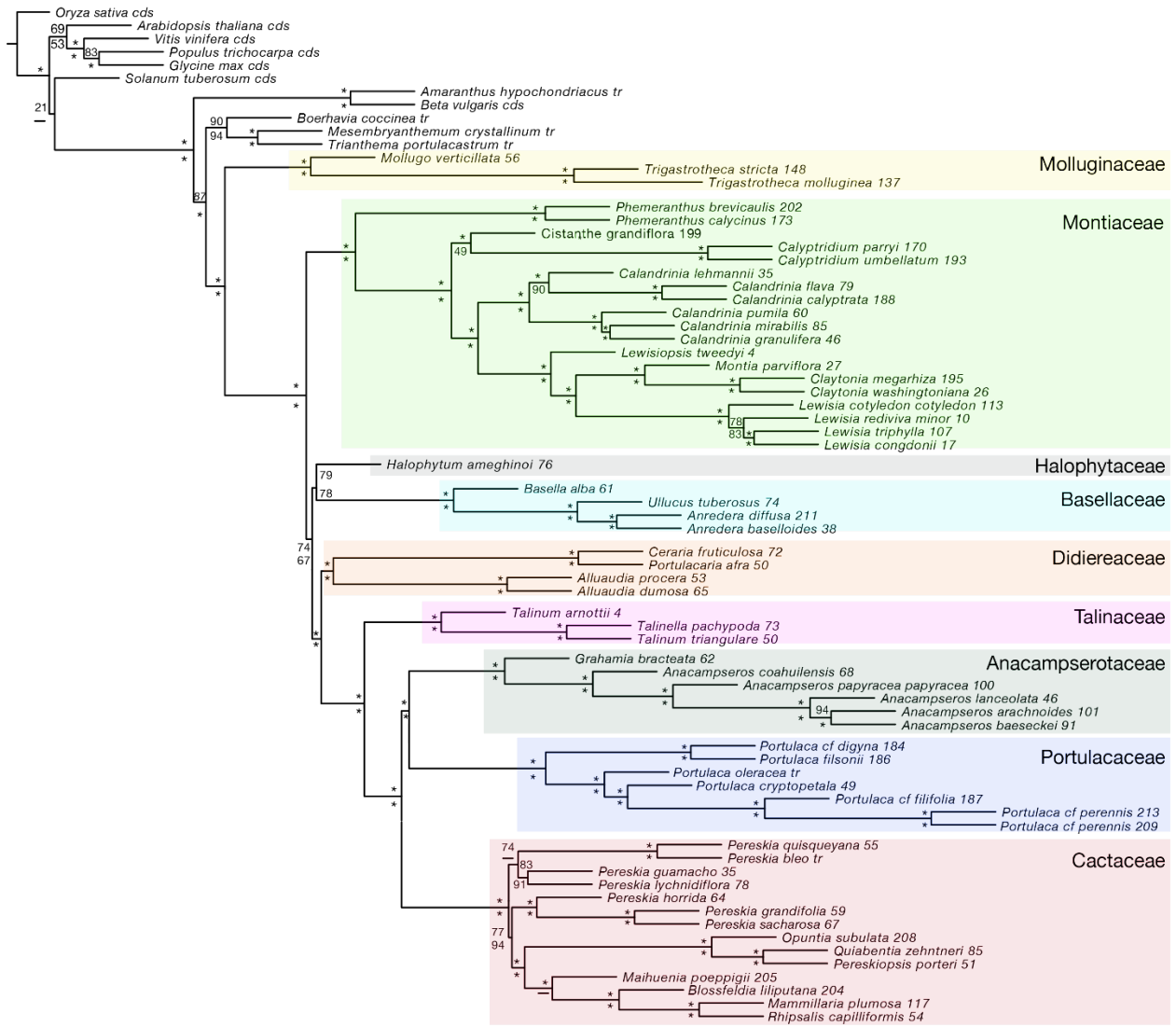
847Figure 2.



848

849

850Figure 3.

