


Chromosome-Scale Genome Assembly of *Gilia yorkii* Enables Genetic Mapping of Floral Traits in an Interspecies Cross

David E. Jarvis ¹, Peter J. Maughan¹, Joseph DeTemple², Veronica Mosquera², Zheng Li³, Michael S. Barker⁴, Leigh A. Johnson², and Clinton J. Whipple^{2,*}

¹Plant and Wildlife Sciences Department, Brigham Young University, USA

²Biology Department, Brigham Young University, USA

³Department of Integrative Biology, University of Texas, Austin, USA

⁴Department of Ecology and Evolutionary Biology, University of Arizona, USA

*Corresponding author: E-mail: whipple@byu.edu.

Accepted: January 26, 2022

Abstract

Substantial morphological variation in land plants remains inaccessible to genetic analysis because current models lack variation in important ecological and agronomic traits. The genus *Gilia* was historically a model for biosystematics studies and includes variation in morphological traits that are poorly understood at the genetic level. We assembled a chromosome-scale reference genome of *G. yorkii* and used it to investigate genome evolution in the Polemoniaceae. We performed QTL (quantitative trait loci) mapping in a *G. yorkii* × *G. capitata* interspecific population for traits related to inflorescence architecture and flower color. The genome assembly spans 2.75 Gb of the estimated 2.80-Gb genome, with 96.7% of the sequence contained in the nine largest chromosome-scale scaffolds matching the haploid chromosome number. *Gilia yorkii* experienced at least one round of whole-genome duplication shared with other Polemoniaceae after the eudicot paleohexaploidization event. We identified QTL linked to variation in inflorescence architecture and petal color, including a candidate for the major flower color QTL—a tandem duplication of flavanol 3',5'-hydroxylase. Our results demonstrate the utility of *Gilia* as a forward genetic model for dissecting the evolution of development in plants including the causal loci underlying inflorescence architecture transitions.

Key words: evolution of development, genome assembly, *Gilia*, inflorescence architecture, QTL mapping, whole-genome duplication.

Significance

Forward genetics with model species has made substantial contributions to our understanding of the genes regulating plant growth and development; however, these model species represent only a small fraction of existing plant diversity. Here, we report the development of genetic and genomic resources in *Gilia yorkii*, including a high-quality genome assembly and interspecies mapping population. These resources will facilitate the identification of genes underlying diversity in important plant traits, including floral and inflorescence architecture.

Introduction

The broad adoption of a few model organisms in genetics and molecular biology has been highly productive for diverse fields of biology. In plants, sustained focus on all aspects of *Arabidopsis thaliana* (*Arabidopsis*) informs research in a diversity of species and problems beyond that

narrow model (Woodward and Bartel 2018). However, *Arabidopsis* lacks important agronomic and ecological traits. Even apparently conserved traits can diverge at the genetic level via developmental system drift (True and Haag 2001), underscoring the need for comparative work outside of these established models. Judiciously

© The Author(s) 2022. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

expanding the suite of plant species used for genetic studies remains imperative, particularly in fields like evolution of development (evo–devo), where understanding the emergence and change of developmental networks requires comparative work (Wagner 2014). Despite efforts to develop phylogenetically diverse land plant models for evo–devo (Rensing et al. 2008; Kramer 2009; Banks et al. 2011; Bowman et al. 2017; Yuan 2019), considerable plant morphological diversity remains inaccessible because no species polymorphic for these traits are currently subjects of genetic analysis.

Historically, genome size limited potential model species, but advances in high-throughput sequencing are making assembly of large complex genomes routine (Zimin et al. 2017; Hufford et al. 2021), opening a much larger suite of possible species. An important remaining constraint is the ability to employ forward genetics to move beyond the limitations of candidate gene studies. In the context of evo–devo, a forward genetic approach involves crossing closely related species that are polymorphic for interesting morphological traits to identify quantitative trait loci (QTL) and eventually the causative polymorphisms underlying morphological evolution (Colosimo et al. 2005; Doebley et al. 2006; Jeong et al. 2008). Harnessing the power of forward genetics in wild species requires that they not only be polymorphic for traits of interest, but also capable of producing a fertile hybrid so that segregating mapping populations can be generated.

The ability to make wide hybrids is common in many plant lineages, and pioneering work examining the genetic basis of ecological and morphological differences among closely related species (Clausen and Hiesey 1958; Grant 1981) can now be revisited with the increased resolution of modern genomic and molecular approaches. Among these early studies, work on the leafy-stemmed gillias (*Gilia* section *Gilia*, Polemoniaceae) stands out (Grant and Grant 1956). Leafy-stemmed gillias are mostly diploid annuals native to western North and South America, grow readily in greenhouse conditions, hybridize easily, and are polymorphic for a diverse array of morphological traits (Grant 1954). We have chosen two closely related leafy-stemmed *Gilia* species that form a fertile hybrid for forward genetic analysis, the broadly dispersed *G. capitata* and the narrow endemic *G. yorkii*. These species differ in multiple morphological features including flower color and inflorescence architecture.

Here, we propose *G. yorkii* and *G. capitata* as a promising new model for the evolution of development. We have generated a chromosome-scale genome assembly for *G. yorkii* which facilitated an analysis of whole-genome duplication (WGD) history in Ericales and Polemoniaceae and identification of QTL for morphological traits that distinguish these species.

Results

Genome Assembly and Annotation

A chromosome squash revealed that metaphase-stage mitotic cells of *G. yorkii* contain 18 chromosomes ($2n = 2x = 18$) (fig. 1a and [supplementary fig. S1, Supplementary Material online](#)). Using a kmer analysis, we estimated the *G. yorkii* genome to be 2.801 Gb, with 75% estimated to be repetitive sequences. The sequenced plant exhibited low heterozygosity (0.133%; [supplementary fig. S2, Supplementary Material online](#)). The genome was sequenced with PacBio to approximately 67× coverage, and sequencing reads were assembled into 3,947 contigs ([table 1](#)). The contig assembly was polished with the PacBio and Illumina reads to correct SNP and InDel errors and then combined into 2,043 scaffolds ([table 1](#)) using Hi-C sequencing reads. The scaffolded assembly is chromosome scale, with 96.7% of the total sequence length assembled into the nine largest scaffolds (fig. 1b), corresponding to the haploid chromosome number of *G. yorkii*. The pattern of chromosomal interactions in the ordered Hi-C heatmap was characteristic of full-length chromosomes in which arms of a given chromosome interact with each other ([supplementary fig. S3, Supplementary Material online](#)) in the previously described Rabl conformation (Cowan et al. 2001). In addition, seven of the nine chromosome-scale scaffolds have the canonical telomeric repeat sequences (TTTAGGG) in at least one end of the chromosome, with three containing the telomeric sequences at both ends, indicative of a largely complete chromosome-scale assembly (fig. 1b). A benchmarking of universal, single-copy orthologs (BUSCO) analysis also indicated that the assembly is largely complete, with 97.6% of BUSCOs being identified in the scaffolded assembly and 96.8% identified as complete ([table 1](#)). As expected for a diploid, most BUSCOs (92.4%) were identified as single copy. Given these results, we refer to the largest nine scaffolds hereafter as chromosomes, numbered from largest (Gy1) to smallest (Gy9).

In total, 75.60% of the assembly was annotated as repetitive ([supplementary table S2, Supplementary Material online](#)), in agreement with the kmer-based prediction. Long-terminal repeat (LTR) retrotransposons of the Class I transposable elements are the most common repetitive sequences in the genome, accounting for 45.81% of the entire genome assembly, with *Copia* LTRs alone representing 29.43% of the genome. A large proportion of the genome (21.92%) was annotated as unknown repetitive sequences, suggesting that many of the repetitive sequences in *G. yorkii* are family-, genus-, or species-specific.

