

<https://helda.helsinki.fi>

---

## Chromosome-level genome assembly of the diploid blueberry *Vaccinium darrowii* provides insights into its subtropical adaptation and cuticle synthesis

Cui, Fuqiang

2022-07-11

---

Cui , F , Ye , X , Li , X , Yang , Y , Hu , Z , Overmyer , K , Brosche , M , Yu , H & Salojärvi , J  
2022 , ' Chromosome-level genome assembly of the diploid blueberry *Vaccinium darrowii*  
provides insights into its subtropical adaptation and cuticle synthesis ' , *Plant  
communications* , vol. 3 , no. 4 , 100307 . <https://doi.org/10.1016/j.xplc.2022.100307>

---

<http://hdl.handle.net/10138/347896>

<https://doi.org/10.1016/j.xplc.2022.100307>

---

cc\_by\_nc\_nd

publishedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

# Chromosome-level genome assembly of the diploid blueberry *Vaccinium darrowii* provides insights into its subtropical adaptation and cuticle synthesis

Fuqiang Cui<sup>1,7,\*</sup>, Xiaoxue Ye<sup>2,3,7</sup>, Xiaoxiao Li<sup>1</sup>, Yifan Yang<sup>1</sup>, Zhubing Hu<sup>6</sup>, Kirk Overmyer<sup>4</sup>, Mikael Brosché<sup>4</sup>, Hong Yu<sup>5</sup> and Jarkko Salojärvi<sup>3,4,\*</sup>

<sup>1</sup>College of Forestry and Biotechnology, State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Lin'an, Hangzhou 311300, China

<sup>2</sup>Key Laboratory of Biology and Genetic Resources of Tropical Crops, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan 571101, China

<sup>3</sup>School of Biological Sciences, Nanyang Technological University, Singapore, Singapore

<sup>4</sup>Organismal and Evolutionary Biology Research Program, Faculty of Biological and Environmental Sciences, and the Viikki Plant Science Centre, University of Helsinki, PO Box 65 (Viikinkaari 1), 00014 Helsinki, Finland

<sup>5</sup>Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China

<sup>6</sup>State Key Laboratory of Cotton Biology, Department of Biology, Institute of Plant Stress Biology, Henan University, Kaifeng, China

<sup>7</sup>These authors contributed equally to this article.

\*Correspondence: Fuqiang Cui ([fuqiang.cui@zafu.edu.cn](mailto:fuqiang.cui@zafu.edu.cn)), Jarkko Salojärvi ([jarkko@ntu.edu.sg](mailto:jarkko@ntu.edu.sg))

<https://doi.org/10.1016/j.xplc.2022.100307>

## ABSTRACT

*Vaccinium darrowii* is a subtropical wild blueberry species that has been used to breed economically important southern highbush cultivars. The adaptive traits of *V. darrowii* to subtropical climates can provide valuable information for breeding blueberry and perhaps other plants, especially against the background of global warming. Here, we assembled the *V. darrowii* genome into 12 pseudochromosomes using Oxford Nanopore long reads complemented with Hi-C scaffolding technologies, and we predicted 41 815 genes using RNA-sequencing evidence. Syntenic analysis across three *Vaccinium* species revealed a highly conserved genome structure, with the highest collinearity between *V. darrowii* and *Vaccinium corymbosum*. This conserved genome structure may explain the high fertility observed during crossbreeding of *V. darrowii* with other blueberry cultivars. Analysis of gene expansion and tandem duplication indicated possible roles for defense- and flowering-associated genes in the adaptation of *V. darrowii* to the subtropics. Putative *SOC1* genes in *V. darrowii* were identified based on phylogeny and expression analysis. Blueberries are covered in a thick cuticle layer and contain anthocyanins, which confer their powdery blue color. Using RNA sequencing, we delineated the cuticle biosynthesis pathways of *Vaccinium* species in *V. darrowii*. This result can serve as a reference for breeding berries whose colors are appealing to customers. The *V. darrowii* reference genome, together with the unique traits of this species, including its diploid genome, short vegetative phase, and high compatibility in hybridization with other blueberries, make *V. darrowii* a potential research model for blueberry species.

**Key words:** *Vaccinium darrowii*, blueberry, genome sequence, cuticle, subtropical, adaptation

Cui F., Ye X., Li X., Yang Y., Hu Z., Overmyer K., Brosché M., Yu H., and Salojärvi J. (2022). Chromosome-level genome assembly of the diploid blueberry *Vaccinium darrowii* provides insights into its subtropical adaptation and cuticle synthesis. *Plant Comm.* **3**, 100307.

## INTRODUCTION

Blueberries are rich in bioactive compounds such as anthocyanins, flavonoids, and other phenolic compounds (Wu et al., 2018; Kalt et al., 2020). Diets rich in blueberries have shown multiple health benefits, as they have antitumor, antibacterial, and anti-inflammatory effects and offer neural, cardiovascular, and retinal protection (Wu et al., 2018; Kalt et al., 2020). Together with their taste, these characteristics have promoted blueberry consumption and thereby steadily increased market demand. World blueberry production has more than doubled, from 300 000 tons in 2008 to 657 000 tons in 2017 (Romo-Muñoz et al., 2020). In China, production increased more than 100-fold between 2006 and 2015 and reached 234 700 tons in 2020 (Li et al., 2017, 2021). The two major cultivar categories of blueberry crop species are northern highbush and southern highbush, which constitute the majority of blueberry production worldwide (Strik, 2006; Retamales et al., 2014; Li et al., 2021). Southern highbush blueberry breeding has made considerable progress over recent years in terms of both newly bred cultivars and growth area (Li et al., 2021). Notably, the southern highbush cultivars have a lower chilling requirement for flower induction, making it possible to produce berries year-round with improved management techniques (Cho et al., 2019; Fang et al., 2020). Therefore, the southern highbush blueberry is a highly attractive emerging crop species for both farmers and researchers.

*Vaccinium darrowii* is an evergreen shrub that originated in subtropical areas of North America (Sharpe and Darrow, 1960). Crossbreeding between *V. darrowii* and the northern highbush blueberry (*Vaccinium corymbosum*) generated the southern highbush cultivars, which have expanded the blueberry-producing region to subtropical/tropical areas (Sharpe and Darrow, 1960; Lobos and Hancock, 2015). In addition to the practical use of *V. darrowii* in breeding, it is also a good model for the study of plant adaptation to subtropical climates. Furthermore, *V. darrowii* has multiple advantages for basic research, such as a short vegetative phase, with seedlings that usually flower within one year in the wild. This species is self-compatible in fertilization: berries develop from up to 20% of flowers after self-pollination and 70% after cross-pollination (Chavez and Lyrene, 2009b), and self-pollinated berries produce 21 plump seeds on average (Chavez and Lyrene, 2009b). In contrast to the tetraploid *V. corymbosum*, *V. darrowii* is mostly diploid ( $2n = 2x = 24$ ) in nature (Sultana et al., 2019). These properties make *V. darrowii* an ideal candidate genetic model for the genus *Vaccinium*, especially for species growing in subtropical regions.

Taxonomically, blueberries make up the *Cyanococcus* section of the genus *Vaccinium*. The northern highbush blueberry, *V. corymbosum*, is the only species in the *Cyanococcus* section with a chromosome-level genome assembly (Colle et al., 2019). Species in other sections of *Vaccinium*, such as cranberry (*Vaccinium macrocarpon*), which shares a similar temperate geographic distribution with *V. corymbosum* in the temperate region, and the boreal bilberry (*Vaccinium myrtillus*) have recently been sequenced (Diaz-García et al., 2021; Wu et al., 2021). These genomes serve as valuable resources for genome-enabled breeding of northern *Vaccinium* species, but

genome information for southern-adapted blueberries is still lacking. Here, we sequenced the genome of wild *V. darrowii*. A high-quality genome assembly will assist studies on southern blueberry cultivars, which have a shorter vegetative phase and a lower chilling requirement than northern cultivars (Wilkie et al., 2008; Song et al., 2013). Another advantage of *V. darrowii* is its compatibility for hybridization with other blueberry cultivars, regardless of their ploidy levels (Chavez and Lyrene, 2009a, 2009b). This enables the easy transfer of desired genetic traits from *V. darrowii* into any targeted cultivar.

## RESULTS

### Genome sequencing and assembly

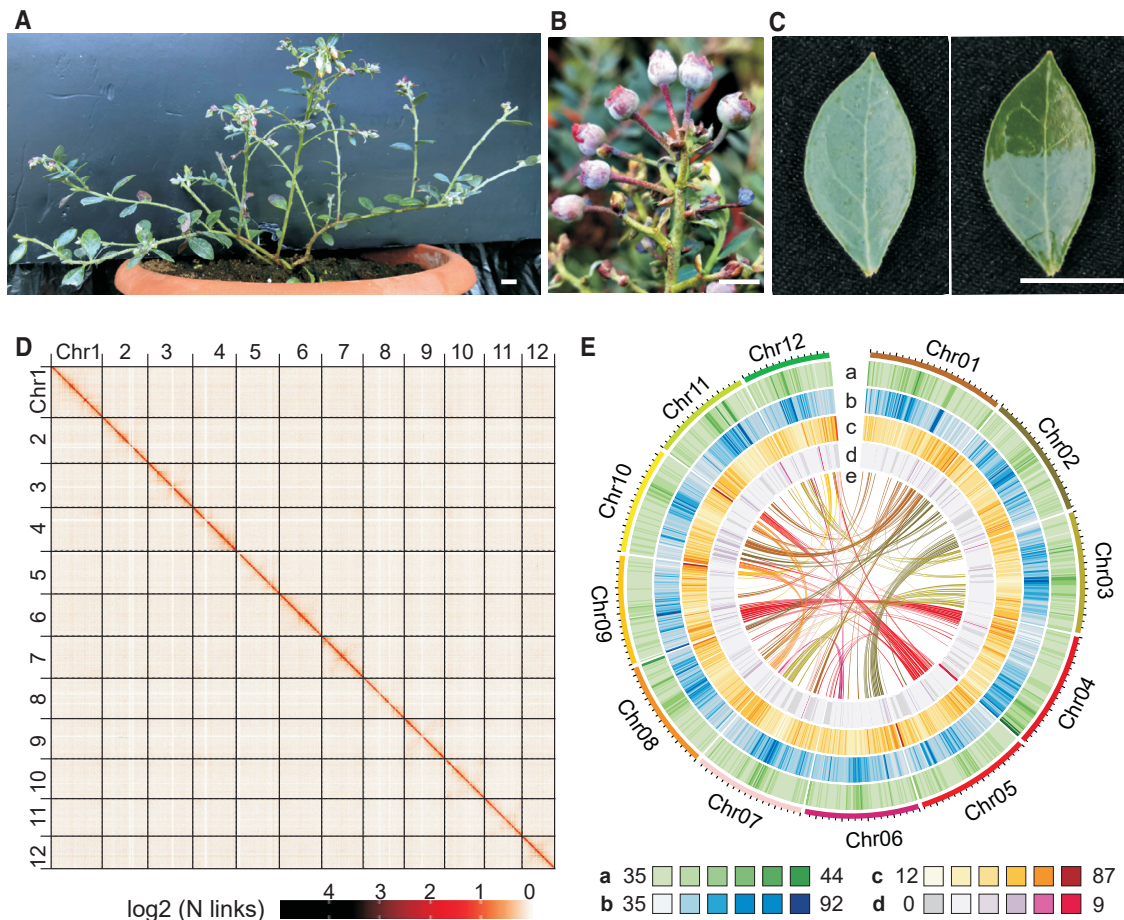
*V. darrowii* is an evergreen dwarf shrub from the subtropics (Sharpe and Darrow, 1960). It exhibits early flowering, usually flowering in the spring of the second year after planting (Figure 1A). Both berries and leaves are covered by visible waxy layers (Figure 1B and 1C). Multiple studies have shown the genome of *V. darrowii* to be ~550 Mb in size with 12 chromosomes (Lyrene et al., 2003; Chavez and Lyrene, 2009a; Vorsa et al., 2009; Sakhanokho et al., 2018). Consistent with these results, our flow cytometry analysis estimated the genome size of *V. darrowii* to be ~560.5 Mb (Supplemental Figure 1). A K-mer-based analysis of Illumina short-read sequencing data yielded 555.38 Mb as the estimated genome size, with 42.92% repetitive sequence. The K-mer-based estimate of heterozygosity was 1.27%, a common level of heterozygosity for wild woody species (Supplemental Figure 2). A similar level of heterozygosity ( $\pi = 0.013$ ) was obtained when the Illumina short-read data were aligned against the reference genome and nucleotide diversity of the sequenced individual was calculated.

A total of 91.37 Gb of clean reads (~166×) were produced using the Oxford Nanopore Technologies kit (ONT; N50 = 21.75 kb), and this was complemented by a total of 35.17 Gb (~63×) of Illumina sequencing data. After genome assembly, polishing, and removal of haplotypic contigs, the final assembly size of the genome was 562.96 Mb, close to the flow cytometry-based estimate (Supplemental Figure 1), and the resulting contig N50 was ~1.60 Mb (Supplemental Table 1). Nearly all of the Illumina sequencing data (97.59%) mapped to the genome assembly, and 90.0% of the mapped reads were properly oriented (Supplemental Table 2). Genome completeness was estimated to be 96.07% and 93.49% using CEGMA v.2.5 and BUSCO v.3.0 (with the embryophyta\_odb10 database), respectively (Supplemental Table 3; Parra et al., 2007; Simão et al., 2015).

Among the 1614 plant-specific single-copy orthologs in the BUSCO database, 1509 (93.49%) were identified in the assembly, and 1400 (86.74%) were complete and single copy (Supplemental Table 3). Thus, a high-quality genome sequence of *V. darrowii* was assembled.

### Chromosome-level assembly with Hi-C

To provide a chromosome-level assembly, we then used high-throughput chromosome conformation capture (Hi-C) sequencing to generate 32.81 Gb (60×) in 117.96 million paired-end Hi-C reads. Duplicate removal, sorting, and quality



**Figure 1. *V. darrowii* phenotype and genome assembly.**

(A) Cutting-propagated plant that flowered at 16 months of age.

(B) Waxy berries set on the pedicels.

(C) A typical waxy leaf of *V. darrowii*. Intact leaf (left); the wax layer on the top part of the leaf was removed by wiping with a finger (right). Scale bar, 1 cm.

(D) Hi-C heatmap of chromosome interactions.

(E) An overview of *V. darrowii* genomic features. Tracks: (a) GC content; (b) repetitive sequence coverage (%); (c) gene density; (d) noncoding RNA density; (e) chromosome collinearity visualized by linkage of syntenic genes. The color scales indicate the abundance in percentage of the corresponding items.

