Check for updates

the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2023) **116,** 989–1002 doi: 10.1111/tpj.16433

RESOURCE

A haplotype resolved chromosome-scale assembly of North American wild apple *Malus fusca* and comparative genomics of the fire blight *Mfu10* locus

Ben N. Mansfeld[1,*], Alan Yocca[2], Shujun Ou[3], Alex Harkess[2], Erik Burchard[4], Benjamin Gutierrez[5], Steve van Nocker[6] and Christopher Gottschalk[4,*]

[1]*Department of Biology, Washington University in St. Louis, St. Louis, Missouri, USA,*
[2]*HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA,*
[3]*Department of Molecular Genetics, The Ohio State University, Columbus, Ohio, USA,*
[4]*USDA ARS, Appalachian Fruit Research Station, Kearneysville, West Virginia, USA,*
[5]*USDA ARS, Plant Genetic Resources Unit, Geneva, New York, USA, and*
[6]*Department of Horticulture, Michigan State University, East Lansing, Michigan, USA*

Received 22 March 2023; revised 8 July 2023; accepted 12 August 2023; published online 28 August 2023.
*For correspondence (e-mail bmansfeld@wustl.edu; christopher.gottschalk@usda.gov).

SUMMARY

The Pacific crabapple (*Malus fusca*) is a wild relative of the commercial apple (*Malus* × *domestica*). With a range extending from Alaska to Northern California, *M. fusca* is extremely hardy and disease resistant. The species represents an untapped genetic resource for the development of new apple cultivars with enhanced stress resistance. However, gene discovery and utilization of *M. fusca* have been hampered by the lack of genomic resources. Here, we present a high-quality, haplotype-resolved, chromosome-scale genome assembly and annotation for *M. fusca*. The genome was assembled using high-fidelity long-reads and scaffolded using genetic maps and high-throughput chromatin conformation capture sequencing, resulting in one of the most contiguous apple genomes to date. We annotated the genome using public transcriptomic data from the same species taken from diverse plant structures and developmental stages. Using this assembly, we explored haplotypic structural variation within the genome of *M. fusca*, identifying thousands of large variants. We further showed high sequence co-linearity with other domesticated and wild *Malus* species. Finally, we resolve a known quantitative trait locus associated with resistance to fire blight (*Erwinia amylovora*). Insights gained from the assembly of a reference-quality genome of this hardy wild apple relative will be invaluable as a tool to facilitate DNA-informed introgression breeding.

Keywords: Malus, genome assembly, crop wild relative, genomic resource, *Erwinia amylovora*, copy number variation.

INTRODUCTION

The Pacific crabapple (*Malus fusca*), one of four native North American species, is found in the Pacific Northwest ranging from Alaska and British Columbia to California (USDA Agricultural Research Service, 2015). These hardy trees routinely grow in conditions in which the vast majority of cultivated apple (*Malus* × *domestica* Borkh.) cultivars cannot survive and reproduce in; *M. fusca* can withstand winters of −46°C or colder (Fiala, 1994), can grow on beach heads, in sandy soils, exposed to brackish water and in waterlogged conditions (USDA Agricultural Research Service, 2015; Volk, 2019). Moreover, *M. fusca* has been found to be resistant to fire blight (*Erwinia amylovora*), a devastating disease that is endemic to North America, but is now found worldwide (Bonn & van der Zwet, 2000; Dougherty et al., 2021; Emeriewen et al., 2014).

Many environmental conditions, including those derived from human-caused climate change, increasingly burden apple production (Volk, Chao, et al., 2015). For example, abiotic stress such as water logging and spring frosts during bloom can lead to reduced yields, disrupted growth patterns, and in extreme situations, loss of trees

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

(Atkinson et al., 2000; Bhusal et al., 2019; Dalhaus et al., 2020; Gottschalk & Van Nocker, 2013; Schrader et al., 2001; Torres et al., 2013, 2016; Way et al., 1991). Furthermore, in addition to abiotic stresses, apple suffers from other devastating diseases apart from fire blight, such as apple scab, bitter pit, and cedar apple rust, as well as other pests (e.g., codling moth). These can additionally reduce yields or the market value of fruit (MacHardy, 1996; Way et al., 1991). To alleviate these production limitations, plant breeders strive to impart genetic resistance or resilience into improved cultivars. However, in apple, this process has been limited by breeding bottlenecks resulting in high interrelatedness of many cultivated varieties and breeding lines (Migicovsky et al., 2021; Muranty et al., 2020).

One approach to overcome this problem is to introduce novel genetics from crop wild relatives (CWRs). One successful example of CWR hybridization in apple was the introduction of resistance to apple scab caused by the fungus *Venturia inaequalis* (Gessler & Pertot, 2012). To that end, *M. floribunda* selection 821 was used to develop hybrids with *M.* × *domestica* which, ultimately, were used to identify a resistance locus named *Vf* (Hough et al., 1953; Williams, 1966). While these opportunities are afforded by the fact that *M.* × *domestica* readily hybridizes with multiple wild relatives, this process is limited by the considerable effort needed to purge undesirable traits and linkage drag from wild introgressions. This is in part due to the lack of genomic resources in wild species, the high heterozygosity due to self-incompatibility, and the long generation time in apple (Migicovsky et al., 2021; Sakurai et al., 2000; Volk, Henk, et al., 2015).

Introgression breeding and genome editing, offer additional opportunities for further improvement of cultivated apples using genetics from wild species, and there has been a concerted effort, in recent years, to develop genomic resources in apple and its wild relatives (Chen et al., 2019; Daccord et al., 2017; Khan et al., 2022; Li, Wang, et al., 2022; Sun et al., 2020; Velasco et al., 2010; Zhang et al., 2019). This has been bolstered by technological advancements that have enabled generating haplotype-resolved genomes of different progenitor species including *M. sieversii* and *M. sylvestris* (Sun et al., 2020). The *Malus* genus consists of 25 to 47 recognized species and additional hybrids (Robinson et al., 2001). Of those species, *M. fusca* and the other North American natives have been isolated from Asian and European gene pools used in the domestication of apple (Volk, 2019). Thus, *M. fusca* and its other North American relatives, offer untapped potential for unique disease resistance and abiotic tolerance traits.

As part of the effort to expand resources for breeding and genetics in apples, we report herein the high-quality, haplotype-resolved genome assembly of the *M. fusca* accession PI 589975. We used high-fidelity (HiFi) long reads, together with high-throughput chromosome conformation capture (Hi-C) to assemble and phase both haplotypes of this coastal Alaskan accession, which is resistant to fire blight, and moderately resistant to apple scab, as well as potentially other abiotic stresses (Dougherty et al., 2021; Fiala, 1994; Khan & Chao, 2017; Papp et al., 2020; USDA Agricultural Research Service, 2015; Way et al., 1991). After gene annotation, we compared the synteny and genome architecture of this assembly to other domesticated and wild *Malus* genomes. Lastly, we explored synteny and presence-absence variance of candidate genes identified within the herein resolved fire blight resistance locus, *FB_MFu10* (Emeriewen et al., 2020), in comparison to other *Malus* genomes.

RESULTS AND DISCUSSION

We selected to sequence PI 589975 (GMAL 2891) as a representative accession for *M. fusca*, since multiple previous evaluations indicated that it was hardy and resistant to both fire blight and apple scab (Khan & Chao, 2017; Papp et al., 2020) (Figure 1a–c). This accession is one of 2349 accessions in the USDA collection that has been evaluated for natural fire blight shoot infection, where it is rated as a "1; Very resistant – no occurrence" (USDA Agricultural Research Service, 2015). Moreover, Khan and Chao (2017) conducted a 2-year artificial inoculation study that rated PI 589975 as "resistant" based on necrosis length measurements and natural blight score as "1 – very resistant". For apple scab, PI 589975 was rated as "moderately resistant" based on visual assessment in a replicated block trial evaluated across three sampling dates for 2 years (Papp et al., 2020). PI 589975 was originally sampled from Scow Bay near Petersburg, Alaska near the edge of woods along the beach (USDA Agricultural Research Service, 2015; Figure 1d). The accession was donated to the USDA collection in August of 1988 by Michael Medalen and is a member of the core collection grown on 'Budagovsky 9' rootstock (USDA Agricultural Research Service, 2015).

