



1-1-2019

Evolution of Portulacineae Marked by Gene Tree Conflict and Gene Family Expansion Associated with Adaptation to Harsh Environments

Ning Wang

University of Michigan, Ann Arbor

Ya Yang

University of Minnesota Twin Cities

Michael J. Moore

Oberlin College

Samuel F. Brockington

University of Cambridge

Joseph F. Walker

University of Michigan, Ann Arbor

Follow this and additional works at: <https://scholarlycommons.pacific.edu/soecs-facarticles>

See next page for additional authors



Part of the [Computer Sciences Commons](#)

Recommended Citation

Wang, N., Yang, Y., Moore, M. J., Brockington, S. F., Walker, J. F., Brown, J. W., Liang, B., Feng, T., Edwards, C., Mikenas, J., Olivieri, J. E., Hutchison, V., Timoneda, A., Stoughton, T., Puente, R., Majure, L. C., Eggli, U., & Smith, S. A. (2019). Evolution of Portulacineae Marked by Gene Tree Conflict and Gene Family Expansion Associated with Adaptation to Harsh Environments. *Molecular Biology and Evolution*, 36(1), 112–126.

DOI: [10.1093/molbev/msy200](https://doi.org/10.1093/molbev/msy200)

<https://scholarlycommons.pacific.edu/soecs-facarticles/303>

This Article is brought to you for free and open access by the All Faculty Scholarship at Scholarly Commons. It has been accepted for inclusion in All Faculty Articles - School of Engineering and Computer Science by an authorized administrator of Scholarly Commons. For more information, please contact mjibney@pacific.edu.

Authors

Ning Wang, Ya Yang, Michael J. Moore, Samuel F. Brockington, Joseph F. Walker, Joseph W. Brown, Bin Liang, Tao Feng, Caroline Edwards, Jessica Mikenas, Julia E. Olivieri, Vera Hutchison, Alfonso Timoneda, Tommy Stoughton, Raúl Puente, Lucas C. Majure, Urs Eggli, and Stephen A. Smith

Evolution of Portulacineae Marked by Gene Tree Conflict and Gene Family Expansion Associated with Adaptation to Harsh Environments

Ning Wang,^{*,1} Ya Yang,² Michael J. Moore,³ Samuel F. Brockington,⁴ Joseph F. Walker,¹ Joseph W. Brown,⁵ Bin Liang,¹ Tao Feng,⁴ Caroline Edwards,³ Jessica Mikenas,³ Julia Olivieri,³ Vera Hutchison,³ Alfonso Timoneda,⁴ Tommy Stoughton,⁶ Raúl Puente,⁷ Lucas C. Majure,^{7,8} Urs Eggli,⁹ and Stephen A. Smith^{*,1}

¹Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI

²Department of Plant and Microbial Biology, University of Minnesota-Twin Cities, St. Paul, MN

³Department of Biology, Oberlin College, Science Center K111, Oberlin, OH

⁴Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom

⁵Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, United Kingdom

⁶Center for the Environment, MSC 63, Plymouth State University, Plymouth, NH

⁷Department of Research, Conservation and Collections, Desert Botanical Garden, Phoenix, AZ

⁸Florida Museum of Natural History, University of Florida, Gainesville, FL

⁹Sukkulanten-Sammlung Zürich, Zürich, Switzerland

***Corresponding authors:** E-mails: ningwum@umich.edu; eebsmith@umich.edu.

Associate editor: Stephen Wright

Abstract

Several plant lineages have evolved adaptations that allow survival in extreme and harsh environments including many families within the plant clade Portulacineae (Caryophyllales) such as the Cactaceae, Didiereaceae, and Montiaceae. Here, using newly generated transcriptomic data, we reconstructed the phylogeny of Portulacineae and examined potential correlates between molecular evolution and adaptation to harsh environments. Our phylogenetic results were largely congruent with previous analyses, but we identified several early diverging nodes characterized by extensive gene tree conflict. For particularly contentious nodes, we present detailed information about the phylogenetic signal for alternative relationships. We also analyzed the frequency of gene duplications, confirmed previously identified whole genome duplications (WGD), and proposed a previously unidentified WGD event within the Didiereaceae. We found that the WGD events were typically associated with shifts in climatic niche but did not find a direct association with WGDs and diversification rate shifts. Diversification shifts occurred within the Portulacaceae, Cactaceae, and Anacampserotaceae, and whereas these did not experience WGDs, the Cactaceae experienced extensive gene duplications. We examined gene family expansion and molecular evolutionary patterns with a focus on genes associated with environmental stress responses and found evidence for significant gene family expansion in genes with stress adaptation and clades found in extreme environments. These results provide important directions for further and deeper examination of the potential links between molecular evolutionary patterns and adaptation to harsh environments.

Key words: gene tree conflict, gene duplication, molecular evolution, paleopolyploidy, Portulacineae, stress adaptation.

Introduction

Temperature and water availability are two major ecological determinants of plant distribution and survival (e.g., Peel et al. 2007). Plants living in harsh environments characterized by low precipitation, extreme temperature ranges, intense sun, and/or dry winds, often exhibit specialized morphological and physiological adaptations to abiotic stresses (Gibson 1996; Kreps et al. 2002). While the morphological adaptations may be easier to identify, long-term selective pressures are expected to leave genetic and genomic signatures in lineages adapted to extreme environments. Researchers have

addressed these questions in several ways, including exploring the gene regions and functions associated with adaptation in certain species or small groups (e.g., Christin et al. 2007).

Over the last decade, genomic data have increased dramatically and provide new ways to address macroevolutionary questions. For example, transcriptomic and genomic data sets provide the necessary means to more accurately identify whole genome duplication events (WGD, e.g., Jiao et al. 2011; Yang et al. 2015; Parks et al. 2018). WGDs have been identified throughout plants and have previously been suggested to be associated with adaptations to extreme environments (e.g., Stebbins 1971; Soltis and Soltis 2000; Brochmann et al. 2004),

speciation/diversification (Stebbins 1971; Wood et al. 2009), and success at colonizing new regions (Soltis and Soltis 2000). Researchers continue to identify more WGDs as taxon sampling increases (Yang et al. 2018). Broadly sampled genomic data also facilitate analyses of lineage specific molecular evolution such as gene family expansion and differential selection among genes and taxa. Finally, genomic data also allow for more thorough and accurate phylogenetic reconstruction. In particular, recent studies have illustrated that phylogenetic conflict is common, but the impact of this conflict on phylogenetic reconstruction varies across the tree of life due to different macroevolutionary processes (e.g., Shen et al. 2017; Walker et al. 2017).

The Portulacineae comprise ~2,200 species in nine families and 160 genera (Nyffeler and Eggli 2010a; Angiosperm Phylogeny Group 2016) that are inferred to have diverged from its sister group, the Molluginaceae, ~54 Ma (Arakaki et al. 2011). Species in this clade are incredibly morphologically diverse, ranging from annual herbs to long-lived stem succulents to trees, and include several iconic plants such as the cacti and Malagasy Didiereaceae (supplementary fig. S1, Supplementary Material online). They also exhibit variable habitat preferences ranging from rainforests to deserts (Smith et al. 2018a). Whereas some species are distributed worldwide, the majority is restricted to seasonally dry areas of the northern and southern hemisphere, under either hot arid or cold arid conditions (Hernández-Ledesma et al. 2015). Many specialized traits such as fleshy or succulent stems, leaves, and/or underground perennating structures, have arisen in association with the adaptation of this clade to xeric, alpine and arctic environments. Portulacineae also include several transitions to Crassulacean acid metabolism (CAM) and C_4 photosynthesis that improve photosynthetic efficiency and hence minimize water loss in hot and dry environments compared with C_3 photosynthesis (Edwards and Walker 1983; Borland et al. 2009; Edwards and Ogburn 2012).

Three clades within the Portulacineae—the Cactaceae, Didiereaceae, and Montiaceae—have particularly notable adaptations to harsh environments. The Cactaceae encompass roughly 80% of species in Portulacineae (~1,850 species; Nyffeler and Eggli 2010b) and represent perhaps the most spectacular radiation of succulent plants (Mauseth 2006; Arakaki et al. 2011). In addition to succulent structures that enable water storage, Cactaceae may also exhibit spines that reduce air flow on surface area for evapotranspiration, and deter herbivores (Anderson 2001). They could also have extensive, but relatively shallow, root systems that facilitate quick absorption of rainfall (Gibson and Nobel 1990). The Didiereaceae that live in the hot and seasonally dry environments of Madagascar have thorns and succulent leaves. Unlike the Cactaceae and Didiereaceae that are typically found in tropical and subtropical areas, the Montiaceae have a more cosmopolitan distribution with most species occurring in colder environments. Some species of Montiaceae, especially in the genera *Claytonia* and *Montia*, are adapted to high alpine zones and/or the high Arctic (Ogburn and Edwards 2015; Stoughton et al. 2017). Repeated evolution into harsh environments suggests the

possibility that the clade as a whole may be predisposed to adaptation to extreme environments.

Recent phylogenetic work has helped to resolve many relationships within the Portulacineae, but the earliest diverging relationships are only moderately or poorly supported (Arakaki et al. 2011; Ogburn and Edwards 2015; Moore et al. 2018; Walker et al. 2018a; Yang et al. 2018). This may reflect a rapid radiation, uncertainty due to gene tree conflict or WGD, or other macroevolutionary processes. Recent studies have inferred several WGD events (i.e., paleopolyploidy) within the Portulacineae, including one at the base of the clade (Yang et al. 2018). Increased transcriptomic sampling may allow for the detection of additional WGD events.

In addition to phylogenetic analyses, genomic data can provide preliminary, but essential, information on the genetic basis of adaptation to abiotic stresses through gene duplication and other evolutionary processes. For example, genes associated with response to drought, cold, and hot conditions through the abscisic acid (ABA) signaling pathway (e.g., ABRE-binding protein/ABRE-binding factors, the dehydration-responsive element binding [DREB] factors, and the NAC transcription factors, e.g., Qin et al. 2004; Nakashima et al. 2012) provide good candidates for detecting common signatures in plant lineages, such as Portulacineae. Of course, plant stress responses can be complicated by multiple stresses (Nakashima et al. 2014). Also, how genes mediate stress responses in most plant taxa are still poorly known. Nevertheless, molecular analyses of potential stress related genes provide essential information for future targeted studies. Because the Portulacineae include many species that occupy high stress environments, they are excellent clades with which to examine the genetic basis of adaptation to abiotic stresses. Using transcriptomic data, we can identify lineage-specific gene family expansions and/or positive selection among taxa that occupy these environments and gain new insights into the evolution of stress responses that may be broadly applicable to plants.

Here, we analyzed 82 transcriptomes, 47 of which were newly generated, to better understand the evolutionary history of the Portulacineae. By closely examining patterns of gene duplication and paleopolyploidy, gene-tree/species-tree conflict, and clade-specific evolutionary patterns of genes associated with stress-related responses, we aimed to 1) identify additional WGDs; 2) examine gene tree conflict as a source of uncertainty in the reconstruction of relationships within the Portulacineae; 3) estimate phylogenetic signal for particularly contentious relationships; and 4) examine whether gene/genome duplications were associated with clade-specific adaptive traits and/or plant diversification rate.