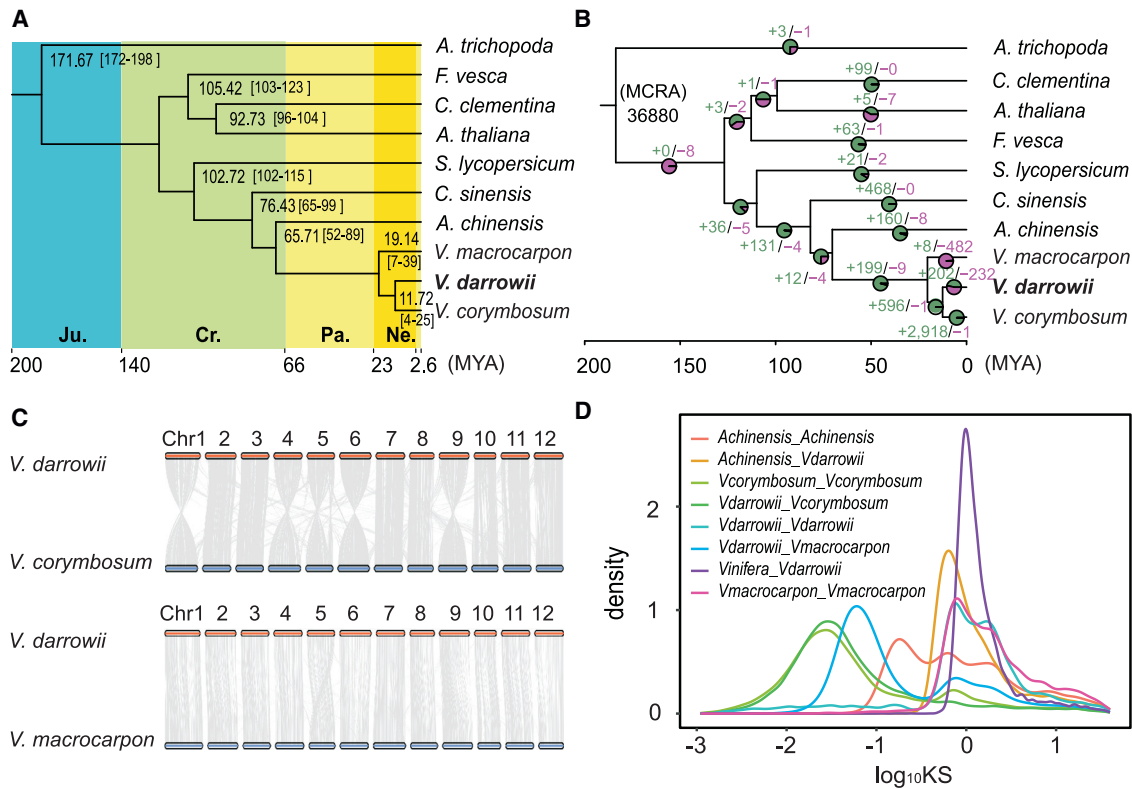
assessment were performed with HiC-Pro (Servant et al., 2015). There were 55.37 million unique paired alignments to the genome, and 46.15 million pairs were valid for Hi-C scaffolding (Supplemental Table 4). After correction of chromosome order and direction, the final assembly of the *V. darrowii* genome was 563.0 Mb in size. After Hi-C scaffolding, 96.48% (543.12 Mb) of the assembly was mapped to 12 pseudomolecules, and 505.56 Mb was properly oriented (Supplemental Table 5). The contig N50 of the final assembly was 1.50 Mb, and the scaffold N50 was 41.99 Mb (Supplemental Table 6). The total amount of gaps was limited to 52.93 kb (Supplemental Table 6). The Hi-C heatmap shows that the 12 pseudochromosomes were well connected along the diagonal line (Figure 1D). Thus a high-quality chromosome-scale genome was assembled.

### Gene prediction and functional annotation

A combination of plant protein homology mapping, transcriptome data, and *ab initio* gene prediction was used to generate gene model predictions (Supplemental Figure 3, Supplemental

Table 7), and the different evidence tracks were combined into consensus gene models using EvidenceModeler (Haas et al., 2008), resulting in 41 815 protein-coding genes (Supplemental Table 8). The average gene length was 5076.09 bp, and the mean exon number of each gene was 5.02 (Supplemental Table 8). Functional annotation of the protein-coding genes was performed using public databases, including the Gene Ontology (GO; 37.81% of genes annotated), KEGG (26.45%), KOG (41.32%), Pfam (63.7%), SwissProt (53.03%), TrEMBL (79.92%), and NCBI nr (80.18%) databases (Supplemental Table 9). About 18.9% of the predicted protein-coding genes were not annotated; this is comparable to annotation levels in other *Vaccinium* species (Diaz-Garcia et al., 2021), as well as *Ericales* (Yang et al., 2020), and probably reflects the evolutionary distance from model organisms whose functional data are available.

In addition to protein-coding genes, 992 noncoding RNAs were predicted, including 119 microRNAs, 257 rRNAs, and 616 tRNAs (Supplemental Table 8). Analyses combining BLAT and



**Figure 2. Genome comparison and evolution analysis.**

**(A)** Phylogenomic analysis with 10 species and 1059 single-copy gene families. The divergence time was estimated with MCMCtree to obtain the confidence intervals (shown in brackets) and maximum likelihood to obtain a point estimate. The divergence time is shown in million years ago (MYA). **(B)** Gene family expansion and contraction in *V. darrowii*. The numbers of contracted gene families are in light green, and those of expanded gene families are in light pink. **(C)** Synteny conservation between *V. darrowii* and two *Vaccinium* species, *V. corymbosum* and *V. macrocarpon*. Altogether, 96 334 genes from *V. corymbosum* and 28 527 genes from *V. macrocarpon* were collinear with those of *V. darrowii*. **(D)** Distribution of the synonymous substitution (Ks) rates for collinear genes from *V. darrowii*, *V. corymbosum*, *V. macrocarpon*, *A. chinensis*, and *V. vinifera*.

GeneWise software (Kent, 2002; Birney et al., 2004) identified 5300 pseudogenes (Supplemental Table 8). The *V. darrowii* genome contained 333.76 Mb of repetitive sequence, accounting for 59.28% of the genome. Long terminal repeat (LTR) retrotransposons comprised 24.47% of the genome, including 9.1% Copia LTRs and 15.14% Gypsy LTRs (Supplemental Table 10).

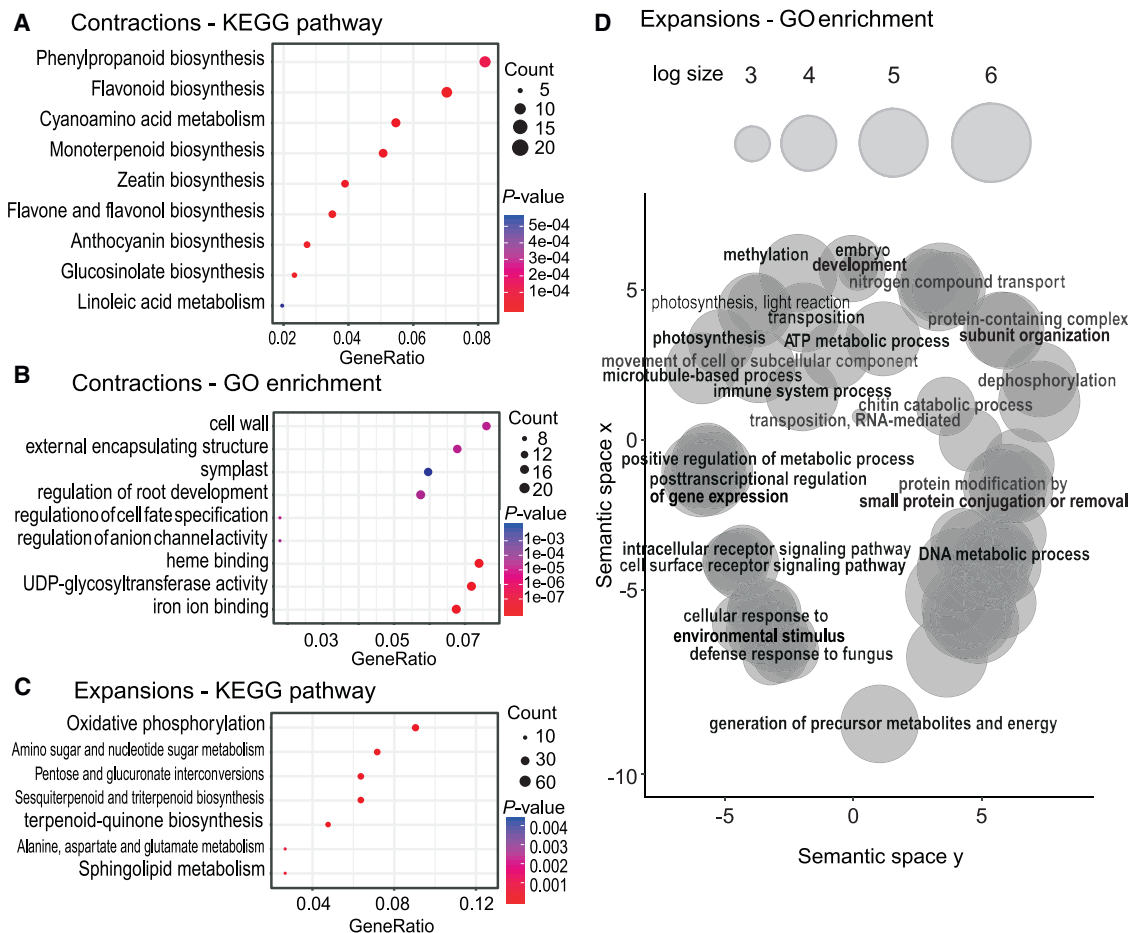
### Comparative genomics analysis

To analyze gene family evolution and obtain a phylogeny, we used OrthoFinder to identify unique and shared gene families among blueberry and nine other species: *V. darrowii*, *V. corymbosum*, *V. macrocarpon*, *Actinidia chinensis*, *Arabidopsis thaliana*, *Amborella trichopoda*, *Citrus clementina*, *Citrus sinensis*, *Fragaria vesca*, and *Solanum lycopersicum*. All species shared 3732 gene families (Supplemental Figure 4), and 347 gene families were specific to *V. darrowii* (Supplemental Table 11). A phylogenetic tree was constructed using 1059 single-copy gene families. It was calibrated using fossil evidence (Hedges et al., 2006; Kumar et al., 2017) and the divergence times of *V. corymbosum* versus *A. chinensis* (52–96 mya), *V. macrocarpon* versus *A. chinensis* (52–96 mya), *V. corymbosum* versus *A. trichopoda* (173–199 mya), *C. clementina* versus *A. thaliana* (96–104 mya),

*S. lycopersicum* versus *C. sinensis* (99–114 mya) as secondary calibration points (Hedges et al., 2006; Kumar et al., 2017). The MCMCtree results indicated that *V. darrowii* and *V. corymbosum* diverged 4–25 mya, preceded by a split from *V. macrocarpon* at 7–39 mya (Figure 2A). A second estimate, using the maximum-likelihood-based method, suggested that *V. darrowii* and *V. corymbosum* diverged at 11.72 mya, preceded by the divergence from *V. macrocarpon* at 19.14 mya (Figure 2A). Using CAFÉ (De Bie et al., 2006), we predicted a total of 232 contracted and 202 expanded gene families (Figure 2B).

### Collinearity and WGD analysis

Syntelogous genes among *V. darrowii*, *V. macrocarpon*, and *V. corymbosum* were identified by genome collinearity analysis. Genome organization was well preserved among these species, as their chromosomes showed high levels of synteny (Figure 3C, Supplemental Tables 12 and 13). In total, 96 334 genes from *V. corymbosum* were collinear with those from *V. darrowii*, whereas 28 527 genes from *V. macrocarpon* were syntenic. Analysis of synonymous substitutions (Ks) using syntelogs in each genome (Supplemental Table 14) revealed two whole-genome duplication (WGD) events in *V. darrowii* (Figure 2D;



**Figure 3. Functional enrichment analysis of expanded and contracted gene families.**

(A) KEGG pathway enrichment of the contracted gene families.

(B) Gene ontology (GO) enrichment of the contracted gene families.

(C) KEGG pathway enrichment of the expanded gene families. Gene ratio represents the ratio of the number of input genes annotated with a GO term to the number of all genes annotated with that term. The size of the dot represents the number of corresponding genes.

(D) The expanded gene families were subjected to GO enrichment analysis. The genes in the enriched GO terms were summarized with Revigo to visualize the relationships between the enriched GO terms.

Supplemental Figure 5). Syntenic comparison against grape (*Vitis vinifera*), a species that has undergone only the paleohexaploid ( $\gamma$ ) event, showed one WGD event after divergence (Supplemental Figure 6), confirming that the older event is the paleohexaploid ( $\gamma$ ) event and that the later event is a duplication. Syntenic analyses of both *V. corymbosum* (Supplemental Figure 7) and *V. macrocarpon* (Supplemental Figure 8) showed evidence of a bimodal nonsynonymous ( $K_s$ ) spectrum associated with two whole-genome multiplication events, and the peak  $K_s$  values coincided in all species. Furthermore, pairwise alignments of *V. darwii* against *V. corymbosum* and *V. macrocarpon* (Supplemental Figure 10) showed that the species divergence occurred more recently than the identified WGD. In conclusion, the three *Vaccinium* species share a common WGD event in addition to the ancient paleohexaploid WGT ( $\gamma$ ) event approximately 130–150 mya (Bowers et al., 2003) shared by all eudicots. Based on the time of the paleohexaploid event and its  $K_s$  peak at 1.58, we estimated that the recent WGD event occurred 51.8–59.7 mya ( $T = K_s/2r$ ). The identified WGD therefore matches the Vm- $\alpha$

duplication reported earlier in *V. macrocarpon* and Ad- $\beta$  in *A. chinensis* (Figure 2D; Diaz-Garcia et al., 2021; Wu et al., 2019).

*V. corymbosum* self-alignment showed four-fold synteny, probably resulting from a second WGD event when the tetraploid formed. Compared with *V. darwii*, its genome fractionation patterns suggested that *V. corymbosum* is an allopolyploid, consistent with earlier analyses (Colle et al., 2019). In addition, we observed many genes with low divergence (in terms of  $K_s$ ; Supplemental Figure 7) between the subgenomes; we interpret this to have resulted from homologous exchange and gene conversion events (Bird et al., 2018), similar to what has been observed in other recent polyploids.