Reference-quality genome assembly

We obtained 95 Gb and 21.7 Gb of sequence for this accession of *M. fusca* from Illumina and PacBio HiFi sequencing, respectively. The Illumina sequencing-based *k*-mer analysis yielded an estimated genome size of 694 Mb and heterozygosity of 0.8% (Figure 1e). Thus, the HiFi sequencing represented 31.3× coverage from 1.4 million reads, which is sufficient for *de novo* haplotype resolved assembly using HiFiasm (Cheng et al., 2021, 2022). High-quality Hi-C libraries for *M. fusca* were extremely difficult to generate and after several attempts with different kits were still of consistently of low concentrations, suggesting issues with extraction and amplification of the proximal DNA contacts. Regardless, we sequenced one Hi-C library and obtained ~130× data with sufficient quality for haplotype phasing. HiFiasm successfully resolved the two haplotype

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

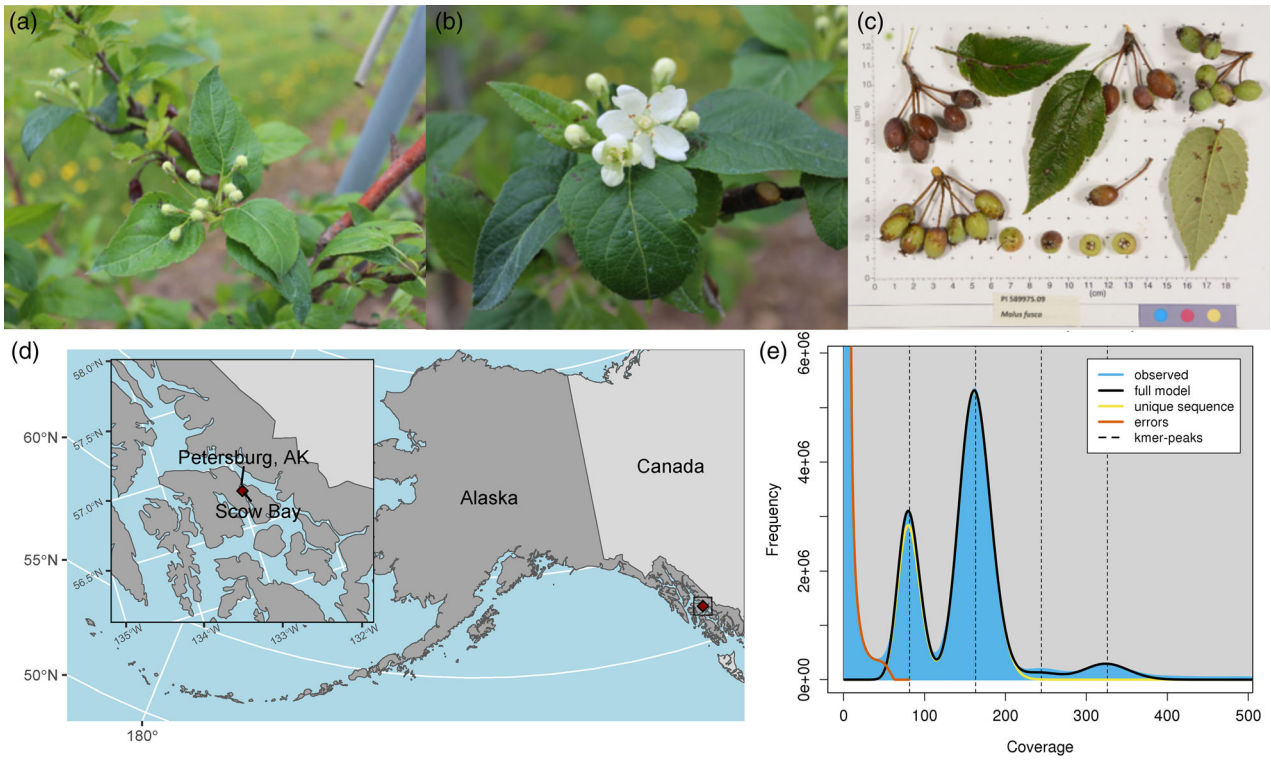

Figure 1. (a) *Malus fusca* accession PI 589975 GRIN-Global identification photograph. (b) Ballon stage blossoms of the living PI 589975 accession in the USDA Plant Germplasm Repository Unit, Geneva, NY. (c) King bloom stage of the living PI 589975 accession in the USDA Plant Germplasm Repository Unit, Geneva, NY. (d) Map of the approximate collection position of the PI 589975 in Alaska. Insert is of the specific location of Scow Bay near Petersburg, AK. (e) *k*-mer count plot from GenomeScope. The haploid genome size was predicted to be 694 Mb with approximately 56% repetitive sequence and an estimated heterozygosity of 0.8%.

assemblies of 682 and 644 Mb in length (Figure S1) and contig N50 of 18.7 and 21.4 Mb, respectively (Table 1). Some contigs were near chromosome length, with the largest contig assembled at 42.3 Mb.

We assessed two different strategies to scaffold this assembly, (1) a Hi-C-based approach using Salsa2 (Ghurye et al., 2019), and (2) using a *Malus* combined linkage map (Bianco et al., 2014; Catchen et al., 2020; Di Pierro et al., 2016). Interestingly, the extremely rapid (elapsed time ~ 30 sec) linkage map-based method achieved better results than the Hi-C-based approach. For example, the Salsa2 pipeline produced a mis-join between two chromosomes resulting in a chromosome of >70 Mb. Moreover, the linkage map + Hi-C approach resulted in a greater number of contigs placed within the 17 chromosome-sized pseudomolecules, thus yielding a lower total number of contigs, greater N50s values, and more contiguous scaffolds of >1 Mb that more closely corresponded to the base chromosome number of 17 for *Malus* (Table S1). We subsequently used the Hi-C data as an orthogonal verification of the map-based scaffolding. We inspected the Hi-C contact maps with Juicebox (Durand et al., 2016), and found they showed high proximal interactions with the scaffolds in contig placement and order and needed only minor manual curation (Figure S2a,b). This suggests that if available, a map from a closely related, albeit different species can be effective in rapidly scaffolding such highly contiguous long-read-based assemblies. Thus, most of the structural variation between the map and assembly is captured within the megabase-sized contigs, and the map's primary role lies in the orientation and placement of only a few contigs per chromosomal pseudomolecule.

The two haplotype assemblies (or haplomes) are 94% and 91% complete based on the estimated genome size. Merqury *k*-mer-based analysis showed a completeness score of 98.2% for the diploid assembly and a QV score of 64.1, the highest to date in any wild *Malus* species. We report a complete genome BUSCO (Simão et al., 2015) score of 98.8 and 98.9% for the two assemblies respectively. The duplicate BUSCO score was 37.2 and 37.4%, comparable to other *Malus* spp. (Sun et al., 2020), and indicative of the relatively recent paleo-duplication event in *Malus* (Velasco et al., 2010). In summary, after scaffolding and decontamination, we obtained two haplotype assembles, which were highly contiguous, nearly complete, and extremely accurate with lengths of 651 Mb and 634 Mb and with scaffold N50s of 36.8 and 36.1 Mb, respectively (Table 1).

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.,

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1 Genome assembly statistics

Assembly	Length (bp)	Number of contigs	N50	Contigs >1 Mb	% of the estimated size	Merqury QV	Merqury QV error rate	Merqury *k*-mer completeness	% BUSCO	Complete single copy Busco	Complete duplicate Busco	Fragmented Busco	Missing Busco
HiFi Contigs													
Haplotype 1	682 253 989	1141	18 869 149	48	98.31%	62.39	5.77E-07	86.82%					
Haplotype 2	644 255 151	273	21 444 431	41	92.83%	67.14	1.93E-07	86.51%					
Combined	1 326 509 140	1414				64.09	3.90E-07	98.21%					
Scaffolds													
Haplotype 1	682 257 189	1109	36 508 756	18	98.31%								
Haplotype 2	644 257 951	245	36 111 528	17	92.84%								
Decontaminated assembly													
Haplotype 1	651 182 355	235	36 779 757	18	93.83%				98.70%	993	600	11	10
Haplotype 2	637 756 398	97	36 111 528	17	91.90%				98.82%	992	603	11	8
Chromosome-only assembly													
Haplotype 1	637 459 684	17	36 779 757	17	91.86%				98.76%	996	598	11	9
Haplotype 2	631 922 950	17	36 111 528	17	91.06%				98.88%	994	602	11	7

To further evaluate the assembly contiguity of repetitive sequences, we computed the LAI metric (Ou et al., 2018) of *M. fusca* haplotypes and other published apple genomes/haplotypes (Chen et al., 2019; Daccord et al., 2017; Khan et al., 2022; Li, Wang, et al., 2022; Linsmith et al., 2019; Sun et al., 2020; Velasco et al., 2010; Zhang et al., 2019; Table S2). Our phased haplotypes have the highest LAI scores and are greater than 20 ("gold" standard based on Ou et al., 2018), indicating that long terminal repeats (LTRs) and TEs were assembled with high quality (Table S2). The use of HiFi reads allowed us to sequence through repeats accurately and enabled assembly of chromosome-scale contigs. As a result, our assemblies represent one of the highest-quality diploid pome fruit genome published to date.

Genome annotation

The availability of multiple *Malus* genomes allowed side-by-side comparisons of their repeat content and contiguity but required consistent TE annotation for this purpose. We collected all published *Malus* genomes and the European pear (*Pyrus Communis*) genome (Chen et al., 2019; Daccord et al., 2017; Khan et al., 2022; Li, Wang, et al., 2022; Linsmith et al., 2019; Sun et al., 2020; Velasco et al., 2010; Zhang et al., 2019). Together with the two haplotypes of *M. fusca* genome, we created a TE annotation for the *Malus* genus with the European pear genome as an outgroup. We found that the *Malus* genome assemblies contained between 48.6% to 62.43% TEs, while the European pear genome contained roughly 45.3% (Table S2). The two *M. fusca* assemblies fell within this range with an average TE content of 58% (Table 2). Long Terminal Repeat (LTRs) retrotransposons were the most abundant TE within these genomes. The two *M. fusca* haplotypes were found to contain 370.54 and 366.02 Mb of annotated transposable elements (TE), equating to 58.13% and 57.92%, respectively, of each haplotype's total length as predicted by the *k*-mer-based approach (Figure 1e, Table 2).

