Check for updates

Received: 26 January 2024 | Accepted: 26 April 2024

DOI: 10.1002/ajb2.16350

American Journal of Botany

SPECIAL ISSUE ARTICLE

The link between ancient whole-genome duplications and cold adaptations in the Caryophyllaceae

Keyi Feng[1] | **Joseph F. Walker**[2] | **Hannah E. Marx**[3] | **Ya Yang**[4] | **Samuel F. Brockington**[5] | **Michael J. Moore**[6] | **Richard K. Rabeler**[1] | **Stephen A. Smith**[1]

[1]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA

[2]Department of Biological Sciences, University of Illinois Chicago, Chicago, IL 60607, USA

[3]Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA

[4]Department of Plant and Microbial Biology, University of Minnesota-Twin Cities, St. Paul, MN 55108, USA

[5]Department of Plant Sciences, University of Cambridge, Cambridge CB2 1TN, UK

[6]Department of Biology, Oberlin College, Oberlin, OH 44074, USA

Correspondence

Keyi Feng and Stephen A. Smith, Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109, USA.
Email: fky@umich.edu and eebsmith@umich.edu

Abstract

Premise: The Caryophyllaceae (the carnation family) have undergone multiple transitions into colder climates and convergence on cushion plant adaptation, indicating that they may provide a natural system for cold adaptation research. Previous research has suggested that putative ancient whole-genome duplications (WGDs) are correlated with niche shifts into colder climates across the Caryophyllales. Here, we explored the genomic changes potentially involved in one of these discovered shifts in the Caryophyllaceae.

Methods: We constructed a data set combining 26 newly generated transcriptomes with 45 published transcriptomes, including 11 cushion plant species across seven genera. With this data set, we inferred a dated phylogeny for the Caryophyllaceae and mapped ancient WGDs and gene duplications onto the phylogeny. We also examined functional groups enriched for gene duplications related to the climatic shift.

Results: The ASTRAL topology was mostly congruent with the current consensus of relationships within the family. We inferred 15 putative ancient WGDs in the family, including eight that have not been previously published. The oldest ancient WGD (ca. 64.4–56.7 million years ago), WGD1, was found to be associated with a shift into colder climates by previous research. Gene regions associated with ubiquitination were overrepresented in gene duplications retained after WGD1 and those convergently retained by cushion plants in *Colobanthus* and *Eremogone*, along with other functional annotations.

Conclusions: Gene family expansions induced by ancient WGDs may have contributed to the shifts to cold climatic niches in the Caryophyllaceae. Transcriptomic data are crucial resources that help unravel heterogeneity in deep-time evolutionary patterns in plants.

KEYWORDS

convergent evolution, cushion plants, gene duplication, gene ontology, paleopolypoidy, phylogenomics, phylogeny, polyploidy, transcriptome

Angiosperms likely originated in mesic tropical forests (Feild et al., 2004; Chaboureau et al., 2014; Willis and McElwain, 2014). The cold (i.e., freezing) biomes currently prevalent at higher latitudes and altitudes have only become increasingly widespread since the start of the Cenozoic (Willis and McElwain, 2014). The cooling and drying trend over the past 66 million years (Myr) led to repeated adaptations to cold climates across angiosperm lineages

---

This article is part of joint special issues of the *American Journal of Botany* and *Applications in Plant Sciences*: “Twice as Nice: New Techniques and Discoveries in Polyploid Biology.”

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Author(s). *American Journal of Botany* published by Wiley Periodicals LLC on behalf of Botanical Society of America.

(Willis and McElwain, 2014; Zanne et al., 2014; Folk et al., 2020). Numerous studies have shed light on the molecular, physiological, and morphological modifications behind plant cold adaptations. Some well-known examples of these modifications are osmolyte synthesis to prevent frost damage, life history changes, and cushion-forming phenotype (Chinnusamy et al., 2004; Beck et al., 2007; Preston and Sandve, 2013; Zanne et al., 2014; Körner, 2016; Folk et al., 2020). However, questions remain regarding the potential macroevolutionary enablers of these modifications.

Ancient whole-genome duplications (WGDs) are common throughout the evolutionary history of angiosperms and are significant contributors to trait innovations and climatic niche shifts (Soltis et al., 2009; te Beest et al., 2012; Moghe and Shiu, 2014; Brockington et al., 2015; Smith et al., 2018b; Baniaga et al., 2020; Wu et al., 2020). The increase in gene copies has been hypothesized to increase the potential for adaptation (Ohno, 1970; Zhang, 2003). Still, the connection between ancient WGDs and cold adaptations in angiosperms, whether causal or correlative, remains an active area of research (Wu et al., 2020; Zhang et al., 2020).

In the last 15 years, the influx of genome-scale data has enabled botanists to further investigate complex deep-time evolutionary questions from functional and molecular angles in addition to a phylogenetic perspective. For example, genomic and transcriptomic data have been used to infer the phylogenetic placement of ancient WGDs (Jiao et al., 2011; Walker et al., 2017; Yang et al., 2018; Wang et al., 2019; Wu et al., 2020). Patterns of gene family evolution from genomic and transcriptomic data sets have provided information on molecular changes underline the trait evolution (Brockington et al., 2015; Walker et al., 2018b; Wang et al., 2019; Wu et al., 2020). Transcriptome data generate long contigs that enable phylogeny-informed methods for distinguishing paralogs. The tens of thousands of loci provided by these data sets have allowed for improved phylogenetic reconstruction and, more importantly, investigations into gene tree conflicts, gene duplications, and losses, and their underlying abiotic and biological causes (Shen et al., 2017; Walker et al., 2018a; Wang et al., 2019).

The family Caryophyllaceae (the carnations) contains 3300 currently accepted species in 101 genera with an estimated crown age of ca. 80 Myr (Magallón et al., 2015; POWO, 2022). Members of the family have diverse floral morphologies, often with clawed and limbed petals. Several species are popular for horticulture (e.g., *Dianthus* species). Until recently, three subfamilies (Paronychioideae, Alsinoideae, and Caryophylloideae) were traditionally recognized (e.g., Bittrich, 1993). Molecular sampling across the family revealed the subfamilies are not monophyletic, which led Harbaugh et al. (2010) to propose that relationships within the Caryophyllaceae are more accurately reflected as 11 tribes. This alignment has been supported by subsequent studies (e.g., Greenberg and Donoghue, 2011; Sadeghian et al., 2015). While most of the diversity in the family is found in the 11 tribes, a twelfth tribe, Thylacospermeae, was proposed to segregate the two species of *Thylacospermum* from tribe Eremogoneae based on morphological and phylogenetic evidence (Pusalkar, 2015). However, support for the phylogenetic placement of this tribe remains ambiguous (Xu et al., 2019).

Most species of Caryophyllaceae are distributed in dry and cold regions in the northern hemisphere. Notably, the Caryophyllaceae is the only eudicot family with a distribution spanning all seven continents. The Caryophyllaceae comprise annuals, herbaceous perennials, and woody perennials (subshrubs). Most of the subshrubs consist of herbaceous stems with a woody base, while some exceptions are mostly woody, such as cushion-forming plants. The cushion growth form has a spherical shape to retain heat and moisture (Aubert et al., 2014) and is associated with the arctic-alpine climate, defined as an average growing season temperature lower than 10°C (Körner, 2011; Boucher et al., 2016; Birkeland et al., 2020). The cushion habit has originated at least 115 times independently among flowering plants, and cushion lineages are phylogenetically clustered, suggesting potential physiological prerequisites for this adaptation (Boucher et al., 2016).

The Caryophyllaceae is well suited for investigating the link between cold climatic shifts and ancient WGDs. Smith et al. (2018b) inferred associations between ancient WGDs and shifts into lower mean annual temperature (MAT) in the Caryophyllales, and one of the associations is within the Caryophyllaceae. Furthermore, the carnation family has multiple shifts to the arctic-alpine biome (Harbaugh et al., 2010; Greenberg and Donoghue, 2011) and contains 164 cushion-forming species among 20 of the 101 genera (Aubert et al., 2014; Boucher et al., 2016). The growth form evolved approximately 16 times independently within the family, displaying remarkable morphological and physiological convergence. The Caryophyllaceae is also rich in genome duplication events, with 11 putative events detected by previous phylogenetic studies (Frajman et al., 2018; Yang et al., 2018; Eroğlu et al., 2020). Further investigation into the molecular evolution of this clade is valuable for a deeper understanding of the connection between genome duplications and cold adaptations.

This study combined 26 newly generated transcriptomes in Caryophyllaceae with 45 previously published transcriptomes. The data set included 11 cushion plant species representing at least seven independent origins of cushion growth form. We used these data to explore the hypothesized association between the shift toward colder climates and ancient WGD in the Caryophyllaceae from a genomic angle (Smith et al., 2018b). We aimed to (1) reconstruct the backbone relationships within the Caryophyllaceae and examine gene tree conflict, (2) identify putative ancient WGDs and gene duplications, and (3) explore the relationship between gene duplications, shifts toward cold climate, and associated life form changes. Because similar gene copies in de novo assembled transcriptome data cannot be distinguished, we used "ancient WGD" to refer to older duplication events that show a significantly elevated number of assembled paralogs.

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

MATERIALS AND METHODS

Taxon sampling and transcriptome generation

We sampled transcriptomes from 59 ingroup Caryophyllaceae species, representing 34 of 101 genera and 11 of the 12 tribes (without Thylacospermeae), including 11 cushion-forming species across seven genera. We included 12 species spanning the families Polygonaceae, Montiaceae, Achatocarpaceae, and Amaranthaceae as outgroups. Then, we cross-referenced the species names with the World Checklist of Vascular Plants (WCVP data set downloaded February 2022; POWO, 2022) to ensure taxonomic accuracy. In addition, the currently accepted name for *Arenaria globiflora* (Fenzl) Wall. ex Edgew. & Hook.f. is *Dolophragma globiflorum* Fenzl (Hernández-Ledesma et al., 2015; Sadeghian et al., 2015), which is not an accepted scientific name in the WCVP data set. Of the 71 assembled transcriptomes, we obtained 45 from a previous study (Walker et al., 2018b) and newly sequenced 26 using the procedures for phylotranscriptomics of Yang et al. (2017). GENEWIZ (South Plainfield, NJ, USA) prepared libraries and sequenced RNA. We used total RNA with poly-A selection (unstranded) for library preparation with the NEBNext Ultra RNA library prep kit for Illumina (NEB, Ipswich, MA, USA) according to the instructions. Sequencing was multiplexed over three lanes on an Illumina HiSeq. 4000 platform for 150-bp paired-end reads with a single index. Detailed information for the transcriptome data sets can be found in Appendix S1 and appendices of Walker et al. (2018).

Phylotranscriptomic data set assembly

We used the pipeline described by Yang and Smith (2014) for the phylotranscriptomic data set assembly. In short, we cleaned and processed the raw sequencing reads by first removing adapters. We identified random sequencing errors with Rcorrector v1.0.4 (Song and Florea, 2015) and removed reads determined to be unfixable using the unfixable_filter.py (Morales-Briones et al., 2021). Then, we trimmed reads with low-quality scores with Trimmomatic v0.39 (Bolger et al., 2014). We used Bowtie2 v2.4.4 (Langmead and Salzberg, 2012) to map the remaining reads against a database of organellar genomes within the Caryophyllaceae and to remove organellar (chloroplast and mitochondrial) reads from the data set. We utilized FastQC v0.11.9 (Andrews, 2010) to remove overrepresented reads (Yang and Smith, 2013, 2014).

Next, we used the cleaned and filtered reads for de novo transcriptome assembly with Trinity v2.5.1 (Grabherr et al., 2011). We employed Transrate v1.0.3 (Smith-Unna et al., 2016) to assess the quality of each assembly and remove low-quality transcripts. We detected and removed chimeras using a BLAST-based method described by Yang and Smith (2013). Subsequently, we performed clustering of transcripts from the same putative gene by mapping the reads to the assembly with Salmon v0.9.1 (Patro et al., 2017) and filtering for the best-supported transcript using Corset v1.09 (Davidson and Oshlack, 2014). We then inferred open reading frames from the filtered transcripts using TransDecoder v5.3.0 (Haas, 2018) guided by a custom blast database of high-quality proteomes for *Arabidopsis thaliana* (L.) Heynh. and *Beta vulgaris* L. (Dohm et al., 2012).