The genome alignments between *V. darwii* and *V. corymbosum* also identified over 1000 syntelogs with very low divergence rates between the two species (in terms of  $K_s$ ; Supplemental Figure 9). Because the highly similar genes are spread in syntenic blocks over the whole genome, the similarity is more likely to be due to extensive, ongoing gene flow between the two species rather

than to a single introgression event into the sequenced individual. The ranges of the two species overlap in nature, for example, in Florida (Kloet, 1980), and their crosses are used extensively in blueberry breeding (Chavez and Lyrene, 2009b), making this scenario quite plausible.

### Genome evolution in *V. darrowii*

As illustrated by the phylogeny, *V. macrocarpon* diverged from blueberry first, followed by the divergence of *V. corymbosum* and *V. darrowii* (Figure 2). *V. macrocarpon* and *V. corymbosum* are both native to temperate regions, whereas only *V. darrowii* is a subtropical plant, suggesting that a temperate habitat may have been the original environment of the common ancestor, although the possibility that *V. corymbosum* and *V. macrocarpon* independently adapted to temperate environments cannot be excluded. To study gene family evolution related to adaptation to a subtropical climate, we analyzed gene family expansions and contractions between *V. darrowii* and *V. corymbosum*. KEGG pathways related to anthocyanin biosynthesis, such as “phenylpropanoid biosynthesis,” “flavonoid biosynthesis,” and “anthocyanin biosynthesis,” were contracted in *V. darrowii* (Figure 3A). Anthocyanins play a protective role in plants during cold temperatures (Archetti et al., 2009), whereas high temperatures repress plant anthocyanin production (Rabino and Mancinelli, 1986; Rowan et al., 2009; Lin-Wang et al., 2011). Thus, the contraction of anthocyanin-related gene families may represent an adaptation of *V. darrowii* to the warmer temperatures of the subtropics. In an individual gene family, the number of genes encoding Sn-2 acyl-lipid omega-3 desaturase was one in *V. darrowii*, eight in *V. corymbosum*, and four in *V. macrocarpon*, suggesting a loss of three genes in *V. darrowii* (Supplemental Table 15). Sn-2 acyl-lipid omega-3 desaturase is a positive regulator of plant tolerance to cold temperatures, and its knockout mutant in *Arabidopsis* exhibits a chilling-sensitive phenotype (Chen and Thelen, 2013; Román et al., 2015). Unsaturated lipids can remain in the liquid phase at lower temperatures than saturated lipids, allowing plants to maintain physiological functions during chilling stress (Chen and Thelen, 2013). Thus, the loss of lipid desaturases in *V. darrowii* is a possible adaptation to the subtropical climate. Other terms enriched in the contracted gene families included “cell wall,” “root development,” and “iron ion binding” (Figure 3B); it is unclear how they are related to adaptation to the subtropics.

Many of the GO terms enriched in the expanded gene families in *V. darrowii* were initially associated with chloroplasts and placed in contigs not mapped to the pseudochromosomes. After removal of this possible plastid contamination, secondary metabolism was enriched in the KEGG annotation, including “amino sugar and nucleotide sugar metabolism,” “sesquiterpenoid and triterpenoid biosynthesis,” and “terpenoid-quinone biosynthesis” (Figure 3C). Plant secondary metabolism is one of the main adaptive mechanisms for coping with altered environmental conditions and biotic stressors (Ramakrishna and Ravishankar, 2011; Aguirre-Becerra et al., 2021). To gain an overall view of the enriched GO terms in the expanded gene families, we summarized the terms using Revigo (Supek et al., 2011) and identified two larger clusters of terms with high similarity. Within the cluster defined by the GO term “DNA metabolic process,” a subterm, “DNA repair,” had the lowest *P* value (Figure 3D;

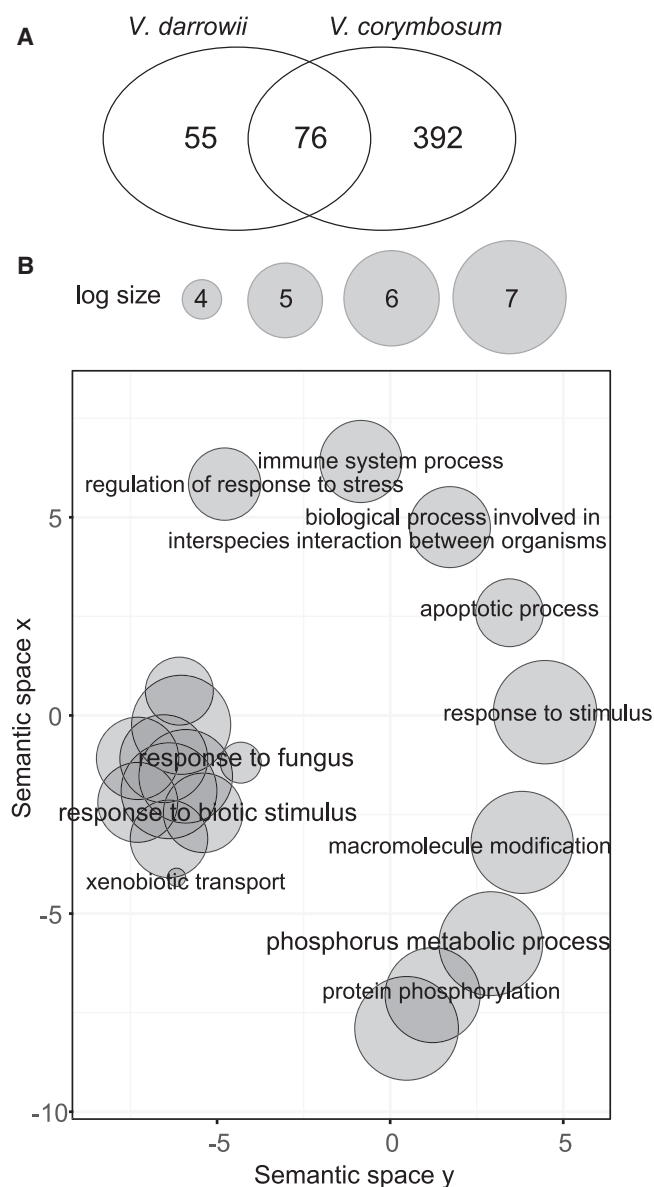
Supplemental Table 15). Genes associated with this term were mainly DNA helicases (Supplemental Table 15). DNA repair systems and DNA helicases play an active role in plant thermotolerance (Han et al., 2020). Therefore, an expansion of the DNA helicases may contribute to adaptation to elevated temperatures. The “blue-light signaling pathway” was among the enriched GO terms associated with the cluster linked to environmental adaptation (Figure 3D). The expanded gene family contained four duplicates of *Maintenance of Meristems* (*MAIN*) and *MAIN-like 1* genes (Supplemental Table 15). *MAIN* and *MAIN-like 1* interact with protein phosphatase PP7-like (Nicolau et al., 2020) to regulate transposable element silencing, as well as chloroplast development and abiotic stress tolerance (Xu et al., 2019). PP7L is closely related to PP7, a component in cryptochrome-mediated blue-light signaling (Møller et al., 2003) and a regulator of red/far-red light perception through control of the phytochrome pathway (Genoud et al., 2008). The expansion of *MAIN* and *MAIN-like 1* genes may therefore be due to altered light conditions in subtropical regions and suppression of transposable element expression resulting from environmental stresses.

Interestingly, *PsbP*, a membrane-extrinsic subunit of photosystem II (Ifuku et al., 2005a), was among the 13 genes under positive selection in *V. darrowii* (Supplemental Table 16). *PsbP* is essential for the stabilization of photosystem II, and decreased expression of *PsbP*, but not another extrinsic subunit gene, *PsbQ*, resulted in pale green leaves and enhanced sensitivity to high light in tobacco (Ifuku et al., 2005b). Enhanced *PsbP* expression in *V. darrowii* may help to maintain photosynthesis in the drought/hot summer of the subtropical region.

To probe for short-term evolutionary adaptations, we performed tandem duplication analysis with both *V. darrowii* and *V. corymbosum* (Supplemental Tables 17 and 18). After *P*-value adjustment with a Bonferroni correction, the tandemly duplicated genes of *V. darrowii* and *V. corymbosum* were enriched for 131 and 468 GO terms, respectively (Figure 4A). In total, 55 terms were specific to *V. darrowii* (Figure 4A). Consistent with tandem expansion analyses in other species, the enrichments were related to secondary metabolism (“xenobiotic transport”), defense responses (“response to fungus”), or adaptation to the environment (“regulation of response to stress”) (Figure 4B; Supplemental Table 19). Overall, the tandem expansions suggested adaptation to a new environment and genomic rewiring of defenses against new pathogens. Interestingly, five *ERdj3B* genes were tandemly duplicated in *V. darrowii* (Supplemental Table 19). *ERdj3B* encodes a DnaJ domain protein that binds to heat shock proteins (Yamamoto et al., 2020). *ERdj3B* is essential for pollen development at high temperatures, and the corresponding *Arabidopsis* mutant produced few seeds at slightly elevated temperatures from 22°C to 29°C (Yamamoto et al., 2020). Fertilization in *Vaccinium* species is sensitive to high temperature (Taulavuori et al., 2013; Yang et al., 2019). The duplicated *ERdj3B* genes could indicate an evolutionary strategy for successful fertilization of *V. darrowii* in warmer regions.

### Flowering-related genes in *V. darrowii*

Except for the florigen gene *FT*, few flowering-regulation genes have been identified and functionally characterized in blueberry



**Figure 4. GO enrichment analysis of tandem duplicated genes.**

**(A)** Venn diagram showing the numbers of GO terms enriched in the tandem duplicated genes of *V. darrowii* and *V. corymbosum*.

**(B)** The *V. darrowii*-specific GO terms were summarized with Revigo to reveal their relationships.

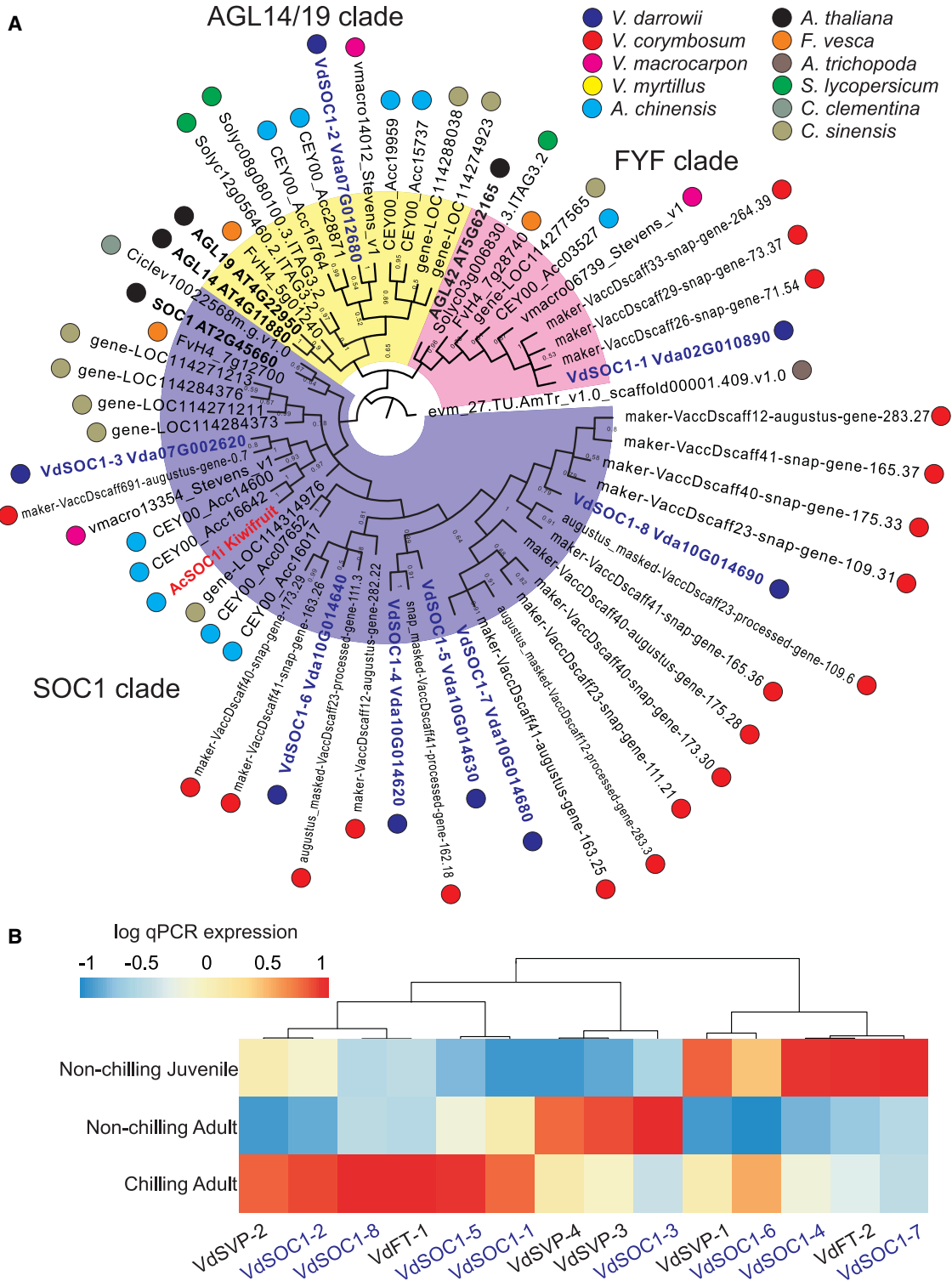
(Walworth et al., 2016; Gao et al., 2016). *V. darrowii* exhibits a short vegetative phase, and flowering can occur as early as 1 year (Wilkie et al., 2008; Song et al., 2013). Four *FPA* and six *Suppressor of Overexpression of CO 1 (SOC1)* genes were among the tandem duplications from the set of 26 genes assigned to the GO term “regulation of flower development” (Supplemental Table 19). In *Arabidopsis*, *FPA* promotes early flowering via regulation of RNA processing, which leads to reduced expression of the flowering-suppressor *FLC* (Schomburg et al., 2001; Veley and Michaels, 2008; Sonmez et al., 2011; Wu et al., 2020). The function of *FPA* in perennial plants has not been reported. *SOC1* is a major player in the promotion of flowering in many annual plants and several perennial plants, such as bamboo (Yoo et al., 2005; Hou et al.,

2021). *SOC1* directly promotes the expression of genes that initiate flower development (Hepworth et al., 2002; Immink et al., 2012). In trees, *SOC1* lowers the chilling requirement and promotes the release of flower buds from dormancy (Wilkie et al., 2008; Horvath, 2009; Trainin et al., 2013). One of the major contributions of *V. darrowii* to blueberry breeding is a reduction in the chilling exposure time required for flowering by northern highbush blueberry, which resulted in the production of southern highbush blueberry (Sharpe and Darrow, 1960; Song et al., 2013). We therefore focused on *SOC1* and its tandem duplications. To confirm the homology relationships, we performed multiple sequence alignment of orthogroups that contained genes annotated as *SOC1* homologs and constructed a gene tree from the alignment. Closer inspection revealed a split into three clades of *AGAMOUS*-like paralogs, the *FYF* clade, the *AGL14/19* clade, and the *SOC1* clade, each of which was named after the *Arabidopsis* homolog within the clade (Figure 5A). The most distant clade, *FYF*, controls flower senescence and abscission in *Arabidopsis* (Chen et al., 2011). In the *AGL14/19* clade, *AGL14* is expressed in roots of *Arabidopsis* and is therefore not likely to be involved in the cold response of flowering, whereas variation in *AGL19* has been found to be associated with the vernalization response and the timing of flowering (Suter et al., 2014). Therefore, blueberry homologs in this clade may also be regulated under cold stress and could be regulators of flowering.