Previous exhaustive sequencing of the *M. fusca* transcriptome (PRJNA267116), including 72 different tissue types and developmental stages, allowed us an unprecedented opportunity to thoroughly annotate the genome of *M. fusca*. In total, we annotated 46 622 and 45 853 genes for each of the haplotype assemblies, respectively (Table 2). Additionally, using pfam functional annotation we classified 4234 and 3912 of those genes as TE-related in the two haplotypes, respectively. Roughly 75% of annotated genes had an Annotation Edit Distance of less than 0.25 indicating high support by transcriptional and protein evidence (Figure S3). A BUSCO analysis of the two annotated transcriptomes indicated 96.8% and 96.6% completeness, respectively. Additionally, we annotated 947 tRNA and 1495 rRNA genes between the two haplotypes (Table 2).

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2 Genome annotation statistics

Assembly	Haplotype 1	Haplotype 2
Intergenic		
LTR - Copia	9.96%	11.16%
LTR - Ty3	15.48%	15.40%
LTR - Unknown	16.17%	14.93%
LINE - L1	0.83%	0.78%
LINE - RTE	0.17%	0.20%
SINE	0.50%	0.49%
TIR-CACTA	2.06%	1.97%
TIR-Mutator	4.57%	5.31%
TIR-PIF Harbinger	2.31%	2.63%
TIR-Tc1 Mariner	0.39%	0.18%
TIR-hAT	3.02%	2.98%
Helitron	2.67%	1.88%
Total %	58.13%	57.92%
Total bp	370.54 Mb	366.02 Mb
Genic		
% Complete BUSCO genes	97.5	96.6
Complete single copy BUSCO	1006	1010
Complete duplicate BUSCO	557	549
Fragmented BUSCO	14	13
Missing BUSCO	37	42
Genes	46 622	45 853
tRNA genes	475	472
rRNA gene	673	822

These resources amount to one of the most thoroughly annotated apple genomes to date (Figure 2a).

Haplotypic polymorphisms and structural variation in *M. fusca*

We aligned the two haplotype assemblies against each other and identified SNPs and large SVs between the two haplotypes. We found a total of 2 454 873 SNPs and 1 969 639 small InDels. However, high heterozygosity in plants manifests not only as single nucleotide polymorphisms, and recent work in other heterozygous crop species (e.g., Mansfeld et al., 2021; Zhou et al., 2019) revealed that large haplotypic structural variation (SV) contributes to differences between the haplotypes in these species. Our highly contiguous *M. fusca* haplotype assemblies thus present an opportunity to evaluate the variation between haplotypes in an outcrossing wild relative of a cultivated fruit tree. Apart from the small haplotypic differences, we identified over 18 000 large SVs ranging in size from 50 to 50 000 bp (File S1). These included large insertions, deletions, tandem duplications/contractions, and repeat expansions/contractions (Figure 2b). Even though this is a conservative, length-limited, estimate of haplotypic SVs, greater than 77 Mb (~10%) of total genomic space was impacted by SV. For comparison to another highly heterozygous crop genome analyzed by similar methods, the volume of haplotypic SVs detected in this assembly was larger than that observed in the African cassava genome (Mansfeld et al., 2021), even though heterozygosity in cassava (~1.4%) was estimated at nearly double that of *M. fusca*. This is likely due to the high completeness, contiguity, and scaffolded nature of both haplotypes in the *M. fusca* assembly, which allows for more accurate and thorough haplotypic comparisons. This suggests that modern haplotype-resolved assembly strategies (Cheng et al., 2021) such as the one used herein, have crucial implications for the ability to detect these important haplotypic SVs.

We were especially interested in the 3774 large haplotypic deletions observed, as these might cause gene hemizygosity (e.g., Zhou et al., 2019) or impact important gene function (Table S3). To further validate these large InDels, we compared short read coverage at the deletion sites to random (non-deletion) regions of the same size (Figure 2c). As expected, we found that mapped read coverage in the deletions was roughly 50%, and significantly different from that of randomly selected genome regions (1000× bootstrap KS test, *P*-value <2.22e-16), suggesting that most haplotypic deletions identified by sequence alignment are accurate. We further explored specific examples of large deletions that overlapped with genes and further validated these cases by examining inflated insert sizes between paired reads. Overall, we identified 3853 unique cases where exons overlapped with a haplotypic deletion (Figure 2d; Table S4). This included instances of complete hemizygosity of some genes due to these large deletions. For example, a 42 kb heterozygous deletion on chromosome 01 causes hemizygosity of *MfusH1_01g00351*, while a second 25 kb deletion on chromosome 15, removes one allele of *MfusH1_15g02509*. Similar structural variation has been shown to have substantive effects on important agricultural traits. For example, berry color in Chardonnay grape is likely altered due to hemizygosity at the *MybA* locus (Zhou et al., 2019). The above two validated deletions also remove the entire upstream regions for *MfusH1_01g00352* and *MfusH1_15g02509*, respectively. Large haplotypic deletions have been also implicated in impacting gene expression profiles by modifying their cis-regulatory landscape (Mansfeld et al., 2021; Sun et al., 2020), as well as have important consequences for the epigenomic landscape (Zhong et al., 2022). Using this assembly, gene hemizygosity can now be taken into account when attempting to introgress traits from *M. fusca* in breeding efforts. Future research should thus explore the impacts of these large haplotypic SVs on allele-specific expression in regard to important traits in *M. fusca* and how these might be useful to breeders.

Comparing apples to apples: Synteny within Malus

The emergence of 3rd generation sequencing technologies and improved scaffolding methods (e.g., Hi-C and Omni-C sequencing), have resulted in numerous high-quality apple genomes to compare against. The first, high-quality apple

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.,

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

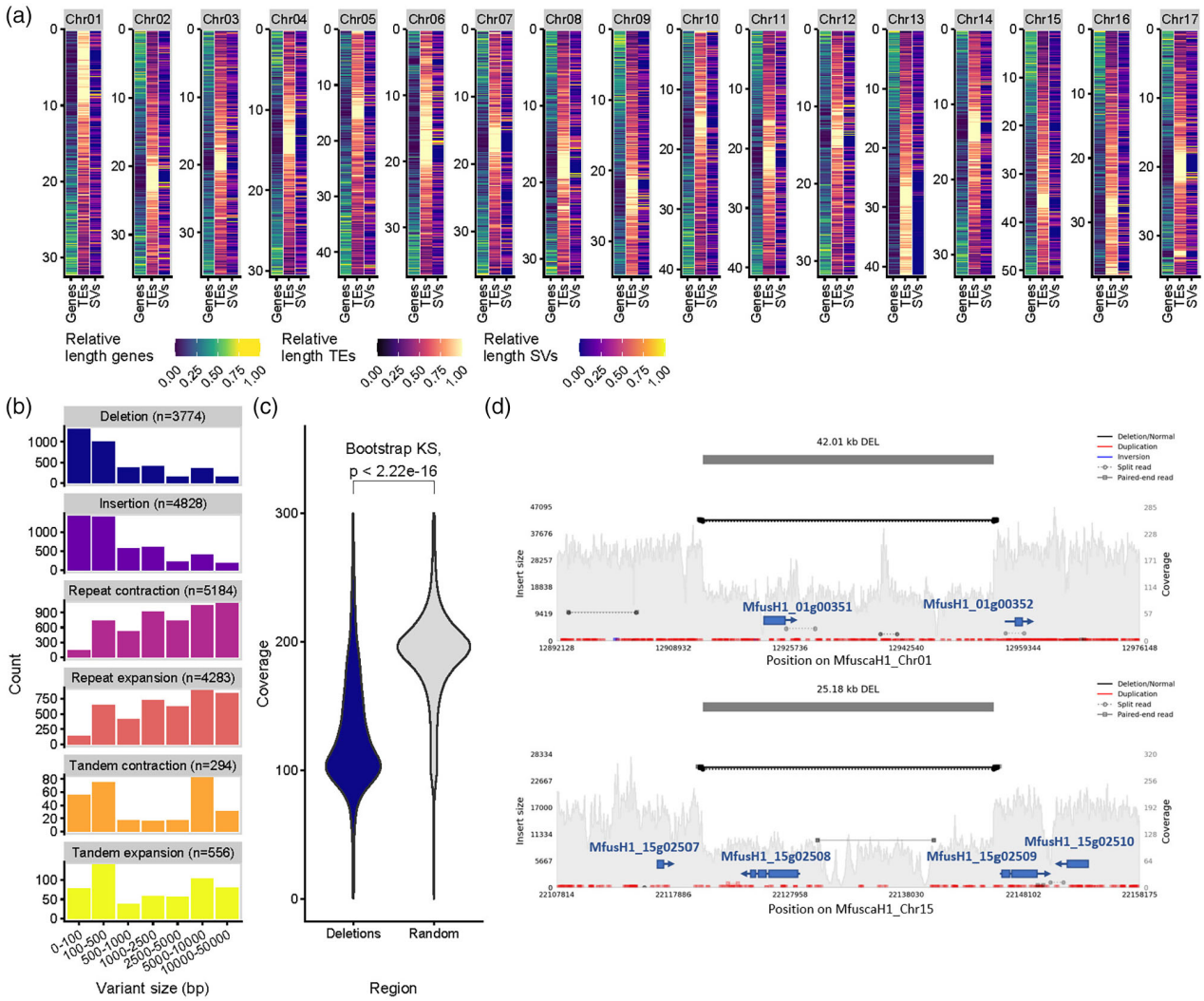

Figure 2. *Malus fusca* genome assembly annotation and haplotype comparisons. (a) Distribution heat maps of annotated features across all 17 chromosomes (genes, TEs-Transposable Elements, and SVs-Structural Variations). The relative cumulative length of each feature type was calculated within a 100 Kbp window. (b) Size distribution of different haplotypic SV within the two haplotypes. (c) WGS read coverage between deletions vs random positions within the genome. The difference in distance distributions was evaluated by a 1000× bootstrapped Kolmogorov–Smirnov (KS) test. (d) Examples of large haplotypic deletions that contain annotated genes within the SV, which result in hemizygous genes. Gray histograms represent depth of coverage (right *y*-axis) of short reads mapped to the haplotype 1 assembly. Reads pairs spanning the identified deletions are highlighted in black and the size of the insert between pairs is denoted by the position vs. the left y-axis. Reads paired with dotted lines indicate split reads that map to both sides of the deletion. Read pairs in red indicate sequence duplications. Gene models are denoted in blue.