Results and Discussion

Phylogeny of Portulacineae

Our concatenated supermatrix contained 841 gene regions and had a total aligned length of 1,239,871 bp. Gene and character occupancy were 95.4% and 84.5%, respectively. Of the ingroup taxa, only *Pereskia aculeata* had a gene occupancy <80%, whereas the majority (67 taxa) had over 90%

of genes present in the final supermatrix. Both the concatenated maximum likelihood tree (hereafter the CML tree) and the maximum quartet support species tree (MQSST) recovered the same topology with similar support values (fig. 1), all of which were largely consistent with previous estimates of the Portulacineae phylogeny (e.g., Arakaki et al. 2011; Moore et al. 2018) despite different data sets and datatypes. For example, our analysis recovered the ACPT clade (Anacampserotaceae, Cactaceae, Portulacaceae and Talinaceae; Nyffeler 2007; Nyffeler and Eggli 2010b; Ocampo and Columbus 2010; Arakaki et al. 2011; Moore et al. 2018), with high support across genes trees for a topology of (((A,P),C),T) (figs. 1 and 2) that generally agrees with that of Moore et al. (2018) using a targeted enrichment approach. However, because bootstrap values may be a poor indicator of support in large phylogenomic data sets (e.g., erroneous topologies can be increasingly supported as more sequence data are added, Alfaro et al. 2003; Phillips et al. 2004), we also conducted gene tree conflict analyses (fig. 2, Smith, Moore, et al. 2015). For example, while we found strong support (100% bootstrap for both MQSST and CML trees) for the sister relationship of Molluginaceae and Portulacineae (fig. 1), consistent with several recent studies (Edwards and Ogburn 2012; Yang et al. 2015, 2018; Moore et al. 2018), gene tree analyses highlighted that most genes were uninformative for or conflicted with this resolution. We discuss the ACPT resolution in more detail below.

Our broad taxon sampling allowed for resolution of relationships among early-diverging lineages within Portulacineae families. We found support for the sister relationship of Didiereaceae subfamilies Portulacarioideae (*Ceraria* and *Portulacaria*) and Didiereoideae (Madagascar Didiereaceae; Applequist and Wallace 2003; fig. 1). Although many gene trees were uninformative at this node, the ML topology was also the most frequent found among the gene trees (fig. 2). Relationships among the genera within the Didiereoideae have been difficult to resolve with low throughput analyses (Applequist and Wallace 2000; Nyffeler and Eggli 2010b). Here, we recovered strong support (100%) in gene tree analyses for a clade including only *Alluaudia* and *Alluaudiopsis* (fig. 1). This result differs from Bruyns et al. (2014) that found *Alluaudiopsis* to be sister to the remaining Didiereoideae.

Within Cactaceae, we recovered three major clades (*Pereskia* s.s., *Leuenbergeria*, and the core cacti, as in Edwards et al. 2005; Bárcenas et al. 2011) with strong support (>97%, fig. 1), but with substantial gene tree discordance (see below, fig. 2). Most genes (>85%) were uninformative or in conflict (fig. 2) among the earliest diverging nodes of Cactaceae. Within the core cacti, we recovered *Maihuenia* as sister to Cactoideae with strong support, and this clade as sister to Opuntioideae (fig. 1) as found by Edwards et al. (2005) and Moore et al. (2018). Nevertheless, the position of *Maihuenia* has been found to be highly unstable (e.g., within Opuntioideae, Butterworth and Wallace 2005; sister to Opuntioideae, Nyffeler 2002; or sister to Opuntioideae + Cactoideae, Hernández-Hernández et al. 2014; Moore et al. 2018). Unlike other areas within the Cactaceae that have high gene tree discordance, the

monophyly of Opuntioideae (e.g., Barthlott and Hunt 1993; Wallace and Dickie 2002; Griffith and Porter 2009; Hernández-Hernández et al. 2011) and Cactoideae (Nyffeler 2002; Bárcenas et al. 2011) had relatively low conflict (95–97% gene trees support these monophyletic groups; fig. 2). Within Opuntioideae, however, there was a high level of discordance among gene trees. These conflicts may have been influenced by limited taxon sampling as we only have one to six species representing each tribe (figs. 1 and 2). The topology of the Cactoideae was largely congruent with previous studies (e.g., Butterworth et al. 2002; Nyffeler 2002; Bárcenas et al. 2011; Hernández-Hernández et al. 2011). Within Cactaceae, the relationships among the five sampled genera were strongly supported and are consistent with relationships recovered by Hernández-Hernández et al. (2011) and Vázquez-Sánchez et al. (2013) using five loci. However, the core Cactoideae (sister to Cactaceae, Hernández-Hernández et al. 2011) were poorly supported in both the CML tree (52%) and the MQSST inference (54.5%, fig. 1). With limited taxon sampling, some relationships (e.g., the monophyly of the Pachycereeae) are largely consistent with previous studies (e.g., Nyffeler and Eggli 2010b), whereas others (e.g., the position of *Gymnocalycium*) are still incongruent among studies (Arakaki et al. 2011; Bárcenas et al. 2011; Hernández-Hernández et al. 2014).

Assessing Conflict at Specific Nodes

Incongruence can result from several biological processes, including incomplete lineage sorting (ILS), hybridization, and horizontal gene transfer (HGT). We found that gene tree conflicts are prevalent within Portulacineae. Several areas with high conflict occurred immediately after inferred WGD events (fig. 2) suggesting a major role of gene duplication and loss.

While incorporating discordance into species tree analyses is important (Liu and Pearl 2007; Liu et al. 2008), a close examination of gene-specific discordance as it pertains to support, or lack thereof, for different clades is also valuable. Displaying concordance/conflicts of gene trees on a given phylogeny (e.g., species tree) facilitates the detection of topological discordance among gene trees (fig. 2). Alternatively, comparing likelihood scores of gene trees built with different constraint topologies provides a more direct means of assessing each genes support, or lack thereof, for specific relationships (fig. 3). Recently, studies have also demonstrated that only a few genes in phylogenomic data sets can drive the resolution of nodes (e.g., Brown and Thomson 2017; Shen et al. 2017; Walker, Brown, et al. 2018). We examined individual genes and their support for specific phylogenetic relationships that had previously been identified to have significant conflicts or low supports, including the earliest divergences of Cactaceae (particularly the relationships among *Pereskia* s.s., *Leuenbergeria*, and the core cacti; Edwards et al. 2005; Moore et al. 2018), the relationships among Portulacaceae, Cactaceae, and Anacampserotaceae, and the positions of Basellaceae and Didiereaceae (e.g., Moore et al. 2018; Walker, Yang, et al. 2018). This focused analysis allowed us to isolate the signal at a particular node without having to accommodate the many diverse processes that may shape

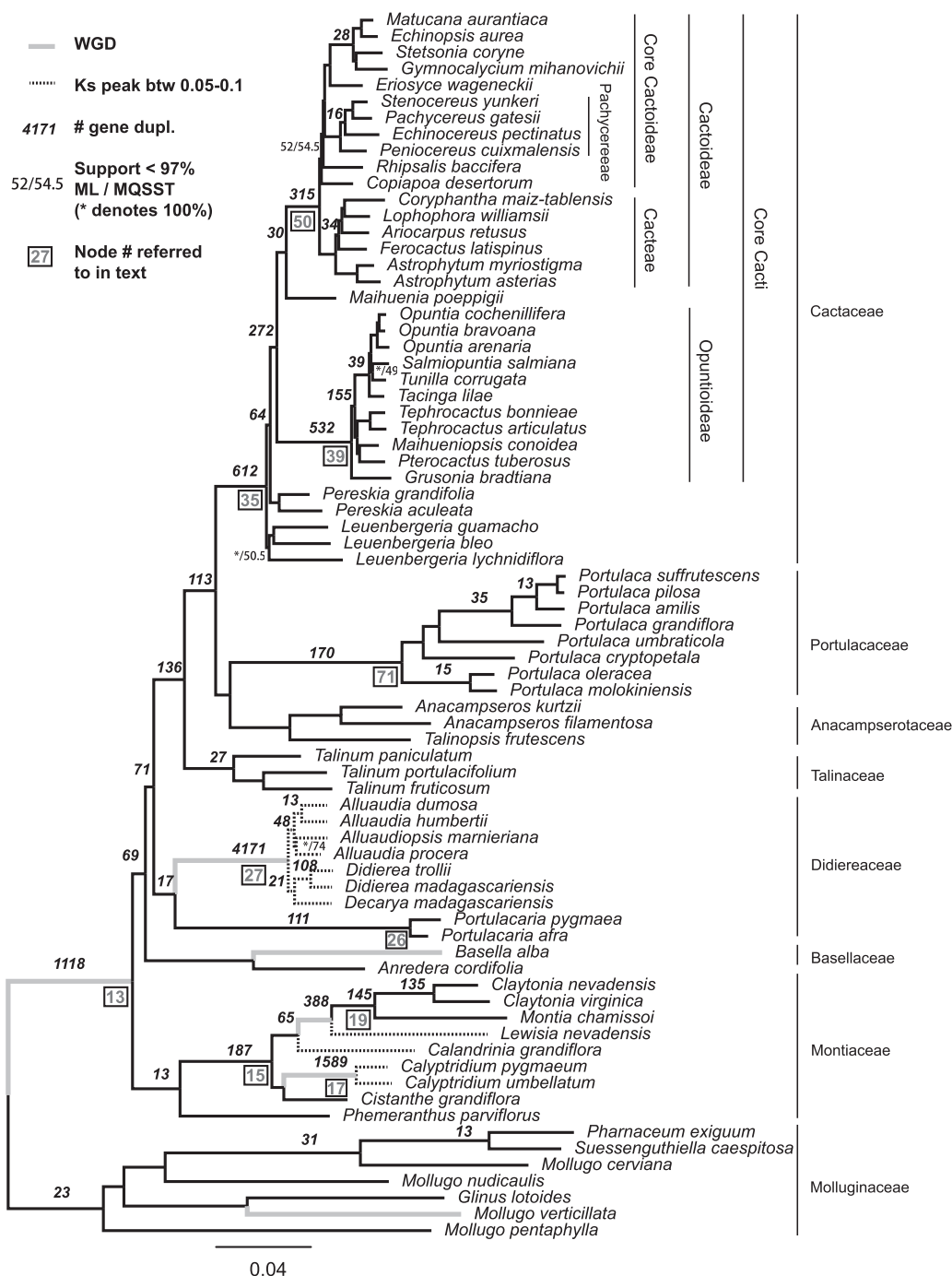


Fig. 1. Species tree from ML analysis of the 841-gene supermatrix. Gene duplication numbers (i.e., the number of duplication events; number > 12 are shown) are calculated only for branches with strong support (SH-like > 80) in the 8,332 rooted clusters.

conflict in other parts of the tree. For the early diverging Cactaceae, we examined the support for *Pereskia* + core cacti versus *Leuenbergeria* + *Pereskia* (fig. 3A₁). When being constrained for these two topologies, 56.2% genes (45.8% have $\Delta \ln L > 2$) supported the former, whereas 43.8% (34.2% have $\Delta \ln L > 2$) supported the latter. We also found that one outlying gene (cluster 4707, with homology to the “NF-X1-type zinc finger protein”) strongly supported *Pereskia* + core cacti (>110 $\ln L$ units), although with no lineage-specific positive selection, gene duplication, or any obvious alignment

problems. It is noteworthy that when relaxing constraints, almost half of the ML gene trees conflicted with both topologies suggesting other possible relationships (fig. 3A₂). However, the strongest signal supported the resolution of *Pereskia* + core cacti.