Within the *SOC1* clade, two typical *SOC1* proteins, *AtSOC1* from *Arabidopsis* and *AcSOC1i* from kiwifruit, exhibited a split corresponding to plant phylogenetic relationships (Figure 5A). *AcSOC1i* is an experimentally verified *SOC1* protein in kiwifruit that promotes dormancy release (Voogd et al., 2015). The phylogenetic tree placed *Vda07G002620* in the same clade as *AcSOC1i*, suggesting that these genes may share a similar function (Figure 5A). The tandem duplicates, including *Vda10G014620*, *Vda10G014630*, *Vda10G014640*, *Vda10G014680*, and *Vda10G014690*, were grouped together, indicating their recent diversification (Figure 5A).

To identify orthologous relationships, we investigated the expression patterns of all putative *SOC1* homologs together with homologs of two other key flowering-regulation genes, the flowering-promotion *Flowering Locus T (FT)* genes and the flowering-suppression *Short Vegetative Phase (SVP)* genes (Horvath et al., 2003; Gregis et al., 2013). We studied a total of eight *SOC1*s, two *FT*s, and four *SVP*s annotated in the *V. darrowii* genome (Supplemental Table 20). With qPCR, we examined leaf samples from 2-year-old adult plants with or without chilling treatment and from 4-month-old juvenile plants. We used the initiation of flower buds as the standard for a successful chilling treatment. The transcript level of *VdFT1* was high only in the chilling-treated plants, whereas the *VdFT2* transcript level was high in the juvenile plantlets (Figure 6B), suggesting that *VdFT1* may be the florigen. *V. darrowii VdSOC1-1* (*Vda02G010890*), -2 (*Vda07G012680*), -5 (*Vda10G014630*), and -8 (*Vda10G014690*) exhibited expression patterns similar to that of *VdFT1*, with the highest transcript levels in plants treated with chilling (Figure 6B). Thus, these four *VdSOC1*s may function in the promotion of flower development. The transcript levels of *VdSVP-3* and -4 were highest in the non-chilling-treated adult plants and were much

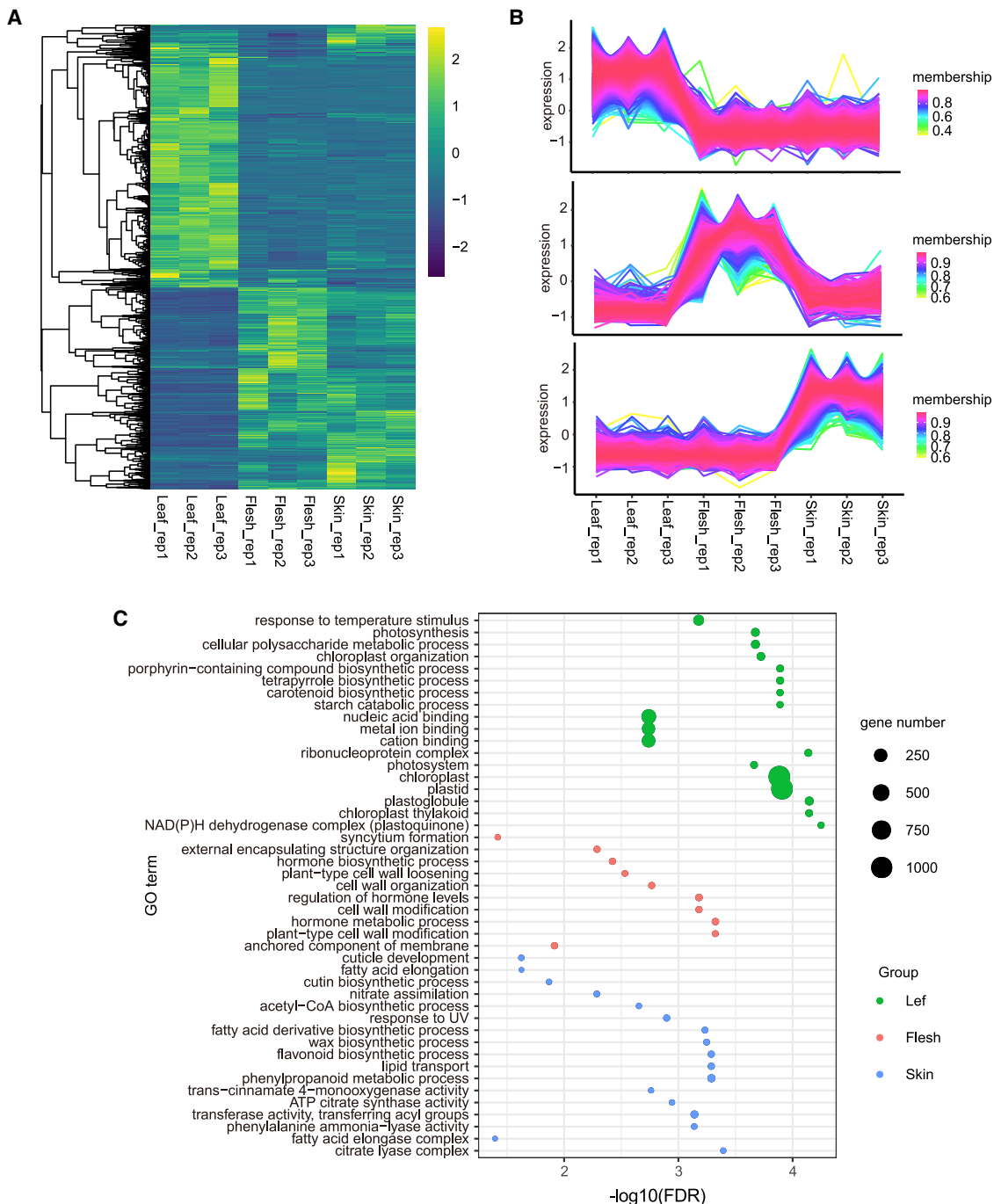




**Figure 5. Phylogeny analysis and expression profile of the SOC1-like genes in *V. darwii*.**

**(A)** The SOC1 proteins were identified from an orthogroup that includes the *Arabidopsis* SOC1 gene. Phylogenetic analysis of SOC1-like genes from *V. darwii*, *V. macrocarpon*, and *V. myrtilillus* was performed using the complete protein sequences, which were clustered into three clades named according to the corresponding *Arabidopsis* genes. The maximum-likelihood tree was reconstructed using MUSCLE alignment and MEGA X software. A total of 500 bootstrap replicates were used to assess tree reliability. The tree was rooted with the *A. trichopoda* SOC1 gene. Species are indicated by color-coded circles.

**(B)** Expression of *VdSOC1*s and other flowering regulatory genes, *VdFT*s and *VdSVP*s, was examined in adult and juvenile clones of *V. darwii*. A chilling treatment was performed under natural conditions to induce flowering in the adult clones. The age and chilling effects on gene expression were examined by qPCR using *VdTub.alpha3*, *VdTub.beta8*, and *VdActin3* as reference genes. The *VdSOC1* genes are highlighted in blue characters.



**Figure 6. Tissue-specific gene expression profiles in *V. darrowii*.**

**(A)** Heatmap of differential gene expression among three tissues: whole leaves (leaf), berry flesh (flesh), and berry skin (skin).

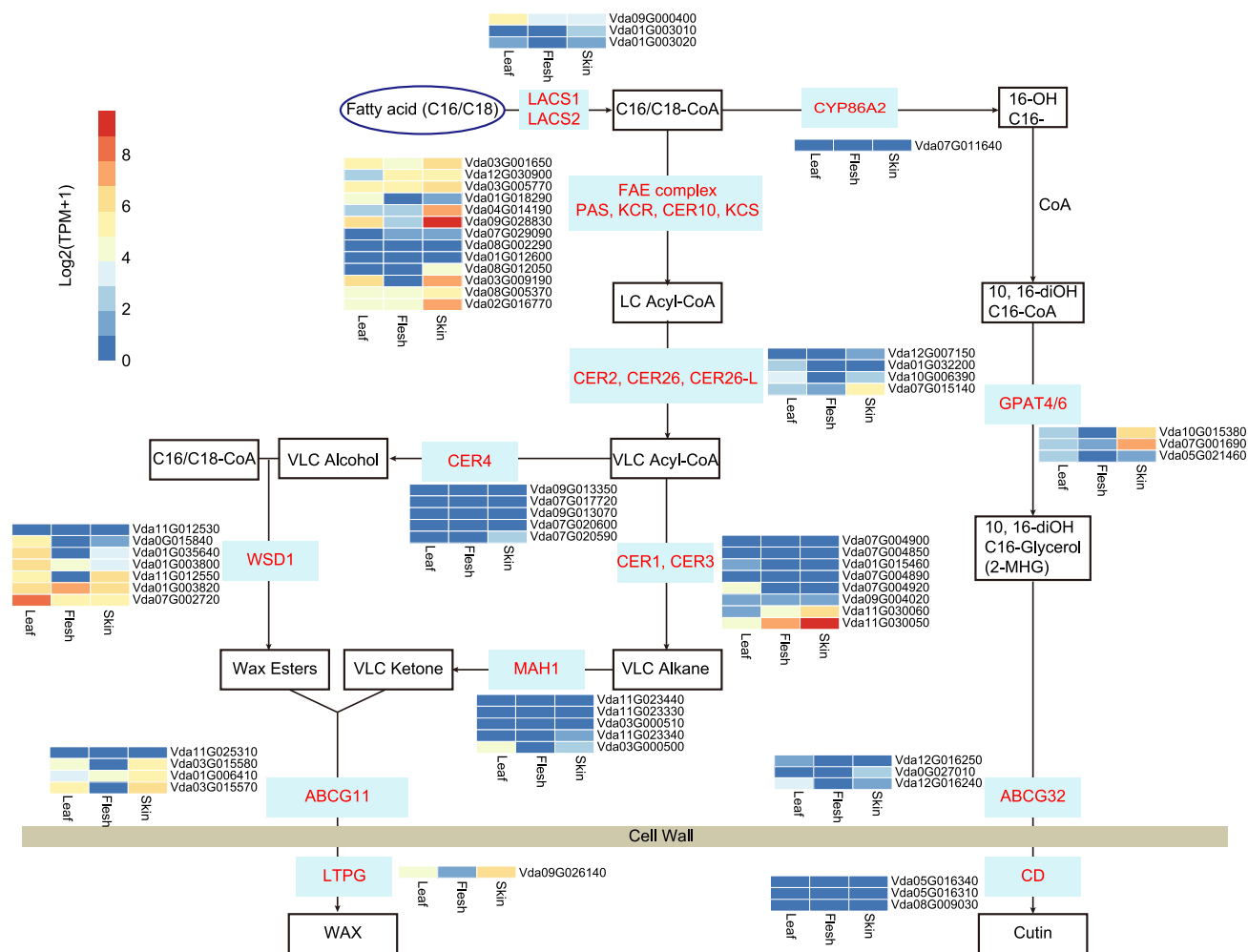
**(B)** Gene expression profile line graphs of genes specifically expressed in leaves, flesh, or skin. Membership values in color indicate the degree to which genes belong to the profile cluster.

**(C)** GO enrichment of the three groups of tissue-specific genes. The size of the dot represents the number of corresponding genes.

lower in the chilling-treated adult plants, indicating that *VdSVP-3* and -4 were more likely to be the functional SVs in *V. darrowii* (Figure 6B). Overall, we identified the key flowering-regulation genes in *V. darrowii*.

To test whether *SOC1* genes are tandemly duplicated among *Vaccinium* species, we analyzed the genomes of two species

from the other section of *Vaccinium*, *V. macrocarpon* and *V. myrtillos*. There were five *SOC1s* in *V. macrocarpon*, three of which were tandemly duplicated, and seven in *V. myrtillos*, four of which were tandemly duplicated (Figure 5A). The numbers of both total and duplicated *SOC1* gene models were comparable to those in *V. darrowii*. Thus, *SOC1* duplication is universal in *Vaccinium*. This finding indicated



**Figure 7. Putative cuticle pathway predicted in *V. darwii* based on transcriptome expression data.**

The genes in this pathway were identified based on BLASTP alignment with corresponding *Arabidopsis* amino acid sequences. Transcript levels from RNA-seq data are visualized as heatmaps and normalized with  $\log_2(\text{TPM} + 1)$ .

that *SOC1* genes may function in an unstudied, complex pattern across the genus.

### Gene expression patterns in leaves, berry flesh, and berry skin

Their powdery blue color is one of the major quality-determining factors for blueberries in the marketplace. The color depends on the content of anthocyanin pigments and wax, both of which are synthesized in the epidermal layer of the berry skin (Wang et al., 2012b; Yeats and Rose, 2013). Interestingly, the leaf surfaces of *V. darwii* are also covered by a thick layer of waxy cuticle, resembling the berry surface (Figure 1). We performed RNA-sequencing (RNA-seq) analysis with samples from three different tissues, berry skin, flesh, and leaves. Three biological replicates showed distinct clustering in the principal component analysis (Supplemental Figure 6). Genes with tissue-specific expression were collected and visualized as a heatmap (Figure 6A). There were 4131 genes expressed specifically in leaves, and according to GO enrichment analysis, they were mostly enriched in photosynthesis-associated processes (Figure 6B and 6C). This result serves as an internal control that

validates the analysis, as photosynthetic processes are known to cease in ripening blueberries. In the skin, 661 genes were specifically expressed (Figure 6B), with GO enrichment in both cuticle- and anthocyanin-associated terms (Figure 6C). The key steps of cuticle biosynthesis were included in the enriched GO terms “citrate lyase complex” and “ATP citrate synthase activity.” These are involved in the conversion of citrate to acetyl-CoA; acetyl-CoA participates in the “fatty acid derivative biosynthetic process” via “transferase activity, transferring acyl groups”; then, fatty acids are subjected to “fatty acid elongation” in a series of “fatty acid elongase complexes,” and finally the “cutin/wax biosynthesis process” is initiated (Figure 7C). Anthocyanins are synthesized from phenylalanine with *trans*-cinnamate and flavonoids as intermediates (Jaakola et al., 2002; Zhang et al., 2014). Accordingly, the GO terms “phenylalanine ammonia-lyase activity,” “*trans*-cinnamate 4-monooxygenase activity,” and “flavonoid biosynthetic process” were significantly enriched among the skin-specific genes (Figure 6C). These results indicate that the strategy of using skins to identify cuticle-related genes was successful. Among 425 flesh-specific genes, common processes involved in berry ripening (Forlani et al., 2019), including “plant-type cell wall

loosening/modification” and “hormone metabolic process” (Figure 7B and 7C), were enriched.