genome developed using long-reads was GDDH13, a doubled haploid of Golden Delicious, and it serves as an inbred reference genome (Daccord et al., 2017). We aligned *M. fusca* haplotype 1 to GDDH13 to identify regions of high sequencing similarity and length. Even though *M. fusca* has been kept separate from the rest of the domestication history of apple, we still observe high sequence similarity and co-linearity between *M. fusca* and *M.* × *domestica* (Figure 3a). We identified 2 474 476 SNPs and 1 400 493 small indels as well as 17 467 large SVs (affecting 114 Mbp) between the species (Figure S4, File S2). Similar species-specific variations have been shown to underlie useful crop improvement traits. In tomato, for example, SV between wild and domesticated material was shown to impact fruit size and volatile composition (Alonge et al., 2020). Thus, the genome assembly herein will support future work to establish the role of SV on similar traits in *Malus*.

We also performed gene synteny analysis with two wild progenitors of domesticated apple, *M. sieversii* and *M. sylvestris* (Sun et al., 2020). Comparisons of chloroplast genomes indicate that *M. fusca* is closely related to these species, which suggests Asiatic origins for *M. fusca* (Nikiforova et al., 2013; Robinson et al., 2001; Volk, Chao, et al.,

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

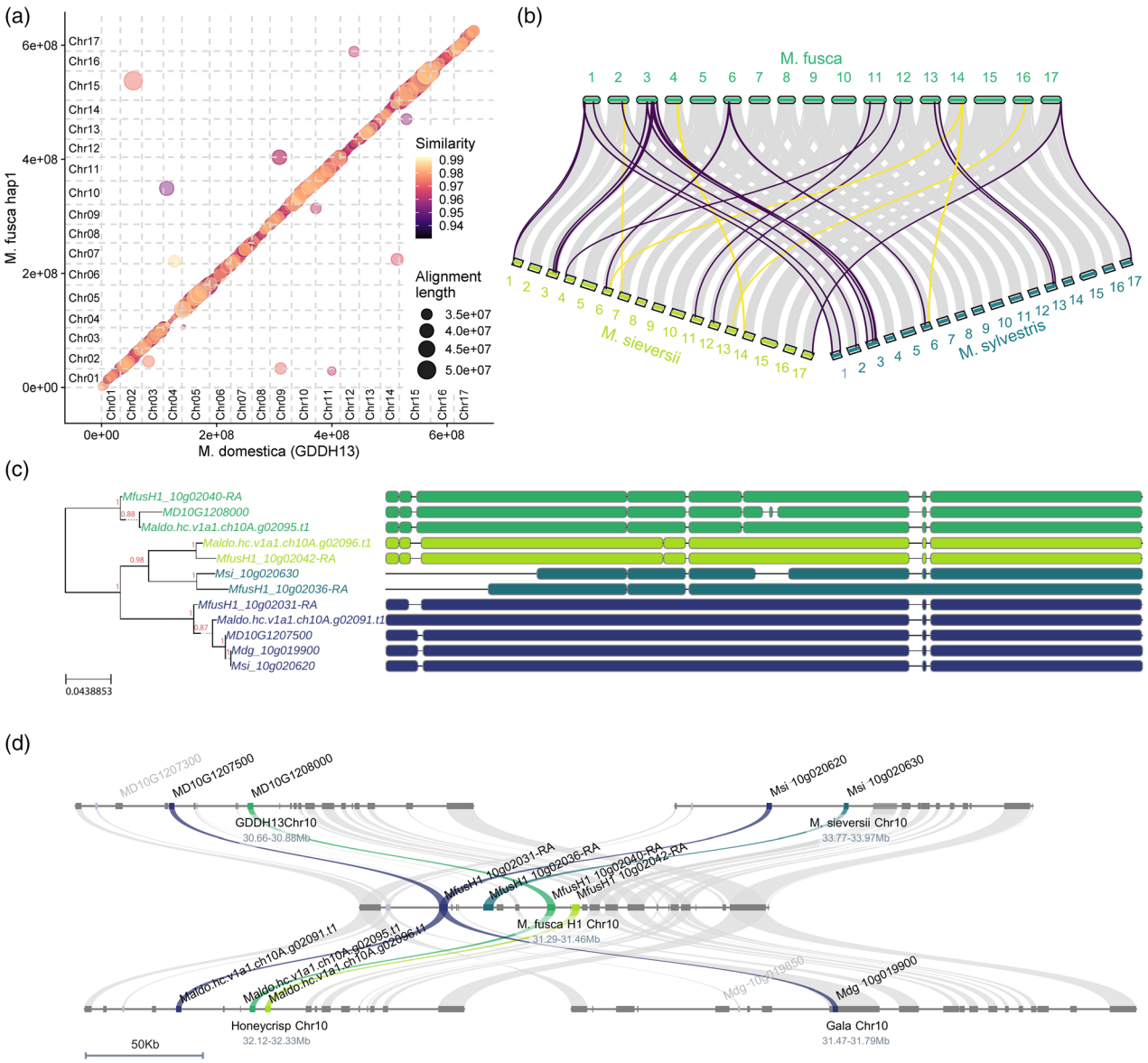

Figure 3. Comparative genomics within *Malus* and the *FB_Mfu10* fire blight resistance locus. (a) Alignment length and sequence similarity between *M. fusca* haplotype 1 and *M.* × *domestica* GDDH13. (b) Gene synteny between the *M. fusca* genome and wild apple species *M. sieversii* and *M. sylvestris* genomes (purple = inversions, yellow = translocations). (c) Phylogenetic tree and peptide alignments of genes encoding G-type lectin S-receptor-like serine/threonine-protein kinases within the *FB_Mfu10* locus in *M. fusca*, *M.* × *domestica* cultivars, and *M. sieversii*. (d) *Microsynteny at the Mfu10 locus*. Copy number variation of G-type lectin S-receptor-like genes within the *FB_Mfu10* locus associated with fire blight resistance in *M. fusca* (center) compared to susceptible *M. sieversii* (top right), and *M.* × *domestica* Golden Delicious (top left), *Honeycrisp* (bottom left), and Gala (bottom right). Orthologs of G-type lectin S-receptor-like serine/threonine-protein kinase genes are annotated by locus id and syntenic genes are annotated by color between the four genomes. Gene colors in panels (c and d) match to show the phylogeny-based synteny inference. Gene fragments with sequence similarity to these receptors are also denoted but shaded in light gray.

2015; Volk, Henk, et al., 2015). Thus perhaps, these wild species were only recently separated geographically by the submersion of the Beringia land bridge, that connected Asia to North America (Routson et al., 2012; Williams, 1982). Indeed, we observed high syntenic relationships between the three species (Figure 3b). However, several macro-scale variations were found, including large inversions and translocations. Most of the variation from *M. fusca* was shared by the other wild species, however, some species-specific translocations and inversions were observed. For example, several inversions were observed on chromosome 03 that were shared between *M. fusca* and the two other wild apples, but the translocation between *M. fusca* chromosome 16 and *M. sieversii* chromosome 13 was specific to that comparison. Taken together, the relatively high whole-genome synteny and limited macro-variations support the Asiatic origins of *M. fusca* (Nikiforova et al., 2013; Robinson et al., 2001;

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.,

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Volk, Chao, et al., 2015; Volk, Henk, et al., 2015). More population genetic work, focused on whole genome evolution, should be performed in the future to better understand the relationship between these wild species. It will be interesting to explore how selection and cultivation of *M. fusca* by indigenous people in the Pacific Northwest of North America (Wyllie & de Echeverria, 2013), and its isolation from the domestication history of *M.* × *domestica*, have impacted traits that may be utilized in future improvement of apple cultivars and rootstocks.