The relationship of Portulacaceae, Anacampserotaceae, and Cactaceae have also been contentious. Portulacaceae (P) has been recovered as sister to Anacampserotaceae (A, Ogburn and Edwards 2015; Walker et al. 2018a; and recovered here), to Cactaceae (C, Arakaki et al. 2011), and to both

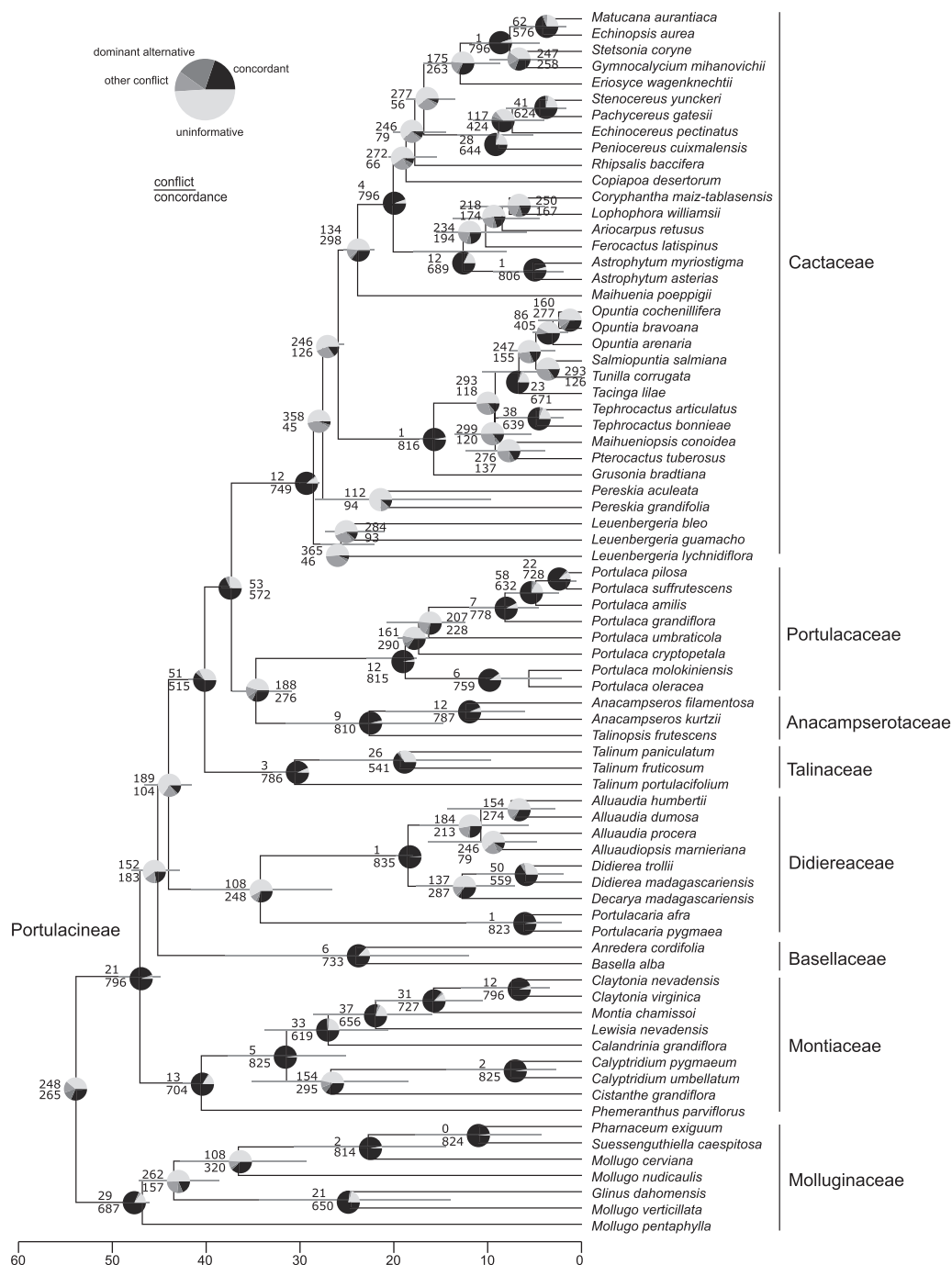


Fig. 2. Gene-tree conflict analyses. The pie chart proportions represent genes supporting congruent relationships (concordant), genes supporting the most common conflict (dominant alternative), all other conflicting gene trees (other conflict), and genes with no support (uninformative).

(Nyffeler 2007, Nyffeler and Eggli 2010a). Among the three constraints, we found that 56.6% genes supported A + P (40.8% with $\Delta\ln L > 2$), 19.5% genes supported C + P (10.8% with $\Delta\ln L > 2$), and 23.9% genes supported C + A (14% with $\Delta\ln L > 2$, fig. 3B₁). While several processes may contribute to this conflict, the distribution of the alternative placements suggest ILS to be a source of conflict (as in Moore et al. 2018) though missing data complicates our ability to test this pattern. The ML gene trees, when requiring bootstrap support $> 70\%$, confirm this as 32.8% (276) genes were congruent with the species tree, 8.1% (68) were congruent with C + P, and

10.1% (85) were congruent with C + A (fig. 3B₂). Extensive gene duplications can confound phylogenetic reconstruction due to “incomplete sorting” of paralogs (see Moore et al. 2018). However, this may be less of a problem given the low level of gene duplication at the origin of PAC (1.3%, fig. 1). We found several outlying genes including one (cluster4488, homologous to the “ARABILLO 1-like” in *Beta vulgaris*) that supported C + P (> 60 $\ln L$ units) and one (cluster7144, homologous to the “UV-B induced protein chloroplastic-like” gene in *Chenopodium quinoa*) that supported C + A strongly (> 100 $\ln L$ units). We did not detect any

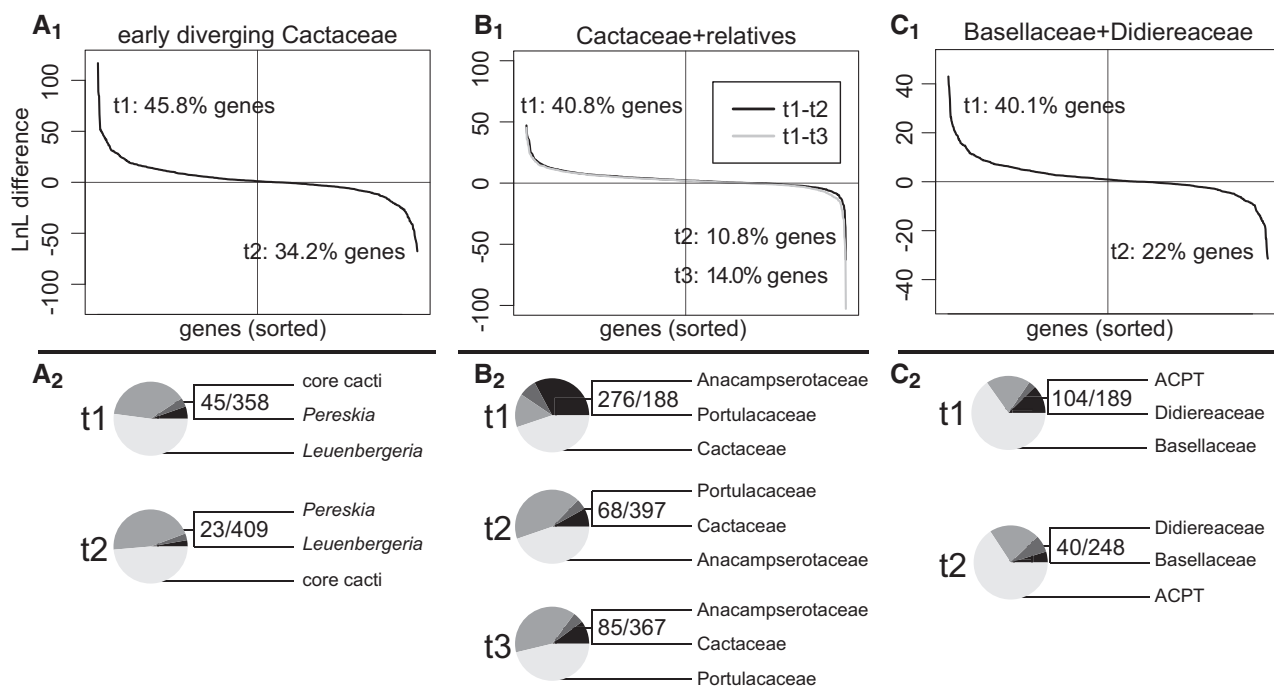


Fig. 3. Support for alternative phylogenetic hypotheses, including (A) The two alternative topologies involving the early diverging Cactaceae. (B) The three alternative topologies involving Cactaceae, Portulacaceae, and Anacampserotaceae. (C) The two alternative topologies involving Basellaceae, Didiereaceae, and ACPT (Anacampserotaceae, Cactaceae, Portulacaceae, and Talinaceae). Upper panel (A₁, B₁, C₁): comparison of lnL (for each gene, 841 genes in total) between gene trees built with alternative constraint topologies (sorted by greatest Δ lnL). Values above the x-axis show the difference in likelihood for those genes supported the resolution found in “t1,” whereas those falling below the axis support “t2” (black) or “t3” (gray). The percentage of genes with Δ lnL > 2 for each topology is shown. Outlying genes (those genes with extremely high support for certain topology) are discussed in Results and Discussion. Lower panel (A₂, B₂, C₂): the number of ML gene trees (built without any constraint) that are concordant/conflict (labeled on each node) with the given topologies for each edge test, with nodes >70% support were considered. Please note that ML genes trees can exhibit many other possible relationships, they were not examined or shown due to focal interest. Pie chart proportions are the same as in figure 2.

obvious errors in orthology inference (e.g., Brown and Thomson 2017), alignment, or lineage-specific positive selection.

We conducted similar analyses for the resolution of Basellaceae and Didiereaceae (fig. 3C₁). Specifically, between the two constraints, 60.2% genes (40.1% with Δ lnL > 2) supported Didiereaceae + ACPT and 39.8% genes (22% with Δ lnL > 2) supported Basellaceae + Didiereaceae as found by Soltis et al. (2011) and Anton et al. (2014). No significant outlying genes (the Δ lnL ranging from −31.41 to 42.98 smoothly) were observed supporting either topology (fig. 3C₁). The dominant signal supported Didiereaceae + ACPT (fig. 3C₂).