We used the coding sequence (CDS) output files from each transcriptome for homology search and utilized cd-hit-est v4.8.1 (Fu et al., 2012) (-c 0.99 -n 10 -r 0 for positive strands) to reduce the redundancy in each CDS data set. We identified initial homology using all-by-all BLASTN v2.12.0 with max targets of 1000. We filtered BLASTN outputs by hit fractions of 0.4 and clustered filtered hits with MCL v14-137 (Van Dongen, 2000) and an inflation value of 2. Then, we aligned homolog clusters with less than 1000 sequences with MAFFT v7.149b with the settings "--genafpair --maxiterate 1000" (Katoh and Standley, 2013), and inferred maximum likelihood gene trees using RAxML 8.2.12 (Stamatakis, 2014) with a rapid bootstrap search. We aligned clusters with 1000 sequences or more with pasta v.1.8.2 (Mirarab et al., 2015) and used fasttree v2.1.7 (Price et al., 2010) for gene tree inference.

Homolog cleaning and trimming

We removed potential errors in assembly and homology inference through a procedure that involves removing spurious branches of a specified length (absolute cutoff) and branches that are longer than a fixed value and at least 10× greater in length than the sister branch (relative length). The specific cutoff values for each round are stated below. We removed alternative splice sites, isoforms, and phylogenetically uninformative in-paralogs, retaining the single most character-rich tip on a clade containing only a single taxon. We separated clades connected by a branch longer than a specified length and containing at least four taxa into two homolog trees, which indicates misidentified or remote homology. We performed these procedures for four rounds, examining the number of homologs that remained after each pass to identify when the cleaning procedure no longer cleaned the homologs.

In the first round, we trimmed the 70,220 homolog trees using a relative tip cutoff of 0.4 and an absolute tip cutoff of 0.5. Then, we separated clades based on 1.0 substitution (subs) per base pair for a connecting branch. For round 2, we re-aligned the fasta sequences corresponding to the resulting 69,322 clades from the first round using mafft v7.490 (Katoh and Standley, 2013) with the settings "—auto –maxiterate 1000" and cleaned for alignment occupancy of at least 10% using the phyx v1.2 program pxclsq (Brown et al., 2017). We inferred maximum likelihood homolog trees using IQtree v1.6.12 with the GTR model of evolution and gamma rate variation. Next, we trimmed the homolog trees using a relative tip cutoff of 0.3 and an absolute tip

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

cutoff of 0.5 and separated clades based on a branch length of 1.0 subs/bp and four taxa on either side. The result was 68,153 homolog trees for downstream analysis. We performed the same procedure as round 2 in round 3; however, we masked the monophyletic tips after trimming absolute and relative branch lengths. The third round produced 68,113 homolog trees, and then we performed a fourth round of cleaning following the same procedure as the third round.

We extracted the rooted gene duplication families from the homolog trees using the extract_clades.py script from Yang and Smith (2014), specifying a minimum of 10 taxa and the 5 species in the family Polygonaceae as outgroups. The extraction procedure resulted in 4137 rooted gene family trees.

Phylogeny and conflict inference

We extracted orthologs using the rooted tree (RT) procedure, designed for large numbers of gene duplications (Yang and Smith, 2014). The extraction procedure resulted in 9051 ortholog trees for downstream analysis. We specified the five species in Polygonaceae as outgroups. Based on these outgroups, the orthology identification procedure cut out and rooted the ortholog trees. Thus, the resulting ortholog trees did not contain the specified outgroups. We re-aligned the ortholog sequences using Prank v.170427 with default settings (Löytynoja, 2014). We removed all columns with less than 30% occupancy from the alignments using the phyx v1.2 program pxclsq (Brown et al., 2017). Eight of the 9051 ortholog sequences were not usable for phylogenetic analysis because they contained insufficient variable sites (parsimony informative sites). We used the 9043 remaining cleaned alignments to infer ortholog gene trees using maximum likelihood as implemented in IQtree, with the GTR + G model of evolution and 1000 ultrafast bootstrap (UFBoot) replicates (Hoang et al., 2018). We then utilized ASTRAL v.5.7.8 (Zhang et al., 2018) to infer a coalescent-based maximum support phylogeny from the resulting gene trees.

To infer a dated phylogeny, we re-estimated the branch lengths of the ASTRAL tree with a supermatrix of 50 ortholog alignments instead of including all extracted orthologs. We selected 50 best orthologs via SortaDate v2 (Smith et al., 2018a). The chosen orthologs have the highest taxon sampling, the most consistent root-to-tip distance, and the least conflict with the ASTRAL topology in the data set, so they are the most suitable genes for divergence time estimation. Next, we re-estimated branch lengths of the ASTRAL topology with a concatenated supermatrix of the 50 orthologs using RaxML-NG v1.2.0 GTR + G model (Kozlov et al., 2019). We also tested the sensitivity of branch length re-estimation with the best 20, 30, and 40 orthologs, finding no significant difference in branch lengths across these analyses. Thus, only using 50 orthologs for branch length estimation should not cast any bias on dating results.

We performed dating using penalized likelihood as implemented in treePL (Smith and O'Meara, 2012). We used a Late Campanian pollen fossil *Periporopollenites polyoratus* (Stover and Partridge, 1973) for constraining the minimum crown age of the Caryophyllaceae (72.1 Myr) and the 95% HPD maximum age (88.56 Myr) from Magallón et al. (2015) as the maximum crown age of the family. Also, we included the Middle-Late Eocene inflorescence fossil *Caryophylloflora paleogenica* (Jordan and Macphail, 2003) for constraining the split between former subfamilies Alsinoideae and Caryophylloideae at a minimum age of 33.9 Myr. This split is the most recent common ancestor (MRCA) of seven currently recognized tribes: Eremogoneae, Sileneae, Caryophylleae, Alsineae, Arenarieae, Sagineae, and Sclerantheae. We set the node's maximum constraint to the Caryophyllaceae's minimum crown age (72.1 Myr). treePL only utilizes the constraints when the estimated dates are out of the range. Thus, although the minimum age of *Caryophylloflora paleogenica* is probably much younger than the appointed node, it would not bias the estimation to a younger age. In addition, treePL only provides point estimates for dates, so we performed bootstrap analysis to infer uncertainty. We generated 1000 bootstrap supermatrix alignments for the 50 orthologs with RaxML-NG v1.2.0 (Kozlov et al., 2019). Then, we performed branch length re-estimation and dating in the same way as the original supermatrix for the 1000 bootstrap replicates. We summarized the 1000 dated trees as the confidence intervals for the node ages.

We inferred gene tree conflict inference using the program bellerophon (Smith et al., 2020, https://git.sr.ht/~hms/bellerophon), which is a time-efficient tool to analyze bipartitions for large data sets. We utilized the 043 ortholog trees and inferred the percentage of concordant and conflict of those gene trees with the ASTRAL topology. We discarded bipartitions with lower than 90% bootstrap support in the analyses. To further explore the complex conflicts, we conducted alternative relationship likelihood analysis (Smith et al., 2020; phyckle: https://github.com/FePhyFoFum/phyckle) and quartet sampling (QS) analysis (Pease et al., 2018). We generated phylogenetic figures, including those with conflict pie charts, with the tree-plotting software gokstad (https://git.sr.ht/~hms/gokstad) and edited them in Illustrator (Adobe, San Jose, CA, USA).

We checked for contamination and misplaced taxa by examining ortholog trees, alignments, and conflict analysis results, and we decided to remove three newly sequenced samples from the study. The sample for *Spergula arvensis* L. was placed in *Silene* in the ASTRAL topology instead of in Sperguleae; the sample for *Silene davidii* (Franch.) Oxelman & Lidén was frequently placed in the genus *Dianthus* in the ortholog gene trees, often sister to *Dianthus caryophyllus* L. Judging from the gene trees and alignments, we were unable to determine the source of gene tree conflicts between contamination during library prep/sequencing or introgression. Although exploring the

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

possibility of potential introgression (Frost et al., 2024) would be interesting, it is outside the scope of this study, and there is no previous support for the possibility of introgression in the two species.

The sample for *Claytonia perfoliata* Donn ex Willd. was frequently placed within Caryophyllaceae in the ortholog trees without a dominant alternative and did not seem to be contaminated based on the alignments. Its sequence length in the ortholog alignments was often much shorter than the other species in the same cluster. The *C. perfoliata* sample also had the lowest QF score from the QS analysis, meaning that its placement was only concordant with the ASTRAL tree 54% of the time. Therefore, *C. perfoliata* was likely a "rogue taxon" (Wilkinson, 1996; Aberer et al., 2014), but we could not determine the cause of this erroneous placement. Thus, we removed *Spergula arvensis*, *Silene davidii*, and *Claytonia perfoliata* from the ortholog gene trees and re-performed all the downstream analyses with the remaining 57 ingroups and six outgroups.

Ancient whole-genome duplications (WGDs), gene duplications, and gene functions

Previous studies (Yang et al., 2015; Tiley et al., 2018) showed that analysis of transcriptomes recovered approximately one third of genes supporting nodes with ancient WGDs, similar to the proportions recovered from whole-genome sequencing data and successfully recovered all ancient WGDs recovered from whole genome sequencing. We inferred putative ancient WGDs from phylogenetic locations of elevated gene duplications, peaks in Ks (synonymous distance between paralogous gene pairs) distribution, and published chromosome counts. As ancient WGDs were inferred from transcriptome data, we define them as duplication events that produce a highly elevated number of paralogs that can be distinguished from de novo assembled transcriptome data. We generated Ks plots for each species following Yang et al. (2015) for Ks distributions, with Nei-Gojobori correction for multiple substitutions. We identified gene duplication events from the 4137 rooted gene family trees and mapped them onto the ASTRAL phylogeny with PhyParts v0.0.1 (Smith et al., 2015). Also, we gathered chromosome counts (median) from the Chromosome Counts Database v1.66 (Rice et al., 2014). In the database, the median monoploid chromosome counts are usually the most reported number in the species, so these counts are the most representative of the chromosome number across the species. We confirmed the phylogenetic positions of newly discovered ancient WGDs by comparing paralog and ortholog Ks distributions of closely related species via ksrates v1.1.3 (Sensalari et al., 2022).

Based on Ks peaks and the estimated number of gene duplications, there was ambiguity in mapping the phylogenetic location of WGD1 between two adjacent nodes on the ASTRAL tree. Thus, we tested whether WGD1 is an allopolyploidy event with GRAMPA v1.4.0 (Thomas et al., 2017) with the 4137 rooted gene family trees. GRAMPA is a parsimony-based method for identifying polyploidy events and their parental lineages in a phylogeny.

We analyzed gene functions for gene duplications associated with WGD1 and gene duplications associated with alpine cushion plant clades. We inferred gene IDs by BLASTing homologous peptide sequences to the *Arabidopsis thaliana* Genome Annotation Official Release Araport11 from the Arabidopsis Information Resource (TAIR; Berardini et al., 2015). Then, we identified gene ontology (GO) terms for biological processes by matching *Arabidopsis* gene IDs with the UniProt database release 2023_03 (Bateman et al., 2023). We also performed a PANTHER overrepresentation test (PANTHER version 18.0 Released 2023-08-01) for duplicated genes based on the corresponding *Arabidopsis thaliana* gene IDs (Thomas et al., 2003). If one gene contains multiple duplications, it was only counted once. We conducted these overrepresentation tests (Fisher's exact) with the 4137 extracted gene family clusters as the reference gene list with GO biological process complete annotation data set and false discovery rate (FDR) correction. We also performed the same overrepresentation test for the duplicated genes and the gene families with *Solanum lycopersicum* L. as the reference gene list.

Climatic niche and life-form categorization

We obtained recent species coordinates (after the year 1950) for all 65 species in the ASTRAL phylogeny from GBIF.org (2022) for preserved specimens and human observation. We removed taxonomic mismatches or invalid coordinates (e.g., zero coordinates, mismatch with geographic region, coordinates in water, etc.) from the data set. We also removed duplicated coordinates and those with less than two decimal places. None of the records of *Mcneillia saxifraga* (Friv.) Dillenb. & Kadereit include usable digital coordinates, so we manually obtained coordinates from herbarium specimens on GBIF. We acquired bioclimatic layers for average growing season temperatures (gst) and mean daily minimum air temperature of the coldest month (bio6) from CHELSA (Karger et al., 2017). We calculated the climatic values in each species' distribution from these coordinates.

Zanne et al. (2014) defined freezing-exposed as having a freezing temperature anywhere in the species distribution, i.e., a bio6 value <0°C. It is too stringent for a cold delimitation for the Caryophyllaceae, a clade with only ca. 1% of strictly tropical species (POWO, 2022). Also, plants that share the same life-form have broad climatic distributions (Woodward et al., 2004). Thus, we defined freezing-exposed as having a mean bio6 value lower than 0°C across their distribution; others as freezing-unexposed. We also characterized species as arctic-alpine if the average gst value across species distribution is lower than 10°C (Körner et al., 2011; Birkeland et al., 2020).

For the arctic-alpine biome, temperature was found to have a positive correlation with precipitation (Boucher

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

et al., 2016). However, we did not observe a pattern for temperature (CHELSA bio6) vs. precipitation (CHELSA bio12) in our sampled species, potentially due to the low sampling fraction. Thus, we did not include a precipitation definition for arctic-alpine species.