Genes were annotated in the *G. yorkii* assembly using transcript evidence from PacBio Iso-Seq data assembled into a final mapped transcriptome data set consisting of 96,691 full-length transcripts. An analysis of BUSCO genes indicated that the transcriptome assembly was largely complete, with only slightly fewer BUSCOs identified in the transcriptome

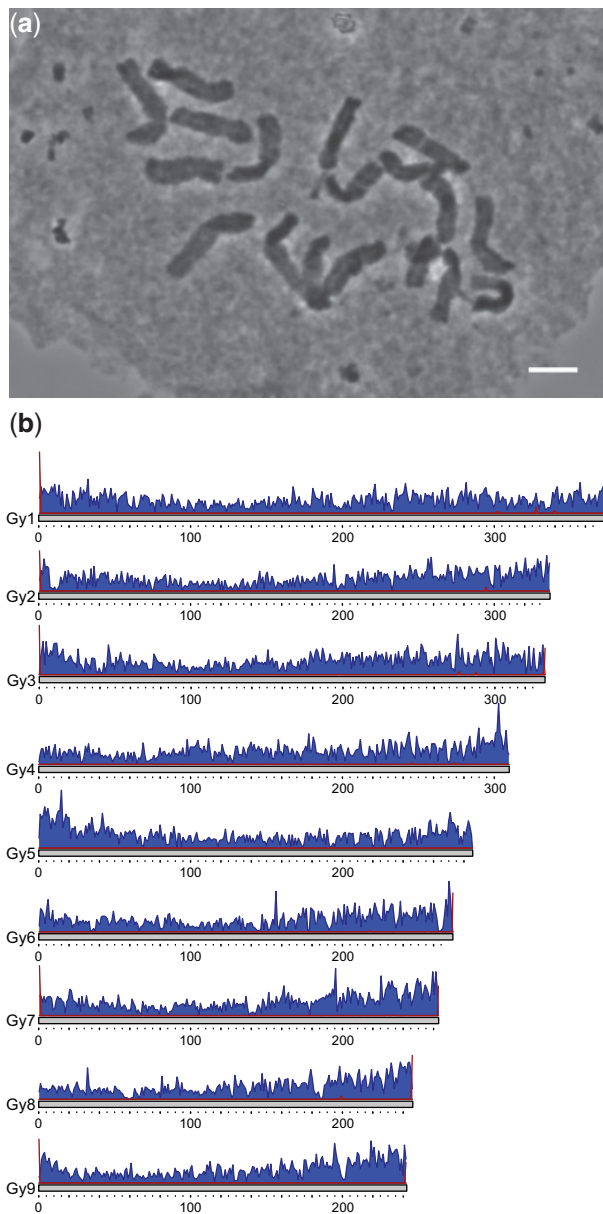


FIG. 1.—Chromosomes of the *Gilia yorkii* genome. (a) Squash of metaphase staged root meristematic cells of *G. yorkii* showing $2n = 18$ chromosomes. Scale bar = 5 μm . (b) Chromosome-scale scaffolds of the genome assembly. The distribution of canonical telomeric repeat sequences (red) and genes (blue) is shown for the nine *G. yorkii* chromosomes; all other scaffolds combined represent only 3.3% of the total sequence length and are not shown here. x axis = Mb.

assembly (96.1%) than in the scaffolded genome assembly (96.8%; table 1). However, unlike in the genome assembly, in which most complete BUSCOs were identified as single copy (92.4%), most BUSCOs in the transcriptome assembly were identified as duplicates (79.2%), indicating the extensive occurrence of alternative gene splicing. This was not unexpected, given the large number of total assembled full-length transcript isoforms. The *G. yorkii* genome was annotated with

43,969 genes. Final gene models were strongly supported by the evidence used for annotation, with 80% of the genes having annotation edit distance values less than 0.4 (supplementary fig. S4, Supplementary Material online). Genes are distributed throughout all nine chromosomes, with high density near the ends of chromosomes that generally decreases with increasing distance from the ends (fig. 1b).

Whole-Genome Duplication in *G. yorkii*

Duplicated, collinear genes are distributed among all chromosomes in the *G. yorkii* genome, in many cases occurring in long stretches of collinear genes (fig. 2a). To assess whether these duplications were consistent with ancient WGDs in *G. yorkii*, we analyzed the age distribution of gene duplications. The mixture model supported a single peak consistent with WGD around $K_s \sim 1.07$ (named GILLx, fig. 2b, supplementary table S3, Supplementary Material online). Comparison of this gene age distribution to transcriptomes of two additional members of the temperate subfamily of the Polemoniaceae, *Saltugilia latimeri* and *Phlox cuspidata* (Larson et al. 2020), revealed that both species show a similar peak of duplication at $K_s \sim 1$ (fig. 2b). The ortholog divergences among *S. latimeri*, *P. cuspidata*, and *G. yorkii*, were $K_s = 0.2\text{--}0.7$ (fig. 2c and supplementary table S4, Supplementary Material online). This suggests that the ancient WGD is likely shared by at least the temperate subfamily of Polemoniaceae. We then compared *G. yorkii* with the more distantly related species *Diospyros oleifera* (Ebenaceae) (Suo et al. 2020) in the same order (Ericales). The *D. oleifera* genome has a peak of duplication at $K_s \sim 0.68$ (fig. 2b), but *G. yorkii* and *D. oleifera* diverged before this duplication event with a median ortholog divergence of $K_s = 0.94$ (fig. 2c). This suggests that the ancient WGD in Polemoniaceae may not be shared with Ebenaceae or other more distantly related families of Ericales.

To test the hypothesized phylogenetic placement of this ancient WGD, we used the Multi-tAxon Paleopolyploidy Search (MAPS) approach (Li et al. 2018) with the *G. yorkii* genome; the *S. latimeri* and *P. cuspidata* transcriptomes; and transcriptomes or genomes of *Fouquieria macdougalii* (Leebens-Mack et al. 2019), *D. oleifera*, *Cornus florida* (Leebens-Mack et al. 2019), and *Solanum lycopersicum* (Hosmani et al. 2019) as outgroups. We found one significant burst of shared gene duplication that was statistically consistent with our simulations of WGD in the ancestry of the three Polemoniaceae species (fig. 2d and supplementary table S5, Supplementary Material online). However, we do not find any significant episodes of gene duplication at other nodes in the phylogeny.

To confirm that *G. yorkii* has undergone a single round of WGD after the ancient eudicot hexaploidization event, we compared the syntenic depth ratio between the genomes of *G. yorkii* and *Vitis vinifera*. We observed an overall two-

Table 1Statistics of the *Gilia yorkii* Contig, Scaffold, and Transcriptome Assemblies

	Genome Contig Assembly	Genome Scaffold Assembly	Transcriptome Assembly
Length (Mb)	2,753	2,754	192.63
% of estimated genome size	98.29	98.32	NA
Contigs/scaffolds	3,947	2,043	96,691
Contig/scaffold N50 (Mb)	2.54	285.77	0.0022
Contig/scaffold L50	317	5	31,532
Longest contig/scaffold (Mb)	18.30	374.98	0.0099
Complete BUSCO genes (%)	1,330 (96.7)	1,331 (96.8)	1,322 (96.1)
Single copy (%)	1,264 (91.9)	1,271 (92.4)	233 (16.9)
Complete, duplicate (%)	66 (4.8)	60 (4.4)	1,089 (79.2)
Fragmented BUSCO genes (%)	14 (1.0)	11 (0.8)	12 (0.9)
Missing BUSCO genes (%)	31 (2.3)	33 (2.4)	41 (3.0)
Total BUSCO genes searched	1,375	1,375	1,375

to-one syntenic depth ratio between *G. yorkii* and *V. vinifera* (supplementary fig. S5, Supplementary Material online). We further compared the *G. yorkii* genome with the other Ericales genomes from *Actinidia eriantha* (Tang et al. 2019), *C. sinensis* (Xia et al. 2020), *D. oleifera*, and *Rhododendron simsii* (Yang et al. 2020). We found a two-to-two syntenic depth ratio to *C. sinensis*, *D. oleifera*, and *R. simsii* (fig. 2e and supplementary fig. S5, Supplementary Material online), but a two-to-four ratio with *A. eriantha* (fig. 2f), consistent with *Actinidia* having experienced an additional round of ancient WGD compared with other Ericales genomes (Shi et al. 2010; Huang et al. 2013; Wu et al. 2019). For comparison to *G. yorkii*, we also analyzed the syntenic depth ratios between *V. vinifera* and *A. eriantha*, *C. sinensis*, *D. oleifera*, or *R. simsii*. In *A. eriantha*, we found a four-to-one syntenic depth ratio to *V. vinifera* (supplementary fig. S5, Supplementary Material online). In *C. sinensis*, *D. oleifera*, and *R. simsii*, we recovered a two-to-one ratio to *V. vinifera* (supplementary fig. S5, Supplementary Material online). Combined with other results above, the syntenic evidence confirms a single round of ancient WGD in *G. yorkii* after the eudicot ancient hexaploidization event.

Gilia yorkii and *G. capitata* Are Morphologically Distinct Species that Form a Fertile F1 Hybrid

We used the chromosome-scale *G. yorkii* genome assembly to investigate the genetic basis of morphological traits that are polymorphic with *G. capitata*. *Gilia yorkii* is a narrow endemic in the Monarch Wilderness (CA) (Shevock and Day 1998). The inflorescence of *G. yorkii* is an open panicle with white flowers, typically producing only one solitary lateral flower before the production of a terminal flower (fig. 3). Unlike *G. yorkii*, *G. capitata* is a morphologically diverse species, with eight described subspecies distributed throughout the Pacific coast of North America (Grant 1950). The tight capitata inflorescence composed of up to 80 blue flowers forms a robust landing pad on which larger pollinating insects can alight (Grant and Grant 1965). The *G. capitata* inflorescence is a

compound raceme which can produce ≥ 30 solitary lateral flowers before the apical meristem is consumed in a terminal flower (fig. 3). The tight capitata structure of this raceme results in part from the failure of internodes and pedicels to elongate (fig. 3b and c).