Resolving the fire blight resistance locus (FB_Mfu10)

Apart from the unique climate and temperature cline that *M. fusca* has adapted to (Routson et al., 2012), the geographic localization of *M. fusca* in North America has potentially allowed for important co-evolution with *Erwinia amylovora*. This native North American bacterium causes the disease fire blight and is the most important constraint on pome production in the world (Norelli et al., 2003; Van der Zwet et al., 2012). Importantly, most *M. fusca* accessions are tolerant or resistant to fire blight (Dougherty et al., 2021), and indeed a noted resistance locus was identified on *M. fusca* chromosome 10 through screening of segregating populations derived from crosses of *M. fusca* with the susceptible *M.* x *domestica* cultivar Idared (Emeriewen et al., 2017). In that research, a genetic map was developed positioning a QTL on Chromosome 10 (*FB_Mfu10*) that explained 66% of the variation (Emeriewen et al., 2017, 2020). Further analysis of that region using Illumina- and Nanopore-sequenced BACs, identified a potential candidate gene, as well as other repetitive fragments of high sequence similarity to the candidate gene (Emeriewen et al., 2018, 2022). We sought to leverage our highly contiguous assembly of this fire blight resistant *M. fusca* accession to help in resolving this important locus and further analyze the genes therein, which likely contribute to this resistance trait.

Within the fine-mapped boundaries identified by Emeriewen et al. (2018) we identified several genes including a tandem duplication array consisting of four copies (*MfusH1_10g02031*, *MfusH1_10g02036*, *MfusH1_10g02040*, and *MfusH1_10g02042*) of the G-type lectin S-receptor-like serine/threonine-protein kinase genes implicated by Emeriewen et al. (2018) and (2022). Since similar resistance receptors (i.e., R-genes) are often part of such tandem duplications; we hypothesized that apart from sequence polymorphism, copy number variation (CNV) within this locus could potentially contribute to the resistance phenotype. We thus performed micro-synteny level comparisons between the resistant *M. fusca* and three susceptible *M.* x *domestica* cultivars (Honeycrisp, Golden Delicious, and Gala) (Daccord et al., 2017; Khan et al., 2022; Sun et al., 2020). We observed CNV in genes encoding these receptor-like genes that correlated with reported resistance phenotypes. While the resistant *M. fusca* carries four copies of this R-genes, 'Honeycrisp' contains three, 'Golden Delicious' has two full and one fragmented gene, and finally, 'Gala' contains two copies of which one is a truncated fragment (Figure 3d). 'Gala' is moderate to highly susceptible, 'Honeycrisp' is moderately susceptible to moderately resistant, and 'Golden Delicious' is moderately resistant (Dougherty et al., 2021; Kostick et al., 2019). However, it should be noted that the GDDH13 assembly is of a doubled haploid of 'Golden Delicious' and thus only represents the haplotype of this cultivar.

We expanded the comparative genomics analysis of *FB_Mfu10* locus to compare *M. fusca* to the other sequenced wild *Malus* relatives – *M. sieversii* and *M. sylvestris* (Sun et al., 2020). The accessions of *M. sieversii* (PI 613981) and *M. sylvestris* (PI 633825) are reportedly susceptible and resistant to fire blight, respectively, presenting an ideal opportunity to test our hypothesis (USDA Agricultural Research Service, 2015; B. Gutierrez personal communication). *M. sieversii* was found to have two copies of the G-type receptor gene within the locus while *M. sylvestris* had one large (>8000 bp) ortholog annotated, likely representing three copies misjoined as one gene (Figure S5). This result lends support to our hypothesis that CNV correlates with the resistance phenotype. However, *M. sylvestris* may only contain three copies which are similar to the susceptible Honeycrisp and Golden Delicious alleles. Thus, sequence variation that affects gene function or expression may also contribute to the resistance phenotype. This type of variation is evident in the sequence alignment and phylogenetic relationship between the genes (Figure 3c). Alternatively, *M. sylvestris* might have other loci contributing to resistance elsewhere in the genome. Similar examples of CNV of R-genes have been reported in maize (Chavan et al., 2015), soybean (Cook et al., 2012; Lee et al., 2015), and R-genes were found to be enriched for CNV in the genome of *M.* × *domestica* (Boocock et al., 2015). Additionally, Linkage Group 10 has been previously implicated in resistance to fire blight within a mapping population of *M.* × *domestica* generated from 'Florina' × 'Nova Easygro' (Le Roux et al., 2010), suggesting some contribution of genes on this chromosome already within the *M.* × *domestica* germplasm.

Previously, Fahrentrapp et al. (2013) and (2018) speculated that the single candidate genes that underlie fire blight resistance loci in *Malus spp.* are not products of co-evolution with the *Erwinia* due to their lack of positioning within clusters of paralogs (i.e., arrays). However, the results presented by Emeriewen et al. (2022) and our genome support a co-evolutionary origin of the fire blight resistance in *M. fusca*. Our assembly demonstrates that the Emeriewen et al. (2022) candidate R-gene was located within a tandem array of similar R-genes, which also span into other *Malus spp.* and domesticated cultivars. Taken

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

together, it can be hypothesized that the *FB_Mfu10* locus underwent selection for increased CNV of the R-gene in response to co-evolution with *Erwinia* in their overlapping habitats. Furthermore, it can be hypothesized that other North American species of *Malus* (*M. angustifolia*, *M. coronaria*, and *M. ioensis*) co-evolved even stronger resistance due to a longer evolutionary history with *Erwinia* in the pathogen's center of origin in the eastern North America (McGhee & Sundin, 2012; Stukenbrock & McDonald, 2008; Van Der Zwet, 2006; Zeng et al., 2018). In support of this hypothesis, *M. angustifolia* was found to exhibit lower susceptibility to natural infection by *Erwinia* than *M. fusca* (Dougherty et al., 2021). Exploring this locus in other native North American *Malus* species should help shed light on this possibility and help identify crucial alleles that confer higher and more durable resistance to infection.

Conclusion

The herein-described haplotype-resolved genome and annotation of *M. fusca* will add to the many recent developments in *Malus* genomics. Moreover, it provides an example of the opportunity that is afforded by increased access to long-read sequencing and improved genome assembly, scaffolding, and annotation methods in generating high-quality genomes for CWRs. Importantly for this work, our high-quality genome provides a valuable resource for understanding the genetic basis of important traits in this species and in the genus at large, such as disease resistance and stress tolerance, which are crucial for apple breeding programs moving forward. Furthermore, this study also highlights the significance of preserving wild apple relatives as a source of genetic diversity for future breeding efforts.

METHODS

Plant materials

For genomic DNA (gDNA) extractions, dormant scion cuttings of PI 589975 were obtained from the USDA Malus Collection at the United States Department of Agriculture (USDA) Plant Genetic Resources Unit (PGRU) located in Geneva, NY, USA. The cut ends of the dormant branches were placed into a beaker with a rehydration solution (Rose 100, Floralife, Waterboro, SC) using the manufacturer's recommended concentration and placed under long-day conditions at 21°C. Once bud-break was achieved and expanded leaves reached 3 cm in length, the branches were transferred to dark conditions for 48 h at 21°C. Leaves were then excised using a razor blade, weighed, and split into 1 g samples. Tissue samples were immediately flash-frozen in liquid N_2 and stored at −80°C. Additional leaves were collected in early sprin, when fresh growth was observed. This tissue was shipped overnight from the USDA Malus Collection, weighed and split into 0.25 g samples, and immediately flash frozen in liquid N_2 and stored at −80°C. Plant material of PI 589975 (seeds, propagation material, and tissue samples) can be obtained through the USDA GRIN-Global U.S. National Plant Germplasm System. Additional biological replicate trees are being established at the USDA Agricultural Research Service (ARS) Appalachian Fruit Research Station (AFRS) in Kearneysville, WV. Future material requests from these trees can be made to the corresponding authors.

DNA extraction

Genomic DNA (gDNA) and high-molecular-weight (HMW) DNA were extracted from frozen dark-treated leaf tissues. For each extraction, replicates of 0.25 g of frozen leaf tissue were ground into a fine powder using a mortar and pestle chilled using liquid N_2. For high-depth short-read whole-genome sequencing (WGS), gDNA was extracted using a Plant Pro DNA extraction kit (Qiagen, Germantown, MD) following the manufacturer's protocol. Following extraction, DNA samples were cleaned and concentrated to improve quality using a Zymo Genomic DNA Clean & Concentration kit (Irvine, CA). Purified gDNA was checked for quality and quantity on a Nanodrop Spectrometer (Thermo Fisher Scientific, Waltham, MA) and Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA). For the extraction of HMW DNA for long-read sequencing, Qiagen Genomic-tip 100/G kit was used following the protocol developed by Driguez et al. (2021). Extracted HMW DNA was checked for quality and quantity on a Nanodrop Spectrometer and Qubit 4 Fluorometer.

Library preparation and sequencing

First, we obtained high-depth (~100×) and high-quality short read sequence data. This sequencing of gDNA was carried out using an Illumina HiSeq platform (Illumina, San Diego, CA) with a read length of 150 bp in a paired-end format. Library preparations and sequencing were performed by GeneWiz (South Plainfield, NJ). For PacBio (Menlo Park, CA) sequencing, we employed the HiFi protocol on a Sequel II system using a single SMRT cell. PacBio HiFi library preparation and sequencing were performed at the University of Maryland Institute of Genome Sciences (Baltimore, MD). Lastly, Hi-C libraries were prepared using the Phase Genomic Proximo Plant Kits (Seattle, WA) and sequenced on a MiSeq and HiSeq Illumina platform by GeneWiz.