We downloaded life-form and biome information for every accepted species of Caryophyllaceae from the World Checklist of Vascular Plants (POWO, 2022). We also acquired counts of accepted species in each climate category from WCVP. The life-form categories are in the Raunkiaer (1934) system, and we scored the Raunkiaer life-forms into three commonly used categories: annual, herbaceous perennial, and woody perennial, following the methods of Humphreys et al. (2019). We obtained cushion life-form data from the most recent documentation of cushion plants by Aubert et al. (2014). All cushion-forming plants are woody perennials, but we listed them as a fourth life-form category to highlight their significance in this study.

RESULTS

Phylogenetic relationships and conflicts in the Caryophyllaceae

The phylogenetic relationships recovered here via ASTRAL (Figure 1) were mostly congruent with recent systematics studies for this family (Sadeghian et al., 2015; Madhani et al., 2018; Arabi et al., 2022). Due to missing taxa in the ortholog trees, each ortholog might be uninformative concerning specific nodes. About one third of the 9043 orthologs contained at least 30 species, so most orthologs were informational to a small subset of nodes in the ASTRAL tree. The sizes of the pie charts in Figure 1 are proportional to the number of gene trees that were informational to the corresponding node. These numbers ranged from 1093 to 4403. We included the number of gene trees that were not informational to each node in the pie charts in Figure S1 (Appendix S2), which essentially displays the same information as Figure 1.

Based on our current species sampling, the 11 tribes represented here were recovered as monophyletic with the highest local posterior support. The placement of *Dolophragma globiflorum* was congruent with the *rps16* phylogeny of Sadeghian et al. (2015). We followed Sadeghian et al. (2015) in not considering it as a member of the Eremogoneae. However, our taxon sampling did not allow testing whether *D. globiflorum* is segregated from Eremogoneae. Corrigioleae, Paronychieae, Polycarpaeae, and Sperguleae were recovered in a well-supported early-diverging grade within the family. The gene tree conflicts were low in these four tribes except for *Paronychia* and Polycarpaeae. Although the proportions of conflicted gene trees were high for two nodes in Polycarpaeae, the actual numbers of conflicted gene trees were relatively low for these nodes.

For the remaining tribes, the level of conflict was elevated compared to other parts of the tree, not only along the backbone (esp. nodes 2 and 4–7) but also within *Eremogone*, Sileneae, Caryophylleae, and *Cerastium* (Figure 1). These elevated conflicts did not encompass a dominant conflicting relationship (Appendix S2). For backbone nodes 2 and 4–7, the conflicts were accompanied by short internal branch lengths. For *Eremogone*, Sileneae, Caryophylleae, and *Cerastium*, the conflicts were all from rearrangements of internal relationships.

Gene duplications and ancient whole-genome duplications (WGDs)

Our analyses with increased transcriptome sampling uncovered 15 putative ancient WGDs in Caryophyllaceae: eight were newly detected by our study; seven were previously published by Yang et al. (2018). Ks plots and ksrates results are provided in Appendix S3 (Figures S2–S4). The ancient WGDs were identified by (1) having the same Ks peak in all descendant taxa, e.g., WGD1, or (2) having ambiguous Ks peaks but confirmed by chromosome counts, e.g., WGD10. Ambiguous Ks peaks without chromosome counts were not recognized as ancient WGDs, e.g., *Herniaria latifolia* Lapeyr, because Ks values below 0.5 can be difficult to assess (Tiley et al., 2018). The numbers of gene duplications and chromosome counts were mapped onto the phylogeny along with the ancient WGDs in Figure 2. For the eight newly detected events, all were located on one terminal branch given our sampling, except WGD6 was shared by *Spergularia marina* (L.) Besser and *S. bocconei* (Scheele) Graebn.

The treePL analysis placed the crown age of the Caryophyllaceae at ca. 88.56 Myr, which was the maximum constraint for the node (Figure 3). The bootstrap analysis for the crown node hit the maximum constraint every time, so it appears as no uncertainty in Figure 3. Date ranges of some putative ancient WGDs could be estimated from the dated tree, as noted in Figure 2. WGD1 occurred ca. 64.4–56.7 Myr ago (Mya), followed by the diversification of the seven more nested tribes. WGD6 was shared by *Drymaria subumbellata* I.M. Johnst. and *D. cordata* (L.) Willd. ex Schult., ca. 45.4–29.3 Mya; WGD7 was shared by *Spergularia bocconei* and *S. marina*, ca. 8.8–2.3 Mya; WGD12 was shared by two arctic-alpine cushion species *Colobanthus subulatus* Hook.f. and *C. quitensis* (Kunth) Bartl., ca. 26.7–3.5 Mya; WGD15 was shared by *Schiedea membranacea* H.St.John and *S. globosa H.Mann*, with a date range of ca. 12.3–3.5 Mya based on orthologs.

Many of the detected ancient WGDs were exclusive to one species in this study, thus difficult to date (Figure 2). Although ages of ancient WGDs can be estimated from Ks plots with the assumption of a strict clock, it was deemed unreliable since Caryophyllaceae has a variety of life histories and generation times (Clark and Donoghue, 2017). Ancient WGDs detectable by Ks plots were old enough that paralogs were significantly differentiated and were assembled separately in the de novo assembly process. The

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

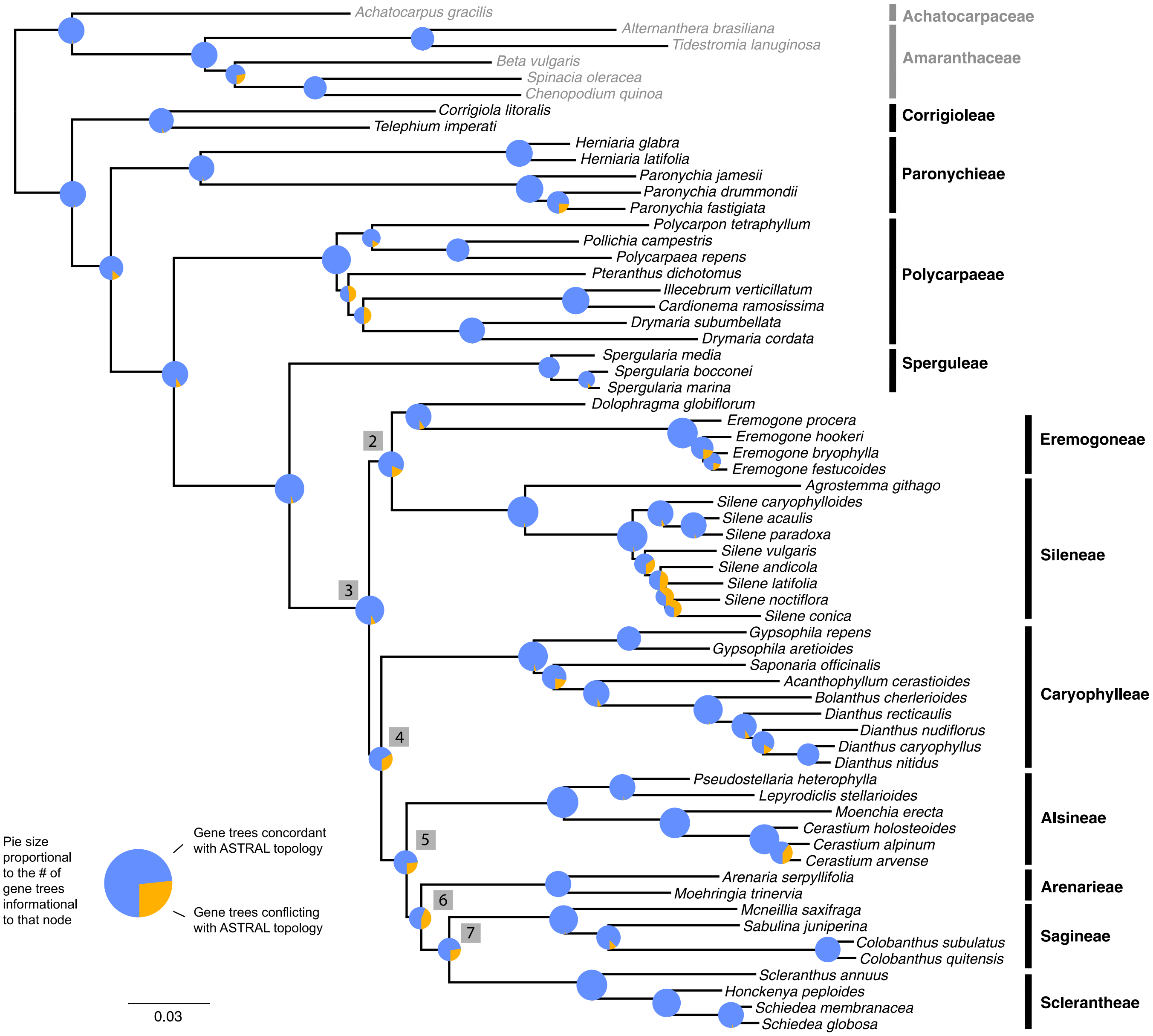

FIGURE 1 ASTRAL topology with molecular branch lengths estimated from 50 ortholog supermatrix. The size of the pie charts is proportional to the number of gene trees that are informational to the corresponding node. The color slices of the pies represent the proportion of ortholog gene trees in concordance (blue) and conflict (yellow) with the topology. The local posterior support from ASTRAL is 1 for all nodes. Numbers in gray squares are node numbers used in the main text. Taxa in gray indicate outgroups. Corresponding tribes for the ingroups are labeled on the right.

Ks peak for core eudicot genome triplication (~117 Mya, Jiao et al., 2012) had a synonymous substitution (Ks) of around 2.0 substitutions per synonymous site between paralog pairs; the peak for WGD1 (64.42–56.69 Mya) had a Ks peak at around 0.5 (Appendix S3: Figure S2). The remaining ancient WGDs had Ks peaks around 0.02 to 0.2.

GRAMPA results remain inconclusive about the nature of WGD1—being an autopolyploidy, allopolyploidy, or triploidy event (Appendix S4: Figure S5). The heterogeneity in gene family trees and missing data might be the reason that GRAMPA delivered ambiguous results for WGD1 (Appendix S5: Figure S6). The singly labeled tree (no polyploidy at node 3) was the most parsimonious reconciliation (parsimony score = 385905; Appendix S4: Figure S5), which was against the strong evidence for a polyploidy event at node 3 from Ks plots (Appendix S3: Figures S2 and S3). The second most parsimonious reconciliation (parsimony score = 386241) presented WGD1 as an allopolyploidy event with Sperguleae as the other parental lineage, while the sixth most parsimonious reconciliation (parsimony score = 390963) presented WGD1 as an autopolyploidy event. Therefore, GRAMPA results supported that WGD1 was more likely to be an alloploidy event rather than autopolyploidy because the parsimony score was much higher for autopolyploidy.