### Cuticle biosynthesis pathway

Among the two important functions of blueberry skins, the anthocyanin synthesis pathway has been well documented (Gupta et al., 2015; Colle et al., 2019), but blueberry wax biosynthesis pathways are not well understood, although several berry genomes have been published. The cuticle, which consists of wax and cutin, not only directly influences the color of blueberries but also has a direct influence on storage losses (Sapers and Phillips, 1985; Wang et al., 2012b; Trivedi et al., 2019; Li et al., 2020). We identified cuticle-synthesis genes in *V. darrowii* from OrthoFinder clustering to their known *Arabidopsis* orthologs (Supplemental Table 21). Interestingly, many genes related to wax synthesis showed small-scale tandem duplications (Supplemental Tables 18 and 21). Notably, four to seven duplicates were found for *Eceriferum 1* (*CER1*) and *CER3*, two key synthesis enzymes of the wax monomer alkane (Bourdenx et al., 2011; Bernard et al., 2012), and *Mid-chain Alkane Hydroxylase 1* (*MAH1*) and *Wax Ester Synthase/Acyl-CoA:Diacylglycerol Acyltransferase* (*WSD1*), which function in the last two steps of plant wax biosynthesis (Supplemental Tables 18 and 21; Greer et al., 2007; Li et al., 2008b). Overall, of the 75 wax-associated genes in *V. darrowii*, 36 were tandemly duplicated, suggesting significant enrichment ( $P = 3.6 \times 10^{-9}$ ; Fisher's exact test with a total of 7603 tandem duplications). These duplications may explain the enhanced wax accumulation on blueberries.

We next explored the complete pathway of cuticle biosynthesis using genes predicted from their relevant *Arabidopsis* homologs (Supplemental Table 21) and plotted them together with their expression levels in samples of leaves, berry flesh, and berry skins (Figure 7). Cuticle biosynthesis in model plants is considered to be specific to the epidermis (Yeats and Rose, 2013). Consistently, most genes listed in the pathway exhibited higher expression in the epidermis samples, including leaves and berry skins (Figure 6). Cutin Deficient (*CD*) is the key cutin synthase at the cell wall that catalyzes the final step of cutin formation in many land plants (Yeats et al., 2014). None of the homologs of the *CD* gene in *V. darrowii* were expressed. The possibility of the *CD* gene being missed from the annotation was excluded by querying the *CD* genes against the *V. darrowii* genome sequence using BLAST, and no other homologs were found. Thus, cutin synthesis in the cell wall of *V. darrowii* may differ from that of other plants studied previously (Yeats et al., 2014).

## DISCUSSION

*Vaccinium* species are widely distributed from tropical regions to the polar circle and tundra. *V. darrowii* is a wild blueberry species of practical and economic value. Because it was used to breed blueberry cultivars adapted to subtropical regions, studies on its unique traits, including lower chilling requirement, early flowering, and heat tolerance, are increasing (Sharpe and Darrow, 1960). The phylogenetic tree and parsimony principle imply that *V. darrowii* may have originated in temperate regions and subsequently adapted to subtropical regions (Figure 1D). We

identified several genome characteristics that possibly contributed to the adaptation of *V. darrowii* to subtropical climates. Tandem duplicates contribute to gene family expansions and enable alternative transcriptional regulation, and they have therefore been considered a common strategy for adaptive evolution to altered environments (Hanada et al., 2008; Wang et al., 2021). A clear functional bias in tandemly duplicated genes has been observed in plants, making tandem genes a good index for environmental adaptation (Hanada et al., 2008; Freeling, 2009; Rody et al., 2017). GO enrichment analysis of tandemly duplicated genes in *V. darrowii* provides an overview of possible adaptive traits. Defense-related GO terms were specifically abundant in the tandem duplicates of *V. darrowii* (Figure 4; Supplemental Table 19). The same terms were also enriched in the gene family expansion analysis, indicating that the retention of these genes may assist the growth of *V. darrowii* in subtropical environments. In addition, duplications were observed in multiple genes that have been reported to participate in the regulation of plant tolerance to elevated temperatures, such as genes involved in secondary metabolism and stress responses (Supplemental Tables 18 and 19). Notably, genes that mediate plant fertilization in warmer temperatures were also present among the tandemly duplicated genes (Supplemental Tables 18 and 19). It is common on farms for high temperatures to reduce the fertilization of blueberry cultivars (Yang et al., 2019). The tandemly duplicated genes related to fertilization may promote *V. darrowii* propagation in subtropical regions. In addition, 26 genes were predicted within the GO term “regulation of flower development” in *V. darrowii* (Supplemental Tables 18 and 19). Because *V. darrowii* originated in high latitudes (Figure 1), successful flowering and fertilization under conditions of reduced chilling and relatively hot temperatures may be essential for its propagation at low latitudes.

SOC1 plays a key role in promoting the vegetative to reproductive phase transition in annual plants (Yoo et al., 2005; Immink et al., 2012). In perennial species, SOC1 is a major regulator of bud dormancy release (Wilkie et al., 2008). A role for SOC1 in regulating the requirement for an extended cold treatment to release flower bud dormancy has been reported in kiwifruit, Japanese apricot, and sweet cherry (Voogd et al., 2015; Kitamura et al., 2016; Wang et al., 2020). The SOC1 genes in *V. darrowii* were highly duplicated, and six of eight VdSOC1s were tandem duplicates (Supplemental Table 18). Although the eight VdSOC1s were split into three clades, six of them were clustered together with the *Arabidopsis* SOC1 and functional SOC1 of kiwifruit (Figure 5A). This suggested that the VdSOC1s may function in promoting flowering and breaking dormancy. However, the timing and location of gene expression largely determine gene function. Vda07G002620 (VdSOC1-3) shared the same subclade with the kiwifruit AcSOC1i, but chilling treatment suppressed its expression, suggesting it may not be the functional SOC1 in *V. darrowii* (Figure 5). However, Vda07G012680 (VdSOC1-2), situated in another clade with *Arabidopsis* AGL42, was expressed at the highest level in chilling-treated plants (Figure 5). It may function as a flowering promoter, as AGL42 was reported to promote flowering in *Arabidopsis* (Dorca-Fornell et al., 2011). SOC1 was also reported to suppress flowering in strawberry (Mouhu et al., 2013). Thus, complex or alternative functions may be attributed to SOC1 genes, even those with high sequence identities. In *Arabidopsis*, the transcript levels of SOC1 in leaves

first increased and then decreased during development (Hepworth et al., 2002). This pattern is mainly due to the complex function of *SOC1* in both the vegetative and reproductive phases (Melzer et al., 2008; Chen et al., 2017; Tyagi et al., 2019). The single *SOC1* in *Arabidopsis* integrates complex signals associated with thermosensory, autonomous, gibberellin, vernalization, and photoperiod pathways (Borner et al., 2000; Shen et al., 2011). In *V. darrowii*, four tandemly duplicated *VdSOC1*s in the same cluster exhibited different expression patterns (Figure 5B), implying functional differentiation between these highly similar genes. This hypothesis requires further examination *in vivo* and is beyond the scope of this paper. Future studies may reveal the individual function of each *VdSOC1*. Our genome information here can facilitate the mining of these functional genes in the future.

A waxy surface is a very typical characteristic of *Vaccinium* berries. Multiple health-promoting activities have been attributed to wax components, such as triterpenoids, which play roles as anticancer, anti-inflammatory, and antimicrobial agents (Szakiel et al., 2012). Wax monomers vary among *Vaccinium* species and have been well documented both morphologically and chemically (Chu et al., 2017, 2018; Trivedi et al., 2019). However, the genes associated with wax biosynthesis have not been well defined in published *Vaccinium* genomes (Colle et al., 2019; Diaz-Garcia et al., 2021). Here, we identified all candidate cuticle biosynthesis genes in *V. darrowii* (Figure 7), which can facilitate breeding of the desired powdery color of blueberries. Compared with cutin-related genes, wax synthesis genes of *V. darrowii* exhibited multiple tandem duplications (Supplemental Table 18; Figure 7). The tandem duplication of wax synthesis genes was also observed in *V. corymbosum* (Colle et al., 2019). This suggests that enhanced wax synthesis is a common feature in all blueberries, consistent with the universal waxy berries of all blueberry cultivars (Chu et al., 2017, 2018; Trivedi et al., 2019). One unique character of *V. darrowii* is its waxy leaves, which phenotypically resemble the powdery surface of blueberries (Figure 1B and 1C). Some cuticle synthesis genes were highly expressed in the leaf compared with the berry skin (Figure 7). Therefore, using leaves to study the wax biosynthesis pathway instead of berries is possible in *V. darrowii*. This would dramatically shorten the time for the functional examination of wax-associated genes, as leaves are available much earlier than berries. Thus, *V. darrowii* may serve as a model plant for wax-related studies in the genus *Vaccinium*.

Blueberry is a relatively young crop species with very little breeding, due in part to the lack of genome-level information on the species. Similar to many other commercially important crop species, blueberry cultivars are of polyploid origin, being mostly tetraploids or hexaploids. Polyploid genomes may confer blueberries with heterosis-like advantages, yielding improved growth, stress resistance, and berry size. However, the resulting complex genome structure makes polyploid species less amenable to basic research compared with diploids. One way to obtain molecular information on polyploid cultivars is to study their diploid progenitors. For example, in the study of strawberry, the diploid woodland strawberry (*F. vesca*) has been widely used as a model because of its simple genome and easy genome-editing characteristics (Oosumi et al., 2006; Shulaev et al., 2011; Li et al., 2019). However, no diploid blueberry

species has been proposed as a model to date. *V. darrowii* is naturally diploid. With the completion of its genome sequence, it could emerge as a blueberry model, especially for subtropical species. Other advantages of *V. darrowii* as a model blueberry include its early flowering, self-compatibility, and smaller plant size. Recently, birch has been promoted as a promising model tree with a small genome size and accelerated flowering within 1 year through high CO<sub>2</sub> treatment (Longman and Wareing, 1959; Salojärvi et al., 2017). Likewise, “Mini-Citrus,” a model promoted for orange species, can bear one or two fruits in the first year (Zhu et al., 2019). These two woody perennial species show significant advantages compared with the most commonly used model tree, poplar, which is dioecious and has a very long vegetative phase (Bradshaw et al., 2000; Taylor, 2002). We propose that *V. darrowii* could be a model plant not only for blueberry species but also for subtropical woody species.

## METHODS

### Growth conditions and plant material

*V. darrowii* Camp was obtained from the Stephen F. Austin State University, Texas, USA by Prof. Hong Yu, Nanjing Botanical Garden Mem Sun Yat-sen and was cultivated in a mixture of peat, vermiculite, and bark pieces (1:1:1). Growth room conditions were 150–200  $\mu\text{mol m}^{-2} \text{s}^{-2}$  light, 60% humidity, 12/12 h (light/dark) photoperiod, and 23/18°C (day/night). The outdoor growing area was in Hangzhou, a subtropical city in China. Seedlings from cuttings of a single *V. darrowii* clone were used for this study.

### Flow cytometry examination

Nuclei of *V. darrowii* were extracted from young leaf tissues as previously described (Galbraith et al., 1983) with modifications according to Robinson and Grégori (2007). In brief, leaves submerged in ice-cold nucleus isolation buffer were chopped with a sharp razor, and the nuclei were then collected and stained with DNA fluorochrome (propidium iodide, 50 mg/ml). Flow cytometric examination was performed on a BD FACSCalibur flow cytometer (Becton Dickinson, San Jose, CA) with tomato (*S. lycopersicum*, genome size 950 Mb) as a reference.

### DNA preparation and sequencing

Young leaves from a 3-year-old single cutting of *V. darrowii* were collected for DNA extraction with the CTAB DNA extraction method (Porebski et al., 1997). DNA was further purified using VAHTS DNA Clean Beads (N411-03, Vazyme, China) and then used for DNA sequence preparation (Berensmeier, 2006).

For Illumina PE150 sequencing, purified DNA was fragmented by ultrasonication for library construction (350 bp). A total of 35.17 Gb of clean reads (around 60×) were obtained with Q20 > 97% and Q30 > 92%. The possibility of DNA contamination was excluded by alignment examination. BLAST alignment of 10 000 randomly selected reads returned best matches only to the closely related plant species *V. macrocarpon* (51.37%) and *Vitis vinifera* (3.82%) (Altschul et al., 1990). The proportion of plastid sequence was less than 0.02%, which was confirmed by comparison with the blueberry chloroplast genome (KJ773964.1; 1320 bp). The genome characteristics were assessed by K-mer analysis (Li et al., 2008a). The estimated genome size was 555.38 Mb, with 42.92% repetitive sequence and 1.27% heterozygosity.

For Nanopore sequencing, genome DNA was repaired with the NEBNext FFPE DNA Repair Mix kit (M6630; USA), and the library was constructed with the ONT template prep kit (SQK-LSK109; Nanopore, UK). Large fragments were collected with the BluePippin system (Sage Science; USA)

and pipetted into an R9 flow cell. Sequencing was performed with Nanopore sequencing technology at Biomarker using the ONT PromethION platform with the corresponding R9 cell and the ONT sequencing reagent kit (EXP-FLP001.PRO.6, UK). Base calling was performed using Guppy with the “flipflop” algorithm v.1.5 (Wick et al., 2019). A total of 98.1 Gb of high quality reads, corresponding to a depth of 166×, were obtained (Deamer et al., 2016).