The phylogenetic relationships of leafy-stemmed gilies are poorly resolved, but current work suggests *G. yorkii* and the morphologically distinct *Gilia achilleifolia* are closely allied to a subset of *G. capitata* which is likely a polyphyletic taxon (Johnson 2007; Johnson and Porter 2017). We crossed *G. yorkii* to *G. capitata* from three different populations spanning multiple subspecies (see methods). All crosses produced vigorous hybrid F1 progeny, but two of these crosses resulted in sterile hybrids. One isolate, *G. capitata* (Cow Canyon), produced a vigorous and fertile F1 progeny. Nine chromosome pairs ($2n = 2x = 18$) were observed in meiotic pollen from these fertile hybrids (supplementary fig. S6, Supplementary Material online). No obvious lagging or bridge chromosomes were observed.

In addition to inflorescence architecture and flower color, *G. capitata* (Cow Canyon) was polymorphic with *G. yorkii* for many additional morphological features including leaf shape, floral shape (corolla size and shape, stamen exertion), and calyx pubescence (wooly vs. stipitate glandular, not shown), as well as breeding system, with *G. capitata* being self-incompatible whereas *G. yorkii* is self-compatible. The F1 hybrid was self-compatible, and largely intermediate in all polymorphic characters (fig. 3).

QTL Mapping Reveals Candidate Genes for Major Effect Flower Color Locus and Complex Inheritance of Inflorescence Architecture

A population of 189 F2 individuals was generated from the selfing of a single F1 *G. yorkii* \times *G. capitata* hybrid. Petal color (PC) was scored on a scale of 1–7, with one having no color (white) and seven having saturated blue/purple petals. All *G. yorkii* individuals displayed a score of 1, whereas all *G. capitata* individuals displayed a score of 7. PC of six F1

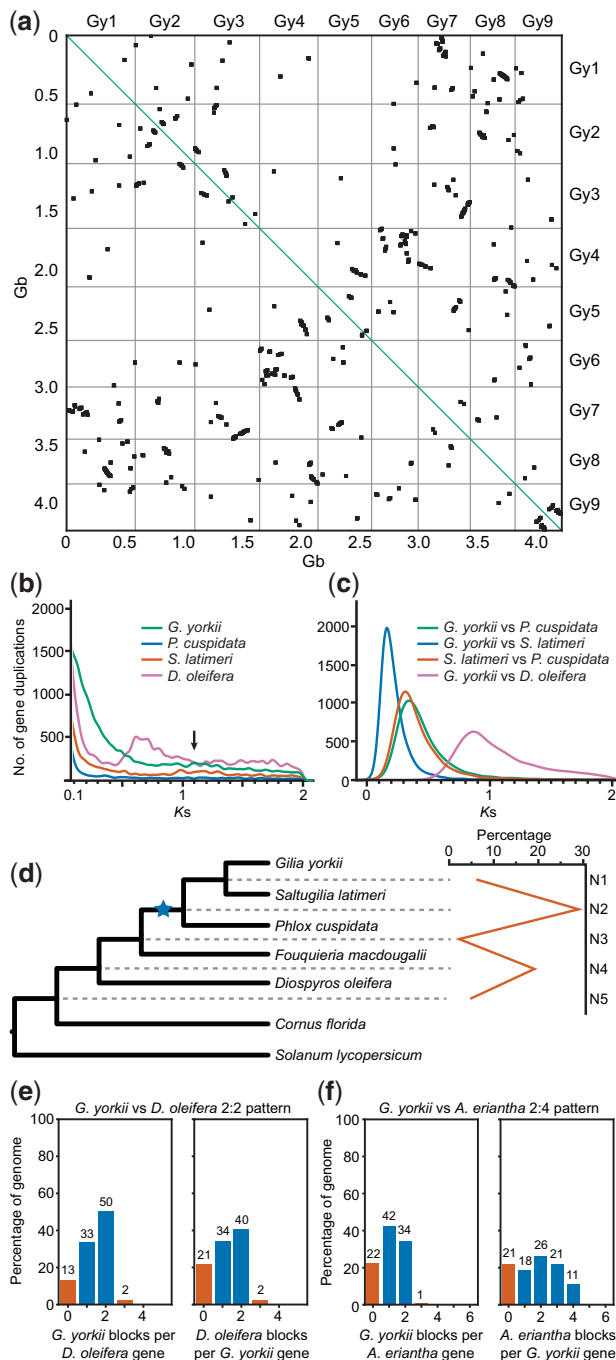


FIG. 2.—Whole-genome duplication in the *Gilia yorkii* genome. (a) Dotplot visualization of blocks of collinear duplicate genes (black dots) in chromosomes of the *G. yorkii* genome. (b, c) Number of gene duplications as a function of divergence (K_s) for gene duplications identified within (b) and between (c) species. Black arrow in (b) indicates the K_s peak in *G. yorkii*. (d) MAPS analysis on the species tree including the percentage of subtrees that contain a gene duplication on nodes N1–N5. Blue star indicates the Polemoniaceae-specific WGD on node N2. (e, f) Syntenic depth ratios between *G. yorkii* and *Diosphyros oleifera* (e) or *Actinidia eriantha* (f).

individuals was 5, intermediate between the parents. The F2 distribution of PC had distinct peaks at 1 and 5 (fig. 4a). This bimodal distribution is consistent with one to a few major loci controlling PC, and in combination with the intermediate F1 phenotype suggests that the *G. capitata* blue allele(s) are semidominant to the white *G. yorkii*.

Solitary flower number (SFN) refers to the number of solitary flowers below the terminal flower of the inflorescence and is a quantitative measure of inflorescence type with racemes having a high SFN and panicles a low SFN. Given the presence of both terminal and lateral inflorescences in all parental and F2 progeny, we scored SFN only for the terminal inflorescence in all individuals. In the inflorescence of *G. yorkii*, SFN was always 1, but ranged between 20 and 70 in *G. capitata*, with an average of 36. The F1 was intermediate between the parents, though strongly skewed toward *G. yorkii*, with a SFN average of 8. The F2 distribution was normal and centered around 8 or 9, with a range from 1 to 15 (fig. 4b). No F1 or F2 individuals approached the average SFN for *G. capitata*. The normal distribution of phenotypes suggests that SFN is controlled by multiple loci of smaller effect.

We used all 189 F2 individuals to generate a linkage map comprised 5,335 SNP markers on nine linkage groups. Positions of markers on the genetic map were highly colinear with their physical positions (supplementary fig. S7a, Supplementary Material online). A single marker on Gy8 was incongruent between the physical and genetic maps and was removed from further analyses. Although the majority of the physical map was represented in the genetic map, large regions of Gy3, Gy4, and Gy5 had no corresponding markers on the genetic map (supplementary fig. S7b, Supplementary Material online). The nonrandom filtering of these linked markers could be consistent with segregation distortion associated with either the self-incompatibility system of *G. capitata* or other genetic incompatibilities generated by the wide cross.

We performed a preliminary single-QTL analysis for the traits PC and SFN. PC is inherited as a relatively simple trait where candidate genes from the anthocyanin biosynthesis pathway are well characterized (Saigo et al. 2020). On the other hand, SFN, a quantitative measure of inflorescence type (panicle vs. raceme), correlates with an important morphological transition that is less well studied and likely involves a more complex set of loci.