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

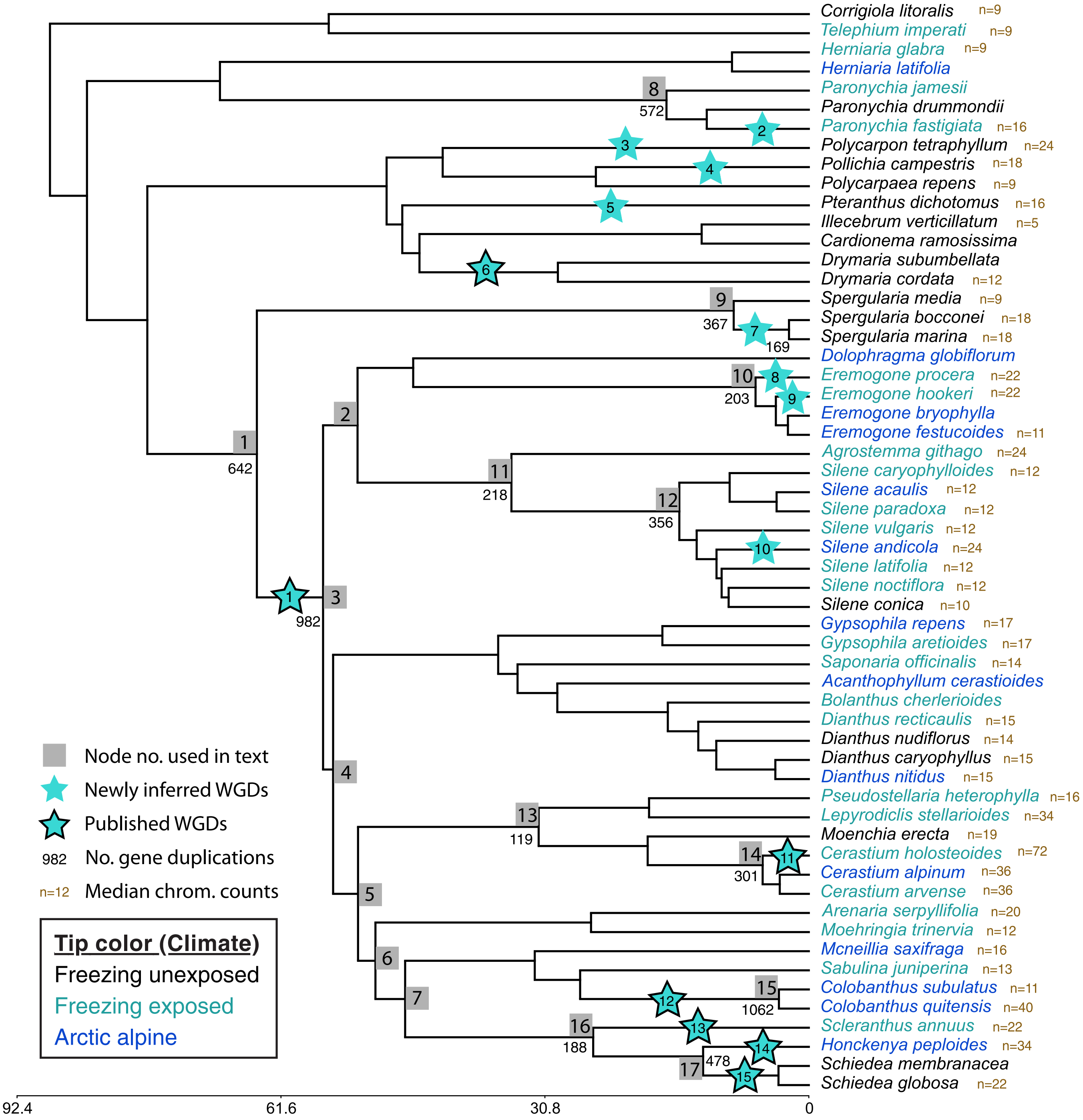

FIGURE 2 Dated tree (only ingroups) inferred from ASTRAL phylogeny with treePL. Tip branches and tip labels are colored based on climate: freezing-unexposed (black), freezing-exposed (turquoise), and arctic-alpine (blue). All ancient WGDs are labeled as stars. Ks plots supporting the inferences are included in Appendix S3. The seven ancient WGDs published by Yang et al. (2018) have black outlines. The relative position of the ancient WGDs along the branch does not indicate the time of the events. Nodes with more than 100 gene duplications are labeled with the number of gene duplications. Numbers in gray squares are node numbers used in the main text. Chromosome counts (median) from the Chromosome Counts Database v1.66 (Rice et al., 2014) are listed next to the tip labels. Some species do not have a chromosome count.

Linking genomic changes to climatic shift and the cushion life-form

Of the 57 Caryophyllaceae species included in this analysis, 20 species were characterized as not freezing-exposed and the rest as freezing-exposed (Figure 3). Of the 37 freezing-exposed species, 14 were arctic-alpine. There were no freezing-unexposed arctic-alpine species. All 11 cushion-forming species were identified either as freezing-exposed or arctic-alpine. All arctic-alpine species were herbaceous or woody perennials (including cushion forms). Most of the cold-tolerant taxa sampled in this study were within the

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

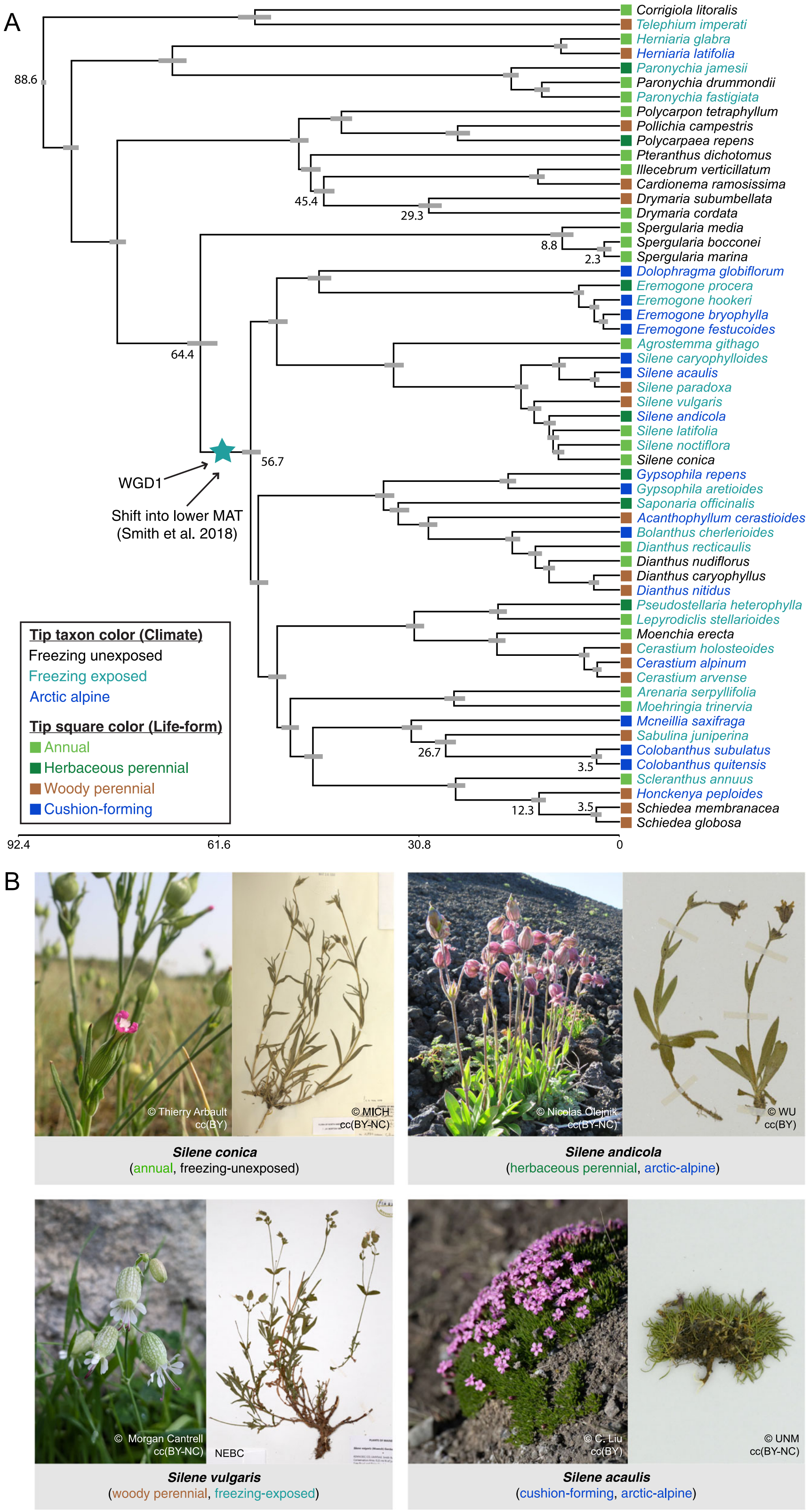

FIGURE 3 Dated tree and life-form diversity in sampled species. (A) Dated tree (only ingroups) inferred from ASTRAL phylogeny with treePL. Tip branches and tip taxa are colored based on climate. The squares before tip taxa are colored based on life-forms. Dates that are referenced in the manuscript are labeled next to the nodes. Error bars (gray) were calculated based on bootstrapping ortholog alignments, branch length re-estimation with RaxML-NG, and dating with treePL. The position of WGD1 (star) and a shift into lower mean annual temperature (MAT) inferred by Smith et al. (2018b) are labeled. (B) Life-form diversity in four sampled *Silene* species: *Silene conica* L., *Silene andicola* Gillies ex Hook. & Arn., *Silene vulgaris* (Moench) Garcke, and *Silene acaulis* (L.) Jacq.

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

seven tribes descending from the oldest ancient WGD in Caryophyllaceae (WGD1). All 11 cushion plants were also members of these seven tribes. However, there was no apparent co-occurrence between younger ancient WGDs and arctic-alpine lineages in this study. Additional and broader sampling would be required to address this adequately.

WGD1 coincided with a shift in climatic niche (Smith et al., 2018b), so we looked closer at the gene duplications retained after the duplication events. Given that duplicated genes retained at nodes 1 and 3 had similar GO term (biological processes) compositions and WGD1 was putatively allopolyploidy, we could reasonably assume that these duplications were all associated with WGD1. Thus, we combined these duplications in the GO term analysis. Among the 931 genes duplicated at nodes 1 and 3, 21 genes were labeled with the GO term response to cold (Table 1). There was no statistically significant result for the PANTHER overrepresentation test when accounting for the false discovery rate (FDR), a conservative method that corrects for false positive results. Yet, based on raw *P*-values, 54 were overrepresented GO biological processes. The top positively enriched processes are listed in Table 2; the full list is in Appendix S6 (Table S1). Most of the overrepresented processes were involved in phosphorylation, ubiquitination, and vesicle-mediated transport.

Only two clades in the ASTRAL phylogeny mainly comprised arctic-alpine cushion species and had large numbers of gene duplications at the crown node of the clade: *Eremogone* and *Colobanthus* (Figure 2, nodes 10 and 15). The two clades shared 68 of the 829 duplicated genes for *Colobanthus* and 187 duplicated genes for *Eremogone*. Similarly, there was no significant result based on FDR, but there were 61 overrepresented GO biological processes based on raw *P*-values. The top 30 processes are listed in Table 3; the full list is in Appendix S6 (Table S2). The overrepresented processes involved carbohydrate homeostasis, ubiquitination, and trichome morphogenesis.

We performed the gene overrepresentation tests with gene family clusters as reference instead of with a reference genome because the gene families obtained from transcriptomes tend to be the large gene families and are consecutively expressed (Yang et al., 2015). To evaluate the potential bias, we ran the PANTHER overrepresentation test for the gene family clusters with *Solanum lycopersicum* as the reference genome, and ~1700 GO biological processes were significantly overrepresented based on FDR correction (Appendix S6: Table S3). Essential biological processes, such as various organ development and metabolic processes, were overrepresented in the gene families. They would bias our functional analyses with duplicated genes if we did not use the gene families as the background reference for the tests. Supportively, a similar set of organ development processes was shown to be overrepresented with the duplicated genes associated with WGD1 with *S. lycopersicum* as the reference gene list (Appendix S6: Table S4).

DISCUSSION

Interpreting the nature of putative ancient whole-genome duplications (WGDs) in the Caryophyllaceae

An excess number of gene duplications detected at the node ancestral to the position of an ancient WGD inferred from Ks plots suggests an allopolyploidy event (Yang et al., 2018). For WGD1, there were 982 duplications (744 genes) at the same node (node 3, Figure 2) and 624 duplications (574 genes) at the closest ancestral node (node 1, Figure 2). Therefore, WGD1 might be the result of a hybridization event with an ancestor of the tribe Sperguleae or an unsampled lineage. We examined the gene family trees of the gene duplications to explore the nature of WGD1 (e.g., autopolyploidy, allopolyploidy, or triploidy; Appendix S5: Figure S6). Most gene family trees that showed duplications at node 1 only had Sperguleae in one paralog clade, indicating allopolyploidy. Most gene family trees that showed duplications at node 3 did not involve Sperguleae in the paralogs, which supports autopolyploidy. Additionally, some gene family trees showed a gene triplication with Sperguleae in one paralog, but the number was too low to support genome triplication for WGD1 confidently. All these gene families had different levels of missing gene copies.

Given the heterogeneity in the gene family trees, it is not surprising that GRAMPA could not determine the nature of WGD1. According to GRAMPA, the most parsimonious outcome was no polyploidy (score = 385905), which is against the strong evidence for an ancient WGD from Ks and gene duplication analyses. This result may be because GRAMPA is designed for genome sequencing data and is sensitive to missing data (Thomas et al., 2017). Although allopolyploidy (score = 386241) appeared to be much more parsimonious than autopolyploidy (score = 390963), the fact that GRAMPA did not rank polyploidy first makes the parental lineage test inconclusive. Due to the short branch lengths between nodes 1 and 3 relative to their age, between-species Ks peaks largely overlapped with within-species Ks peaks (data not shown). Thus, it was not feasible to infer allopolyploidy with Ks peaks either. More genomic sampling of the Caryophyllaceae and improved methods for inferring allopolyploidy are needed to confidently conclude the origin and nature of WGD1.