K-mer-based genome size estimation

Adapters on the raw Illumina short-reads were removed by the sequencing provider, and duplicate reads were removed using beam-dedupe software (Dai & Guan, 2020). To estimate genome size and heterozygosity, *k*-mers were counted using the Illumina WGS dataset with the software KMC 3 (v3.0) (Kokot et al., 2017) using the parameters '-k21 -t10 -m64 -ci1 -cs1000000'. We then exported the resulting *k*-mer count histogram into GenomeScope (v2.0) following default instructions (Ranallo-Benavidez et al., 2020).

Assembly and scaffolding

HiFi FASTQ files were analyzed for adapter contamination using the HiFiAdapterFilt (v2.0.0) (Sims et al., 2022) and quality using FASTQC (v0.11.9) (Andrews, 2010).

HiFi reads were then assembled using the HiFiasm (v0.16.1) assembler with the Hi-C paired-read option enabled to facilitate phasing of the contigs (Cheng et al., 2021).

Once assembly was complete, each individual haplotype assembly was scaffolded using Chromonomer (v.1.13) with linkage map markers from *M.* × *domestica* (Bianco et al., 2014; Catchen et al., 2020; Di Pierro et al., 2016). Prior to scaffolding, marker sequences were aligned to the haplotype FASTAs using bwa mem (v0.7.17-r1188) (Li & Durbin, 2009), and a pre-scaffold agp file was produced using the fasta2agp.py script from

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.,

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Chromonomer. The output agp files from Chromonomer were then converted to a Juicebox assembly file using Phase Genomics agp2assembly script (https://github.com/phasegenomics/juicebox_scripts). Assembly statistics were assessed using Merqury (v1.3), Busco (v.5.3) using the embryophyta odb10 database, and gt seqstat (v1.6.2) (Gremme et al., 2013; Rhie et al., 2020; Simão et al., 2015).

Draft scaffolds were then validated by Hi-C contact matrixes. For this validation, Hi-C reads were mapped and filtered following the Arima Pipeline (https://github.com/ArimaGenomics/mapping_pipeline) (Ghurye et al., 2017). Mapped reads were then processed using Phase Genomics Juicebox utility Matlock (https://github.com/phasegenomics/matlock) and sorted following Phase Genomics suggested protocols. The sorted links file and Chromonomer assembly files were then passed through the 3D-DNA run-assembly-visualizer script to generate a Juicebox editable assembly file (https://github.com/aidenlab/3d-dna/blob/master/visualize/run-assembly-visualizer.sh). Any mis-joins or inversions in the scaffolds were then corrected in Juicebox (v1.11.08) by manual curation (Durand et al., 2016). Finalized scaffold assembly files were then used to generate a new representative FASTA using the contig sequences using the juicebox_assembly_converter script (https://github.com/phasegenomics/juicebox_scripts). As an additional, independent validation of our scaffolding process, we compared scaffolds generated using Arima pipeline processed Hi-C reads and Salsa2 (v2.3) scaffolding pipeline (Ghurye et al., 2019). Scaffolding success was evaluated visually by Juicebox (v2.20.00) and by generating assembly statistics with gt seqstat.

The haplotype FASTA was then aligned to GDDH13 v1.1 assembly using nucmer from the MUMmer package (v4.0.0beta2) (Marçais et al., 2018), to assign scaffolds with known chromosome numbers. Chromosome05 was reverse-complemented to allow for easy comparisons between apple genomes. Plastid decontamination of the phased assemblies was done using the blobtools (v1.1.1) following the published protocols (Laetsch & Blaxter, 2017). Additionally, during upload to the NCBI database, the system flagged 10 scaffolds in haplotype 1 as having additional mitochondrial contamination from foreign biological origins and thus were removed from the final assembly.

Genome annotation

The annotation of transposable elements (TE) for each genome was first conducted using EDTA (v1.9.6) (Ou et al., 2019) under default parameters except the '--species others' option. In order to better compare repeat landscape contiguity across different *Malus* species (for the purpose of quality assessment) we annotated TEs, filtered, and consolidated the results using panEDTA (Ou et al., 2022). To further annotate non-LTR retrotransposons in *M. fusca*, we identified exemplar LINE elements using RepeatModeler2 (v2.0.1) (Flynn et al., 2020), and SINE elements using AnnoSINE (Li, Jiang, & Sun, 2022). The resulting exemplar LINE and SINE elements were combined to become the nonLTR library, which was supplied to EDTA via the --curatedlib parameter to reannotate the *M. fusca* genome. The quality and contiguity of repeat assembly were determined using the LTR Assembly Index (LAI) metric (Ou et al., 2018) from the LTR_retriever software (v2.9.0) (Ou & Jiang, 2018).

For annotation of the gene space, a comprehensive approach using MAKER (v2.31.10) was applied (Holt & Yandell, 2011). Transcript evidence was assembled from publicly available RNA-seq libraries generated for a gene expression atlas (Rogers & Van Nocker, 2013) of *Malus fusca* (NCBI Bioproject PRJNA267116). This transcriptional data was generated from a USDA accession of *Malus fusca* PI 589941. As this dataset contains more than 200 Gb of data, to provide EST evidence for annotation, 100 M pairs of reads were semi-randomly pulled from the BioProject's corresponding SRAs using VARUS (v1.0.0) (Stanke et al., 2019). This approach ensures sufficiently high coverage across the genome while limiting data transfer. Reads were then mapped to each chromosome-only haplotype using STAR (v2.7.8a) aligner with the options '--outSAMstrandField intronMotif' and '--alignIntronMaxenabled 10 kb' (Dobin et al., 2013). Transcripts for each haplotype were then assembled from the read alignments using StringTie2 (v1.3.5) (Kovaka et al., 2019).

FASTA sequences of the StringTie2 assembled transcripts were used as input into the first round of evidence-based annotation using MAKER2. Next, first-round annotations were extracted using the extract_anno_evi.sh script and used to train *ab initio* gene predictors SNAP (v2006-07-28) and Augustus (v3.3.2) (Hoff & Stanke, 2019; Korf, 2004). A second round of *ab initio* prediction was conducted using optimized transcripts from the first round of prediction and used as input into a final run of MAKER2. Annotations of rRNA and tRNA features used RNAmmer (v1.2) and tRNAscan-SE (v2.0.7) with default options (Chan & Lowe, 2019; Lagesen et al., 2007). Additionally, any genes with 80% of mRNA sequence overlapping with an annotated TE that did not have any pfam functional annotations, were labeled as TE-related in the gene annotation.

Synteny analysis

Comparison of gene synteny between the haplotype 1 assembly and two wild species (*M. sieversii* and *M. sylvestris*, Sun et al., 2020) as well as three *M.* × *domestica* genomes (GDDH13 v1.1, Gala, and Honeycrisp; Daccord et al., 2017; Khan et al., 2022; Sun et al., 2020) was performed with the Python MCScanX pipeline (v1.1.12) (Tang et al., 2008; Wang et al., 2012). Briefly, annotation gff files were downloaded from the Genome Database for Rosaceae (GDR) (Jung et al., 2019) and converted to bed format using jcvi.formats.gff. A pairwise synteny search was performed. To mitigate the impact of the recent whole genome duplication in *Malus*, only the reciprocal best hits ('--cscore = 0.99') were used for establishing the high-quality synteny blocks utilized in syntenic depth comparisons and plotting of karyotypes and macrosynteny and microsynteny plots, as well as syntenic block depths.

Haplotypic structural variation

SVs between the two haplotypes were detected by aligning the two FASTA files using nucmer with the settings '--maxmatch -l 100 -c 500' (Marçais et al., 2018). The resulting gziped delta file was uploaded to the Assemblytics web interface (http://www.assemblytics.com/; Nattestad & Schatz, 2016). Due to the difficulty of distinguishing extremely large SV from translocations or potential assembly errors, Assemblytics was originally designed to identify variants of approximately 10 kilobases in size, but in contiguous assemblies, the range may be extended confidently. We thus used a maximum variant size of 50 kbp. Larger, macro-level SV between *Malus* genotypes was also evaluated using synteny approaches. The results were exported as a bed file and imported into R (v4.2.2) for plotting. Deletions overlapping genes and exons were identified using Bedtools intersect (Quinlan & Hall, 2010). Haplotypic deletions were validated by mapping the short Illumina reads. First, reads were aligned to the haplotype 1 assembly using bwa mem with default settings (Li & Durbin, 2009). The average coverage for deletions was then compared to random non-deletion regions. Mean coverage across each deletion was calculated by Bedtools (v2.30.0) 'coverage -mean' for each predicted deletion. Bedtools shuffle was used to collect 100× random

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

non-deletion regions of the same size of each deletion. The coverage distributions of deletion and non-deletion regions were compared using a 1000× bootstrapped Kolmogorov–Smirnov test from the 'matching' R package (Sekhon, 2011). The sorted bam file was then loaded into samplot (v1.3.0) (Belyeu et al., 2021) to visualize selected deletions, and plot the read coverage and identify discordant mapping and long insert sizes.