With a threshold LOD score of 4.16 ($P = 0.05$) PC showed a single strong peak on Gy1 (LOD 30.0, $P = 4.3e-32$), with a minor peak on Gy4 (LOD 5.4, $P = 3.2e-7$), suggesting a single major QTL with a possible modest effect modifier (fig. 4c). The presence of a single major QTL regulating PC is consistent with the bimodal distribution of PC in the F2 population. An investigation of phenotypes conditioned on the most significant markers for these two peaks showed no evidence of epistasis (supplementary fig. S8, Supplementary Material online). We investigated annotated genes near the large-effect

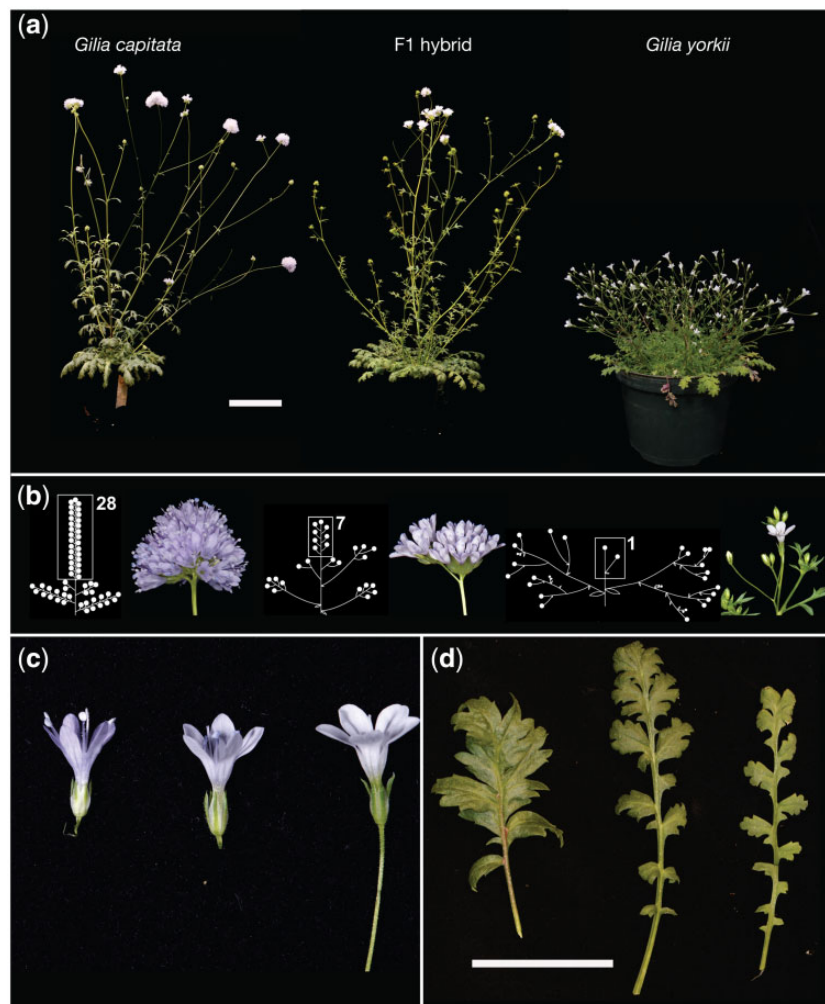


FIG. 3.—Morphological polymorphism in *Gilia capitata* and *Gilia yorkii*. (a) Plant habit for *G. capitata* (left), *G. yorkii* (right), and the F1 hybrid of these two parental species (center). Scale bar = 10 cm. (b) Terminal inflorescence of *G. capitata* (left), *G. yorkii* (right), and the F1 hybrid (center), with branching diagram to the left of each inflorescence. The box and number indicate the terminal SFN which is a count of all solitary axillary flowers produced before the terminal flower, a quantitative measure of racemose (high SFN) versus paniculate (low SFN) inflorescence architecture. (c) Flower and (d) basal leaf morphology of *G. capitata* (left), *Gilia yorkii* (right) and the F1 (center). Scale bar in (d) = 5 cm.

QTL on Gy1 and found a tandem duplication of a flavanoid 3',5'-hydroxylase (F3'5'H) gene, an essential enzyme in the anthocyanin biosynthesis pathway shown previously to regulate flower color in other species (Holton et al. 1993; Zabala and Vodkin 2007; Hopkins and Rausher 2011; Smith and Rausher 2011; Moreau et al. 2012; Wessinger and Rausher 2015). No obvious deleterious mutations were observed in these genes, although a promoter or other regulatory mutation could not be ruled out. Although the causative polymorphism is still unclear, one or both members of this F3'5'H duplication are strong candidates for the loss of anthocyanin in *G. yorkii*.

In contrast to PC, SFN had a much more complex inheritance pattern, with significant peaks on several chromosomes, including Gy1, Gy3, Gy5, and Gy9 (fig. 4d),

suggesting that the shift in inflorescence architecture is likely to be controlled by multiple, smaller effect QTL. Further analysis will be needed to determine relative effect sizes of these loci and to identify candidate genes. Nevertheless, these results confirm that a forward genetic approach is capable of identifying loci necessary to dramatically alter inflorescence architecture and thus highlight the promise of *Gilia* as a model for evo–devo.

Discussion

Genome Assembly and WGDs

The *G. yorkii* genome assembly reported here is the first for Polemoniaceae and adds to a growing number of assemblies within Ericales. Like many flowering plants, evidence indicates

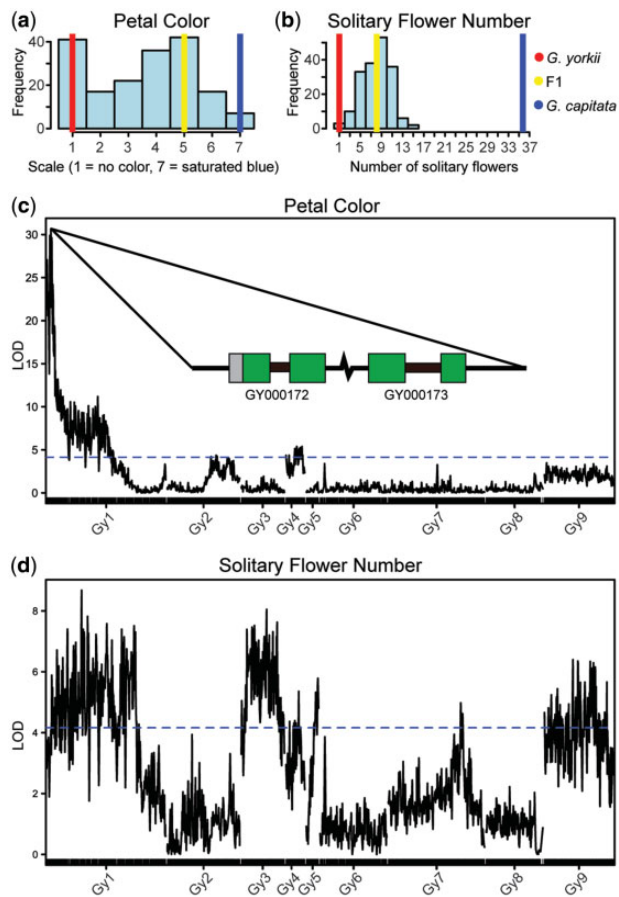


FIG. 4.—Mapping of QTL regulating flower color and inflorescence architecture. (a, b) F2 phenotypic distribution for flower color (a) and SFN (b), including average values for the parental and F1 genotypes as colored bars. (c, d) Plot of LOD scores for all markers in the genetic map for PC (c) and SFN (d). Dotted blue line indicates the significant LOD threshold. Gene models in (c) are a tandem duplicated pair of F3'5'H genes likely involved in anthocyanin biosynthesis that are strong candidate genes for the blue-white color polymorphism.

that WGDs occurred throughout the history of the Ericales. Putative WGD peaks in gene age distributions are found in all Ericales transcriptomes (Shi et al. 2010; Leebens-Mack et al. 2019; Larson et al. 2020), and multiple ancient WGDs have been confirmed by syntenic analyses (Hosmani et al. 2019; Tang et al. 2019; Larson et al. 2020; Suo et al. 2020; Xia et al. 2020; Yang et al. 2020). Recent phylogenomic analyses also suggest a WGD occurred near the origin of Ericales, although the exact timing remains unsolved (Leebens-Mack et al. 2019; Larson et al. 2020; Wang et al. 2021). Resolving WGDs in Ericales has been difficult due to substantial rate heterogeneity and rapid radiation (Leebens-Mack et al. 2019; Larson et al. 2020; Wang et al. 2021). Using our *G. yorkii* genome, we found evidence for one WGD based on synteny, K_s plot, and phylogenomic approaches. Our syntenic evidence for WGD in the Polemoniaceae confirms the putative WGD previously

found in Polemoniaceae based on transcriptomes (Leebens-Mack et al. 2019; Larson et al. 2020). Although a recent phylogenomic study suggests an ancestral Ericales WGD (Wang et al. 2021), our syntenic and phylogenomic analyses indicate that the Polemoniaceae WGD was not shared with other Ericales and the similar K_s peaks in Polemoniaceae, Fouquieriaceae, and Ebenaceae may be three independent WGDs. This conflicting evidence suggests that the exact phylogenetic placement of a single ancestral Ericales WGD or multiple independent ancient WGDs remains to be resolved and will require additional genome assemblies from key families, such as from Primulaceae and Sapotaceae.

Gilia as a Promising Evo–Devo Model

The chromosome-scale genome assembly for *G. yorkii* provides a crucial reference for functional dissection of polymorphic traits in *Gilia*. Although the Polemoniaceae includes some popular horticultural species such as *Phlox*, it lacks any food or forage species which likely accounts for the previous lack of attention at the genomic level. The diverse ecology and morphology represented by members of the genus *Gilia* have made it an attractive model for multiple evolutionary studies (Schoen 1982, 1983; Morrell and Rieseberg 1998; Morrell et al. 2000; Takebayashi et al. 2006), and the leafy-stemmed gilies in particular were a classic model for Verne and Alva Grant's pioneering work on plant evolution and speciation (Grant 1950, 1952, 1953, 1954, 1956, 1966, 1981; Grant and Grant 1956). The complete *Gilia yorkii* genome reference will facilitate re-analysis of multiple questions in this group with modern genetic approaches. Additionally, *Gilia yorkii* has many traits commonly sought in a model genetic system including relatively short annual life cycle (~3 months), self-compatibility, simple crossing, copious seed production, and minimal culturing requirements.