All the tests discussed demonstrated significant conflict among genes. To address whether strong signal for hybridization contributed to the conflicting signal, we conducted several analyses. The results from them suggested that we cannot rule out ancient hybridization (supplementary table S1 and fig. S2, Supplementary Material online). However, there is no other evidence (such as species history and geographical distribution) for hybridization between these lineages, and so we consider this to be further evidence of complex gene family evolution within the clade. Nevertheless, future analyses of synteny or other evidence

may provide more direct evidence for or against hybridization as a source of conflict.

Multiple WGD Events in Portulacineae

WGD have had a profound influence on the evolutionary history of plants (e.g., Cui et al. 2006; Soltis and Soltis 2009; Jiao et al. 2011; Wendel 2015; Yang et al. 2015; Smith et al. 2018a). Previous analyses have identified WGD events within the Caryophyllales (Yang et al. 2015, 2018; Walker et al. 2017) including in the ancestor of Portulacineae, in the lineage of *Basella alba*, and within Montiaceae. Improved taxon sampling enabled us to confirm these events and infer additional putative WGD events as well as several instances of large-scale gene duplications (fig. 1). Ks plots have been widely applied to infer WGDs but are not without limitations (Parks et al. 2018). Higher Ks values (e.g., >2) are associated with increasingly large error (Vanneste et al. 2013) and smaller Ks values (e.g., <0.25) can be difficult to interpret due to the errors in assembly, splice variants, or other anomalies. We used both gene tree and Ks analyses to identify WGDs more confidently (fig. 1).

We found evidence for a previously identified WGD event in the ancestor of Portulacineae (Yang et al. 2015, 2018) given the high percentage of gene duplications (13.4%) and a Ks

peak (~ 0.4 – 1.0) shared by all members in this clade. However, we did not detect a similar pattern of gene duplications at the ancestor of Portulacineae + Molluginaceae (1%), confirming that Yang et al. (2018)'s estimation could result from phylogenetic uncertainty. We found a high percentage of gene duplications at the base of the Didiereoideae (50.1%) accompanied by a very recent Ks peak (between 0.05 and 0.10) for all the species within this clade. We note that the Didiereoideae is a relatively slow evolving lineage with a recent Ks peak. Therefore, we rely heavily on corroborating both the Ks analyses and gene tree analyses in order to confidently identify the WGD (i.e., Ks peak occurred < 0.1 , [supplementary fig. S3, Supplementary Material](#) online). We also identified genome duplications within the Montiaceae. Yang et al. (2018) identified WGD in the ancestor of *Claytonia* species and, with expanded sampling, we placed this WGD event at the ancestor of *Lewisia* and *Claytonia* as all four species in this clade exhibited Ks peaks between 0.2 and 0.4 ([supplementary figs. S3 and S4, Supplementary Material](#) online). We also identified a high number of gene duplications in the two *Calyptidium* species, accompanied by a very recent Ks peak (between 0.05 and 0.10; [fig. 1](#)).

While we identified many gene duplications (7.3%) at the origin of the Cactaceae, this was not corroborated by Ks analyses. This is consistent with other recent studies (Walker et al. 2017; Yang et al. 2018). Several other nodes (e.g., nodes 50 and 39) exhibited similar patterns with high numbers of duplications (3.8–6.4%) but no corresponding Ks peak. It is unclear what events may be associated with these duplication events or whether gene loss, gene tree conflict, life history shifts, etc., may be obscuring the detection of gene/genome duplications, but further investigation will help shed light on these patterns and processes.

Although gene-tree-based methods can help identify the phylogenetic location of WGD events, they are unable to infer WGD events that have occurred on terminal branches. To identify these, we examined Ks peaks that are unique to one taxon and absent from close relatives. We inferred two such cases: on the branch leading to *Basella alba* ($K_s = 0.45$) and the branch leading to *Mollugo verticillata* ($K_s = 0.25$; [supplementary fig. S3, Supplementary Material](#) online). Both species also exhibit higher chromosome numbers relative to close allies and are consistent with a previous study (Yang et al. 2018).

WGD, Diversification, and Climatic Niche Shifts

Smith et al. (2018a) examined the potential associations between WGDs, diversification rate shifts, and climatic niche shifts within the Caryophyllales and found that some WGD events were associated with shifts in climatic niche but not diversification rate shifts. Here, we conducted similar analyses with a focus on the Portulacineae. Specifically, we examined mean annual temperature, mean annual precipitation, and the first axis of a principal component analysis on bioclimatic variables 1–19 (PCA1) among clades with WGDs (Portulacineae, Didiereoideae, and Montiaceae), as well as the Cactaceae using a species-level phylogeny. The Montiaceae, except for *Phemeranthus*, were treated together

as the species-level tree conflicted with the transcriptomic tree regarding internal resolution of clades within the Montiaceae. Consistent with Smith et al. (2018a), WGD events were associated with shifts in climate tolerance ([fig. 4A and B](#)). Montiaceae was associated with movement into colder environments, as found in Smith et al. (2018a); Didiereoideae occupied a wetter climate than the sister clade but with both clades occupying relatively seasonally dry areas ([supplementary fig. S5, Supplementary Material](#) online). The Cactaceae exhibited more variable rates of climatic niche evolution ([fig. 4B](#)), occupied a slightly hotter environment, but did not exhibit a strong pattern of different climate occupancy. This does not negate climate expansion within the clade, but suggests it may not be experienced by the entire clade.

While WGDs may not always be associated with shifts in niche occupancy, their association appears to be common within the Portulacineae and the broader Caryophyllales. The Cactaceae did not exhibit a WGD but instead extensive gene duplication and the climatic niche results for that clade are less clear. Some have suggested that WGDs may be associated with diversification rate increases (Landis et al. 2018). We examined this question on both the species-level phylogeny and a reduced phylogeny of major lineages using the estimated diversity. In the species-level phylogeny, we found diversification rate shifts within Portulacaceae, Anacampserotaceae, and Cactaceae ([fig. 4](#)). In the reduced phylogeny, we found shifts at the origin of core Cactaceae ([supplementary fig. S6, Supplementary Material](#) online). The difference between these results suggests that the species-level phylogeny may better place all potential shifts whereas the reduced phylogeny may better represent the overall diversity. Nevertheless, in either case we did not find a correspondence between diversification shifts and WGDs.

Gene Families Show Broad Expansions across Portulacineae

Examining gene family expansion in depth has been instrumental in understanding functional trait evolution in Caryophyllales. For example, gene expansion and neofunctionalization have been implicated in the evolution of betalain pigmentation in Caryophyllales (Brockington et al. 2015; Yang et al. 2015). Here, we explored gene family expansion in Portulacineae with respect to stress tolerance. The top 20 most expanded gene families in Portulacineae ([table 1](#)) included genes encoding transporters, proteases, cytoskeletal proteins and enzymes that are involved in photosynthetic pathways. Several genes were associated with responses to drought: the gene encoding plasma membrane intrinsic protein (PIP) that can regulate the transport of water through membranes (Vandeleur et al. 2009), heat shock proteins (Kiang and Tsokos 1998), and ubiquitin (Hochstrasser 2009). But perhaps the most interesting gene expansions involved those encoding phosphoenol-pyruvate carboxylase (PEPC, Silvera et al. 2014) and NADP-dependent malic enzyme catalyses (Ferreira et al. 2003), which are key enzymes involved in CAM/C₄ photosynthesis (Smith and Winter 1996; Cushman 2001). Lineage-specific selection analyses on these

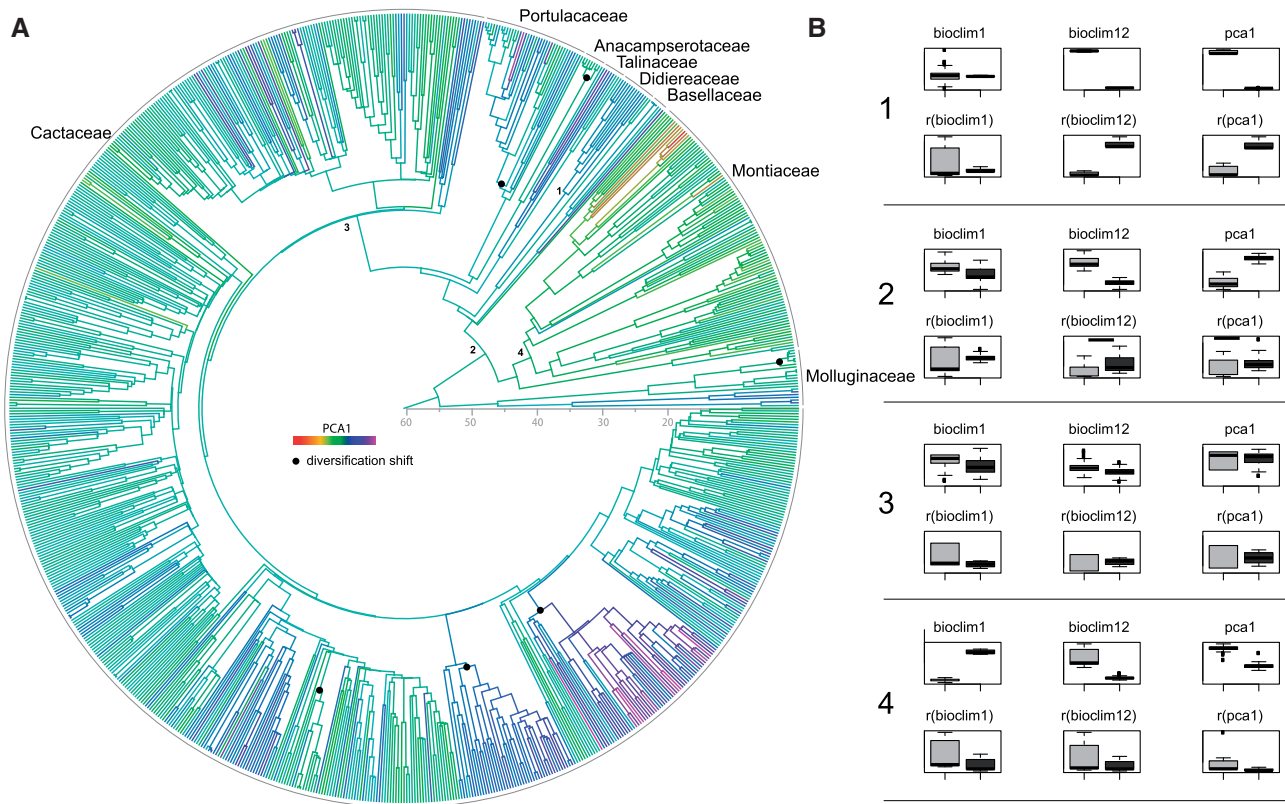


FIG. 4. (A) Reconstruction of first axis of principal component analyses (PCA1) on the species level tree of Portulacineae. Black nodes indicate the diversification rate shifts. (B) Comparison of MAT (mean annual temperature), MAP (mean annual precipitation), and PCA1 between clades with WGD/extensive gene duplication and their corresponding sister clades (i.e., 1: Didiereoideae vs. *Portulacaria*, 2: the whole Portulacineae vs. Molluginaceae, 3: the whole Cactaceae vs. Portulacaceae + Anacampserotaceae, 4: Montiaceae except for *Phemeranthus* vs. *Phemeranthus*). The upper panel is the comparison of reconstructed ancestral states. The lower panel is the comparison of evolutionary rate (i.e., Brownian motion rate that was used for ancestral state reconstruction). Box plots are constructed from bootstrapped phylogenies.