### RNA preparation and sequencing

Mature berries were separated into flesh and skin by manual peeling. RNA from flesh, skin, and fully expanded young leaves was used for tissue-specific expression assays, and RNA from branches, leaves, flowers, and berries of the sequenced *V. darrowii* individual was used for gene annotation. RNA extraction was performed according to Jaakola et al. (2001). In brief, the isolated RNA (5 µg) was purified with VAHTS mRNA Capture Beads (N401-01; Vazyme, China) for two rounds and then subjected to fragmentation. The cDNA libraries were constructed with the NEBNext Ultra RNA Library Prep Kit (E7530L; NEB, USA) and then sequenced in paired-end mode on an Illumina HiSeq 4000 platform (LC Sciences, USA) at LC-Biotechnologies (Hangzhou).

### Genome assembly

Long read error correction was performed with Canu (<https://github.com/marbl/canu>, v.1.5; Koren et al., 2017) using the criteria “genomeSize = 1000000000” and “corOutCoverage = 50”. Next, overlapping was performed with the highly sensitive overlapper MHAP (mhap-2.1.2, option “corMhapSensitivity = low/normal/high”), and subsequent error correction was performed using the Falcon\_sense method. In addition, WTDBG2 and SMARTdenovo assemblies were performed, followed by three rounds of error correction using Racon and adjustment via Pilon (Walker et al., 2014; Vaser et al., 2017). The assembly quality was evaluated by alignment of the reads from the short-read sequencing. Altogether, 97.59% of the reads were mapped, with 90% mapping uniquely. Genome completeness was estimated as 96.07% and 93.49% with CEGMA v.2.5 and BUSCO v.3.0, respectively (Parra et al., 2007; Simão et al., 2015).

### Hi-C assembly

Hi-C fragment libraries were constructed in sizes from 300 to 700 bp and sequenced with the Illumina platform (Rao et al., 2014). Raw reads were then subjected to removal of adapters and low-quality reads and trimming of low-quality bases. The trimmed reads, which together provided 60× coverage of the draft genome, were truncated at the putative Hi-C junctions and then aligned to the assembly with the BWA aligner (Li and Durbin, 2009). Uniquely aligned reads with quality  $\geq$  Q20 were retained for further analysis. Invalid read pairs were filtered out using HiC-Pro v.2.8.1 (Servant et al., 2015). Uniquely mapping read pairs totaled 46.94% of the clean reads, and 83.35% of them were valid interaction pairs. The valid interaction pairs were clustered, ordered, and oriented onto pseudochromosomes with LACHESIS (Burton et al., 2013). Twelve pseudochromosomes were constructed manually, with 96.48% of valid interaction pairs anchored and 93.08% properly ordered and oriented.

### Repetitive sequences, noncoding RNA, and pseudogene analysis

LTR\_Finder (v.1.05) and RepeatScout (v.1.0.5) were used with default parameters to build a primary database of repetitive sequences in *V. darrowii* (Price et al., 2005; Xu and Wang, 2007). The database was classified with PASTECClassifier and then combined with Repbase databases (Jurka et al., 2005; Hoede et al., 2014). RepeatMasker v.4.0.5 was used to screen repetitive sequences that included simple repeats, satellites, and low-complexity repeats with the parameters -no-low -no\_is -norna -engine wublast -qq -frag 20000 (Tarailo-Graovac and Chen, 2009). For noncoding RNAs, rRNAs and microRNAs were predicted with the Rfam database, and tRNAs were predicted with

tRNAscan-SE v.1.3.1 with the option “-E -H” (Lowe and Eddy, 1997; Griffiths-Jones et al., 2005). Pseudogene homolog sequences were aligned with genBlastA v.1.0.4, and frameshifts and internal stop codons were screened using GeneWise v.2.4.1 (Birney et al., 2004; She et al., 2009).

### Protein-coding gene prediction and functional annotation

Gene prediction was performed with the EVM v.1.1.1 pipeline (Haas et al., 2008). In brief, *de novo* gene prediction was performed with GENSCAN (Burge and Karlin, 1997), Augustus v.2.4 (Stanke and Waack, 2003), GlimmerHMM v.3.0.4 (Majoros et al., 2004), geneid v.1.4 (Blanco et al., 2007), and SNAP v.2006-07-28 (Korf, 2004) with default parameters. GeMoMa v.1.3.1 was used for homolog prediction (Keilwagen et al., 2018). The RNA reads were assembled using HISAT v.2.0.4 (Kim et al., 2015) and StringTie v.1.2.3 (Pertea et al., 2015), and then TransDecoder v.2.0 and GeneMark-ET v.5.1 were used for gene prediction (Tang et al., 2015). RNA transcript assembly and unigene sequence prediction were performed with PASA v.2.0.2 (Campbell et al., 2006). The prediction methods above were combined in EVM v.1.1.1 and revised using PASA v.2.0.2. The annotation of protein-coding genes was performed via BLAST v.2.2.31 (Altschul et al., 1990) (-evalue 1e-5) against the nr (Marchler-Bauer et al., 2011), KOG (Koonin et al., 2004), GO (Dimmer et al., 2012), KEGG (Ogata et al., 1999), and TrEMBL (Boeckmann et al., 2003) databases.

### Heterozygosity estimation

The Illumina raw reads were trimmed with fastp (Chen et al., 2018) and, after quality control, mapped to the genome using BWA-MEM. Subsequently, GATK v.4.1.9 (O'Connor and Van der Auwera, 2020) was used for SNP calling: adding read groups, sorting the reads, removing duplicates, haplotype calling, and finally genotype calling. The variant calls were filtered for SNPs, and low-quality calls were removed using (QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0 || AF < 0.05). A windowed nucleotide diversity estimate was calculated with VCFtools (Danecek et al., 2011) using a 100-kb window with a 10-kb step size. A genome-wide nucleotide diversity estimate was calculated by taking the mean of all windows.

### Genome comparison and evolution analysis

Protein data from representative plant species, including *V. darrowii*, *V. corymbosum*, *V. macrocarpon*, *A. chinensis*, *A. thaliana*, *A. trichopoda*, *C. clementina*, *C. sinensis*, *F. vesca*, and *S. lycopersicum*, were retrieved for gene family clustering and phylogenetic analysis. The protein sequences were clustered into orthogroups using OrthoFinder (version 2.4) with default parameter settings. Pairwise similarities between proteins from all species were searched in an all-versus-all manner with DIAMOND using an e value  $\leq$  0.001. Single-copy families in the analyzed genomes were subjected to multiple alignment using MAFFT v.7.205, and the alignments were trimmed using Gblocks v.0.91b. The aligned gene family sequences were then concatenated, and a phylogenetic tree was constructed using the maximum-likelihood method in IQ-TREE v.1.6.11 with 1000 bootstraps. *A. trichopoda* was used as the root for the tree.

Positive selection analysis was performed with MAFFT (-localpair -maxiterate 1000) for alignments of each protein sequence in each gene family. The alignments were then converted to codon alignments using PAL2NAL. CodeML (F3x4 model of codon frequencies), based on the Branch-site model, was performed. The chi2 program in PAML was used to analyze model A ( $\omega > 1$ ) and the null model ( $\omega \leq 1$ ) to perform a likelihood ratio test. Significantly different ( $P < 0.01$ ) selection sites were obtained and further tested with the empirical Bayes method. Positive selection sites (>0.95) were used to identify genes under positive selection.

Divergence times were estimated using MCMCtree incorporated into PAML v.4.9i software (Yang, 2007). We also obtained a maximum-likelihood point

estimate using R with chronos in the ape library. The fossil times used for tree calibration were obtained from TimeTree (Hedges et al., 2006; Kumar et al., 2017) (<http://www.timetree.org/>), and angiosperm fossil pollen records were used to calibrate the root of the tree to 125.9–247.3 mya (Barba-Montoya et al., 2017). Gene family expansions and contractions were analyzed using CAFÉ (v.4.2) with default settings (De Bie et al., 2006). Gene families were selected with both family-wide *P* values and Viterbi *P* values <0.05.

### Synteny and Ks analysis

All-versus-all BLASTP analyses of proteins were performed between and within the selected species, and the results were filtered using a cutoff value of  $e < 10e^{-3}$ . BLASTP results and genome annotation files for each species were then used to identify syntenic blocks in MCScanX with default parameters (Wang et al., 2012a). The synonymous mutation rate (*Ks*) values of homologs within collinear blocks were calculated using wgd v.1.1.0 software. The values of all gene pairs were plotted to identify putative WGD events. The duplication time was estimated using the formula  $t = Ks/2r$ , which represents the neutral substitution rate, and was used to estimate the divergence times between blueberry and other species. Tandem duplications were obtained as part of the MCScanX analysis.

### Transcriptome analysis

cDNA libraries of leaf, flesh, and skin samples were constructed for RNA-seq, and three biological replicates were performed for each tissue. Trimmomatic v.0.39 was used to remove low-quality and adapter sequences to obtain clean data. HISAT2 v.2.1.0 (Kim et al., 2015) was used to align the clean reads to the reference genome. Gene model prediction and calculation of gene expression were performed with StringTie v.1.3.4 (Pertea et al., 2015). DESeq2 was used for differential expression analysis between tissues. C-means was performed for the set of screened differentially expressed genes to find groups of genes with specific expression patterns in tissues. For the detection of differentially expressed genes between tissues, fold change >2 and false discovery rate <0.01 were used as cutoff values. To understand the biological functions of gene clusters, GO enrichment analysis was performed using GOATOOLS (Klopfenstein et al., 2018). The enrichments were corrected by false discovery rate for *P* values, and then the *q* values were calculated. Gene expression in the samples was plotted using the R package pheatmap.

### Chilling treatments and qPCR procedures

Two-year-old cuttings from one clone of *V. darrowii* were moved outdoors from a growth room in mid-September 2019. This allowed *V. darrowii* to grow naturally in the subtropical city of Hangzhou. All the cuttings were moved into growth rooms (23°C) in the beginning of November, except those left outside for chilling treatment. Chilling treatment was performed outdoors until the middle of December. The average outdoor temperature was 10°C–18°C in November and 5°C–12°C in December. The cuttings were then moved into a growth room for 3 weeks for flowering induction. Individual cuttings that successfully developed flower buds were used for leaf collection. Non-chilling-treated 2-year-old and 4-month-old cuttings of *V. darrowii* were collected at the same time. RNA was extracted using the RNAPrep Pure kit (DP441, Tiangen, China) for RNA isolation, and then the cDNA was synthesized with the HiScript III 1st Strand cDNA Synthesis Kit (R312, Vazyme, China). Primers were designed with the Real-Time PCR (TaqMan) Primer and Probes Design Tool from the GenScript website ([www.genscript.com/tools/real-time-pcr-taqman-primer-design-tool](http://www.genscript.com/tools/real-time-pcr-taqman-primer-design-tool)) and are listed in Supplemental Table 22. qPCR was performed as described in Cui et al., (2016) on a Bio-Rad CFX Touch system. The raw cycle threshold values were processed in Qbase+ 2.1 (Hellemans et al., 2007) using *Tubulin α2*, *Tubulin α3*, and *Tubulin β8* as reference genes.

### ACCESSION NUMBERS

The genome sequence and analysis-associated data were deposited in CoGe (<https://genomevolution.org/coge/GenomeInfo.pl?gid=62137>). Clean reads from RNA-seq were uploaded to the NCBI Sequence Read Archive under project number PRJNA758745 (<https://dataview.ncbi.nlm.nih.gov/object/PRJNA758745?reviewer=67gtdiagbfut5pmvf1vjgeu5v>).

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at *Plant Communications Online*.

### FUNDING

These studies were supported by the Natural Science Foundation of Zhejiang Province (grant LY22C160005), the National Natural Science Foundation of China (grant 31700224), Innovative Scientific and Technological Talents in Henan Province (no. 20HASTIT041), and the Outstanding Youth Foundation of Henan Province (no. 202300410041), as well as a Nanyang Technological University startup grant and the Academy of Finland (decisions 318288 and 319947) to J.S.

### AUTHOR CONTRIBUTIONS

F.C. designed the experiments. F.C. and J.S. conceived and interpreted the data. X.Y. and J.S. performed the bioinformatic analysis. F.C. and X.L. performed the experiments. F.C. and Z.H. acquired funding. F.C., X.Y., and J.S. wrote the first manuscript, which all authors edited and approved.

### ACKNOWLEDGMENTS

We gratefully acknowledge Prof. Wenwu Wu (Zhejiang A and F University) for help with RNA-seq analysis and Dr. Xiaohong Zhou for all the help with plant cultivation. No conflict of interest is declared.

Received: September 15, 2021

Revised: February 9, 2022

Accepted: February 24, 2022

Published: February 26, 2022

### REFERENCES

- Aguirre-Becerra, H., Vazquez-Hernandez, M.C., Saenz de la O, D., Alvarado-Mariana, A., Guevara-Gonzalez, R.G., Garcia-Trejo, J.F., and Feregrino-Perez, A.A. (2021). Role of stress and defense in plant secondary metabolites production. In *Bioactive Natural Products for Pharmaceutical Applications*, D. Pal and A.K. Nayak, eds. (Cham: Springer International Publishing), pp. 151–195.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- Archetti, M., Döring, T.F., Hagen, S.B., Hughes, N.M., Leather, S.R., Lee, D.W., Lev-Yadun, S., Manetas, Y., Ougham, H.J., Schaberg, P.G., et al. (2009). Unravelling the evolution of autumn colours: an interdisciplinary approach. *Trends Ecol. Evol.* **24**:166–173.
- Barba-Montoya, J., dos Reis, M., and Yang, Z. (2017). Comparison of different strategies for using fossil calibrations to generate the time prior in Bayesian molecular clock dating. *Mol. Phylogenet. Evol.* **114**:386–400.
- Berensmeier, S. (2006). Magnetic particles for the separation and purification of nucleic acids. *Appl. Microbiol. Biotechnol.* **73**:495–504.
- Bernard, A., Domergue, F., Pascal, S., Jetter, R., Renne, C., Faure, J.-D., Haslam, R.P., Napier, J.A., Lessire, R., and Joubès, J. (2012). Reconstitution of plant alkane biosynthesis in yeast demonstrates that *Arabidopsis* ECERIFERUM1 and ECERIFERUM3 are core components of a very-long-chain alkane synthesis complex. *Plant Cell* **24**:3106–3118.