We show that despite striking morphological divergence, *G. yorkii* and *G. capitata* form a fertile F1 hybrid and F2 progeny that segregate the traits of their parents. This positions *Gilia* alongside other plant evo–devo models (e.g., *Erythranthe*, formerly *Mimulus*, Wu et al. 2008; *Aquilegia*, Kramer 2009; *Penstemon*, Wessinger et al. 2014; *Petunia*, Klahre et al. 2011) in which a forward genetic QTL approach can be used to identify the molecular causes of species differences. Similar approaches have proven especially informative in animal evo–devo (Shapiro et al. 2004; Jeong et al. 2008). Plants hybridize much more frequently and promiscuously than animals, opening up a host of diverse morphologies to dissect at the genetic level. With the growing ease of genome assembly and high-throughput genotyping, many other morphologically divergent but interfertile species pairs are similarly poised for development.

Preliminary Mapping of QTL for Flower Color and Inflorescence Architecture

Although hybridization is common, not all hybrids are fertile and among those that are, structural rearrangements can inhibit crossing over and thus limit the resolution of genetic mapping (Fishman and Sweigart 2018). Our preliminary interspecies genetic map recovered the majority of the physical map of *G. yorkii*. This result, in combination with the apparently normal segregation of meiotic chromosomes during microsporogenesis, suggests that crossing over is largely undisrupted between these genomes, although we cannot rule out local interference due to smaller scale rearrangements. Even with a small mapping population, the major PC locus on Gy1 mapped precisely, allowing the identification of a strong pair of candidate genes. Despite the high resolution of most chromosomes, large regions of Gy3, Gy4, and Gy5 were absent from our genetic map. These regions contained a large proportion of markers with segregation distortion that were filtered before creating the genetic map. There are many possible causes of segregation distortion (Hall and Willis 2005) and it is still unclear which of these are contributing, although loci linked to the self-incompatibility locus (or loci) from *G. capitata* were possibly selected against. In addition to distorted markers, many markers were filtered because they did not cosegregate consistently with adjacent markers. Considering the wide cross and highly repetitive nature of the *G. yorkii* genome, a genotype by sequencing (GBS) approach to identify polymorphic markers may not have been ideal. The random genome reduction resulting from GBS is necessarily biased to intergenic and repetitive regions, and thus many of the polymorphic SNPs are likely to be from repetitive regions and may not be homologous across these species. A different marker technology, such as targeted capture (Grover et al. 2012) of genes across the *G. yorkii* genome may help alleviate these concerns in future QTL mapping experiments.

We identified a tandem gene duplication as candidates for the major PC QTL. It is still unclear if this duplication is present in both *G. capitata* and *G. yorkii*, and the lack of obvious deleterious mutations makes it difficult to speculate how they may cause the loss of anthocyanin in *G. yorkii*. Thus, either gene, or perhaps even both genes together, are potentially causative. These genes encode orthologs of F3'5'H that catalyzes hydroxylation of dihydroquercetin (DHQ) to dihydromyricetin (DHM) required for the production of blue/purple anthocyanidins. In multiple studies of PC change in wild species, loss of F3'5'H activity results in a blue to red color shift (Hopkins and Rausher 2011; Smith and Rausher 2011; Wessinger and Rausher 2015), unlike the shift from blue to white seen in *G. yorkii*. It is not immediately clear why loss of F3'5'H would create a white as opposed to red flower in *Gilia*, but the PC response to mutations in the anthocyanin biosynthesis pathway can be complex (Berardi et al. 2021) and an F5'3'H ortholog in soybean mediates a purple to white

transition (Zabala and Vodkin 2007). More work will be necessary to confirm the molecular nature of the flower color change. Beyond this specific cross, blue/white flower polymorphism occurs in native accessions of *G. capitata* (Grant 1950), as well as more broadly in the leafy-stemmed gillias (Grant 1954) representing an attractive system to investigate molecular parallelism versus convergence in flower color.

Unlike flower color polymorphism, which has been addressed in multiple wild species, inflorescence architecture has received little attention (Prusinkiewicz et al. 2007; Lippman et al. 2008). Angiosperm inflorescence architecture is highly diverse (Rickett 1944; Stebbins 1973; Weberling 1989) and contributes to fecundity, making it important both ecologically and agronomically (Wyatt 1982; Fishbein and Venable 1996; Friedman and Harder 2004; Jiao et al. 2010; Bommert et al. 2013; Gao et al. 2015). Models of inflorescence architecture (Liljegren et al. 1999; Prusinkiewicz et al. 2007; Park et al. 2012) involve competitive interaction of genes that promote or delay acquisition of floral meristem identity and are based on induced mutations or natural variants in crop species that typically generate quantitative as opposed to qualitative shifts in inflorescence type. The inflorescence shift from *G. capitata* to *G. yorkii* is remarkable in that it involves changes from a raceme to a panicle, which along with cymes characterize the majority of inflorescence diversity in the angiosperms (Prusinkiewicz et al. 2007). Developmental modeling and surveys of inflorescence architecture in the angiosperms show that transitions from racemes or panicles to cymes are particularly rare, whereas transitions from racemes to panicles appear to be more common (Prusinkiewicz et al. 2007; Harder and Prusinkiewicz 2013). Unlike cymes, both panicles and racemes have an indeterminate apical meristem that produces multiple lateral meristems. Although definitions of raceme and panicle are not uniform (Rickett 1944; Prenner et al. 2009; Endress 2010), a simplified distinction can be reduced to the fate of lateral meristems. Specifically, in a raceme the lateral meristems take on a determinate floral fate immediately, whereas in a panicle, the lateral meristems maintain an indeterminate inflorescence meristem identity, only producing a determinate lateral flower after several rounds of branching. Developmental modeling indicates that the raceme panicle transition is quantitative rather than qualitative with multiple intermediate states (Prusinkiewicz et al. 2007; Harder and Prusinkiewicz 2013). In *Gilia*, SFN is a robust measure of the quantitative raceme-panicle continuum and is normally distributed in F2 segregating progeny showing a range of intermediate morphologies. SFN is regulated by at least four QTL, suggesting that multiple loci of smaller effect mediate the shift from panicle to raceme. Identifying these genes will require the development of new genetic mapping populations such as recombinant inbred and near isogenic lines.

Beyond inflorescence architecture, *G. yorkii* and *G. capitata* differ in other morphological traits that could be dissected

with a QTL-cloning approach including leaf, floral organ, and trichome morphology, and presence or absence of a self-incompatibility system. Many of these traits are also polymorphic more broadly in the leafy-stemmed gillias as well. The *G. yorkii* genome will provide a platform to dissect the genetic basis of these traits in future studies.

Materials and Methods

Plant Material and Mapping Populations

Gilia yorkii Shevock and A.G. Day was grown from seed collected from an herbarium sheet (BRY-614948). Plants were self-fertilized and an inbred line produced by single-seed descent for eight generations. *Gilia capitata* Sims was grown from seed collected in Cow Canyon near Mt. Baldy, San Bernardino Co., CA (BRY-642206), and propagated by sibling crossing. F1 hybrids were generated by emasculating and crossing *G. yorkii* × *G. capitata*. A single seed source of *G. yorkii* was used to which *G. capitata* from three different populations (*G. capitata* 9228, *G. capitata* Baja, and *G. capitata* Cow Canyon) were crossed. F1 hybrid progeny (a minimum of three) for each *G. yorkii* × *G. capitata* cross were grown to maturity. Although all F1 hybrids were vigorous, only the cross with *G. capitata* Cow Canyon produced a fertile hybrid, and these F1 plants were self-fertilized to generate an F2 population. PC was determined by averaging by eye more than five flowers from each plant and using a printed scale of graded white to purple shades as a standard. SFN was determined for only the apical inflorescence of each plant. All traits were measured on five plants each of the parent species and F1 as well as 189 F2 progeny of a single F1 parent. These same 189 F2 progeny were used for DNA isolation and QTL mapping. All plants were grown in Sungro soilless potting mix supplemented with (18 g/l) osmocote in 3- or 6-in pots using supplemental lights with a 16-h day, at 25°C day and 20°C night.