Table 1. Portulacineae Gene Families with the Highest Total Number of Tips.

Cluster ID	Ntip	Ntaxa	Annotation
cluster1_2rr_1rr_2rr	949	82	Tubulin beta chain [<i>Spinacia oleracea</i>]
cluster2_1rr_1rr_1rr	871	82	Heat shock cognate 70 kDa protein-like [<i>Nicotiana tabacum</i>]
cluster4_1rr_1rr_1rr	698	82	Actin-11, NBD sugar-kinase HSP70 actin [<i>Chenopodium quinoa</i>]
cluster3_3rr_1rr_1rr	670	81	Polyubiquitin-A [<i>Triticum urartu</i>]
cluster5_1rr_1rr_1rr	655	82	Plasma membrane H ⁺ ATPase 9 [<i>Sesuvium portulacastrum</i>]
cluster8_1rr_1rr_1rr	582	82	Tubulin alpha-3 chain [<i>Vitis vinifera</i>]
cluster12_1rr_1rr_1rr	565	82	Phosphoenolpyruvate carboxylase [<i>Suaeda aralocaspica</i>]
cluster43_1rr_1rr_1rr	488	81	Ubiquitin-conjugating enzyme E2 28 [<i>Solanum lycopersicum</i>]
cluster18_1rr_1rr_1rr	472	82	Serine/threonine-protein kinase tricornet isoform X1 [<i>Beta vulgaris vulgaris</i>]
cluster22_1rr_1rr_1rr	467	82	NADP-dependent malic enzyme [<i>Hylocereus undatus</i>]
cluster15_1rr_1rr_1rr	463	82	Transmembrane 9 superfamily member 3 [<i>Chenopodium quinoa</i>]
cluster17_1rr_1rr_1rr	462	82	Auxin transporter-like protein 2 [<i>Chenopodium quinoa</i>]
cluster19_1rr_1rr_1rr	461	82	Plasma membrane intrinsic protein 2 [<i>Sesuvium portulacastrum</i>]
cluster7_1rr_1rr_1rr	456	81	Uridine-cytidine kinase C isoform X1 [<i>Beta vulgaris vulgaris</i>]
cluster26_1rr_1rr_1rr	435	82	S-adenosylmethionine synthase 1 [<i>Beta vulgaris vulgaris</i>]
cluster14_1rr_1rr_1rr	467	82	Probable xyloglucan endotransglucosylase/hydrolase protein 23 [<i>Beta vulgaris vulgaris</i>]
cluster6_2rr_1rr_1rr	530	81	Homeobox-leucine zipper protein ATHB-15 isoform X1 [<i>Beta vulgaris vulgaris</i>]
cluster9_1rr_1rr_1rr	527	81	Expansin-A10-like [<i>Chenopodium quinoa</i>]
cluster16_1rr_1rr_1rr	438	82	Ubiquitin carboxyl-terminal hydrolase 12 [<i>Beta vulgaris vulgaris</i>]
cluster37_1rr_1rr_1rr	435	82	14-3-3 protein 10-like [<i>Chenopodium quinoa</i>]

NOTE.—Ntip, number of tips; Ntaxa, number of taxa.

two genes revealed that 10–13 internal branches are likely under positive selection, with the most occurring in the Cactaceae (supplementary table S2, Supplementary Material

online). The role of gene duplication in the evolution of *C₄* photosynthesis has been contentious, and some authors have proposed that neo-functionalization of genes following

duplication has not played a major role in the evolution of C_4 photosynthesis (e.g., Williams et al. 2012; van den Bergh et al. 2014). However, recent analyses provide evidence that duplication and subsequent retention of genes are associated with the evolution of C_4 photosynthesis (Emms et al. 2016), specifically including the NADP-dependent malic enzyme. Moreover, convergent evolution in several key amino acid residues of PEPC has been suggested to be associated with the origin of both C_4 and CAM (Christin et al. 2007, 2014; Goolsby et al. 2018). A recent study has also confirmed the occurrence of multiple rounds of duplication within the major PEPC paralog (*ppc-1E1*) in the ancestral Portulacineae (Christin et al. 2014). The results presented here provide additional evidence that gene family expansion may play an important role in some aspects of photosynthetic pathway evolution.

Lineage-Specific Gene Expansions Associated with Adaptive Traits

In addition to investigating genes that experienced significant expansion, we also conducted Gene Ontology (GO) overrepresentation analyses on genes duplicated at biologically significant nodes (nodes 15, 17, 19, 26, 27, 35, 39, 50, and 71; fig. 1). Of genes duplicated at each clade, 30–55% had corresponding *Arabidopsis* locus IDs and most could be mapped in PANTHER (supplementary table S3, Supplementary Material online). We identified overrepresented GO terms at three nodes (15, 27, and 35) with the Didiereoideae (node 27) exhibiting the most GO overrepresentation. This may be the result of a more recent WGD. Overrepresented GOs reflected diverse functional classes: genes belonging to “calcium ion binding” (GO: 0005509) at the origins of the Montiaceae (except *Phemeranthus parviflorus*) and the Didiereoideae, genes belonging to “sulfur compound metabolic process” (GO: 0006790) at the origin of Didiereoideae, and genes involved in “Hydrolase activity, acting on ester bonds” (i.e., esterase, GO: 0016788, supplementary table S4, Supplementary Material online) at the origin of Cactaceae. Of particular note are the genes belonging to the “sulfur compound metabolic process” (e.g., *AtAHL*, which belongs to the *Arabidopsis* HAL2-like gene family, Gil-Mascarell et al. 1999; and *NIFS1*, which belongs to the cysteine desulfurase family, Schmidt 2005) as sulfur-bearing evaporite compounds (principally gypsum) are commonly found in soils in areas with low rainfall and high evaporation rates (Watson 1979) where many Portulacineae lineages have diversified. Previous studies have also connected primary sulfur metabolism (e.g., sulfate transport in the vasculature, its assimilation in leaves, and the recycling of sulfur-containing compounds) with drought stress responses (e.g., Chan et al. 2013). Consequently, duplication and overrepresentation of these sulfur metabolic genes could potentially be evidence of adaptation to corresponding harsh environments of hot and cold deserts (e.g., Lee et al. 2011). While it is tempting to speculate that these duplications have been maintained through adaptation under selective pressure, we are aware that such

“duplication” events may simply be the remnants of ancient polyploidy or segmental duplication. In the absence of functional characterization, it is necessary to interpret all GO analyses with caution.

Targeted Analyses of Drought and Cold Associated Genes

We examined 29 functionally annotated genes identified to be involved in drought and cold tolerance (table 2), as several lineages have repeatedly colonized cold and dry environments. Our results suggest that these genes duplicated more than expected within the Didiereoideae, Montiaceae, and Cactaceae (supplementary figs. S7 and S8, Supplementary Material online), where adaptations to harsh environment are prevalent. Several of these genes may be good targets for future study. For example, WIN1 (SHN1) proteins, transcription factors associated with epicuticular wax biosynthesis that increase leaf surface wax content, showed duplications in both Didiereoideae and Montiaceae (table 2). Six homologs that are part of the calcium-dependent protein kinase (CDPK or CPK) gene family are known to be involved in drought stress regulation (Geiger et al. 2010; Brandt et al. 2012) and experienced several duplications: four at the origin of Didiereoideae, one within Montiaceae, one at Portulacaceae, and two within the Cactaceae.

These 29 genes also exhibited variable lineage-specific positive selection within Portulacineae (table 2). The strongest signals of positive selection were all found in genes associated with drought and/or cold tolerance in the ABA signaling pathway (such as the NAC10/29 and WRKY33 TFs, the CDPK18 and the PP2Cs, e.g., Golladack et al. 2014; Nakashima et al. 2014; Huang et al. 2015; Li et al. 2017), a central regulator of abiotic stress resistance. Interestingly, the *ABI1* and *PP2CA* genes, which encode two proteins of the PP2Cs family, were among the genes with the highest number of lineages under positive selection. As negative regulators, PP2Cs can regulate many ABA responses, such as stomatal closure, osmotic water permeability, drought-induced resistance, seed germination, and cold acclimation (Gosti et al. 1999; Merlot et al. 2001; Mishra et al. 2006). In addition to PP2Cs, the positive regulators SNF1-related protein kinase 2s (SnRK2s) are also related to ABA signaling (Hubbard et al. 2010). While we did not detect many lineages under positive selection in the three SnRK2 genes (i.e., *SnRK2.4*, 2.5, 2.6, table 2), we found that they experienced ancient duplications at the origin of Portulacineae or earlier and recent duplications within Montiaceae, Didiereaceae, or Cactaceae (nodes 17, 27, 39; table 2). Positive selection and gene family expansion in these gene families within Cactaceae, Didiereaceae, and Montiaceae suggests that these genes warrant further investigation for their potential role in the evolution of adaptations to challenging environmental conditions.

Whereas several of these genes experienced duplications within clades of the Portulacineae, many occurred at the origin of Portulacineae or earlier (table 2, supplementary table S2 and fig. S8, Supplementary Material online). This suggests that the ability for the Caryophyllales, and the Portulacineae, to repeatedly adapt to harsh

Table 2. Selection Analyses on 29 Genes Associated with Cold and Drought Responses.