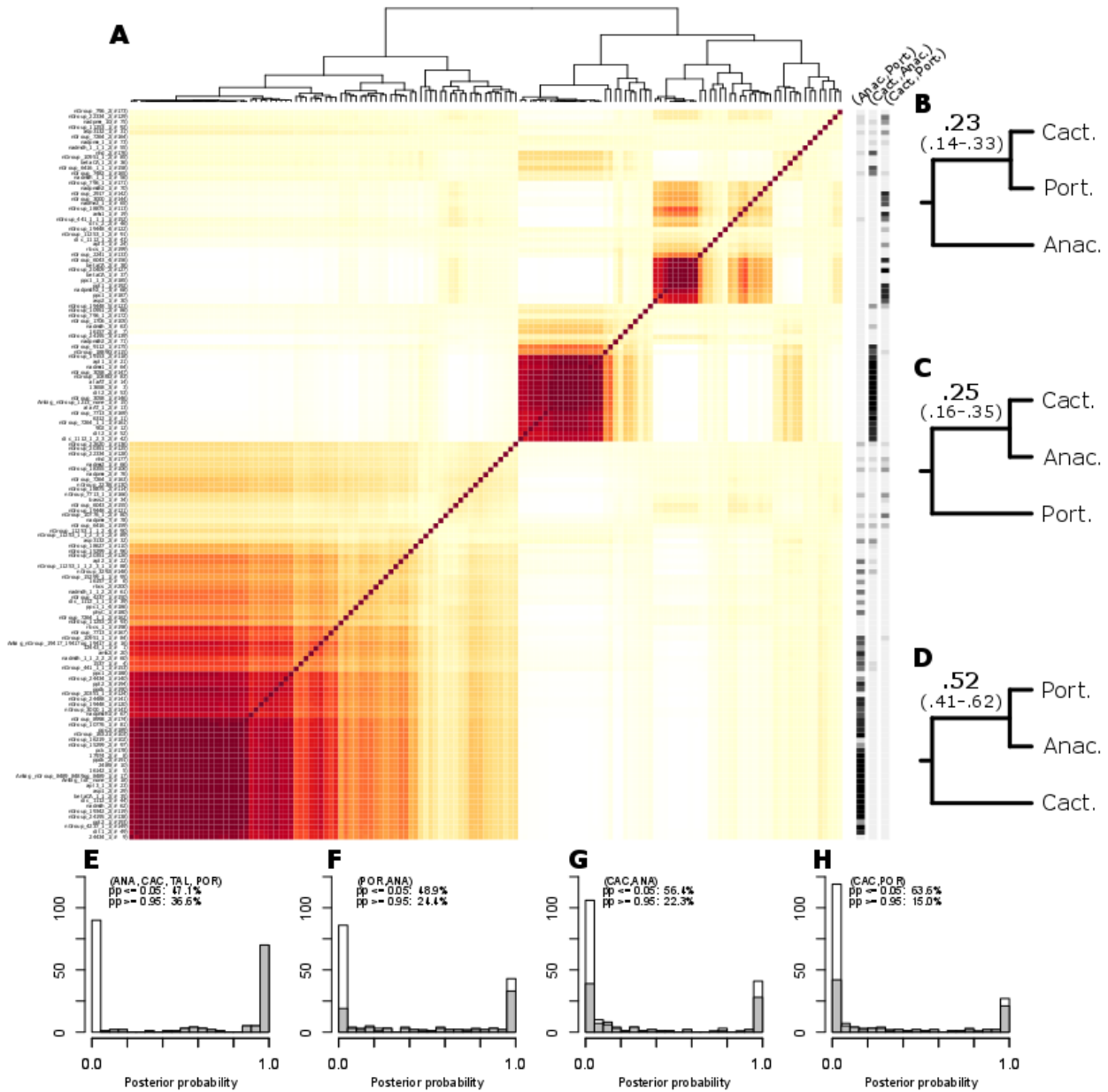


851

852

2.0

853Figure 4.



854

855