Chromosome Squashes

Mitotic root tips (1 cm) were isolated from aseptically germinated seedlings grown on Murashigie and Skoog (MS) agar, placed in ice-cold water for 24 h, fixed in Farmers Solution (ethanol:acetic acid 3:1 v/v) for 24 h (4 °C), then transferred to 70% ethanol. Meiotic pollen mother cells from anthers of immature floral buds were fixed identically to root tips. All samples were dissected under ethanol and transferred to a microscope slide with two drops of propionocarmine (1% w/v carmine in propionic acid 45% v/v) and a small amount of anhydrous iron chloride (FeCl₃) powder applied with a forceps tip. Tissue in propionocarmine was overlaid with a coverslip and gently squashed with a matchstick, heated over a flame for 2–5 s. After cooling, the cover glass was squashed by hand against filter paper. Stained chromosomes were observed with an Axioplan 2 Zeiss microscope with a 100× oil

immersion objective and imaged with an Axiocam 712 Zeiss monochromatic camera.

DNA and RNA Isolation

DNA extraction for Illumina sequencing was performed using a modified CTAB protocol (Gallavotti and Whipple 2015). High-molecular weight (HWM) DNA for PacBio sequencing was extracted from leaf tissue using the Circulomics Nanobind Plant Nuclei Big DNA Kit, following the manufacturer's protocol. RNA was extracted by grinding <1 g of flash-frozen tissue followed by extraction in 1 ml Trizol (Invitrogen). Trizol homogenate was extracted with 0.2 ml chloroform, centrifuged, and 0.2 ml of the aqueous phase was purified with a Qiagen RNEasy spin column after mixing with 0.7 ml RLT buffer and 0.7 ml ethanol. The column was washed with 0.5 ml RPE buffer, and RNA was eluted with 50 µl RNase-free water.

Sequencing

Illumina sequencing of *G. yorkii* was performed on HiSeq X by Novogene using 550-bp insert DNA libraries prepared at Brigham Young University's DNA Sequencing Center (BYUDNASC), generating a total of 161.5 Gb of 150-bp paired-end sequencing data. PacBio library preparation and sequencing of HWM *G. yorkii* DNA was performed on PacBio Sequel by the BYUDNASC. Three SMRT cells were used to generate 198 Gb of continuous long read (CLR) sequencing data.

PacBio Iso-Seq library preparation and sequencing were performed at the BYUDNASC using a single RNA sample created by pooling equal amounts of RNA extracted individually from leaf primordia, mature leaves, shoot apex, mature flowers, flower buds, and roots. One SMRT cell was used to generate 2,341,411 circular consensus sequencing reads (193.3 Mb of data).

Whole-Genome Assembly and Annotation

Illumina sequencing reads were trimmed using Trimmomatic to remove leading and trailing bases with a quality score below 20 or average per-base quality of 20 over a four-nucleotide sliding window; trimmed reads shorter than 75 nucleotides in length were removed. The frequency of kmers in the sequencing data was calculated using Jellyfish (Marçais and Kingsford 2011), and the genome size was estimated based on kmer frequency using GenomeScope (Ranallo-Benavidez et al. 2020) with kmer length = 21, read length = 150, and kmer max coverage = 1,000,000. PacBio CLR reads were assembled into contigs using Canu v1.9 (Koren et al. 2017) with coMhapSensitivity = normal, corOutCoverage = 40, and ovMerThreshold = 500. SNP and Indel correction were performed with the PacBio CLR reads in two successive rounds using the Arrow algorithm in the

PacBio SMRT Tools package v7.0, and additional polishing was performed with the Illumina reads (Indel correction only) in one round using Pilon v1.9 (Walker et al. 2014). The assembly was scaffolded using in vivo Hi-C data by Phase Genomics as previously described (Lightfoot et al. 2017), using 228,955,144 pairs of Illumina sequencing reads. Juicebox (Rao et al. 2014; Durand et al. 2016) was used to manually correct scaffolding errors.

PacBio Iso-Seq reads were assembled into a transcriptome using the IsoSeq v3 pipeline in the PacBio SMRT Tools package, including using pbmm2 to align the Iso-Seq reads to the scaffolded genome and IsoSeq v3 to collapse the transcripts. Completeness of the Hi-C scaffolded genome and transcriptome assembly were assessed using BUSCO v4 (Simão et al. 2015) with the embryophyta_odb10 data set.

Repetitive sequences were identified and classified using RepeatModeler (Smit and Hubley 2008) and RepeatMasker (Smit et al. 2013), respectively, using RepBase database version 20181026. Genes were annotated using MAKER2 (Holt and Yandell 2011) using the Iso-Seq assembly as transcript evidence. Alternative EST and protein evidence were provided by sequences from *Actinidia chinensis* (Wu et al. 2019) and *Camellia sinensis* (Wei et al. 2018). The uniprot-sprot database (downloaded September 25, 2018) was also used as alternative protein evidence. Ab initio gene annotation was performed using *Gilia*-specific AUGUSTUS (Stanke et al. 2004) gene prediction models and *Arabidopsis thaliana* SNAP (Korf 2004) gene models provided to MAKER, and tRNA genes were predicted with tRNAscan-SE (Lowe and Eddy 1997). Telomeres were identified by performing a BLAST search of four repeats of the canonical telomere sequence (TTTAGGG) (Richards and Ausubel 1988). The chromosomal positions of BLAST hits with 100% identity, as well as the chromosomal positions of annotated genes, were plotted using karyoploteR (Gel and Serra 2017).

Analysis of Ancient WGDs

The DupPipe pipeline was used to construct gene families and estimate the age distribution of gene duplications (Barker et al. 2008, 2010). DNA sequences were translated and open reading frames (ORFs) identified by comparing the Genewise (Birney et al. 2004) alignment to the best-hit protein from a collection of proteins from 25 plant genomes from Phytozome (Goodstein et al. 2012). Protein-guided DNA alignments were used to align nucleic acid sequences while maintaining the ORFs. Synonymous divergence (K_s) was estimated using PAML with the F3X4 model (Yang 2007) for nodes in the gene trees. Mixture modeling was used to identify peaks consistent with a potential WGD and estimate their median paralog K_s values. Significant peaks were identified using a likelihood ratio test in the boot.comp function of the package mixtools in R (Benaglia et al. 2009).

Estimating Orthologous Divergence

Synonymous divergence of orthologs was estimated among pairs of species that may share a WGD based on their phylogenetic position and evidence from the within-species K_s plots. The mean and median synonymous divergence of orthologs were estimated with The RBH Orthologue pipeline (Barker et al. 2010) and compared with the synonymous divergence of inferred paleopolyploid peaks. Orthologs were identified as reciprocal best BLAST hits in pairs of transcriptomes. Using protein-guided DNA alignments, the pairwise synonymous divergence for each pair of orthologs was estimated using PAML with the F3X4 model (Yang 2007).

Genome Synteny Analyses

Genomic collinearity blocks for intra- and interspecies comparisons for *G. yorkii* and other genomes were identified by MCScan (Tang et al. 2008). All-against-all LAST (Kielbasa et al. 2011) hits with a distance cutoff of ten genes were chained, requiring \geq five gene pairs per synteny block. The syntenic “depth” function in MCScan was applied to estimate the duplication history in respective genomes. Genomic synteny was visualized by the Python version of MCScan (Tang et al. 2008).

Multi-tAxon Paleopolyploidy Search Analyses

Multi-tAxon Paleopolyploidy Search (Li et al. 2015, 2018) was used to determine the timing of WGD in *G. yorkii*. Orthologous groups were obtained from OrthoFinder (Emms and Kelly 2015) using default parameters, retaining only gene families that with ≥ 1 gene copy from each taxon. Trees for 8,093 gene families constructed by PASTA (Mirarab et al. 2015) were analyzed by MAPS. Both null and positive simulations of the background gene birth and death rates were performed to compare with the observed number of duplications at each node, using WGDgc (Rabier et al. 2014) to generate gene birth (λ) and death (μ) rates for null simulations. Gene count data of each gene family were obtained from OrthoFinder (Emms and Kelly 2015). The estimated parameters ($\lambda = 0.007$; $\mu = 0.014$) were configured in MAPS, and the gene trees were simulated within the species tree using “GuestTreeGen” from GenPhyloData (Sjöstrand et al. 2013). For each species tree, 3,000 gene trees were simulated with at least one tip per species: 1,000 gene trees respectively at 0.5 \times , 1 \times , and 3 \times the estimated λ and μ , according to the 1KP analyses (Leebens-Mack et al. 2019; Li and Barker 2020). Random resampling of 1,000 trees without replacement from the total pool of gene trees 100 times provided a measure of uncertainty on the percentage of subtrees at each node. Fisher’s exact test was used to identify significant increases in gene duplication compared with a null simulation. For positive simulations, gene trees were simulated using the same methods described above. However, WGDs

were incorporated at the location in the MAPS phylogeny resulting in significantly larger numbers of gene duplications at this phylogenetic position compared with the null simulation. At least 20% of the genes were retained following the simulated WGD (Leebens-Mack et al. 2019; Li and Barker 2020).