Autopolyploidy should correspond with many gene-tree-inferred duplications at the same node (Yang et al., 2018). Deviation from this correspondence may be caused by uninformative gene trees, noise, incomplete lineage sorting, accumulation of small-scale gene duplications, allopolyploidy, or gene loss due to diploidization (Yang et al., 2015; Li et al., 2021). Aside from WGD1, there were also other such deviations in this study. Nodes 8–14 and 16–17 all had elevated numbers of gene duplications that did not correspond with an ancient WGD at the same node (Figure 2). All these nodes except node 8 were

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 The full list of duplicated genes retained after WGD1 annotated with the gene ontology (GO) biological process response to cold. The *Arabidopsis* IDs that were used to infer GO terms are listed. We obtained gene names and protein functions from the UniProt database Release 2023_03 (Bateman et al., 2023). We determined the cold-tolerance-related processes in which the proteins are involved using the referenced molecular studies. The entry "Induced by cold stress" means that the study found a change in gene expression during cold stress but did not suggest any potential processes.

***Arabidopsis* ID**	**Gene name**	**Protein**	**Processes related to cold tolerance**	**Reference**
At2g19520	*MSI4/FVE*	WD-40 repeat-containing protein	Sensing cold stress and delaying flowering time	Kim et al. (2004)
At5g53470	*ACBP1*	Acyl-CoA-binding domain-containing protein 1	Maintaining phosphatidic acid in membrane	Du et al. (2010)
At5g04590	*SIR*	Chloroplastic assimilatory sulfite reductase (ferredoxin)	Modulating sulfite reduction	Xia et al. (2018)
At1g74520	*HVA22*	HVA22-like protein a	Induced by cold stress	Chen et al. (2002)
At2g43790	*MPK6*	Mitogen-activated protein kinase 6	Cold stress signaling	Teige et al. (2004)
At5g67320	*HOS15*	WD40 repeat-containing protein	Regulating gene expression through histone deacetylation	Zhu et al. (2008)
At5g63770	*DGK2*	Diacylglycerol kinase 2	Generating phosphatidic acid	Gómez-Merino et al. (2004)
At2g19450	*DGAT1*	Diacylglycerol O-acyltransferase 1	Germination and seedling growth	Lu and Hills (2002)
At4g03430	*STA1*	STABILIZED1	Splicing and turnover of unstable transcripts	Lee et al. (2006)
At1g01140	*CIPK9*	CBL-interacting serine/threonine-protein kinase 9	Regulation of potassium	Pandey et al. (2007)
At1g77120	*ADH1*	Alcohol dehydrogenase class-P	Induced by cold stress	Jarillo et al. (1993)
At3g55280	*RPL23AB*	Large ribosomal subunit protein uL23y	Maintaining active translation of cryoprotective gene transcripts	Degenhardt and Bonham-Smith (2008)
At3g53110	*RH38*	DEAD-box ATP-dependent RNA helicase 38	Early regulator of transcription factors	Gong et al. (2002)
At5g59010	*BSK5*	Serine/threonine-protein kinase	Induced by cold stress	Li et al. (2012)
At2g37220	*CP29B*	chloroplastic RNA-binding protein	Induced by cold stress	Amme et al. (2006)
At1g15520	*ABCG40*	ABC transporter G family member 40	Transporting abscisic acid	Baron et al. (2012)
At2g36530	*ENO2/LOS2*	Bifunctional enolase 2/transcriptional activator	Cold-responsive gene transcription	Lee et al. (2002)
At2g35040	-	phosphoribosylaminoimidazolecarboxamide formyltransferase	Induced by cold stress	Goulas et al. (2006)
At1g27320	*AHK3*	Histidine kinase 3	Negatively regulating cold stress signaling via inhibition of ABA response	Jeon et al. (2010)
At1g74960	*KAS2*	Chloroplastic 3-oxoacyl-[acyl-carrier-protein] synthase II	Synthesizing fatty acids and maintaining chloroplast membrane integrity	Carlsson et al. (2002)
At3g04740	*MED14*	Mediator of RNA polymerase II transcription subunit 14	Regulating cold-responsive gene expression	Hemsley et al. (2014)

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Top 30 positively enriched GO biological processes for duplicated genes associated with WGD1, ordered by *P*-value, with gene family clusters as the reference gene list. FDR = false discovery rate. The full list is in Table S1 (Appendix S6).

	Frequency of genes in:				
GO biological process complete	**Gene family clusters**	**Duplicated genes**	**Fold enrichment**	***P***	**FDR**
Phosphorylation (GO:0016310)	0.03976339	0.06702703	1.68	0.000902	0.606
Vesicle-mediated transport (GO:0016192)	0.04797897	0.07783784	1.62	0.000936	0.472
Signaling (GO:0023052)	0.09069997	0.12648649	1.39	0.00207	0.641
Phosphorus metabolic process (GO:0006793)	0.10647387	0.14378378	1.35	0.0027	0.68
Signal transduction (GO:0007165)	0.08807098	0.12216216	1.38	0.00277	0.656
Protein phosphorylation (GO:0006468)	0.03384818	0.05621622	1.66	0.00351	0.745
Cell communication (GO:0007154)	0.10778837	0.14378378	1.33	0.00408	0.748
Phosphate-containing compound metabolic process (GO:0006796)	0.10154453	0.13621622	1.34	0.00471	0.792
Protein modification process (GO:0036211)	0.12980611	0.16756757	1.29	0.00476	0.767
Regulation of biological process (GO:0050789)	0.32237923	0.37189189	1.15	0.00689	0.959
Peptidyl-serine modification (GO:0018209)	0.01018732	0.0227027	2.22	0.00733	0.985
Peptidyl-serine phosphorylation (GO:0018105)	0.01018732	0.0227027	2.22	0.00733	0.953
Localization (GO:0051179)	0.18895826	0.23027027	1.21	0.00739	0.932
Regulation of cellular process (GO:0050794)	0.27374302	0.32	1.16	0.00884	1
Vesicle organization (GO:0016050)	0.01150181	0.02378378	2.06	0.0107	1
Golgi vesicle transport (GO:0048193)	0.01807427	0.03243243	1.79	0.0131	1
Modification-dependent protein catabolic process (GO:0019941)	0.03319093	0.05189189	1.56	0.0132	1
Modification-dependent macromolecule catabolic process (GO:0043632)	0.0341768	0.05297297	1.54	0.0145	1
Transport (GO:0006810)	0.16201117	0.19783784	1.22	0.0145	1
Establishment of localization (GO:0051234)	0.16398291	0.2	1.22	0.0149	1
Regulation of transport (GO:0051049)	0.00690108	0.01621622	2.34	0.0157	1
Cell differentiation (GO:0030154)	0.06309563	0.08648649	1.37	0.0174	1
Protein autophosphorylation (GO:0046777)	0.01413079	0.02594595	1.83	0.0195	1
Ubiquitin-dependent protein catabolic process (GO:0006511)	0.03253368	0.04972973	1.52	0.021	1
Localization within membrane (GO:0051668)	0.01478804	0.02702703	1.82	0.0215	1
Post-translational protein modification (GO:0043687)	0.03976339	0.05837838	1.46	0.022	1
Nucleoside monophosphate biosynthetic process (GO:0009124)	0.00657246	0.01513514	2.29	0.0228	1
Nucleoside monophosphate metabolic process (GO:0009123)	0.00657246	0.01513514	2.29	0.0228	1
Cytoskeleton organization (GO:0007010)	0.02760434	0.04324324	1.56	0.023	1
Proteolysis involved in protein catabolic process (GO:0051603)	0.03779165	0.05513514	1.45	0.0247	1

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

descendant lineages of WGD1, so the elevated number of gene duplications without a corresponding Ks peak can be due to phylogenetic uncertainty, differential gene loss, or noise. Nodes 9, 10, 14, 16, and 17 had ancient WGDs at adjacent nodes on the phylogeny, so they could also be caused by uninformative gene trees, incomplete lineage sorting, allopolyploidy, or noise. Yang et al. (2018) inferred WGD15 (on the *Schiedea* child node of node 17) as an allopolyploidy event, with one parental lineage closely related to *Honckenya peploides* (L.) Ehrh. and one unsampled parental lineage. Thus, the gene duplications at node 17 could likely reflect allopolyploidy. For node 8, there might be ancient WGDs in addition to WGD2 (Ks~0.1) in *Paronychia*. However, there was not enough information from chromosome counts, and Ks peaks in *Paronychia jamessii* and *P. drummondii* were ambiguous.

WGD1 coincided with high levels of gene tree conflict in the Caryophyllaceae

Several biological processes may cause gene tree conflicts, such as incomplete lineage sorting (ILS), gene duplications and losses, horizontal gene transfer, and hybridization (Maddison, 1997; Degnan and Rosenberg, 2009). Polyploidy leads to an influx of gene duplications, many of which are lost during diploidization (Li et al., 2021). Therefore, ancient WGDs can be a significant source of gene tree conflicts in phylogenomic analyses. Our conflict results align with this claim as most bipartitions that contain elevated gene tree conflicts were descendant lineages of WGD1 (Figure 1). These bipartitions can be categorized into those that derived soon after WGD1 (nodes 2 and 4–7) and those within more recently diversified clades in the family (*Eremogone*, Sileneae, Caryophylleae, and *Cerastium*).

The conflicts in nodes 2 and 4-7 were accompanied by incongruence among these bipartitions compared to previous phylogenetic works (Harbaugh et al., 2010; Greenberg and Donoghue, 2011; Sadeghian et al., 2015; Walker et al., 2018; Yang et al., 2018). The phylogenies based on a few markers showed low support for the nodes connecting the seven descending tribes of WGD1 (Eremogoneae, Sileneae, Caryophylleae, Alsineae, Arenarieae, Sagineae, and Sclerantheae). The transcriptomic studies did not explore the phylogenetic relationships or address incongruence. There was no consensus for the relationships among the seven tribes. However, the relationships among the early-diverging tribes (Corrigioleae, Paronychieae, Polycarpaeae, and Sperguleae) were congruent across all studies.

Although all nodes in the ASTRAL topology had the highest local posterior support, the branch length separating backbone nodes 2 and 4–7 were relatively short with elevated levels of conflicts (Figure 1). There were two potential connections between WGD1 and the conflicts in the backbone nodes. First, as mentioned previously, the high conflicts can be caused by differential gene loss from diploidization (Degnan and Rosenberg, 2009; Li et al., 2021). The genome content that WGD1 increased experienced extensive loss or silencing, which is evidenced by deeper ancient WGDs that had shorter Ks peaks (Appendix S3: Figure S2), the lower number of informational gene trees in the backbone compared to descending clades (Figure 1), and missing paralog copies in the example gene family trees (Appendix S5). Second, the short branch lengths provided little information to resolve the relationship, even with a large amount of data. This lack of information in genomic data can result from the rapid divergence of lineages in a short time frame, increasing the possibilities for conflict due to ILS, among other causes (Sanderson and Shaffer, 2002; Felsenstein, 2004; Suh et al., 2015). Ancient WGDs are hypothesized to promote speciation by increasing genome size and the potential for adaptation or by diploidization that contributes to reproductive isolation (Soltis et al., 2009; Mandáková and Lysak, 2018).

The elevated conflicts from descending clades of WGD1 may also be attributed to diploidization and differential gene loss. The conflicts in *Eremogone*, Sileneae, Caryophylleae, and *Cerastium* all came from rearrangements of lineages within the clade (Appendix S2). While these conflicts might result from ILS, it is essential to consider the effect of differential gene loss because several descendant clades of WGD1 had high gene tree conflicts.

Although having a large amount of data complicated phylogenetic reconstruction, it provided a more comprehensive picture of the evolutionary history of the Caryophyllaceae beyond resolving bifurcating phylogenetic relationships. Our results highlight the importance of genomic data for angiosperm phylogenomics, given how common polyploidy events are in flowering plants (Clark and Donoghue, 2018). As additional genomes are sequenced, there will be opportunities to gain better insights into the relationships in question.

Gene family expansion and convergent evolution in arctic-alpine lineages

Ancient WGDs are often associated with trait innovations and climatic shifts (te Beest et al., 2012; Moghe and Shiu, 2014; Brockington et al., 2015; Smith et al., 2018b; Baniaga et al., 2020; Wu et al., 2020). One of the mechanisms by which ancient WGDs contribute to trait innovations is gene family expansion (Brockington et al., 2015; Wang et al., 2019; Jablonski, 2022).

Smith et al. (2018b) tested climatic shifts in the Caryophyllales with increased sampling using Sanger-sequencing-based markers from GenBank and inferred a shift into colder climates (lower mean annual temperature) at the node corresponding to WGD1 (Figure 3). Although Smith et al. (2018b) only sampled 27% of the species for the Caryophyllaceae, the relationship between the shift to cold climatic and WGD1 was also supported when tallying all accepted Caryophyllaceae species in the World Checklist of Vascular Plants (WCVP) data set (POWO, 2022). Before the

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Top 30 overrepresented GO biological processes for duplicated genes convergently retained by *Colobanthus* and *Eremogone*, ordered by *P*-value with gene family clusters as the reference gene list. The full list is in Table S2 (Appendix S6).