- Bird, K.A., VanBuren, R., Puzey, J.R., and Edger, P.P.** (2018). The causes and consequences of subgenome dominance in hybrids and recent polyploids. *New Phytol.* **220**:87–93.
- Birney, E., Clamp, M., and Durbin, R.** (2004). GeneWise and genomewise. *Genome Res.* **14**:988–995.
- Blanco, E., Parra, G., and Guigó, R.** (2007). Using geneid to identify genes. *Curr. Protoc. Bioinformatics* **Chapter 4**, Unit 4.3.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., et al.** (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**:365–370.
- Borner, R., Kampmann, G., Chandler, J., Gleißner, R., Wisman, E., Apel, K., and Melzer, S.** (2000). A MADS domain gene involved in the transition to flowering in *Arabidopsis*. *Plant J.* **24**:591–599.
- Bourdenx, B., Bernard, A., Domergue, F., Pascal, S., Léger, A., Roby, D., Pervent, M., Vile, D., Haslam, R.P., Napier, J.A., et al.** (2011). Overexpression of *arabidopsis ECERIFERUM1* promotes wax very-long-chain alkane biosynthesis and influences plant response to biotic and abiotic stresses. *Plant Physiol.* **156**:29–45.
- Bowers, J.E., Chapman, B.A., Rong, J., and Paterson, A.H.** (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**:433–438.
- Bradshaw, H.D., Ceulemans, R., Davis, J., and Stettler, R.** (2000). Emerging model systems in plant biology: poplar (*Populus*) as a model forest tree. *J. Plant Growth Regul.* **19**:306–313.
- Burge, C., and Karlin, S.** (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**:78–94.
- Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J.** (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**:1119–1125.
- Campbell, M.A., Haas, B.J., Hamilton, J.P., Mount, S.M., and Buell, C.R.** (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis*. *BMC Genomics* **7**:327.
- Chavez, D.J., and Lyrene, P.M.** (2009a). Interspecific crosses and backcrosses between diploid *Vaccinium darrowii* and tetraploid southern highbush blueberry. *J. Am. Soc. Hortic. Sci.* **134**:273–280.
- Chavez, D.J., and Lyrene, P.M.** (2009b). Effects of self-pollination and cross-pollination of *Vaccinium darrowii* (Ericaceae) and other low-chill blueberries. *HortScience* **44**:1538–1541.
- Chen, M., and Thelen, J.J.** (2013). ACYL-LIPID DESATURASE2 is required for chilling and freezing tolerance in *arabidopsis*[C][W]. *Plant Cell* **25**:1430–1444.
- Chen, M.-K., Hsu, W.-H., Lee, P.-F., Thiruvengadam, M., Chen, H.-I., and Yang, C.-H.** (2011). The MADS box gene, FOREVER YOUNG FLOWER, acts as a repressor controlling floral organ senescence and abscission in *Arabidopsis*. *Plant J.* **68**:168–185.
- Chen, J., Zhu, X., Ren, J., Qiu, K., Li, Z., Xie, Z., Gao, J., Zhou, X., and Kuai, B.** (2017). Suppressor of overexpression of CO 1 negatively regulates dark-induced leaf degreening and senescence by directly repressing pheophytinase and other senescence-associated genes in *arabidopsis*. *Plant Physiol.* **173**:1881–1891.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J.** (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**:i884–i890.
- Cho, H.Y., Kadowaki, M., Che, J., Takahashi, S., Horiuchi, N., Ogiwara, I., et al.** (2019). Influence of Light Quality on Flowering Characteristics, Potential for Year-Round Fruit Production and Fruit Quality of Blueberry in a Plant Factory. *Fruits* **74**:3–10.
- Chu, W., Gao, H., Cao, S., Fang, X., Chen, H., and Xiao, S.** (2017). Composition and morphology of cuticular wax in blueberry (*Vaccinium* spp.) fruits. *Food Chem.* **219**:436–442.
- Chu, W., Gao, H., Chen, H., Wu, W., and Fang, X.** (2018). Changes in cuticular wax composition of two blueberry cultivars during fruit ripening and postharvest cold storage. *J. Agric. Food Chem.* **66**:2870–2876.
- Colle, M., Leisner, C.P., Wai, C.M., Ou, S., Bird, K.A., Wang, J., Wisecaver, J.H., Yocca, A.E., Alger, E.I., Tang, H., et al.** (2019). Haplotype-phased genome and evolution of phytonutrient pathways of tetraploid blueberry. *Gigascience* **8**:giz012.
- Cui, F., Brosché, M., Lehtonen, M.T., Amirouf, A., Xu, E., Punkkinen, M., Valkonen, J.P.T., Fujii, H., and Overmyer, K.** (2016). Dissecting abscisic acid signaling pathways involved in cuticle formation. *Mol. Plant* **9**:926–938.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.** (2011). The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158.
- De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W.** (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**:1269–1271.
- Deamer, D., Akeson, M., and Branton, D.** (2016). Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**:518–524.
- Diaz-Garcia, L., Garcia-Ortega, L.F., González-Rodríguez, M., Delaye, L., Iorizzo, M., and Zalapa, J.** (2021). Chromosome-level genome assembly of the american cranberry (*Vaccinium macrocarpon* Ait.) and its wild relative *Vaccinium microcarpum*. *Front Plant Sci.* **12**:633310.
- Dimmer, E.C., Huntley, R.P., Alam-Faruque, Y., Sawford, T., O'Donovan, C., Martin, M.J., Bely, B., Browne, P., Mun Chan, W., Eberhardt, R., et al.** (2012). The UniProt-GO annotation database in 2011. *Nucleic Acids Res.* **40**:D565–D570.
- Dorca-Fornell, C., Gregis, V., Grandi, V., Coupland, G., Colombo, L., and Kater, M.M.** (2011). The *Arabidopsis* SOC1-like genes AGL42, AGL71 and AGL72 promote flowering in the shoot apical and axillary meristems. *Plant J.* **67**:1006–1017.
- Fang, Y., Nunez, G.H., Silva, M. N. da, Phillips, D.A., and Munoz, P.R.** (2020). A review for southern highbush blueberry alternative production systems. *Agronomy* **10**:1531.
- Forlani, S., Masiero, S., and Mizzotti, C.** (2019). Fruit ripening: the role of hormones, cell wall modifications, and their relationship with pathogens. *J. Exp. Bot.* **70**:2993–3006.
- Freeling, M.** (2009). Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **60**:433–453.
- Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P., and Firoozabady, E.** (1983). Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* **220**:1049–1051.
- Gao, X., Walworth, A.E., Mackie, C., and Song, G.** (2016). Overexpression of blueberry FLOWERING LOCUS T is associated with changes in the expression of phytohormone-related genes in blueberry plants. *Hortic. Res.* **3**:1–9.
- Genoud, T., Cruz, M.T.S., Kulisic, T., Sparla, F., Fankhauser, C., and Métraux, J.-P.** (2008). The protein phosphatase 7 regulates phytochrome signaling in *arabidopsis*. *PLoS One* **3**:e2699.
- Greer, S., Wen, M., Bird, D., Wu, X., Samuels, L., Kunst, L., and Jetter, R.** (2007). The cytochrome P450 enzyme CYP96A15 is the midchain alkane hydroxylase responsible for formation of secondary alcohols and ketones in stem cuticular wax of *Arabidopsis*. *Plant Physiol.* **145**:653–667.
- Gregis, V., Andrés, F., Sessa, A., Guerra, R.F., Simonini, S., Mateos, J.L., Torti, S., Zambelli, F., Prazzoli, G.M., Bjerkan, K.N., et al.** (2013). Identification of pathways directly regulated by SHORT



- VEGETATIVE PHASE during vegetative and reproductive development in Arabidopsis. *Genome Biol.* **14**:R56.
- Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., and Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**:D121–D124.
- Gupta, V., Estrada, A.D., Blakley, I., Reid, R., Patel, K., Meyer, M.D., Andersen, S.U., Brown, A.F., Lila, M.A., and Loraine, A.E. (2015). RNA-Seq analysis and annotation of a draft blueberry genome assembly identifies candidate genes involved in fruit ripening, biosynthesis of bioactive compounds, and stage-specific alternative splicing. *GigaScience* **4**:5.
- Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**:R7.
- Han, S.-H., Park, Y.-J., and Park, C.-M. (2020). HOS1 activates DNA repair systems to enhance plant thermotolerance. *Nat. Plants* **6**:1439–1446.
- Hanada, K., Zou, C., Lehti-Shiu, M.D., Shinozaki, K., and Shiu, S.-H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* **148**:993–1003.
- Hedges, S.B., Dudley, J., and Kumar, S. (2006). TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**:2971–2972.
- Hellems, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**:R19.
- Hepworth, S.R., Valverde, F., Ravenscroft, D., Mouradov, A., and Coupland, G. (2002). Antagonistic regulation of flowering-time gene SOC1 by CONSTANS and FLC via separate promoter motifs. *EMBO J.* **21**:4327–4337.
- Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., and Quesneville, H. (2014). PASTEC: an automatic transposable element classification tool. *PLoS One* **9**:e91929.
- Horvath, D. (2009). Common mechanisms regulate flowering and dormancy. *Plant Sci.* **177**:523–531.
- Horvath, D.P., Anderson, J.V., Chao, W.S., and Foley, M.E. (2003). Knowing when to grow: signals regulating bud dormancy. *Trends Plant Sci.* **8**:534–540.
- Hou, D., Li, L., Ma, T., Pei, J., Zhao, Z., Lu, M., Wu, A., and Lin, X. (2021). The SOC1-like gene BoMADS50 is associated with the flowering of *Bambusa oldhamii*. *Hortic. Res.* **8**:1–13.
- Ifuku, K., Nakatsu, T., Shimamoto, R., Yamamoto, Y., Ishihara, S., Kato, H., and Sato, F. (2005a). Structure and function of the PsbP protein of Photosystem II from higher plants. *Photosynth Res.* **84**:251–255.
- Ifuku, K., Yamamoto, Y., Ono, T., Ishihara, S., and Sato, F. (2005b). PsbP protein, but not PsbQ protein, is essential for the regulation and stabilization of photosystem II in higher plants. *Plant Physiol.* **139**:1175–1184.
- Immink, R.G.H., Posé, D., Ferrario, S., Ott, F., Kaufmann, K., Valentim, F.L., de Folter, S., van der Wal, F., van Dijk, A.D.J., Schmid, M., et al. (2012). Characterization of SOC1's central role in flowering by the identification of its upstream and downstream regulators. *Plant Physiol.* **160**:433–449.
- Jaakola, L., Pirttilä, A.M., Halonen, M., and Hohtola, A. (2001). Isolation of high quality RNA from bilberry (*Vaccinium myrtillus* L.) fruit. *Mol. Biotechnol.* **19**:201–203.
- Jaakola, L., Määttä, K., Pirttilä, A.M., Törrönen, R., Kärenlampi, S., and Hohtola, A. (2002). Expression of genes involved in anthocyanin biosynthesis in relation to anthocyanin, proanthocyanidin, and flavonol levels during bilberry fruit development. *Plant Physiol.* **130**:729–739.
- Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**:462–467.
- Kalt, W., Cassidy, A., Howard, L.R., Krikorian, R., Stull, A.J., Tremblay, F., and Zamora-Ros, R. (2020). Recent research on the health benefits of blueberries and their anthocyanins. *Adv. Nutr.* **11**:224–236.
- Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S.O., and Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics* **19**:189.
- Kent, W.J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* **12**:656–664.
- Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**:357–360.
- Kitamura, Y., Takeuchi, T., Yamane, H., and Tao, R. (2016). Simultaneous down-regulation of DORMANCY-ASSOCIATED MADS-box6 and SOC1 during dormancy release in Japanese apricot (*Prunus mume*) flower buds. *J. Hortic. Sci. Biotechnol.* **91**:476–482.
- Kloet, S.P.V. (1980). The taxonomy of the highbush blueberry, *Vaccinium corymbosum*. *Can. J. Bot.* **58**:1187–1201.
- Klopfenstein, D.V., Zhang, L., Pedersen, B.S., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C.J., Yunes, J.M., Botvinnik, O., Weigel, M., et al. (2018). GOATOOLS: a Python library for Gene Ontology analyses. *Sci. Rep.* **8**:10872.
- Koonin, E.V., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Krylov, D.M., Makarova, K.S., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., et al. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**:R7.
- Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**:722–736.
- Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics* **5**:59.
- Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**:1812–1819.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760.
- Li, R., Li, Y., Kristiansen, K., and Wang, J. (2008a). SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**:713–714.
- Li, F., Wu, X., Lam, P., Bird, D., Zheng, H., Samuels, L., Jetter, R., and Kunst, L. (2008b). Identification of the Wax Ester Synthase/Acyl-Coenzyme A:Diacylglycerol Acyltransferase WSD1 required for stem wax ester biosynthesis in arabidopsis. *Plant Physiol.* **148**:97–107.
- Li, Y., Sun, H., and Chen, L. (2017). The blueberry industry of China: the past 10 years and the future. *Acta Hortic.* **2017**. <https://doi.org/10.17660/ActaHortic.2017.1180.75>.
- Li, Y., Pi, M., Gao, Q., Liu, Z., and Kang, C. (2019). Updated annotation of the wild strawberry *Fragaria vesca* V4 genome. *Hortic. Res.* **6**:1–9.
- Li, R., Sun, S., Wang, H., Wang, K., Yu, H., Zhou, Z., Xin, P., Chu, J., Zhao, T., Wang, H., et al. (2020). FIS1 encodes a GA2-oxidase that regulates fruit firmness in tomato. *Nat. Commun.* **11**:5844.
- Li, Y., Pei, J., Chen, L., and Sun, H. (2021). China blueberry industry report 2020. *J. Jilin Agric. Univ.* **43**:1–8.
- Lin-Wang, K., Micheletti, D., Palmer, J., Volz, R., Lozano, L., Espley, R., Hellens, R.P., Chagnè, D., Rowan, D.D., Troggo, M., et al.