Genotyping-by-Sequencing

Approximately 30 mg of fresh leaf tissue was lyophilized for *G. yorkii*, *G. capitata*, a newly generated F1 hybrid, and each of the 189 F2 progeny in the *G. yorkii* × *G. capitata* population. Samples were sent to Freedom Markers, and genotyped using the modified tGBS protocol (Ott et al. 2017). Libraries were prepared with the restriction enzyme Bsp1286I and sequenced using Illumina HiSeq X, generating a total of 377,694,355 pairs of reads. Reads were trimmed by removing bases with PHRED quality scores less than 15 (out of 40) using Lucy (Chou and Holmes 2001). Trimmed reads were mapped to the polished, scaffolded reference *G. yorkii* genome using GSNAP (Wu and Nacu 2010). Reads with more than two mismatches every 36 bp or more than five mismatches every 75 bp were removed prior to genotype calling.

Genetic Map

Polymorphic sites were identified as any site that differed from the reference in at least one sample. Genotypes at each site were identified using the following criteria: if the most common allele was supported by $\geq 80\%$ of aligned reads and ≥ 5 unique reads, the SNP was classified as homozygous. SNPs were considered heterozygous if each of the two most common alleles were supported by $\geq 30\%$ of aligned reads and ≥ 5 unique reads. Polymorphisms in the first and last 3 bp of reads were ignored, and a PHRED quality score of ≥ 20 was required for each SNP.

These SNPs were then filtered accordingly: minimum calling rate $\geq 50\%$, allele number = 2, number of genotypes ≥ 2 , minor allele frequency $\geq 10\%$, and heterozygosity rate range of 10–65%, resulting in 73,482 SNPs used to create a linkage map. Linkage groups were formed using the R/qtl package in R version 3.6.2 (Broman et al. 2003; Broman and Sen 2009), and markers with excessive linkage disequilibrium or recombination with adjacent markers were removed, reducing the SNPs to 5,335 in the final linkage map.

QTL Analysis

The 5,335-marker data set (supplementary table S1, Supplementary Material online) was used for single-QTL analyses using an extended Haley–Knott regression analysis in RStudio using R/qtl in R version 3.6.2. The LOD threshold of 4.16 was determined using the 95th percentile of LOD scores for all markers across all chromosomes, with 100,000 permutations.

Supplementary Material

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

Acknowledgments

This work was supported by funding from a BYU MEG award and ORCA grant to C.J.W. and the BYU Department of Biology. Jason Ross assisted with a preliminary analysis of PC in *Gilia*. Rick Jellen assisted with the chromosome squashes. Ed Wilcox and the BYU DNASC (RRID: SCR_017781) provided sequencing support. Madelaine Bartlett provided helpful comments on the manuscript.

Data Availability

The data that support the findings of this study are openly available in GenBank at <https://www.ncbi.nlm.nih.gov/genbank/> (last accessed March 3, 2022), BioProject PRJNA729797. Additional information is provided in supplementary table S6, Supplementary Material online. The *Gilia yorkii* genome assembly and annotation are available on CoGe (<https://genomeevolution.org/coge/>, last accessed March 3, 2022) under Genome ID 62042.

Literature Cited

- Banks JA, et al. 2011. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332(6032):960–963.
- Barker MS, et al. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Mol Biol Evol.* 25(11):2445–2455.
- Barker MS, et al. 2010. EvoPipes.net: bioinformatic tools for ecological and evolutionary genomics. *Evol Bioinform Online.* 6:143–149.
- Benaglia T, Chauveau D, Hunter DR, Young D. 2009. mixtools: an R package for analyzing finite mixture models. *J Stat Softw.* 32(6):1–29.
- Berardi AE, et al. 2021. Complex evolution of novel red floral color in *Petunia*. *Plant Cell* 33(7):2273–2295.
- Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. *Genome Res.* 14(5):988–995.
- Bommert P, Nagasawa N, Jackson D. 2013. Quantitative variation in maize kernel row number is controlled by the *FASCIATED EAR2* locus. *Nat Genet.* 45(3):334–337.
- Bowman JL, et al. 2017. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* 171(2):287–304.e15.
- Broman KW, Sen S. 2009. A guide to QTL mapping with R/qtl. New York: Springer.
- Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19(7):889–890.
- Chou H-H, Holmes MH. 2001. DNA sequence quality trimming and vector removal. *Bioinformatics* 17(12):1093–1104.
- Clausen J, Hiesey WM. 1958. Experimental studies on the nature of species. IV. Genetic structure of ecological races. Washington (DC): Carnegie Institution.
- Colosimo PF, et al. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307(5717):1928–1933.

- Cowan CR, Carlton PM, Cande WZ. 2001. The Polar arrangement of telomeres in interphase and meiosis. Rabl organization and the bouquet. *Plant Physiol.* 125(2):532–538.
- Doebley JF, Gaut BS, Smith BD. 2006. The molecular genetics of crop domestication. *Cell* 127(7):1309–1321.
- Durand NC, et al. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 3(1):99–101.
- Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157.
- Endress PK. 2010. Disentangling confusions in inflorescence morphology: patterns and diversity of reproductive shoot ramification in angiosperms. *J Syst Evol.* 48(4):225–239.
- Fishbein M, Venable DL. 1996. Evolution of inflorescence design: theory and data. *Evolution* 50(6):2165–2177.
- Fishman L, Sweigart AL. 2018. When two rights make a wrong: the evolutionary genetics of plant hybrid incompatibilities. *Annu Rev Plant Biol.* 69:707–731.
- Friedman J, Harder LD. 2004. Inflorescence architecture and wind pollination in six grass species. *Funct Ecol.* 18(6):851–860.
- Gallavotti A, Whipple CJ. 2015. Positional cloning in maize (*Zea mays* subsp. *mays*, Poaceae). *Appl Plant Sci.* 3(1):1400092.
- Gao F, et al. 2015. Blocking miR396 increases rice yield by shaping inflorescence architecture. *Nat Plants.* 2:15196.
- Gel B, Serra E. 2017. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* 33(19):3088–3090.
- Goodstein DM, et al. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40(Database issue):D1178–D1186.
- Grant V. 1950. Genetic and taxonomic studies in *Gilia*: I. *Gilia capitata*. *Aliso* 2:239–316.
- Grant V. 1952. Genetic and taxonomic studies in *Gilia*: II. *Gilia capitata abrotanifolia*. *Aliso* 2:361–373.
- Grant V. 1953. The role of hybridization in the evolution of the leafy-stemmed *Gilias*. *Evolution* 7:51–64.
- Grant V. 1954. Genetic and taxonomic studies in *Gilia*: VI. Interspecific relationships in the leafy-stemmed *Gilias*. *Aliso* 3(1):35–49.
- Grant V. 1956. The genetic structure of races and species in *Gilia*. *Adv Genet.* 8:55–87.
- Grant V. 1966. The selective origin of incompatibility barriers in the plant genus *Gilia*. *Am Nat.* 100(911):99–118.
- Grant V. 1981. *Plant speciation*. New York: Columbia University Press.
- Grant V, Grant A. 1956. Genetic and taxonomic studies in *Gilia*: X. Conspectus of the subgenus *Gilia*. *Aliso* 3(3):297–300.
- Grant V, Grant KA. 1965. *Flower pollination in the Phlox family*. New York: Columbia University Press.
- Grover CE, Salmon A, Wendel JF. 2012. Targeted sequence capture as a powerful tool for evolutionary analysis. *Am J Bot.* 99(2):312–319.
- Hall MC, Willis JH. 2005. Transmission ratio distortion in intraspecific hybrids of *Mimulus guttatus*: implications for genomic divergence. *Genetics* 170(1):375–386.
- Harder LD, Prusinkiewicz P. 2013. The interplay between inflorescence development and function as the crucible of architectural diversity. *Ann Bot.* 112(8):1477–1493.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491.
- Holton TA, et al. 1993. Cloning and expression of cytochrome P450 genes controlling flower colour. *Nature* 366(6452):276–279.
- Hopkins R, Rausher MD. 2011. Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*. *Nature* 469(7330):411–414.
- Hosmani PS, et al. 2019. An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. *Biorxiv.* 767764.
- Huang S, et al. 2013. Draft genome of the kiwifruit *Actinidia chinensis*. *Nat Commun.* 4:2640.
- Hufford MB, et al. 2021. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* 373:655–662.
- Jeong S, et al. 2008. The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell* 132(5):783–793.
- Jiao Y, et al. 2010. Regulation of *OsSPL14* by *OsMir156* defines ideal plant architecture in rice. *Nat Genet.* 42(6):541–544.
- Johnson LA. 2007. Transfer of the Western North American species *Gilia splendens* to *Saltugilia* (Polemoniaceae), and the taxonomic affinities of *Gilia scopulorum*, *Gilia stellata*, and *Gilia yorkii*. *Novon J Bot Nomencl.* 17(2):193–197.
- Johnson LA, Porter JM. 2017. Fates of angiosperm species following long-distance dispersal: examples from American amphitropical Polemoniaceae. *Am J Bot.* 104(11):1729–1744.
- Klahre U, et al. 2011. Pollinator choice in *Petunia* depends on two major genetic loci for floral scent production. *Curr Biol.* 21(9):730–739.
- Koren S, et al. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27(5):722–736.
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59.
- Kramer EM. 2009. *Aquilegia*: a new model for plant development, ecology, and evolution. *Annu Rev Plant Biol.* 60:261–277.
- Larson DA, Walker JF, Vargas OM, Smith SA. 2020. A consensus phylogenomic approach highlights paleopolyploid and rapid radiation in the history of Ericales. *Am J Bot.* 107(5):773–789.
- Leebens-Mack JH, et al. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 574:679–685.
- Li Z, Barker MS. 2020. Inferring putative ancient whole-genome duplications in the 1000 Plants (1KP) initiative: access to gene family phylogenies and age distributions. *GigaScience* 9(2):giaa004.
- Li Z, et al. 2015. Early genome duplications in conifers and other seed plants. *Sci Adv.* 1(10):e1501084.
- Li Z, et al. 2018. Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc Natl Acad Sci U S A.* 115(18):4713–4718.
- Lightfoot DJ, et al. 2017. Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biol.* 15(1):74.
- Liljgren SJ, Gustafson-Brown C, Pinyopich A, Ditta GS, Yanofsky MF. 1999. Interactions among *APETALA1*, *LEAFY*, and *TERMINAL FLOWER1* specify meristem fate. *Plant Cell* 11(6):1007–1018.
- Lippman ZB, et al. 2008. The making of a compound inflorescence in tomato and related nightshades. *PLoS Biol.* 6(11):e288.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25(5):955–964.
- Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6):764–770.
- Mirab S, et al. 2015. PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. *J Comput Biol.* 22(5):377–386.
- Moreau C, et al. 2012. The *b* gene of pea encodes a defective flavonoid 3',5'-hydroxylase, and confers pink flower color. *Plant Physiol.* 159(2):759–768.
- Morrell PL, Porter JM, Friar EA. 2000. Intercontinental dispersal: the origin of the widespread South American plant species *Gilia laciniata*