Gene	Ntip	TS	Cac35	Did27	Mont15	Port71	SH80 (duplication)
<i>ABI1</i>	151	17	6	2	4	0	27
<i>NAC10</i>	158	16	6	4	3	1	27, 39
<i>WRKY33</i>	165	13	4	2	3	1	17, 27
<i>NAC29</i>	163	13	7	2	2	0	13
<i>CDPK18</i>	196	13	6	1	0	0	2, 13, 21, 35, 57
<i>PP2C37 (PP2CA)</i>	179	11	6	1	2	0	13, 17, 19, 35
<i>HSP70</i>	270	10	3	1	1	2	0, 19, 27, 33, 49, 51, 52
<i>ICE1</i>	194	10	4	1	3	0	17, 20, 27, 38
<i>HSP70</i>	221	10	2	0	2	0	0, 27, 35
<i>ERF5</i>	86	9	1	2	3	0	17, 27, 30
<i>HSP90-5</i>	213	8	2	2	1	1	26, 27, 34
<i>CDPK1</i>	85	7	4	2	1	0	—
<i>SNRK2.5</i>	156	5	2	1	0	1	39
<i>MYB124 (FLP)</i>	178	5	2	1	0	1	39
<i>HSP83</i>	305	5	1	1	0	0	0, 13, 17, 27, 34, 76
<i>CRPK1</i>	177	5	3	2	0	0	13, 20, 27
<i>AREB1</i>	92	4	1	1	1	0	17, 27
<i>NAC2</i>	157	4	2	0	0	1	27, 39
<i>MYB60</i>	87	4	2	0	0	1	27
<i>WIN1 (SHN1)</i>	79	4	0	0	0	1	17, 27
<i>CDPK13</i>	156	3	0	0	0	2	27
<i>CDPK26</i>	117	3	1	1	0	1	0, 17, 27
<i>PLC2</i>	108	3	1	0	1	0	17, 19, 35
<i>CDPK8</i>	184	2	1	0	0	0	13, 27, 71
<i>CRLK1</i>	170	2	2	0	0	0	27
<i>SNRK2.6</i>	180	2	2	0	0	0	27, 39
<i>SNRK2.4</i>	125	1	0	0	0	0	17, 27
<i>CDPK10</i>	85	1	1	0	0	0	27
<i>ERF2</i>	84	0	0	0	0	0	17, 27

NOTE.—Ntip, number of tips after additional homolog filtering for selection analyses; TS, total number of branches under positive selection; Number of branches under positive selection within certain clades, including Cactaceae (Cac35), Didiereoideae (Did27), Montiaceae (Mont15), and Portulacaceae (Port71). SH80, Duplication nodes with SH-like support > 80, Node number can be referred to in [fig. 1](#) and [supplementary fig. S1, Supplementary Material](#) online. Two genes (*ABI1* and *PP2CA*) that encode important negative regulators in the ABA pathway are mainly discussed in the text. More information about references and gene duplication patterns can be found in [supplementary table S2, Supplementary Material](#) online.

environments may have arisen early in the evolution of the clade and that the predisposition to become adapted to these environments may be the result of the early diversification of the gene families.

Conclusions

WGDs (i.e., paleopolyploidy) have played critical roles in major events of plant evolution, but the nature and scale of their influence on macroevolutionary patterns is still debated. Our results suggest a complex relationship between WGDs in the Portulacineae and several evolutionary patterns and processes. We identify significant gene-tree/species-tree discordance immediately after extensive gene duplication events, suggesting an association between rapid speciation and gene duplication and loss. We also find evidence for the association of WGDs with shifts into different environments ([Smith et al. 2018a](#)). While this alone does not imply a direct cause, it suggests one means by which WGD can impact adaptation in plants and perhaps why some clades exhibit diversification shifts after WGDs.

We also identify potential molecular evolutionary patterns that corresponded to adaptations to extreme environments. While these results provide essential resources for further and deeper examination with more detailed methods, they also

highlight the major limitations in resources for nonmodel organisms. They rely on annotations based on model organisms that may be distantly related to the species of interest. While the gene itself may exhibit interesting patterns, the associated GO label may or may not carry much meaning and may have different functions in these different lineages. Despite these limitations, however, these analyses identified gene families that should be explored further, regardless of the accuracy of the GO label, as the patterns of selection and duplication alone suggest their importance.

Finally, our findings contribute to a growing literature suggesting the importance of gene tree examination along with species tree construction ([Smith, Moore, et al. 2015; Brown and Thomson 2017; Shen et al. 2017; Walker et al. 2018b](#)). Gene tree conflict is common and needs to be incorporated into our analyses of molecular and macro-evolution. The conflicts likely hold interesting evolutionary answers that should be analyzed in more detail.

Materials and Methods

Taxon Sampling, Transcriptome Generation, and Homology Inference

Sixty-eight ingroup species were included in this study, representing all families within the Portulacineae

(Anacampserotaceae, Cactaceae, Basellaceae, Didiereaceae, Montiaceae, Talinaceae, and Portulacaceae) sensu APG IV (Angiosperm Phylogeny Group, 2016) except for Halophytaceae. We also included 14 Caryophyllales species as outgroups, including seven taxa of Molluginaceae (supplementary tables S5 and S6, Supplementary Material online). Of the 82 transcriptomes included in the study, 47 were newly generated following the protocol of Yang et al. (2017). We followed the pipeline of Yang and Smith (2014) with minor modifications (available at <https://bitbucket.org/ningwang83/Portulacineae>) to infer homologs. In total, we obtained 8,592 final homolog clusters, which are used in orthology inference, gene duplication analyses and GO annotation. More details about transcriptome generation and homology inference can be found in Supplementary Material and Methods.

Orthology Inference and Species Tree Estimation

We used the rooted tree (RT) method in Yang and Smith (2014) to extract orthologs from homolog trees with *Beta vulgaris* as the outgroup. Orthologous clades with >30 ingroup taxa were retained (supplementary fig. S9, Supplementary Material online). We extracted orthologs with >75 taxa, aligned and cleaned them with MAFFT v7 (genafpair -maxiterate 1000, Katoh and Standley 2013) and Phyutility (-clean 0.3, Smith and Dunn 2008), and inferred a ML tree for each using RAXML v8.1.22 (Stamatakis 2015) with GTR + Γ model. After removing taxa with a terminal branch length longer than 0.1 and 10 times greater than sister clade from 58 orthologs, 841 orthologs including at least 77 taxa and 500 aligned DNA characters were re-aligned with PRANK v.140110 (Löytynoja and Goldman 2008) using default settings and trimmed in Phyutility (-clean 0.3).

The species tree was inferred using two methods. First, a concatenated matrix was built with the 841 orthologs and a ML tree was estimated by using RAXML with the GTR + Γ model partitioned by gene. Node support was evaluated by 200 fast bootstrap replicates. Second, an MQSST tree was estimated in ASTRAL 4.10.12 (Mirarab et al. 2014) using ML gene trees, with uncertainty evaluated by 200 bootstrap replicates using a two-stage multilocus bootstrap strategy (Seo 2008).

Assessing Conflicts among Gene Trees

We assessed conflict by mapping the 841 rooted ortholog trees onto the species tree topology with PhyParts (Smith, Moore, et al. 2015). Trees were rooted on *Beta vulgaris* if present, using *Limeum aethiopicum* and/or *Stegnosperma halimifolium* if *Beta* was not present, and with *Sesuvium portulacastrum*, *Delosperma echinatum*, *Anisomeria littoralis*, and *Guapira obtusata*, if the first three species were not present. We only considered nodes with >70% bootstrap support in the gene trees. We conducted likelihood comparisons in more detail as in Walker et al. (2018b) for three areas of the tree: 1) Cactaceae and relatives, 2) the early diverging branch within Cactaceae (i.e., *Leuenbergeria* clade), and 3) with the placement of Basellaceae and Didiereaceae. We

calculated likelihood scores for each gene tree constrained to alternative resolutions and compared scores. We considered a higher lnL for one topology as a sign of support, and $|\Delta \ln L|$ (the difference in lnL value) >2 as a sign of significant support (as in Walker et al. 2018b). Potential outlying genes that exhibited extreme deviations in likelihood were examined for lineage specific positive selection. We analyzed potential hybridization with PhyloNet (Than et al. 2008) and D_{FOIL} (Pease and Hahn 2015). PhyloNet does not run well on large data sets so we pruned the gene trees to include five taxa from Cactaceae, one from each ingroup family, with the *L. aethiopicum* as an outgroup. Species were selected based on their gene occupancy. We conducted Maximum Pseudo Likelihood analyses with 10 maximum reticulations. D_{FOIL} analyses can only analyze 5-taxon trees, so we reduced gene trees to four species with the highest gene occupancy for the clade of interest and *Pharnaceum exiguum* as an outgroup. For example, we choose two species from Cactaceae and one from each of Portulacaceae and Anacampserotaceae, to test the controversial relationships among these families.

Inference of Gene and Genome Duplication

To infer gene duplications, we considered nodes in homolog gene trees with SH-like support >80, calculated using the SH-like test (Anisimova et al. 2011) in RAXML. We extracted 8,332 rooted clusters with ≥ 30 taxa, and conducted PhyParts analyses with gene duplication recorded when sister clades sharing two or more taxa based on Yang et al. (2015). We also conducted Ks plot analyses following the same process of Yang et al. (2015). Briefly, we reduced highly similar PEP (peptide) sequences (by CD-HIT: -c 0.99 -n 5, Fu et al. 2012) for each of the 82 species, and conducted an all-by-all BLASTP (-evalue = 10, -max_target_seq = 20) and removed highly divergent hits with <20% similarity (pident) or <50 aligned amino acids (nident). Sequences with ten or more hits were also removed to eliminate large gene families. We then used the PEP and corresponding CDS (coding DNA sequences) from paralogous pairs to calculate Ks values using the pipeline https://github.com/tanghaibao/biopipeline/tree/master/synonymous_calculation. Peaks were identified by eye as in Yang et al. (2015).

To determine whether a potential WGD event occurred before or after a speciation event, we calculated between-species Ks distribution using orthologous gene pairs. The procedure is similar as above, except that a reciprocal BLASTP was carried out between two species instead of an all-by-all BLASTP within one taxon. Between-species Ks plots were compared with within-species Ks plots to determine the relative timing of the WGD and speciation event.

Reconstruction of Climate Occupancy and Diversification Rate

To test whether WGD were associated with climate niche shift and/or diversification rate shift, we reconstructed ancestral climate occupancy and diversification on a species level phylogeny of Portulacineae. First, sequence data from NCBI were gathered using PyPHLAWD (Smith and Walker 2018). To increase the accuracy of phylogenetic inferences, major

clades were constrained based on the transcriptomic results here. Phylogenies of Molluginaceae and Portulacineae were also built individually and then combined. ML and 100 bootstrap trees were constructed with RAXML. For divergence time estimation, we first selected five ortholog genes from our transcriptomic data set using SortaDate (Smith et al. 2018b). A time tree was then built with them in BEAST 1.8.3 using 13 secondary time calibrations (supplementary table S7, Supplementary Material online) from Arakaki et al. (2011). This tree included 82 taxa (hereafter “small tree”) as compared with the species-level tree (hereafter “big tree”). The “big tree” was dated using treePL (Smith and O’Meara 2012) using the dates from the “small tree” as constraints. In addition, 93 species-level bootstrap trees that exhibited consistent family level relationships with our transcriptomic tree were also dated using treePL.

Reconstruction of ancestral climate occupancy state and diversification rate followed the methods of Smith et al. (2018a). Briefly, we extracted the climate occupancy data (COD) for the Portulacineae and Molluginaceae species from GBIF based on the taxa from the “big tree.” Mean climate values of Bioclim 1 (mean annual temperature), 12 (mean annual precipitation), and principal component axis 1 (based on a PCA of the set of full bioclimatic variables) were calculated for each taxon. Both ancestral states and Brownian motion rates of evolution were compared between WGD nodes and their sister clades. We reconstructed diversification using MEDUSA (Alfaro et al. 2009; Pennell et al. 2014) on the ‘small tree’ with the estimated number of species for each taxa group and the “big tree.”