Synteny at the FB_Mfu10 resistance locus

To identify the location of the previously reported fire blight resistance locus, the primers for the markers flanking the fine-mapped region of fire blight resistance on chromosome 10 were extracted from Emeriewen et al. (2018) and their sequence was aligned to haplotype 1 using BLASTN (v2.12.0+). Specifically, markers FR39G5T7xT7y and FR46H22 which were shown to delimit the locus to 0.33 cM were used to locate the narrowest region. As multiple BLAST hits were reported, the location was also confirmed by BLASTN of a candidate gene reported by Emeriewen et al. (2022), however the broader region between the markers was considered for further analysis. The region was then interrogated for microsynteny between the *M. fusca* haplotype 1 and the respective regions on chromosome 10 from the susceptible *M.* × *domestica* genomes, including GDDH13 v1.1 (Daccord et al., 2017), Gala (Sun et al., 2020), Honeycrisp (Khan et al., 2022), *M. sieversii* and *M. sylvestris* (Sun et al., 2020). Because MCScanX (Wang et al., 2012) excludes most tandem gene arrays in synteny analysis due to the difficulty of assigning true matches in these arrays, matches between candidate genes in the region were analyzed in a phylogenetic manner. Peptide sequences were extracted from each ortholog candidate gene and aligned using Custal Omega (v1.2.4) within the ETE3 toolkit (v.3.1.2) (Huerta-Cepas et al., 2016; Sievers & Higgins, 2018). ETE3 then executed the standard FastTree (v2.1) workflow for generation of a phylogenetic tree (Price et al., 2010). Candidate gene orthologs *Msy10g019590*, *Mdg10g019850*, and *MD10G1207300* from *M. sylvestris*, Gala, and GDDH13, respectively, were removed from the analysis due to potential mis-joined annotations or fragmented structure/pseudogene identity to facilitate more clear phylogenetic clade membership. Syntenic orthologs were then classified by the clade membership.

ACKNOWLEDGEMENTS

The authors acknowledge Dr. Becky Bart at the Donald Danforth Plant Science Center for their donation of computational time and Dr. Chris Dardick for critical input on the manuscript. Additionally, the authors thank Dr. Sean Rogers for his efforts in developing the transcriptomic resources used for annotation. This project was made possible by USDA ARS in-house appropriated funding through project number 8080-21000-029-00-D. Partial support for this work was also provided by the National Science Foundation (Plant Genome Research Program grant 1907077 and 2204717) to BNM.

AUTHOR CONTRIBUTIONS

BNM and CG conceptualized the study, collected data, analyzed results, and wrote the manuscript. AY, SO, AH, EB, BG, and SVN contributed data and/or analysis, writing, and review of the manuscript.

CONFLICT OF INTEREST

The authors declare they have no conflicts of interest in association with this work.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Merqury *k*-mer multiplicity analysis of the phased chromosome-only assembly.

Figure S2. Hi-C contact maps with the scaffolded genomes. (a) Haplotype 1. (b) Haplotype 2.

Figure S3. Annotation Edit Distance (AED) of the gene annotation for Haplome 1 and 2.

Figure S4. Assemblytics structural variation between *M. fusca* and *M.* x *domestica* GDDH13.

Figure S5. Synteny within the *Mfu10* locus associated with fire blight resistance in *M. fusca* (top) compared to *M. sylvestris* (bottom). Orthologs of G-type lectin S-receptor-like serine/threonine-protein kinases genes (R-genes) are annotated by locus id and syntenic genes are annotated by color between the four genomes.

Table S1. Scaffolding statistics between the linkage map + Hi-C validated and Salsa2 Hi-C-only pipelines.

Table S2. LTR and TE annotations and LTR assembly index (LAI) scores of *Malus* and *Pyrus* genomes.

Table S3. Structural variant deletions between *M. fusca* haplotype 1 and 2.

Table S4. Structural variants deletions between *M. fusca* haplotype 1 and 2 that overlap with exons.

File S1. *M. fusca* Haplotype 1 vs *M. fusca* Haplotype 2 structural variation BED file.

File S2. *M. fusca* Haplotype 1 vs GDDH13v1.1 structural variation BED file.

OPEN RESEARCH BADGES

This article has earned Open Data and Open Materials badges. Data and materials are available at [https://zenodo.org/record/7304709, https://www.rosaceae.org/Analysis/15540543, https://www.ncbi.nlm.nih.gov/bioproject/903862 and https://www.ncbi.nlm.nih.gov/biosample/SAMN31658743/, https://github.com/bmansfeld/mfusca_figs].

DATA AVAILABILITY STATEMENT

The raw sequence data generated from this project can be retrieved from the NCBI SRA database under BioSample SAMN31658743 and phased assemblies under BioProjects PRJNA899490 and PRJNA899491. Genome assemblies and annotation files can be retrieved from the GDR accession number tfGDR1066 and at a Zenodo repository DOI 10.5281/zenodo.7304709. Code for producing figures for the manuscript is available at github.com/bmansfeld/mfusca_figs/.

REFERENCES

Alonge, M., **Wang, X.**, **Benoit, M.**, **Soyk, S.**, **Pereira, L.**, **Zhang, L.** ***et al.*** (2020) Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell*, **182**, 145–161.e23.

Andrews, S. (2010) FastQC: A quality control tool for high throughput sequence data. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Atkinson, C.J., **Policarpo, M.**, **Webster, A.D.** & **Kingswell, G.** (2000) Drought tolerance of clonal *Malus* determined from measurements of stomatal conductance and leaf water potential. *Tree Physiology*, **20**, 557–563.

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.,

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Belyeu, J.R., **Chowdhury, M.**, **Brown, J.**, **Pedersen, B.S.**, **Cormier, M.J.**, **Quinlan, A.R.** ***et al.*** (2021) Samplot: a platform for structural variant visual validation and automated filtering. *Genome Biology*, **22**, 161.

Bhusal, N., **Han, S.-G.** & **Yoon, T.-M.** (2019) Impact of drought stress on photosynthetic response, leaf water potential, and stem sap flow in two cultivars of bi-leader apple trees (Malus × domestica Borkh.). *Scientia Horticulturae*, **246**, 535–543.

Bianco, L., **Cestaro, A.**, **Sargent, D.J.**, **Banchi, E.**, **Derdak, S.**, **di Guardo, M.** ***et al.*** (2014) Development and validation of a 20 K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh). *PLoS One*, **9**, e110377.

Bonn, W.G. & **Van Der Zwet, T.** (2000) Distribution and economic importance of fire blight. In: Vanneste, J.L. (Ed.) *Fire blight: the disease and its causative agent, Erwinia amylovora*. Wallingford, UK: CABI Publishing, pp. 37–53.

Boocock, J., **Chagné, D.**, **Merriman, T.R.** & **Black, M.A.** (2015) The distribution and impact of common copy-number variation in the genome of the domesticated apple, Malus × domestica Borkh. *BMC Genomics*, **16**, 848.

Catchen, J., **Amores, A.** & **Bassham, S.** (2020) Chromonomer: a tool set for repairing and enhancing assembled genomes through integration of genetic maps and conserved synteny. *G3*, **10**, 4115–4128.

Chan, P.P. & **Lowe, T.M.** (2019) tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods in Molecular Biology*, **1962**, 1–14.

Chavan, S., **Gray, J.** & **Smith, S.M.** (2015) Diversity and evolution of Rp1 rust resistance genes in four maize lines. *Theoretical and Applied Genetics*, **128**, 985–998.

Chen, X., **Li, S.**, **Zhang, D.**, **Han, M.**, **Jin, X.**, **Zhao, C.** ***et al.*** (2019) Sequencing of a wild apple (*Malus baccata*) genome unravels the differences between cultivated and wild apple species regarding disease resistance and cold tolerance. *G3*, **9**, 2051–2060.

Cheng, H., **Concepcion, G.T.**, **Feng, X.**, **Zhang, H.** & **Li, H.** (2021) Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nature Methods*, **18**, 170–175.

Cheng, H., **Jarvis, E.D.**, **Fedrigo, O.**, **Koepfli, K.-P.**, **Urban, L.**, **Gemmell, N.J.** ***et al.*** (2022) Haplotype-resolved assembly of diploid genomes without parental data. *Nature Biotechnology*, **40**, 1332–1335.

Cook, D.E., **Lee, T.G.**, **Guo, X.**, **Melito, S.**, **Wang, K.**, **Bayless, A.M.** ***et al.*** (2012) Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science*, **338**, 1206–1209.

Daccord, N., **Celton, J.-M.**, **Linsmith, G.**, **Becker, C.**, **Choisne, N.**, **Schijlen, E.** ***et al.*** (2017) High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics*, **49**, 1099–1106.

Dai, H. & **Guan, Y.** (2020) The Nubeam reference-free approach to analyze metagenomic sequencing reads. *Genome Research*, **30**, 1364–1375.

Dalhaus, T., **Schlenker, W.**, **Blanke, M.M.**, **Bravin, E.** & **Finger, R.** (2020) The effects of extreme weather on apple quality. *Scientific Reports*, **10**, 7919.

Di Pierro, E.A., **Gianfranceschi, L.**, **Di Guardo, M.**, **Koehorst-van Putten, H.J.**, **Kruisselbrink, J.W.**, **Longhi, S.** ***et al.*** (2016) A high-density, multi-parental SNP genetic map on apple validates a new mapping approach for outcrossing species. *Horticulture Research*, **3**, 1–13.

Dobin, A., **Davis, C.A.**, **Schlesinger, F.**, **Drenkow, J.**, **Zaleski, C.**, **Jha, S.** ***et al.*** (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

Dougherty, L., **Wallis, A.**, **Cox, K.**, **Zhong, G.-Y.** & **Gutierrez, B.** (2021) Phenotypic evaluation of fire blight outbreak in the USDA Malus collection. *Agronomy*, **11**, 144.