	Frequency of genes in:				
GO biological process complete	**Gene family clusters**	**Duplicated genes**	**Fold enrichment**	***P***	**FDR**
Trichome branching (GO:0010091)	0.00361485	0.04411765	12.06	0.00316	1
Cytoplasmic translation (GO:0002181)	0.0085442	0.05882353	6.8	0.00389	1
Glucose homeostasis (GO:0042593)	0.00098587	0.02941176	29.49	0.00461	1
Intracellular glucose homeostasis (GO:0001678)	0.00098587	0.02941176	29.49	0.00461	1
Cellular response to glucose stimulus (GO:0071333)	0.00098587	0.02941176	29.49	0.00461	1
Cellular response to hexose stimulus (GO:0071331)	0.00098587	0.02941176	29.49	0.00461	1
Cellular response to monosaccharide stimulus (GO:0071326)	0.00098587	0.02941176	29.49	0.00461	1
Positive regulation of proteasomal ubiquitin-dependent protein catabolic process (GO:0032436)	0.00131449	0.02941176	22.12	0.00682	1
Positive regulation of ubiquitin-dependent protein catabolic process (GO:2000060)	0.00164312	0.02941176	17.69	0.00942	1
Regulation of proteasomal ubiquitin-dependent protein catabolic process (GO:0032434)	0.00164312	0.02941176	17.69	0.00942	1
Carbohydrate homeostasis (GO:0033500)	0.00164312	0.02941176	17.69	0.00942	1
Defense response to insect (GO:0002213)	0.00164312	0.02941176	17.69	0.00942	1
Translation (GO:0006412)	0.05159382	0.13235294	2.54	0.00987	1
Peptide biosynthetic process (GO:0043043)	0.05192244	0.13235294	2.52	0.0102	1
Peptide metabolic process (GO:0006518)	0.05356556	0.13235294	2.44	0.0123	1
Regulation of chlorophyll biosynthetic process (GO:0010380)	0.00197174	0.02941176	14.74	0.0124	1
Regulation of ubiquitin-dependent protein catabolic process (GO:2000058)	0.00197174	0.02941176	14.74	0.0124	1
Positive regulation of proteasomal protein catabolic process (GO:1901800)	0.00197174	0.02941176	14.74	0.0124	1
Positive regulation of catabolic process (GO:0009896)	0.00690108	0.04411765	6.32	0.015	1
Regulation of protein metabolic process (GO:0051246)	0.02037463	0.07352941	3.57	0.0153	1
Positive regulation of protein catabolic process (GO:0045732)	0.00230036	0.02941176	12.64	0.0157	1
Positive regulation of proteolysis involved in protein catabolic process (GO:1903052)	0.00230036	0.02941176	12.64	0.0157	1
Regulation of tetrapyrrole biosynthetic process (GO:1901463)	0.00230036	0.02941176	12.64	0.0157	1
Regulation of chlorophyll metabolic process (GO:0090056)	0.00230036	0.02941176	12.64	0.0157	1
Glycolytic process (GO:0006096)	0.00755833	0.04411765	5.77	0.0187	1
Positive regulation of proteolysis (GO:0045862)	0.00262898	0.02941176	11.06	0.0193	1
Trichome morphogenesis (GO:0010090)	0.00788695	0.04411765	5.53	0.0207	1
Regulation of proteasomal protein catabolic process (GO:0061136)	0.00295761	0.02941176	9.83	0.0233	1
Trichome differentiation (GO:0010026)	0.00887282	0.04411765	4.91	0.0274	1
Regulation of proteolysis involved in protein catabolic process (GO:1903050)	0.00328623	0.02941176	8.85	0.0276	1

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

shift, 143 of the total 490 species (~29%) in the four early-diverging tribes (Corrigioleae, Paronychieae, Polycarpaeae, and Sperguleae) occur in the temperate or subarctic-subalpine biome, including 24 cushion species. After the shift, 2270 of 2810 species (~81%) among the seven more recently diverged tribes occur in temperate or subarctic-subalpine habitats, and 140 are cushion species.

With increased transcriptomic sampling, we further explored the genes associated with WGD1 to gain insights into the cold climatic shift. Because transcriptomic data sets provide genome-wide evidence on protein-coding sequences and gene duplications, they are a powerful tool to examine the specific functions of gene duplications associated with ancient WGDs. Although extensive gene loss occurred after WGD1, some of its extant descendant species still retain a large number of gene duplications (Figure 2). We examined the gene ontology (GO) terms for the gene duplications associated with WGD1 and performed gene overrepresentation tests.

An often-overlooked aspect of overrepresentation tests is the reference genes used for comparison. Gene family clusters extracted from transcriptomic data sets are already enriched in specific biological processes (Yang et al., 2015). Various organ development and metabolic processes were overrepresented in the gene families we used to infer gene duplications (Appendix S6: Table S3). Therefore, we used the gene families as the reference gene list for gene overrepresentation tests; otherwise, the results may be biased toward the biological processes enriched in the gene families (Appendix S6: Table S4).

There were 21 genes associated with the GO term response to cold in the 931 duplicated genes retained after WGD1 (Table 1). During cold stress, the level of gene expression of these genes changed in *Arabidopsis thaliana*. Some genes were involved in stabilizing membrane and transcription during stress (*ACBP1*: Du et al., 2010; *STA1*: Lee et al., 2006; *RPL23AB*; Degenhardt and Bonham-Smith, 2008; *KAS2*: Carlsson et al., 2002); some were involved in regulating cold stress signaling and cold-responsive gene expression (*MSI4/FVE*: Kim et al., 2004; *MPK6*: Teige et al., 2004; *HOS15*: Zhu et al., 2008; *RH38*: Gong et al., 2002; *ENO2/LOS2*: Lee et al. 2002; *AHK3*: Jeon et al., 2010; *MED14*: Hemsley et al., 2014). However, response to cold is not an overrepresented GO biological process among the genes retained after WGD1, although the number of genes under that label so far could be limited due to the amount of cold tolerance molecular research.

Phosphorylation, ubiquitination, and vesicle-mediated transport were the major overrepresented GO biological processes associated with WGD1, with the gene family clusters as the reference gene list (Table 2; Appendix S6: Table S1). Protein phosphorylation is involved in signaling pathway regulation for cold tolerance (Praat et al., 2021), while the ubiquitin-proteasome system plays a critical role in plant response to environmental stress (Xu and Xue, 2019). It is also notable that phosphorylation, ubiquitination, and vesicle-mediated transport are all processes involved in protein regulation, which can be a response to low temperature (Chinnusamy et al., 2004; Beck et al., 2007; Preston and Sandve, 2013; Janmohammadi et al., 2015). However, protein regulation is not a specific corresponding response for cold climates, and those GO biological processes were statistically significant only without false discovery rate (FDR) corrections (Appendix S6: Table S1). These genes and biological processes are worth further exploration, but they cannot serve as strong evidence for the effect of gene family expansion on cold adaptation in the Caryophyllaceae.

While WGD1 led to retained gene duplications that might aid in cold tolerance in the descendant lineages, *Colobanthus* and *Eremogone* shared 66 duplicated genes that were overrepresented in carbohydrate homeostasis, trichome development, and ubiquitination (Table 3; Appendix S6: Table S2). *Colobanthus* and *Eremogone* both had highly elevated gene duplications at each of their crown nodes, 1062 duplications in 829 genes and 203 duplications in 187 genes, respectively (nodes 10 and 15 in Figure 2), and both were predominantly arctic-alpine cushion plants in our taxon sampling. Gene tree topologies supported that species of *Colobanthus* and *Eremogone* converged on the 68 duplicated genes instead of retaining homologous gene duplications.

The cushion phenotype is a specialized life history strategy in cold-adapted plants (Aubert et al., 2014). Carbohydrate accumulation supports cellular hydration, membrane stability, and other cellular functions and is a commonly observed response to cold (Beck et al., 2007; Preston and Sandve, 2013; Folk et al., 2020). Although trichomes can help retain heat and moisture (Wang et al., 2021) and are often found in arctic-alpine plants, cushion plants maintain their temperature and water by their spherical shape and smooth surface and are usually not found to be remarkably hairy (Boucher et al., 2016; Lee, 2020). Therefore, it is interesting that trichome developmental processes were overrepresented here because the sampled cushion plants in *Colobanthus* and *Eremogone* are not extensively hairy. As discussed previously, ubiquitination is involved in protein regulation and can potentially contribute to plants' response to cold stress (Xu and Xue, 2019). Because these GO biological processes were only statistically significant without FDR corrections (Appendix S6: Table S2), they can serve as candidates for future molecular research, but they do not provide direct evidence for the molecular mechanisms of cushion plant adaptations.

Ancient WGD1 as a potential macroevolutionary enabler in the Caryophyllaceae

WGD1 is associated with a climatic niche shift in the Caryophyllaceae (Smith et al., 2018b). WGD1 putatively took place ca. 64.4–56.7 Mya (Figure 3) during the substantial cooling of the climate in the Early Cenozoic (Willis and McElwain, 2014). Our results showed that many gene duplications retained after WGD1 have a potential link to cold adaptation, coinciding with independent instances of cold

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

adaptation and a shift into colder climates. Similarly, Wu et al. (2020) uncovered the same pattern that ancient WGDs close to the K-Pg boundary retained gene duplications related to cold tolerance in descendant lineages. While a direct link between the duplications and adaptation is beyond the scope of this study, further investigation can help determine whether candidate genes retained after WGD1 contributed to cold adaptations in descendant lineages.

WGD1 was followed by a divergence of lineages into the seven tribes that comprise 85% of the extant species diversity in the family. Based on our results, WGD1 is potentially a critical enabler that allowed its descendants to move into the expanding cold climatic niche in the Early Cenozoic and aided in the diversification of lineages.

CONCLUSIONS

Our results provide evidence for a putative ancient whole-genome duplication (WGD1) followed by the diversification of major lineages in the Caryophyllaceae with adaptations to a newly available cold environment in the Early Cenozoic. Our study showed that genes that are potentially associated with cold adaptation were preferentially retained after WGD1 and convergently retained in additional gene duplications associated with more recent arctic-alpine cushion plant lineages. These results provide evidence for a potential link between ancient WGDs and shifts to cold climates. WGD1-induced gene family expansion and differential gene loss complicated the phylogenomic reconstruction, but these genomic changes also provided evidence about the biological processes in the evolutionary history of the Caryophyllaceae. While further investigations will be necessary to understand better the link between these genomic events and the evolution of cold tolerance, our analyses yielded important insights into molecular evolution associated with cold tolerance and the iconic cushion plants in arctic-alpine environments.

AUTHOR CONTRIBUTIONS

K.F., J.F.W., H.E.M., and S.A.S designed the research. Y.Y., S.F.B., and M.J.M., contributed to sampling; H.E.M. conducted laboratory work; J.F.W. and H.E.M. processed sequences; K.F. and S.A.S. analyzed the data and led the writing; and Y.Y., J.F.W., H.E.M., and R.K.R. assisted with the data interpretation and the writing.

ACKNOWLEDGMENTS

We thank Tom Carruthers, Joey Charboneau, and Avinash Subramanian for thoughtful comments on the manuscript; Jake Berv, Will Weaver, Diana Medellin, Raquel Pizzardo, and Adam Lauring for helpful discussion on the manuscript; two anonymous reviewers for constructive feedback on the manuscript. We also thank Cambridge University Botanic Gardens for help with growing material and access to living collections. The molecular work of this study was conducted in the Genomic Diversity Laboratory of the Department of Ecology and Evolutionary Biology, University of Michigan. This work was supported by NSF DEB awards 1938969 and 1354048 and start-up funding from the University of Illinois Chicago.

DATA AVAILABILITY STATEMENT

We deposited the newly generated raw reads for transcriptomes generated for this study in the NCBI sequence read archive (Bioproject PRJNA1019755: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1019755). Assemblies, orthologous gene clusters, alignments, trees, and data files are available in the Zenodo Digital Repository (https://zenodo.org/doi/10.5281/zenodo.8393420).

ORCID

Keyi Feng http://orcid.org/0000-0002-9773-4320
Joseph F. Walker http://orcid.org/0000-0003-2928-8899
Hannah E. Marx http://orcid.org/0000-0002-3526-7878
Ya Yang http://orcid.org/0000-0001-6221-0984
Samuel F. Brockington http://orcid.org/0000-0003-1216-219X
Michael J. Moore http://orcid.org/0000-0003-2222-8332
Richard K. Rabeler http://orcid.org/0000-0002-6765-0353
Stephen A. Smith http://orcid.org/0000-0003-2035-9531

REFERENCES

Aberer, A. J., K. Kobert, and A. Stamatakis. 2014. ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution* 31: 2553–2556.