- (2011). High temperature reduces apple fruit colour via modulation of the anthocyanin regulatory complex. *Plant Cel Environ.* **34**:1176–1190.
- Lobos, G.A., and Hancock, J.F.** (2015). Breeding blueberries for a changing global environment: a review. *Front. Plant Sci.* **6**:782.
- Longman, K.A., and Wareing, P.F.** (1959). Early induction of flowering in birch seedlings. *Nature* **184**:2037–2038.
- Lowe, T.M., and Eddy, S.R.** (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
- Lyrene, P.M., Vorsa, N., and Ballington, J.R.** (2003). Polyploidy and sexual polyploidization in the genus *Vaccinium*. *Euphytica* **133**:27–36.
- Majoros, W.H., Pertea, M., and Salzberg, S.L.** (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**:2878–2879.
- Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., et al.** (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**:D225–D229.
- Melzer, S., Lens, F., Gennen, J., Vanneste, S., Rohde, A., and Beeckman, T.** (2008). Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. *Nat. Genet.* **40**:1489–1492.
- Møller, S.G., Kim, Y.-S., Kunkel, T., and Chua, N.-H.** (2003). PP7 is a positive regulator of blue light signaling in *Arabidopsis*. *The Plant Cell* **15**:1111–1119.
- Mouhu, K., Kurokura, T., Koskela, E.A., Albert, V.A., Elomaa, P., and Hytönen, T.** (2013). The *Fragaria vesca* homolog of SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 represses flowering and promotes vegetative growth. *The Plant Cell* **25**:3296–3310.
- Nicolau, M., Picault, N., Descombin, J., Jami-Alahmadi, Y., Feng, S., Bucher, E., Jacobsen, S.E., Deragon, J.-M., Wohlschlegel, J., and Moissiard, G.** (2020). The plant mobile domain proteins MAIN and MAIL1 interact with the phosphatase PP7L to regulate gene expression and silence transposable elements in *Arabidopsis thaliana*. *PLoS Genet.* **16**:e1008324.
- O'Connor, B.D., and Van der Auwera, G.A.** (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra, First Edition. (O'Reilly Media).
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M.** (1999). KEGG: Kyoto Encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**:29–34.
- Osumi, T., Gruszewski, H.A., Blischak, L.A., Baxter, A.J., Wadl, P.A., Shuman, J.L., Veilleux, R.E., and Shulaev, V.** (2006). High-efficiency transformation of the diploid strawberry (*Fragaria vesca*) for functional genomics. *Planta* **223**:1219–1230.
- Parra, G., Bradnam, K., and Korf, I.** (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23**:1061–1067.
- Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L.** (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**:290–295.
- Porebski, S., Bailey, L.G., and Baum, B.R.** (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* **15**:8–15.
- Price, A.L., Jones, N.C., and Pevzner, P.A.** (2005). De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–358.
- Rabino, I., and Mancinelli, A.L.** (1986). Light, temperature, and anthocyanin production. *Plant Physiol.* **81**:922–924.
- Ramakrishna, A., and Ravishankar, G.A.** (2011). Influence of abiotic stress signals on secondary metabolites in plants. *Plant Signal Behav.* **6**:1720–1731.
- Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al.** (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**:1665–1680.
- Retamales, J.B., Palma, M.J., Morales, Y.A., Lobos, G.A., Moggia, C.E., and Mena, C.A.** (2014). Blueberry production in Chile: current status and future developments. *Rev. Bras. Frutic.* **36**:58–67.
- Robinson, J.P., and Grégori, G.** (2007). Principles of flow cytometry. In *Flow Cytometry with Plant Cells*, J. Doležel, J. Greilhuber, and J. Suda, eds. (John Wiley & Sons, Ltd), pp. 19–40. <https://onlinelibrary.wiley.com/doi/10.1002/9783527610921.ch2>.
- Rody, H.V., Baute, G.J., Rieseberg, L.H., and Oliveira, L.O.** (2017). Both mechanism and age of duplications contribute to biased gene retention patterns in plants. *BMC genomics* **18**:1–10.
- Román, Á., Hernández, M.L., Soria-García, Á., López-Gomollón, S., Lagunas, B., Picorel, R., Martínez-Rivas, J.M., and Alfonso, M.** (2015). Non-redundant contribution of the Plastidial FAD8  $\omega$ -3 desaturase to glycerolipid unsaturation at different temperatures in *Arabidopsis*. *Mol. Plant* **8**:1599–1611.
- Romo-Muñoz, R., Dote-Pardo, J., Garrido-Henríquez, H., Araneda-Flores, J., and Gil, J.M.** (2020). Blueberry consumption and healthy lifestyles in an emerging market. *Span J. Agric. Res.* **17**:e0111.
- Rowan, D.D., Cao, M., Lin-Wang, K., Cooney, J.M., Jensen, D.J., Austin, P.T., Hunt, M.B., Norling, C., Hellens, R.P., Schaffer, R.J., et al.** (2009). Environmental regulation of leaf colour in red 35S:PAP1 *Arabidopsis thaliana*. *New Phytol.* **182**:102–115.
- Sakhanokho, H.F., Rinehart, T.A., Stringer, S.J., Islam-Faridi, M.N., and Pounders, C.T.** (2018). Variation in nuclear DNA content and chromosome numbers in blueberry. *Scientia Horticulturae* **233**:108–113.
- Salojärvi, J., Smolander, O.-P., Nieminen, K., Rajaraman, S., Safronov, O., Safdari, P., Lamminmäki, A., Immanen, J., Lan, T., Tanskanen, J., et al.** (2017). Genome sequencing and population genomic analyses provide insights into the adaptive landscape of silver birch. *Nat. Genet.* **49**:904–912.
- Sapers, G.M., and Phillips, J.G.** (1985). Leakage of anthocyanins from skin of raw and cooked highbush blueberries (*Vaccinium corymbosum* L.). *J. Food Sci.* **50**:437–439.
- Schomburg, F.M., Patton, D.A., Meinke, D.W., and Amasino, R.M.** (2001). FPA, a gene involved in floral induction in *Arabidopsis*, encodes a protein containing RNA-recognition motifs. *The Plant Cell* **13**:1427–1436.
- Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E.** (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**:259.
- Sharpe, R.H., and Darrow, G.M.** (1960). Breeding blueberries for the Florida climate. *Proc. Fla. State Hortic. Soc.* **72**:308–311.
- She, R., Chu, J.S.-C., Wang, K., Pei, J., and Chen, N.** (2009). GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res.* **19**:143–149.
- Shen, L., Kang, Y.G.G., Liu, L., and Yu, H.** (2011). The J-domain protein J3 mediates the integration of flowering signals in *Arabidopsis*. *The Plant Cell* **23**:499–514.
- Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S.P., et al.** (2011). The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* **43**:109–116.

- Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M.** (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212.
- Song, G., Walworth, A., Zhao, D., Jiang, N., and Hancock, J.F.** (2013). The *Vaccinium corymbosum* FLOWERING LOCUS T-like gene (VcFT): a flowering activator reverses photoperiodic and chilling requirements in blueberry. *Plant Cell Rep* **32**:1759–1769.
- Sonmez, C., Bäurle, I., Magusin, A., Dreos, R., Laubinger, S., Weigel, D., and Dean, C.** (2011). RNA 3' processing functions of Arabidopsis FCA and FPA limit intergenic transcription. *PNAS* **108**:8508–8513.
- Stanke, M., and Waack, S.** (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215–225.
- Strik, B.** (2006). Blueberry production and research trends in North America. *Acta Hortic.* <https://doi.org/10.17660/ActaHortic.2006.715.25>.
- Sultana, N., Pascual-Díaz, J.P., Gers, A., Ilga, K., Serçe, S., Vitales, D., and Garcia, S.** (2019). Contribution to the knowledge of genome size evolution in edible blueberries (genus *Vaccinium*). *J. Berry Res.* **10**:1–15.
- Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T.** (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**:e21800.
- Suter, L., Rüegg, M., Zemp, N., Hennig, L., and Widmer, A.** (2014). Gene regulatory variation mediates flowering responses to vernalization along an altitudinal gradient in Arabidopsis. *Plant Physiol.* **166**:1928–1942.
- Szakiel, A., Pączkowski, C., Pensec, F., and Bertsch, C.** (2012). Fruit cuticular waxes as a source of biologically active triterpenoids. *Phytochem. Rev.* **11**:263–284.
- Tang, S., Lomsadze, A., and Borodovsky, M.** (2015). Identification of protein coding regions in RNA transcripts. *Nucleic Acids Res.* **43**:e78.
- Tarailo-Graovac, M., and Chen, N.** (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **Chapter 4**, Unit 4.10.
- Taulavuori, K., Laine, K., and Taulavuori, E.** (2013). Experimental studies on *Vaccinium myrtillus* and *Vaccinium vitis-idaea* in relation to air pollution and global change at northern high latitudes: a review. *Environ. Exp. Bot.* **87**:191–196.
- Taylor, G.** (2002). *Populus*: arabidopsis for forestry. Do we need a model tree? *Ann. Bot.* **90**:681–689.
- Trainin, T., Bar-Ya'akov, I., and Holland, D.** (2013). ParSOC1, a MADS-box gene closely related to Arabidopsis AGL20/SOC1, is expressed in apricot leaves in a diurnal manner and is linked with chilling requirements for dormancy break. *Tree Genet. Genomes* **9**:753–766.
- Trivedi, P., Karppinen, K., Klavins, L., Kviesis, J., Sundqvist, P., Nguyen, N., Heinonen, E., Klavins, M., Jaakola, L., Väänänen, J., et al.** (2019). Compositional and morphological analyses of wax in northern wild berry species. *Food Chem.* **295**:441–448.
- Tyagi, S., Sri, T., Singh, A., Mayee, P., Shivaraj, S.M., Sharma, P., and Singh, A.** (2019). SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 influences flowering time, lateral branching, oil quality, and seed yield in Brassica juncea cv. Varuna. *Funct. Integr. Genomics* **19**:43–60.
- Vaser, R., Sović, I., Nagarajan, N., and Šikić, M.** (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**:737–746.
- Veley, K.M., and Michaels, S.D.** (2008). Functional redundancy and new roles for genes of the autonomous floral-promotion pathway. *Plant Physiol.* **147**:682–695.
- Voogd, C., Wang, T., and Varkonyi-Gasic, E.** (2015). Functional and expression analyses of kiwifruit SOC1-like genes suggest that they may not have a role in the transition to flowering but may affect the duration of dormancy. *J. Exp. Bot.* **66**:4699–4710.
- Vorsa, N., Johnson-Cicalese, J., and Polashock, J.** (2009). A blueberry by cranberry hybrid derived from a *Vaccinium darrowii* × (*V. macrocarpon* × *V. oxycoccos*) intersectional cross. *Acta Hortic.* <https://doi.org/10.17660/ActaHortic.2009.810.24>.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al.** (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**:e112963.
- Walworth, A.E., Chai, B., and Song, G.** (2016). Transcript profile of flowering regulatory genes in VcFT-overexpressing blueberry plants. *PLoS One* **11**:e0156993.
- Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., et al.** (2012a). MScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**:e49.
- Wang, S.Y., Camp, M.J., and Ehlenfeldt, M.K.** (2012b). Antioxidant capacity and  $\alpha$ -glucosidase inhibitory activity in peel and flesh of blueberry (*Vaccinium* spp.) cultivars. *Food Chem.* **132**:1759–1768.
- Wang, J., Gao, Z., Li, H., Jiu, S., Qu, Y., Wang, L., Ma, C., Xu, W., Wang, S., and Zhang, C.** (2020). Dormancy-associated MADS-box (DAM) genes influence chilling requirement of sweet cherries and co-regulate flower development with SOC1 gene. *Int. J. Mol. Sci.* **21**:921.
- Wang, X., Gao, Y., Wu, X., Wen, X., Li, D., Zhou, H., Li, Z., Liu, B., Wei, J., Chen, F., et al.** (2021). High-quality evergreen azalea genome reveals tandem duplication-facilitated low-altitude adaptability and floral scent evolution. *Plant Biotechnol. J. n/a* **19**:2544–2560.
- Wick, R.R., Judd, L.M., and Holt, K.E.** (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**:129.
- Wilkie, J.D., Sedgley, M., and Olesen, T.** (2008). Regulation of floral initiation in horticultural trees. *J. Exp. Bot.* **59**:3215–3228.
- Wu, T., Gao, Y., Guo, X., Zhang, M., and Gong, L.** (2018). Oxidative Medicine and Cellular Longevity. <https://www.hindawi.com/journals/omcl/2018/4051232/>.
- Wu, H., Ma, T., Kang, M., Ai, F., Zhang, J., Dong, G., and Liu, J.** (2019). A high-quality *Actinidia chinensis* (kiwifruit) genome. *Hortic. Res.* **6**:1–9.
- Wu, Z., Fang, X., Zhu, D., and Dean, C.** (2020). Autonomous pathway: FLOWERING LOCUS C repression through an antisense-mediated chromatin-silencing Mechanism1 [CC-BY]. *Plant Physiol.* **182**:27–37.
- Wu, C., Deng, C., Hilario, E., Albert, N.W., Lafferty, D., Grierson, E.R.P., Plunkett, B.J., Elborough, C., Saei, A., Günther, C.S., et al.** (2021). A chromosome-scale assembly of the bilberry genome identifies a complex locus controlling berry anthocyanin composition. *Mol. Ecol. Resources* **22**:345–360.
- Xu, Z., and Wang, H.** (2007). LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**:W265–W268.
- Xu, D., Marino, G., Klingl, A., Enderle, B., Monte, E., Kurth, J., Hiltbrunner, A., Leister, D., and Kleine, T.** (2019). Extrachloroplastic PP7L functions in chloroplast development and abiotic stress tolerance. *Plant Physiol.* **180**:323–341.
- Yamamoto, M., Uji, S., Sugiyama, T., Sakamoto, T., Kimura, S., Endo, T., and Nishikawa, S.** (2020). ERdj3B-Mediated quality control maintains anther development at high Temperatures1 [OPEN]. *Plant Physiol.* **182**:1979–1990.
- Yang, Z.** (2007). Paml 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**:1586–1591.

- Yang, Q., Liu, E., Fu, Y., Yuan, F., Zhang, T., and Peng, S.** (2019). High temperatures during flowering reduce fruit set in rabbiteye blueberry. *J. Am. Soc. Hortic. Sci.* **144**:339–351.
- Yang, F.-S., Nie, S., Liu, H., Shi, T.-L., Tian, X.-C., Zhou, S.-S., Bao, Y.-T., Jia, K.-H., Guo, J.-F., Zhao, W., et al.** (2020). Chromosome-level genome assembly of a parent species of widely cultivated azaleas. *Nat. Commun.* **11**:5269.
- Yeats, T.H., and Rose, J.K.C.** (2013). The formation and function of plant cuticles. *Plant Physiol.* **163**:5–20.
- Yeats, T.H., Huang, W., Chatterjee, S., Viart, H.M.-F., Clausen, M.H., Stark, R.E., and Rose, J.K.C.** (2014). Tomato Cutin Deficient 1 (CD1) and putative orthologs comprise an ancient family of cutin synthase-

like (CUS) proteins that are conserved among land plants. *Plant J.* **77**:667–675.

- Yoo, S.K., Chung, K.S., Kim, J., Lee, J.H., Hong, S.M., Yoo, S.J., Yoo, S.Y., Lee, J.S., and Ahn, J.H.** (2005). CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in Arabidopsis. *Plant Physiol.* **139**:770–778.
- Zhang, Y., Butelli, E., and Martin, C.** (2014). Engineering anthocyanin biosynthesis in plants. *Curr. Opin. Plant Biol.* **19**:81–90.
- Zhu, C., Zheng, X., Huang, Y., Ye, J., Chen, P., Zhang, C., Zhao, F., Xie, Z., Zhang, S., Wang, N., et al.** (2019). Genome sequencing and CRISPR/Cas9 gene editing of an early flowering Mini-Citrus (*Fortunella hindsii*). *Plant Biotechnol. J.* **17**:2199–2210.