- (Polemoniaceae) from a rare California and Oregon coastal endemic. *PL Syst Evol.* 224(1–2):13–32.
- Morrell PL, Rieseberg LH. 1998. Molecular tests of the proposed diploid hybrid origin of *Gilia achilleifolia* (Polemoniaceae). *Am J Bot.* 85(10):1439–1453.
- Ott A, et al. 2017. tGBS[®] genotyping-by-sequencing enables reliable genotyping of heterozygous loci. *Nucleic Acids Res.* 45(21):e178.
- Park SJ, Jiang K, Schatz MC, Lippman ZB. 2012. Rate of meristem maturation determines inflorescence architecture in tomato. *Proc National Acad Sci U S A.* 109(2):639–644.
- Prenner G, Vergara-Silva F, Rudall PJ. 2009. The key role of morphology in modelling inflorescence architecture. *Trends Plant Sci.* 14(6):302–309.
- Prusinkiewicz P, Erasmus Y, Lane B, Harder LD, Coen E. 2007. Evolution and development of inflorescence architectures. *Science* 316(5830):1452–1456.
- Rabier C-E, Ta T, Ané C. 2014. Detecting and locating whole genome duplications on a phylogeny: a probabilistic approach. *Mol Biol Evol.* 31(3):750–762.
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun.* 11(1):1432.
- Rao S, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159(7):1665–1680.
- Rensing SA, et al. 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319(5859):64–69.
- Richards EJ, Ausubel FM. 1988. Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* 53(1):127–136.
- Rickett HW. 1944. The classification of inflorescences. *Bot Rev.* 10(3):187–231.
- Saigo T, Wang T, Watanabe M, Tohge T. 2020. Diversity of anthocyanin and proanthocyanin biosynthesis in land plants. *Curr Opin Plant Biol.* 55:93–99.
- Schoen DJ. 1982. The breeding system of *Gilia achilleifolia*: variation in floral characteristics and outcrossing rate. *Evolution* 36(2):352–360.
- Schoen DJ. 1983. Relative fitnesses of selfed and outcrossed progeny in *Gilia achilleifolia* (Polemoniaceae). *Evolution* 37(2):292–301.
- Shapiro MD, et al. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428(6984):717–723.
- Shevock JR, Day AG. 1998. A new *Gilia* (Polemoniaceae) from limestone outcrops in the southern Sierra Nevada of California. *Madrone* 45:137–140.
- Shi T, Huang H, Barker MS. 2010. Ancient genome duplications during the evolution of kiwifruit (*Actinidia*) and related Ericales. *Ann Bot.* 106(3):497–504.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
- Sjöstrand J, Arvestad L, Lagergren J, Sennblad B. 2013. GenPhyloData: realistic simulation of gene family evolution. *BMC Bioinformatics* 14:209.
- Smit AFA, Hubley R. 2008. RepeatModeler Open-1.0. Available from: <http://www.repeatmasker.org>. Accessed March 3, 2022.
- Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0. Available from: <http://www.repeatmasker.org>. Accessed March 3, 2022.
- Smith SD, Rausher MD. 2011. Gene loss and parallel evolution contribute to species difference in flower color. *Mol Biol Evol.* 28(10):2799–2810.
- Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* 32(Web Server issue):W309–W312.
- Stebbins GL. 1973. Evolutionary trends in the inflorescence of angiosperms. *Flora* 162(6):501–528.
- Suo Y, et al. 2020. A high-quality chromosomal genome assembly of *Diospyros oleifera* Cheng. *GigaScience* 9(1):giz164.
- Takebayashi N, Wolf DE, Delph LF. 2006. Effect of variation in herkogamy on outcrossing within a population of *Gilia achilleifolia*. *Heredity* 96(2):159–165.
- Tang H, et al. 2008. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* 18(12):1944–1954.
- Tang W, et al. 2019. Chromosome-scale genome assembly of kiwifruit *Actinidia eriantha* with single-molecule sequencing and chromatin interaction mapping. *GigaScience* 8(4):giz027.
- True JR, Haag ES. 2001. Developmental system drift and flexibility in evolutionary trajectories. *Evol Dev.* 3(2):109–119.
- Wagner GP. 2014. Homology, genes, and evolutionary innovation. Princeton (NJ): Princeton University Press.
- Walker BJ, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
- Wang Y, et al. 2021. An ancient whole-genome duplication event and its contribution to flavor compounds in the tea plant (*Camellia sinensis*). *Hortic Res.* 8(1):176.
- Weberling F. 1989. Morphology of flowers and inflorescences. Cambridge: Cambridge University Press.
- Wei C, et al. 2018. Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. *Proc Natl Acad Sci U S A.* 115(18):E4151–E4158.
- Wessinger CA, Hileman LC, Rausher MD. 2014. Identification of major quantitative trait loci underlying floral pollination syndrome divergence in *Penstemon*. *Philos Trans R Soc B.* 369(1648):20130349.
- Wessinger CA, Rausher MD. 2015. Ecological transition predictably associated with gene degeneration. *Mol Biol Evol.* 32(2):347–354.
- Woodward AW, Bartel B. 2018. Biology in bloom: a primer on the *Arabidopsis thaliana* model system. *Genetics* 208(4):1337–1349.
- Wu CA, et al. 2008. *Mimulus* is an emerging model system for the integration of ecological and genomic studies. *Heredity* 100(2):220–230.
- Wu H, et al. 2019. A high-quality *Actinidia chinensis* (kiwifruit) genome. *Hortic Res.* 6:117.
- Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26(7):873–881.
- Wyatt R. 1982. Inflorescence architecture: how flower number, arrangement, and phenology affect pollination and fruit-set. *Am J Bot.* 69(4):585–594.
- Xia E, et al. 2020. The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mol Plant.* 13(7):1013–1026.
- Yang F-S, et al. 2020. Chromosome-level genome assembly of a parent species of widely cultivated azaleas. *Nat Commun.* 11(1):5269.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.
- Yuan Y. 2019. Monkeyflowers (*Mimulus*): new model for plant developmental genetics and evo-devo. *New Phytol.* 222(2): 694–700.
- Zabala G, Vodkin LO. 2007. A rearrangement resulting in small tandem repeats in the F3'5'H gene of white flower genotypes is associated with the soybean *W1* locus. *Crop Sci.* 47(S2):S113–S124.
- Zimin AV, et al. 2017. An improved assembly of the loblolly pine megagenome using long-read single-molecule sequencing. *GigaScience* 6(10):1–4.

Associate editor: Yves Van De Peer