Amme, S., A. Matros, B. Schlesier, and H. P. Mock. 2006. Proteome analysis of cold stress response in *Arabidopsis thaliana* using DIGE-technology. *Journal of Experimental Botany* 57: 1537–1546.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data [online]. Website: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arabi, Z., F. Ghahremaninejad, R. K. Rabeler, I. Sokolova, M. Weigend, and S. Zarre. 2022. Intergeneric relationships within the tribe Alsineae (Caryophyllaceae) as inferred from nrDNA ITS and cpDNA rps16 sequences: a step toward a phylogenetically based generic system. *Taxon* 71: 608–629.

Aubert, S., F. Boucher, S. Lavergne, J. Renaud, and P. Choler. 2014. 1914-2014: A revised worldwide catalogue of cushion plants 100 years after Hauri and Schröter. *Alpine Botany* 124: 59–70.

Baniaga, A. E., H. E. Marx, N. Arrigo, and M. S. Barker. 2020. Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. *Ecology Letters* 23: 68–78.

Baron, K. N., D. F. Schroeder, and C. Stasolla. 2012. Transcriptional response of abscisic acid (ABA) metabolism and transport to cold and heat stress applied at the reproductive stage of development in *Arabidopsis thaliana*. *Plant Science* 188: 48–59.

Bateman, A., M. J. Martin, S. Orchard, M. Magrane, S. Ahmad, E. Alpi, E. H. Bowler-Barnett, et al. 2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* 51: D523–D531.

Beck, E. H., S. Fettig, C. Knake, K. Hartig, and T. Bhattarai. 2007. Specific and unspecific responses of plants to cold and drought stress. *Journal of Biosciences* 32: 501–510.

Berardini, T. Z., L. Reiser, D. Li, Y. Mezheritsky, R. Muller, E. Strait, and E. Huala. 2015. The *Arabidopsis* information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis* 53: 474–485.

Birkeland, S., A. L. S. Gustafsson, A. K. Brysting, C. Brochmann, M. D. Nowak, and M. Purugganan. 2020. Multiple genetic trajectories to extreme abiotic stress adaptation in Arctic Brassicaceae. *Molecular Biology and Evolution* 37: 2052–2068.

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Bittrich, V. 1993. Caryophyllaceae. *In* J. Kubitzki [ed.], The families and genera of vascular plants, 2, 206–236. Springer, Berlin, Germany.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Boucher, F. C., S. Lavergne, M. Basile, P. Choler, and S. Aubert. 2016. Evolution and biogeography of the cushion life form in angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics* 20: 22–31.

Brockington, S. F., Y. Yang, F. Gandia-Herrero, S. Covshoff, J. M. Hibberd, R. F. Sage, G. K. S. Wong, et al. 2015. Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytologist* 207: 1170–1180.

Brown, J. W., J. F. Walker, and S. A. Smith. 2017. Phyx: phylogenetic tools for unix. *Bioinformatics* 33: 1886–1888.

Carlsson, A. S., S. T. LaBrie, A. J. Kinney, Von Wettstein-Knowles, P., and J. Browse. 2002. A *KAS2* cDNA complements the phenotypes of the *Arabidopsis fab1* mutant that differs in a single residue bordering the substrate binding pocket. *Plant Journal* 29: 761–770.

Chaboureau, A. C., P. Sepulchre, Y. Donnadieu, and A. Franc. 2014. Tectonic-driven climate change and the diversification of angiosperms. *Proceedings of the National Academy of Sciences, USA* 111: 14066–14070.

Chen, C. N., C. C. Chu, R. Zentella, S. M. Pan, and D. Ho. 2002. AtHVA22 gene family in *Arabidopsis*: phylogenetic relationship, ABA and stress regulation, and tissue-specific expression. *Plant Molecular Biology* 49: 631–642.

Chinnusamy, V., K. Schumaker, and J. K. Zhu. 2004. Molecular genetic perspectives on cross-talk and specificity in abiotic stress signaling in plants. *Journal of Experimental Botany* 55: 225–236.

Clark, J. W., and P. C. Donoghue. 2017. Constraining the timing of whole genome duplication in plant evolutionary history. *Proceedings of the Royal Society, B, Biological Sciences* 284: 20170912.

Clark, J. W., and P. C. Donoghue. 2018. Whole-genome duplication and plant macroevolution. *Trends in Plant Science* 23: 933–945.

Davidson, N. M., and A. Oshlack. 2014. Corset: Enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biology* 15: 1–14.

Degenhardt, R. F., and P. C. Bonham-Smith. 2008. Transcript profiling demonstrates absence of dosage compensation in *Arabidopsis* following loss of a single RPL23a paralog. *Planta* 228: 627–640.

Degnan, J. H., and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution* 24: 332–340.

Dohm, J. C., C. Lange, Holtgräwe, D., Sörensen, T. R., D. Borchardt, B. Schulz, H. Lehrach, et al. 2012. Palaeohexaploid ancestry for Caryophyllales inferred from extensive gene-based physical and genetic mapping of the sugar beet genome (*Beta vulgaris*). *Plant Journal* 70: 528–540.

Du, Z. Y., S. Xiao, Q. F. Chen, and M. L. Chye. 2010. Depletion of the membrane-associated acyl-coenzyme A-binding protein ACBP1 enhances the ability of cold acclimation in *Arabidopsis*. *Plant Physiology* 152: 1585–1597.

Eroğlu, E. H., D. Altay, Ü. Budak, and M. Esra. 2020. Karyotypic phylogeny and polyploidy variations of Paronychia (Caryophyllaceae) taxa in Turkey. *Turkish Journal of Botany* 44: 5.

Feild, T. S., Crystal N. Arens, J. A. Doyle, T. E. Dawson, and M. J. Donoghue. 2004. Dark and disturbed: a new image of early angiosperm ecology. *Paleobiology* 30: 82–107.

Felsenstein, J. 2004. Inferring phylogenies. Sinauer, Sunderland, MA, USA.

Folk, R. A., C. M. Siniscalchi, and D. E. Soltis. 2020. Angiosperms at the edge: Extremity, diversity, and phylogeny. *Plant, Cell and Environment* 43: 2871–2893.

Frajman, B., Schönswetter, P., Weiss-Schneeweiss, H., and B. Oxelman. 2018. Origin and diversification of South American polyploid *Silene* sect. *Physolychnis* (Caryophyllaceae) in the Andes and Patagonia. *Frontiers in Genetics* 9: 422259.

Frost, L., A. M. Bedoya, and L. Lagomarsino. 2024. Artifactual orthologs and the need for diligent data exploration in complex phylogenomic datasets: a museomic case study from the Andean flora. *Systematic Biology* syad076.

Fu, L., B. Niu, Z. Zhu, S. Wu, and W. Li. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28: 3150–3152.

GBIF.org. 2022. GBIF occurrence download, 31 May 2022. Website: https://doi.org/10.15468/dl.kb4zsf

Gómez-Merino, F. C., C. A. Brearley, M. Ornatowska, Abdel-Haliem, M. E., M. I. Zanor, and Mueller-Roeber, B. 2004. AtDGK2, a novel diacylglycerol kinase from *Arabidopsis thaliana*, phosphorylates 1-stearoyl-2-arachidonoyl-*sn*-glycerol and 1, 2-dioleoyl-*sn*-glycerol and exhibits cold-inducible gene expression. *Journal of Biological Chemistry* 279: 8230–8241.

Gong, Z., H. Lee, L. Xiong, A. Jagendorf, B. Stevenson, and J. K. Zhu. 2002. RNA helicase-like protein as an early regulator of transcription factors for plant chilling and freezing tolerance. *Proceedings of the National Academy of Sciences, USA* 99: 11507–11512.

Goulas, E., M. Schubert, T. Kieselbach, L. A. Kleczkowski, P. Gardeström, W. Schröder, and V. Hurry. 2006. The chloroplast lumen and stromal proteomes of *Arabidopsis thaliana* show differential sensitivity to short- and long-term exposure to low temperature. *Plant Journal* 47: 720–734.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* 29: 644.

Greenberg, A. K., and M. J. Donoghue. 2011. Molecular systematics and character evolution in Caryophyllaceae. *Taxon* 60: 1637–1652.

Haas, B. J. 2018. TransDecoder. Website: https://github.com/TransDecoder/TransDecoder

Harbaugh, D. T., M. Nepokroeff, R. K. Rabeler, J. McNeill, E. A. Zimmer, and W. L. Wagner. 2010. A new lineage-based tribal classification of the family Caryophyllaceae. *International Journal of Plant Sciences* 171: 185–198.

Hemsley, P. A., C. H. Hurst, E. Kaliyadasa, R. Lamb, M. R. Knight, De E. A. Cothi, J. Steele, et al. 2014. The *Arabidopsis* mediator complex subunits MED16, MED14, and MED2 regulate mediator and RNA polymerase II recruitment to CBF-responsive cold-regulated genes. *Plant Cell* 26: 465–484.

Hernández-Ledesma, P., W. G. Berendsohn, T. Borsch, Von S. Mering, H. Akhani, S. Arias, Castañeda-Noa, I., et al. 2015. A taxonomic backbone for the global synthesis of species diversity in the angiosperm order Caryophyllales. *Willdenowia* 45: 281–383.

Humphreys, A. M., R. Govaerts, S. Z. Ficinski, Nic E. Lughadha, and M. S. Vorontsova. 2019. Global dataset shows geography and life form predict modern plant extinction and rediscovery. *Nature Ecology and Evolution* 3: 1043–1047.

Jablonski, D. 2022. Evolvability and Macroevolution: Overview and Synthesis. *Evolutionary Biology* 49: 265–291.

Janmohammadi, M., L. Zolla, and S. Rinalducci. 2015. Low temperature tolerance in plants: changes at the protein level. *Phytochemistry* 117: 76–89.

Jarillo, J. A., A. Leyva, J. Salinas, and Martínez-Zapater, J. M. 1993. Low temperature induces the accumulation of alcohol dehydrogenase mRNA in *Arabidopsis thaliana*, a chilling-tolerant plant. *Plant Physiology* 101: 833–837.

Jeon, J., N. Y. Kim, S. Kim, N. Y. Kang, Novák, O., S. J. Ku, Cho, C, et al. 2010. A subset of cytokinin two-component signaling system plays a role in cold temperature stress response in *Arabidopsis*. *Journal of Biological Chemistry* 285: 23371–23386.

Jiao, Y., Leebens-Mack, J., S. Ayyampalayam, J. E. Bowers, M. R. McKain, J. McNeal, Rolf, M. et al. 2012. A genome triplication associated with early diversification of the core eudicots. *Genome Biology* 13: 1–14.

Jiao, Y., N. J. Wickett, S. Ayyampalayam, A. S. Chanderbali, L. Landherr, P. E. Ralph, L. P. Tomsho, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.

Jordan, G. J., and M. K. Macphail. 2003. A Middle-Late Eocene inflorescence of Caryophyllaceae from Tasmania, Australia. *American Journal of Botany* 90: 761–768.

Karger, D. N., O. Conrad, Böhner, J., T. Kawohl, H. Kreft, Soria-Auza, R. W., N. E. Zimmermann, et al. 2017. Climatologies at high resolution for the earth's land surface areas. *Scientific Data* 4: 1–20.

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Katoh, K., and D. M. Standley. 2013. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–780.

Kim, H. J., Y. Hyun, J. Y. Park, M. J. Park, M. K. Park, M. D. Kim, and J. Kim. 2004. A genetic link between cold responses and flowering time through FVE in *Arabidopsis thaliana*. *Nature Genetics* 36: 167–171.

Körner, C. 2011. Coldest places on earth with angiosperm plant life. *Alpine Botany* 121: 11–22.

Körner, C. 2016. Plant adaptation to cold climates. *F1000Research* 5.

Kozlov, A. M., D. Darriba, T. Flouri, B. Morel, and A. Stamatakis. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35: 4453–4455.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357–359.

Lee, H., Y. Guo, M. Ohta, L. Xiong, B. Stevenson, and J. K. Zhu. 2002. LOS2, a genetic locus required for cold-responsive gene transcription encodes a bi-functional enolase. *EMBO Journal* 21: 2692–2702.

Lee, B. H., A. Kapoor, J. Zhu, and J. K. Zhu. 2006. STABILIZED1, a stress-upregulated nuclear protein, is required for pre-mRNA splicing, mRNA turnover, and stress tolerance in *Arabidopsis*. *Plant Cell* 18: 1736–1749.

Lee, Y. K. 2020. Arctic plants in Svalbard. Springer, Berlin, Germany.