Annotation of Expanded Gene Families

To investigate gene function and GO term overrepresentation, we conducted several analyses. We conducted BLASTX analyses against the nonredundant protein NCBI database by using a PEP sequence from each of the top 20 genes with the highest total number of tips. We also identified clades with large number of gene duplications (i.e., > 160 [2%] gene duplications), and conducted BLASTN analyses (-evalue 10) of these against the database of *A. thaliana* (release-34 EnsemblPlants). GO terms and IDs were recorded. We conducted GO overrepresentation test (i.e., overrepresentation of genes in a specific functional category, <http://www.pantherdb.org>) using genes duplicated at the origin of Portulacineae (Node 13) as background to detect potential overrepresentation of genes that duplicated more recently. We corrected for multiple tests using the Bonferroni correction.

Identification of Lineage-Specific Selection on Stress Response Genes

We carried out selection analyses on a targeted set of 29 genes that were present in our data and are known to be associated with cold and/or drought adaptation. We repeated the homolog filtering as indicated above three times to reduce potential assembly or clustering errors. CDS sequences were aligned based on the corresponding peptide alignments using

phyx (Brown et al. 2017). We further cleaned CDS alignments by combining sequences with identical overlapping segments to eliminate the potential assembly redundancy, which, although uncommon, may cause error in these analyses. After trimming the CDS alignment (Phyutility -clean 0.1), a ML tree was built for each homolog in RAXML with GTR + Γ model and rooted with outgroup species. Lineage specific selection was calculated for each homologous gene in HyPhy v2.2.4 (Pond and Muse 2005) using an adaptive branch-site random effects likelihood test (aBSREL, Smith, Wertheim, et al. 2015), which automatically infers an appropriate model among branches. The optimal number of ω categories was selected according to AICc scores. The branches with episodic positive selection that show significant proportion of sites with $\omega > 1$ were chosen with $P < 0.05$ after applying the Holm–Bonferroni multiple testing correction.

To test whether the 29 targeted genes associated with cold/drought adaptation experienced significantly more duplications at certain nodes, we randomly selected 29 genes from the final homologous clusters (with an additional overlap checking to be compatible with the treatment of the targeted 29 genes) and calculated the number of genes duplicated. We repeated this process for 1000 times and plotted the frequency distribution of the number of duplicated genes for each node (supplementary fig. S7, Supplementary Material online).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

The authors thank Wynn Anderson, the Bureau of Land Management, the US Forest Service, and the staff of Desert Botanical Garden, Sukkulenten-Sammlung Zürich, Cambridge University Botanic Garden, Missouri Botanical Garden, and the Oberlin College Greenhouse for permission to collect specimens, and thank Hilda Flores, Helga Ochoterena, and Norman Douglas for help with collecting. The authors also thank Jeet Sukumaran for constructive discussion on gene tree discordance. Special thanks to the constructive comments from the editors and three anonymous reviewers. Oscar Vargas provided a thorough review of the first draft of the paper. This work was supported by NSF DEB 1354048 to S.A.S. and NSF DEB 1352907 to M.J.M.

References

- Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G, Harmon LJ. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc Natl Acad Sci U S A*. 106(32):13410–13414.
- Alfaro ME, Zoller S, Lutzoni F. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Mol Biol Evol*. 20(2):255–266.
- Anderson EF. 2001. The Cactus family. Pentland (OR): Timber Press.
- Anton AM, Hernández-Hernández T, De-Nova A, Sosa V. 2014. Evaluating the phylogenetic position of the monotypic family

- Halophytaceae (Portulacinae, Caryophyllales) based on plastid and nuclear molecular data sets. *Bot Sci.* 92(3):351–361.
- Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O. 2011. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol.* 60(5):685–699.
- Applequist WL, Wallace RS. 2000. Phylogeny of the Madagascan endemic family Didiereaceae. *Plant Syst Evol.* 221(3–4):157–166.
- Applequist WL, Wallace RS. 2003. Expanded circumscription of Didiereaceae and its division into three subfamilies. *Adansonia* 25(1):13–16.
- Arakaki M, Christin PA, Nyffeler R, Lendel A, Eggli U, Ogburn RM, Spriggs E, Moore MJ, Edwards EJ. 2011. Contemporaneous and recent radiations of the world's major succulent plant lineages. *Proc Natl Acad Sci U S A.* 108(20):8379–8384.
- Bárcenas RT, Yesson C, Hawkins JA. 2011. Molecular systematics of the Cactaceae. *Cladistics* 27(5):470–489.
- Barthlott W, Hunt DR. 1993. Cactaceae. In: Flowering plants dicotyledons. Berlin/Heidelberg (Germany): Springer. p. 161–197.
- Borland AM, Griffiths H, Hartwell J, Smith JA. 2009. Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. *J Exp Bot.* 60(10):2879–2896.
- Brandt B, Brodsky DE, Xue S, Negi J, Iba K, Kangasjärvi J, Ghassemian M, Stephan AB, Hu H, Schroeder JI. 2012. Reconstitution of abscisic acid activation of SLAC1 anion channel by CPK6 and OST1 kinases and branched ABI1 PP2C phosphatase action. *Proc Natl Acad Sci USA.* 109(26):10593–10598.
- Brochmann C, Brysting AK, Alsos IG, Borgen L, Grundt HH, Scheen AC, Elven R. 2004. Polyploidy in arctic plants. *Bio J Linnean Soc.* 82(4):521–536.
- Brockington SF, Yang Y, Gandia-Herrero F, Covshoff S, Hibberd JM, Sage RF, Wong GK, Moore MJ, Smith SA. 2015. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytol.* 207(4):1170–1180.
- Brown JM, Thomson RC. 2017. Bayes factors unmask highly variable information content, bias, and extreme influence in phylogenomic analyses. *Syst Biol.* 66(4):517–530.
- Brown JW, Walker JF, Smith SA. 2017. Phyx: phylogenetic tools for unix. *Bioinformatics* 33(12):1886–1888.
- Bruyns PV, Oliveira-Neto M, Melo-de-Pinna GF, Klak C. 2014. Phylogenetic relationships in the Didiereaceae with special reference to subfamily Portulacarioideae. *Taxon* 63(5):1053–1064.
- Butterworth CA, Wallace RS. 2005. Molecular phylogenetics of the leafy cactus genus *Pereskia* (Cactaceae). *Syst Bot.* 30(4):800–808.
- Butterworth CA, Cota-Sanchez JH, Wallace RS. 2002. Molecular systematics of tribe Cacteae (Cactaceae: cactoideae): a phylogeny based on rpl16 intron sequence variation. *Syst Bot.* 27(2):257–270.
- Chan KX, Wirtz M, Phua SY, Estavillo GM, Pogson BJ. 2013. Balancing metabolites in drought: the sulfur assimilation conundrum. *Trends Plant Sci.* 18(1):18–29.
- Christin PA, Arakaki M, Osborne CP, Bräutigam A, Sage RF, Hibberd JM, Kelly S, Covshoff S, Wong GK, Hancock L, et al. 2014. Shared origins of a key enzyme during the evolution of C4 and CAM metabolism. *J Exp Bot.* 65(13):3609–3621.
- Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr Biol.* 17(14):1241–1247.
- Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A, et al. 2006. Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 16(6):738–749.
- Cushman JC. 2001. Crassulacean acid metabolism. A plastic photosynthetic adaptation to arid environments. *Plant Physiol.* 127(4):1439–1448.
- Edwards EJ, Nyffeler R, Donoghue MJ. 2005. Basal cactus phylogeny: implications of *Pereskia* (Cactaceae) paraphyly for the transition to the cactus life form. *Am J Bot.* 92(7):1177–1188.
- Edwards EJ, Ogburn RM. 2012. Angiosperm responses to a low-CO₂ world: CAM and C4 photosynthesis as parallel evolutionary trajectories. *Int J Plant Sci.* 173(6):724–733.
- Edwards G, Walker D. 1983. C3, C4: mechanisms, and cellular and environmental regulation, of photosynthesis. Oakland, California, USA: University of California Press.
- Emms DM, Covshoff S, Hibberd JM, Kelly S. 2016. Independent and parallel evolution of new genes by gene duplication in two origins of C4 photosynthesis provides new insight into the mechanism of phloem loading in C4 species. *Mol Biol Evol.* 33(7):1796–1806.
- Ferreira MLF, Andreo CS, Podestá FE. 2003. Purification and physical and kinetic characterization of a photosynthetic NADP-dependent malic enzyme from the CAM plant *Aptenia cordifolia*. *Plant Sci.* 164(1):95–102.
- Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23):3150–3152.
- Geiger D, Scherzer S, Mumm P, Marten I, Ache P, Matschi S, Liese A, Wellmann C, Al-Rasheid KA, Grill E, et al. 2010. Guard cell anion channel SLAC1 is regulated by CDPK protein kinases with distinct Ca²⁺ affinities. *Proc Natl Acad Sci USA.* 107(17):8023–8028.
- Gibson AC. 1996. Structure–function relations of warm desert plants. New York: Springer.
- Gibson AC, Nobel PS. 1990. The cactus primer. Cambridge, Massachusetts, USA: Harvard University Press.
- Gil-Mascarell R, López-Coronado JM, Bellés JM, Serrano R, Rodríguez PL. 1999. The *Arabidopsis* HAL2 like gene family includes a novel sodium-sensitive phosphatase. *Plant J.* 17(4):373–383.
- Golldack D, Li C, Mohan H, Probst N. 2014. Tolerance to drought and salt stress in plants: unraveling the signaling networks. *Front Plant Sci.* 5:151.
- Goolsby EW, Moore AJ, Hancock LP, De Vos JM, Edwards EJ. 2018. Molecular evolution of key metabolic genes during transitions to C4 and CAM photosynthesis. *Am J Bot.* 105(3):602–613.
- Gosti F, Beaudoin N, Serizet C, Webb AA, Vartanian N, Giraudat J. 1999. ABI1 protein phosphatase 2C is a negative regulator of abscisic acid signaling. *Plant Cell.* 11(10):1897–1909.
- Griffith MP, Porter JM. 2009. Phylogeny of Opuntioideae (Cactaceae). *Int J Plant Sci.* 170(1):107–116.
- Group AP. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: aPG IV. *Bot J Linnean Soc.* 181(1):1–20.
- Hernández-Hernández T, Brown JW, Schlumberger BO, Eguiarte LE, Magallón S. 2014. Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytol.* 202(4):1382–1397.
- Hernández-Hernández T, Hernández HM, De-Nova JA, Puente R, Eguiarte LE, Magallón S. 2011. Phylogenetic relationships and evolution of growth form in Cactaceae (Caryophyllales, Eudicotyledoneae). *Am J Bot.* 98(1):44–61.
- Hernández-Ledesma P, Berendsohn WG, Borsch T, Mering SV, Akhiani H, Arias S, Castañeda-Noa I, Eggli U, Eriksson R, Flores-Olvera H, et al. 2015. A taxonomic backbone for the global synthesis of species diversity in the angiosperm order Caryophyllales. *Willdenowia* 45(3):281–383.
- Hochstrasser M. 2009. Origin and function of ubiquitin-like proteins. *Nature* 458(7237):422–429.
- Huang Q, Wang Y, Li B, Chang J, Chen M, Li K, Yang G, He G. 2015. TaNAC29, a NAC transcription factor from wheat, enhances salt and drought tolerance in transgenic *Arabidopsis*. *BMC Plant Biol.* 15(1):268.
- Hubbard KE, Nishimura N, Hitomi K, Getzoff ED, Schroeder JI. 2010. Early abscisic acid signal transduction mechanisms: newly discovered components and newly emerging questions. *Genes Dev.* 24(16):1695–1708.
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473(7345):97–100.

- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Kiang JG, Tsokos GC. 1998. Heat shock protein 70 kDa: molecular biology, biochemistry, and physiology. *Pharmacol Ther.* 80(2):183–201.
- Kreps JA, Wu Y, Chang HS, Zhu T, Wang X, Harper JF. 2002. Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. *Plant Physiol.* 130(4):2129–2141.
- Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. *Am J Bot.* 105(3):348–363.
- Lee EK, Cibrian-Jaramillo A, Kolokotronis S-O, Katari MS, Stamatakis A, Ott M, Chiu JC, Little DP, Stevenson DW, McCombie WR, et al. 2011. A functional phylogenomic view of the seed plants. *PLoS Genet.* 7(12):e1002411.
- Li H, Chang J, Zheng J, Dong Y, Liu Q, Yang X, Wei C, Zhang Y, Ma J, Zhang X. 2017. Local melatonin application induces cold tolerance in distant organs of *Citrullus lanatus* L. via long distance transport. *Sci Rep.* 7:40858.
- Liu L, Pearl DK. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Syst Biol.* 56(3):504–514.
- Liu L, Pearl DK, Brumfield RT, Edwards SV. 2008. Estimating species trees using multiple-allele DNA sequence data. *Evolution* 62(8):2080–2091.
- Löytynoja A, Goldman N. 2008. Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* 320(5883):1632–1635.
- Mauseth JD. 2006. Structure–function relationships in highly modified shoots of Cactaceae. *Ann Bot.* 98(5):901–926.
- Merlot S, Gosti F, Guerrier D, Vavasseur A, Giraudat J. 2001. The ABI1 and ABI2 protein phosphatases 2C act in a negative feedback regulatory loop of the abscisic acid signalling pathway. *Plant J.* 25(3):295–303.
- Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* 30(17):i541–i548.
- Mishra G, Zhang W, Deng F, Zhao J, Wang X. 2006. A bifurcating pathway directs abscisic acid effects on stomatal closure and opening in *Arabidopsis*. *Science* 312(5771):264–266.
- Moore A, de Vos JM, Hancock LP, Goolsby E, Edwards EJ. 2018. Targeted enrichment of large gene families for phylogenetic inference: phylogeny and molecular evolution of photosynthesis genes in the Portulacaceae (Caryophyllales). *Syst Biol.* 67(3):367–383.
- Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. 2012. NAC transcription factors in plant abiotic stress responses. *Biochim Biophys Acta.* 1819(2):97–103.
- Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K. 2014. The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front Plant Sci.* 5:170.
- Nyffeler R. 2002. Phylogenetic relationships in the cactus family (Cactaceae) based on evidence from trnK/matK and trnL-trnF sequences. *Am J Bot.* 89(2):312–326.
- Nyffeler R. 2007. The closest relatives of cacti: insights from phylogenetic analyses of chloroplast and mitochondrial sequences with special emphasis on relationships in the tribe Anacampseroteae. *Am J Bot.* 94(1):89–101.
- Nyffeler R, Eggli U. 2010a. Disintegrating Portulacaceae: a new familial classification of the suborder Portulacineae (Caryophyllales) based on molecular and morphological data. *Taxon* 59(1):227–240.
- Nyffeler R, Eggli U. 2010b. A farewell to dated ideas and concepts – molecular phylogenetics and a revised suprageneric classification of the family Cactaceae. *Schumannia* 6:109–149.
- Ocampo G, Columbus JT. 2010. Molecular phylogenetics of suborder Cactineae (Caryophyllales), including insights into photosynthetic diversification and historical biogeography. *Am J Bot.* 97(11):1827–1847.
- Ogburn RM, Edwards EJ. 2015. Life history lability underlies rapid climate niche evolution in the angiosperm clade Montiaceae. *Mol Phylogenet Evol.* 92:181–192.
- Parks MB, Nakov T, Ruck EC, Wickett NJ, Alverson AJ. 2018. Phylogenomics reveals an extensive history of genome duplication in diatoms (Bacillariophyta). *Am J Bot.* 105(3):330–347.
- Pease JB, Hahn MW. 2015. Detection and polarization of introgression in a five-taxon phylogeny. *Syst Biol.* 64(4):651–662.
- Peel MC, Finlayson BL, McMahon TA. 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrol Earth Syst Sci.* 11(5):1633–1644.
- Pennell MW, Eastman JM, Slater GJ, Brown JW, Uyeda JC, FitzJohn RG, Alfaro ME, Harmon LJ. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30(15):2216–2218.
- Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. *Mol Biol Evol.* 21(7):1455–1458.
- Pond SLK, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. In: *Statistical methods in molecular evolution*. New York: Springer. p. 125–181.
- Qin F, Sakuma Y, Li J, Liu Q, Li YQ, Shinozaki K, Yamaguchi-Shinozaki K. 2004. Cloning and functional analysis of a novel DREB1/CBF transcription factor involved in cold-responsive gene expression in *Zea mays* L. *Plant Cell Physiol.* 45(8):1042–1052.
- Schmidt A. 2005. Metabolic background of H₂S release from plants. In: De Kok LJ, Chnug ES, editors. *Proceedings of the 1st Sino-German Workshop on Aspects of Sulfur Nutrition of Plants*, 17th End, Braunschweig, Germany. p. 121–129.
- Seo TK. 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Mol Biol Evol.* 25(5):960–971.
- Shen X, Hittinger CT, Rokas A. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat Ecol Evol.* 1:1–10.
- Silvera K, Winter K, Rodriguez BL, Albion RL, Cushman JC. 2014. Multiple isoforms of phosphoenol-pyruvate carboxylase in the Orchidaceae (subtribe Oncidiinae): implications for the evolution of crassulacean acid metabolism. *J Exp Bot.* 65(13):3623–3636.
- Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS One.* 13(5):e0197433.
- Smith SA, Brown JW, Yang Y, et al. 2018. Disparity, diversity, and duplications in the Caryophyllales. *New Phytol.* 217(2):836–854.
- Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24(5):715–716.
- Smith MD, Wertheim JO, et al. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol.* 32(5):1342–1353.
- Smith SA, Moore MJ, et al. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evol Biol.* 15(1):150.
- Smith SA, O'Meara BC. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28(20):2689–2690.
- Smith SA, Walker JF. 2018. PyPHLAWD: a python tool for phylogenetic dataset construction. *Methods Ecol Evol.* doi:10.1111/2041-210X.13096.
- Smith JAC, Winter K. 1996. Taxonomic distribution of crassulacean acid metabolism. In: Winter K, Smith JAC, editors. *Crassulacean acid metabolism: biochemistry, ecophysiology and evolution*. Berlin/Heidelberg (Germany): Springer-Verlag. p. 427–436.
- Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annu Rev Plant Biol.* 60:561–588.
- Soltis PS, Soltis DS. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proc Natl Acad Sci USA.* 97(13):7051–7057.
- Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB, Moore MJ, Carlswald BS, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *Am J Bot.* 98(4):704–730.

- Stamatakis A. 2015. Using RAxML to infer phylogenies. *Curr Protoc Bioinformatics*. 51:6–14.
- Stebbins GL. 1971. Chromosomal evolution in higher plants. London: Addison Wesley. p. 216.
- Stoughton TR, Jolles DD, O'Quinn RL. 2017. The western spring beauties, *Claytonia lanceolata* (Montiaceae): a review and revised taxonomy for California. *Syst Bot*. 42(2):283–300.
- Than C, Ruths D, Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics*. 9(1):322.
- van den Bergh E, K lahoglu C, Br utigam A, Hibberd JM, Weber AP, Zhu XG, Schranz ME. 2014. Gene and genome duplications and the origin of C4 photosynthesis: birth of a trait in the Cleomaceae. *Curr Plant Biol*. 1:2–9.
- Vandeleur RK, Mayo G, Shelden MC, Gilliam M, Kaiser BN, Tyerman SD. 2009. The role of plasma membrane intrinsic protein aquaporins in water transport through roots: diurnal and drought stress responses reveal different strategies between isohydric and anisohydric cultivars of grapevine. *Plant Physiol*. 149(1):445–460.
- Vanneste K, Van de Peer Y, Maere S. 2013. Inference of genome duplications from age distributions revisited. *Mol Biol Evol*. 30(1):177–190.
- V zquez-S nchez M, Terrazas T, Arias S, Ochoterena H. 2013. Molecular phylogeny, origin and taxonomic implications of the tribe Cacteeae (Cactaceae). *Sys Biodivers*. 11(1):103–116.
- Walker JF, Yang Y, Moore MJ, Mikenas J, Timoneda A, Brockington SF, Smith SA. 2017. Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. *Am J Bot*. 104(6):858–867.
- Walker JF, Brown JW, et al. 2018. Analyzing contentious relationships and outlier genes in phylogenomics. *Syst Biol*. 67(5):916–924.
- Walker JF, Yang Y, et al. 2018. From cacti to carnivores: improved phylo-transcriptomic sampling and hierarchical homology inference provide further insight to the evolution of Caryophyllales. *Am J Bot*. 105(3):446–462.
- Wallace RS, Dickie SL. 2002. Systematic implications of chloroplast DNA sequence variation in subfam. Opuntioideae (Cactaceae). *Succ Pl Res*. 6:9–24.
- Watson A. 1979. Gypsum crusts in deserts. *J Arid Environ*. 2(1):3–21.
- Wendel JF. 2015. The wondrous cycles of polyploidy in plants. *Am J Bot*. 102(11):1753–1756.
- Williams BP, Aubry S, Hibberd JM. 2012. Molecular evolution of genes recruited into C4 photosynthesis. *Trends Plant Sci*. 17(4):213–220.
- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci USA*. 106(33):13875–13879.
- 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.
- Yang Y, Moore MJ, Brockington SF, Soltis DE, Wong GKS, Carpenter EJ, Zhang Y, Chen L, Yan Z, Xie Y, et al. 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Mol Biol Evol*. 32(8):2001–2014.
- Yang Y, Moore MJ, Brockington SF, Timoneda A, Feng T, Marx HE, Walker JF, Smith SA. 2017. An efficient field and laboratory workflow for plant phylotranscriptomic projects. *Appl Plant Sci*. 5(3):1600128.
- Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Mol Biol Evol*. 31(11):3081–3092.