Li, Z. Y., Z. S. Xu, G. Y. He, G. X. Yang, M. Chen, L. C. Li, and Y. Z. Ma. 2012. A mutation in *Arabidopsis BSK5* encoding a brassinosteroid-signaling kinase protein affects responses to salinity and abscisic acid. *Biochemical and Biophysical Research Communications* 426: 522–527.

Li, Z., M. T. W. McKibben, G. S. Finch, P. D. Blischak, B. L. Sutherland, and M. S. Barker. 2021. Patterns and processes of diploidization in land plants. *Annual Review of Plant Biology* 72: 387–410.

Löytynoja, A. 2014. Phylogeny-aware alignment with PRANK. *Methods in Molecular Biology* 1079: 155–170.

Lu, C., and M. J. Hills. 2002. *Arabidopsis* mutants deficient in diacylglycerol acyltransferase display increased sensitivity to abscisic acid, sugars, and osmotic stress during germination and seedling development. *Plant Physiology* 129: 1352–1358.

Madhani, H., R. Rabeler, A. Pirani, B. Oxelman, G. Heubl, and S. Zarre. 2018. Untangling phylogenetic patterns and taxonomic confusion in tribe Caryophylleae (Caryophyllaceae) with special focus on generic boundaries. *Taxon* 67: 83–112.

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523–536.

Magallón, S., S. Gómez-Acevedo, L. L. Sánchez-Reyes, and T. Hernández-Hernández 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* 207: 437–453.

Mandáková, T., and M. A. Lysak. 2018. Post-polyploid diploidization and diversification through dysploid changes. *Current Opinion in Plant Biology* 42: 55–65.

Mirarab, S., N. Nguyen, S. Guo, L. S. Wang, J. Kim, and T. Warnow. 2015. PASTA: Ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. *Journal of Computational Biology* 22: 377–386.

Moghe, G. D., and S. H. Shiu. 2014. The causes and molecular consequences of polyploidy in flowering plants. *Annals of the New York Academy of Sciences* 1320: 16–34.

Morales-Briones, D. F., G. Kadereit, D. T. Tefarikis, M. J. Moore, S. A. Smith, S. F. Brockington, A. Timoneda, et al. 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. *Systematic Biology* 70: 219–235.

Ohno, S. 1970. Evolution by gene duplication. Springer Science and Business Media, NY, NY, USA.

Pandey, G. K., Y. H. Cheong, B. G. Kim, J. J. Grant, L. Li, and S. Luan. 2007. CIPK9: a calcium sensor-interacting protein kinase required for low-potassium tolerance in *Arabidopsis*. *Cell Research* 17: 411–421.

Patro, R., G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* 14: 417–419.

Pease, J. B., J. W. Brown, J. F. Walker, C. E. Hinchliff, and S. A. Smith. 2018. Quartet sampling distinguishes lack of support from conflicting support in the green plant tree of life. *American Journal of Botany* 105: 385–403.

POWO. 2022. Plants of the world online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet. Website: http://www.plantsoftheworldonline.org/. [accessed February 2022].

Praat, M., De I. Smet, and van Zanten, M. 2021. Protein kinase and phosphatase control of plant temperature responses. *Journal of Experimental Botany* 72: 7459–7473.

Preston, J. C., and S. R. Sandve. 2013. Adaptation to seasonality and the winter freeze. *Frontiers in Plant Science* 4: 46453.

Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5: e9490.

Pusalkar, P. K. 2015. Redefining *Thylacospermum* and a new tribe Thylacospermeae (Caryophyllaceae). *Journal of Japanese Botany* 90: 351–355.

Raunkiaer, C. 1934. The life forms of plants and statistical plant geography; being the collected papers of C. Raunkiaer. Oxford University Press, Oxford, UK.

Rice, A., L. Glick, S. Abadi, M. Einhorn, N. M. Kopelman, A. Salman-Minkov, J. Mayzel, O. Chay, and I. Mayrose. 2014. The Chromosome Counts Database (CCDB)—a community resource of plant chromosome numbers. *New Phytologist* 206: 19–26.

Sadeghian, S., S. Zarre, R. K. Rabeler, and G. Heubl. 2015. Molecular phylogenetic analysis of *Arenaria* (Caryophyllaceae: tribe Arenarieae) and its allies inferred from nuclear DNA internal transcribed spacer and plastid DNA rps16 sequences. *Botanical Journal of the Linnean Society* 178: 648–669.

Sanderson, M. J., and H. B. Shaffer. 2002. Troubleshooting molecular phylogenetic analyses. *Annual Review of Ecology and Systematics* 33: 49–72.

Sensalari, C., S. Maere, and R. Lohaus. 2022. ksrates: positioning whole-genome duplications relative to speciation events in KS distributions. *Bioinformatics* 38: 530–532.

Shen, X. X., C. T. Hittinger, and A. Rokas. 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecology and Evolution* 1: 1–10.

Smith, S. A., and O'Meara, B. C. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28: 2689–2690.

Smith, S. A., M. J. Moore, J. W. Brown, and Y. Yang. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* 15: 1–15.

Smith, S. A., J. W. Brown, and J. F. Walker. 2018a. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS One* 13: e0197433.

Smith, S. A., J. W. Brown, Y. Yang, R. Bruenn, C. P. Drummond, S. F. Brockington, J. F. Walker, et al. 2018b. Disparity, diversity, and duplications in the Caryophyllales. *New Phytologist* 217: 836–854.

Smith, S. A., Walker-Hale, N., J. F. Walker, and J. W. Brown. 2020. Phylogenetic conflicts, combinability, and deep phylogenomics in plants. *Systematic Biology* 69: 579–592.

Smith-Unna, R., C. Boursnell, R. Patro, J. M. Hibberd, and S. Kelly. 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Research* 26: 1134–1144.

Soltis, D. E., V. A. Albert, Leebens-Mack, J., C. D. Bell, A. H. Paterson, C. Zheng, D. Sankoff, et al. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* 96: 336–348.

Song, L., and L. Florea. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* 4: 1–8.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Stover, L. E., and A. D. Partridge. 1973. Tertiary and Late Cretaceous spores and pollen from the Gippsland Basin, south-eastern Australia. *Proceedings of the Royal Society of Victoria* 85: 237–286.

Suh, A., L. Smeds, and H. Ellegren. 2015. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biology* 13: e1002224.

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

te Beest, M., J. J. Le Roux, D. M. Richardson, A. K. Brysting, J. Suda, M. Kubešová, and P. Pyšek 2012. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of Botany* 109: 19–45.

Teige, M., E. Scheikl, T. Eulgem, R. Doczi, K. Ichimura, K. Shinozaki, J. Dangl, and H. Hirt. 2004. The MKK2 pathway mediates cold and salt stress signaling in *Arabidopsis*. *Molecular Cell* 15: 141–152.

Thomas, P. D., M. J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, K. Diemer, et al. 2003. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Research* 13: 2129–2141.

Thomas, W. C. G., S. H. Ather, and M. W. Hahn. 2017. Gene-tree reconciliation with MUL-trees to resolve polyploidy events. *Systematic Biology* 66: 1007–1018.

Tiley, G. P., M. S. Barker, and J. G. Burleigh. 2018. Assessing the performance of Ks plots for detecting ancient whole genome duplications. *Genome Biology and Evolution* 10: 2882–2898.

Van Dongen, S. 2000. A cluster algorithm for graphs. National Research Institute for Mathematics and Computer Science, NWO-I, Netherlands.

Walker, J. F., Y. Yang, M. J. Moore, J. Mikenas, A. Timoneda, S. F. Brockington, and S. A. Smith. 2017. Widespread paleopolyploidy, gene tree conflict, and recalcitrant relationships among the carnivorous Caryophyllales. *American Journal of Botany* 104: 858–867.

Walker, J. F., J. W. Brown, and S. A. Smith. 2018a. Analyzing contentious relationships and outlier genes in phylogenomics. *Systematic Biology* 67: 916–924.

Walker, J. F., Y. Yang, T. Feng, A. Timoneda, J. Mikenas, V. Hutchison, C. Edwards, et al. 2018b. From cacti to carnivores: Improved phylotranscriptomic sampling and hierarchical homology inference provide further insight into the evolution of Caryophyllales. *American Journal of Botany* 105: 446–462.

Wang, N., Y. Yang, M. J. Moore, S. F. Brockington, J. F. Walker, J. W. Brown, B. Liang, et al. 2019. Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments. *Molecular Biology and Evolution* 36: 112–126.

Wang, X., C. Shen, P. Meng, G. Tan, and L. Lv. 2021. Analysis and review of trichomes in plants. *BMC Plant Biology* 21: 1–11.

Wilkinson, M. 1996. Majority-rule reduced consensus trees and their use in bootstrapping. *Molecular Biology and Evolution* 13: 437–444.

Willis, K. J., and J. C. McElwain. 2014. The evolution of plants, 2nd ed. Oxford University Press, Oxford, UK.

Woodward, F. I., M. R. Lomas, and C. K. Kelly. 2004. Global climate and the distribution of plant biomes. *Philosophical Transactions of the Royal Society of London, B, Biological Sciences* 359: 1465–1476.

Wu, S., B. Han, and Y. Jiao. 2020. Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Molecular Plant* 13: 59–71.

Xia, Z., M. Wang, and Z. Xu. 2018. The maize sulfite reductase is involved in cold and oxidative stress responses. *Frontiers in Plant Science* 9: 1680.

Xu, B., D. Luo, Z. M. Li, and H. Sun. 2019. Evolutionary radiations of cushion plants on the Qinghai-Tibet Plateau: insights from molecular phylogenetic analysis of two subgenera of *Arenaria* and *Thylacospermum* (Caryophyllaceae). *Taxon* 68: 1003–1020.

Xu, F. Q., and H. W. Xue. 2019. The ubiquitin-proteasome system in plant responses to environments. *Plant, Cell and Environment* 42: 2931–2944.

Yang, Y., M. J. Moore, S. F. Brockington, J. Mikenas, J. Olivieri, J. F. Walker, and S. A. Smith. 2018. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytologist* 217: 855–870.

Yang, Y., M. J. Moore, S. F. Brockington, D. E. Soltis, G. K. S. Wong, E. J. Carpenter, Y. Zhang, et al. 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. *Molecular Biology and Evolution* 32: 2001–2014.

Yang, Y., M. J. Moore, S. F. Brockington, A. Timoneda, T. Feng, H. E. Marx, J. F. Walker, et al. 2017. An efficient field and laboratory workflow for plant phylotranscriptomic projects. *Applications in Plant Sciences* 5: 1600128.

Yang, Y., and S. A. Smith. 2013. Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. *BMC Genomics* 14: 1–11.

Yang, Y., and S. A. Smith. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution* 31: 3081–3092.

Zanne, A. E., D. C. Tank, W. K. Cornwell, J. M. Eastman, S. A. Smith, R. G. Fitzjohn, D. J. McGlinn, et al. 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506: 89–92.

Zhang, J. 2003. Evolution by gene duplication: an update. *Trends in Ecology and Evolution* 18: 292–298.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 15–30.

Zhang, L., S. Wu, X. Chang, X. Wang, Y. Zhao, Y. Xia, R. N. Trigiano, et al. 2020. The ancient wave of polyploidization events in flowering plants and their facilitated adaptation to environmental stress. *Plant, Cell and Environment* 43: 2847–2856.

Zhu, J., J. C. Jeong, Y. Zhu, I. Sokolchik, S. Miyazaki, J. K. Zhu, and R. A. Bressan. 2008. Involvement of *Arabidopsis* HOS15 in histone deacetylation and cold tolerance. *Proceedings of the National Academy of Sciences, USA* 105: 4945–4950.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Table of all newly sequenced transcriptomes. Newly sequenced taxa used for the analyses, number of sequences, RNA extraction method, method of sequencing, collection information, and voucher. Sample ID and scientific names for transcriptomic data sets published by Walker et al. (2018b) are also listed.

Appendix S2. Figure S1 is an alternative version of Figure 1. This dix also includes bellerophon conflict results ordered by highest to lowest gene tree conflicts, only listing the top 5 alternative relationships.

Appendix S3. Figures S2–S4 display all Ks plots and ksrates results.

Appendix S4. Figure S5 displays GRAMPA results.

Appendix S5. Figure S6 shows example gene family trees to discuss the nature of WGD1.

Appendix S6. Tables S1–S4 for full lists of PANTHER overrepresentation test results.

How to cite this article: Feng, K., J. F. Walker, H. E. Marx, Y. Yang, S. F. Brockington, M. J. Moore, R. K. Rabeler, and S. A. Smith. 2024. The link between ancient whole-genome duplications and cold adaptations in the Caryophyllaceae. *American Journal of Botany* 111(8): e16350. https://doi.org/10.1002/ajb2.16350

15372197, 2024, 8, Downloaded from https://bsapubs.onlinelibrary.wiley.com/doi/10.1002/ajb2.16350, Wiley Online Library on [04/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License