Driguez, P., **Bougouffa, S.**, **Carty, K.**, **Putra, A.**, **Jabbari, K.**, **Reddy, M.** ***et al.*** (2021) LeafGo: leaf to genome, a quick workflow to produce high-quality de novo plant genomes using long-read sequencing technology. *Genome Biology*, **22**, 256.

Durand, N.C., **Robinson, J.T.**, **Shamim, M.S.**, **Machol, I.**, **Mesirov, J.P.**, **Lander, E.S.** ***et al.*** (2016) Juicebox provides a visualization system for hi-C contact maps with unlimited zoom. *Cell Systems*, **3**, 99–101.

Emeriewen, O., **Richter, K.**, **Kilian, A.**, **Zini, E.**, **Hanke, M.-V.**, **Malnoy, M.** ***et al.*** (2014) Identification of a major quantitative trait locus for resistance to fire blight in the wild apple species *Malus fusca*. *Molecular Breeding*, **34**, 407–419.

Emeriewen, O.F., **Piazza, S.**, **Cestaro, A.**, **Flachowsky, H.**, **Malnoy, M.** & **Peil, A.** (2022) Identification of additional fire blight resistance candidate genes following MinION Oxford Nanopore sequencing and assembly of BAC clone spanning the *Malus fusca* resistance locus. *Journal of Plant Pathology*, **104**, 1509–1516. Available from: https://doi.org/10.1007/s42161-022-01223-x

Emeriewen, O.F., **Richter, K.**, **Berner, T.**, **Keilwagen, J.**, **Schnable, P.S.**, **Malnoy, M.** ***et al.*** (2020) Construction of a dense genetic map of the *Malus fusca* fire blight resistant accession MAL0045 using tunable genotyping-by-sequencing SNPs and microsatellites. *Scientific Reports*, **10**, 16358.

Emeriewen, O.F., **Richter, K.**, **Hanke, M.-V.**, **Malnoy, M.** & **Peil, A.** (2017) Further insights into *Malus Fusca* fire blight resistance. *Journal of Plant Pathology*, **99**, 45–49.

Emeriewen, O.F., **Richter, K.**, **Piazza, S.**, **Micheletti, D.**, **Broggini, G.A.L.**, **Berner, T.** ***et al.*** (2018) Towards map-based cloning of *FB_Mfu10*: identification of a receptor-like kinase candidate gene underlying the *Malus fusca* fire blight resistance locus on linkage group 10. *Molecular Breeding*, **38**, 106.

Fahrentrapp, J., **Broggini, G.A.L.**, **Kellerhals, M.**, **Peil, A.**, **Richter, K.**, **Zini, E.** ***et al.*** (2013) A candidate gene for fire blight resistance in *Malus* × *robusta* 5 is coding for a CC-NBS-LRR. *Tree Genetics & Genomes*, **9**, 237–251.

Fiala, J.L. (1994) *Flowering Crabapples: The Genus Malus*. Portland, OR: TImber Press.

Flynn, J.M., **Hubley, R.**, **Goubert, C.**, **Rosen, J.**, **Clark, A.G.**, **Feschotte, C.** ***et al.*** (2020) RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, **117**, 9451–9457.

Gessler, C. & **Pertot, I.** (2012) *Vf* scab resistance of *Malus*. *Trees*, **26**, 95–108.

Ghurye, J., **Pop, M.**, **Koren, S.**, **Bickhart, D.** & **Chin, C.-S.** (2017) Scaffolding of long read assemblies using long range contact information. *BMC Genomics*, **18**, 527.

Ghurye, J., **Rhie, A.**, **Walenz, B.P.**, **Schmitt, A.**, **Selvaraj, S.**, **Pop, M.** ***et al.*** (2019) Integrating hi-C links with assembly graphs for chromosome-scale assembly. *PLoS Computational Biology*, **15**, e1007273.

Gottschalk, C. & **Van Nocker, S.** (2013) Diversity in seasonal bloom time and floral development among apple species and hybrids. *Journal of the American Society for Horticultural Science*, **138**, 367–374.

Gremme, G., **Steinbiss, S.** & **Kurtz, S.** (2013) GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **10**, 645–656.

Hoff, K.J. & **Stanke, M.** (2019) Predicting genes in single genomes with AUGUSTUS. *Current Protocols in Bioinformatics*, **65**, e57.

Holt, C. & **Yandell, M.** (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, **12**, 491.

Hough, L., **Shay, J.** & **Dayton, D.** (1953) Apple scab resistance from Malus floribunda Sieb. *Proceedings of the American Society for Horticulture Science*, **62**, 341–347.

Huerta-Cepas, J., **Serra, F.** & **Bork, P.** (2016) ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution*, **33**, 1635–1638.

Jung, S., **Lee, T.**, **Cheng, C.-H.**, **Buble, K.**, **Zheng, P.**, **Yu, J.** ***et al.*** (2019) 15 years of GDR: new data and functionality in the genome database for rosaceae. *Nucleic Acids Research*, **47**, D1137–D1145.

Khan, A., **Carey, S.B.**, **Serrano, A.**, **Zhang, H.**, **Hargarten, H.**, **Hale, H.** ***et al.*** (2022) A phased, chromosome-scale genome of 'Honeycrisp' apple (*Malus domestica*). *Gigabyte*, **2022**, 1–15.

Khan, A. & **Chao, T.** (2017) Wild apple species as a source of fire blight resistance for sustainable productivity of apple orchards. *Fruit Quarterly*, **25**, 13–20.

Kokot, M., **Dlugosz, M.** & **Deorowicz, S.** (2017) KMC 3: counting and manipulating k-mer statistics. *Bioinformatics*, **33**, 2759–2761.

Korf, I. (2004) Gene finding in novel genomes. *BMC Bioinformatics*, **5**, 59.

Kostick, S.A., **Norelli, J.L.** & **Evans, K.M.** (2019) Novel metrics to classify fire blight resistance of 94 apple cultivars. *Plant Pathology*, **68**, 985–996.

Kovaka, S., **Zimin, A.V.**, **Pertea, G.M.**, **Razaghi, R.**, **Salzberg, S.L.** & **Pertea, M.** (2019) Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology*, **20**, 278.

Laetsch, D.R. & **Blaxter, M.L.** (2017) BlobTools: interrogation of genome assemblies. Available at: https://f1000research.com/articles/6-1287

Lagesen, K., **Hallin, P.**, **Rødland, E.A.**, **Staerfeldt, H.-H.**, **Rognes, T.** & **Ussery, D.W.** (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, **35**, 3100–3108.

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Le Roux, P.-M.F., **Khan, M.A.**, **Broggini, G.A.L.**, **Duffy, B.**, **Gessler, C.** & **Patocchi, A.** (2010) Mapping of quantitative trait loci for fire blight resistance in the apple cultivars "Florina" and "Nova Easygro". *Genome*, **53**, 710–722.

Lee, T.G., **Kumar, I.**, **Diers, B.W.** & **Hudson, M.E.** (2015) Evolution and selection of *Rhg1*, a copy-number variant nematode-resistance locus. *Molecular Ecology*, **24**, 1774–1791.

Li, H. & **Durbin, R.** (2009) Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li, Y., **Jiang, N.** & **Sun, Y.** (2022) AnnoSINE: a short interspersed nuclear elements annotation tool for plant genomes. *Plant Physiology*, **188**, 955–970.

Li, Z., **Wang, L.**, **He, J.**, **Li, X.**, **Hou, N.**, **Guo, J.** ***et al.*** (2022) Chromosome-scale reference genome provides insights into the genetic origin and grafting-mediated stress tolerance of *Malus prunifolia*. *Plant Biotechnology Journal*, **20**, 1015–1017.

Linsmith, G., **Rombauts, S.**, **Montanari, S.**, **Deng, C.H.**, **Celton, J.M.**, **Guérif, P.** ***et al.*** (2019) Pseudo-chromosome-length genome assembly of a double haploid "Bartlett" pear (*Pyrus communis* L.). *Gigascience*, **8**, giz138.

MacHardy, W.E. (1996) *Apple scab: biology, epidemiology, and management*. St. Paul, MN: American Phytopathological Society (APS Press).

Mansfeld, B.N., **Boyher, A.**, **Berry, J.C.**, **Wilson, M.**, **Ou, S.**, **Polydore, S.** ***et al.*** (2021) Large structural variations in the haplotype-resolved African cassava genome. *The Plant Journal*, **108**, 1830–1848.

Marçais, G., **Delcher, A.L.**, **Phillippy, A.M.**, **Coston, R.**, **Salzberg, S.L.** & **Zimin, A.** (2018) MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, **14**, e1005944.

McGhee, G.C. & **Sundin, G.W.** (2012) *Erwinia amylovora* CRISPR elements provide new tools for evaluating strain diversity and for microbial source tracking. *PLoS One*, **7**, e41706.

Migicovsky, Z., **Gardner, K.M.**, **Richards, C.**, **Thomas Chao, C.**, **Schwaninger, H.R.**, **Fazio, G.** ***et al.*** (2021) Genomic consequences of apple improvement. *Horticulture Research*, **8**, 1–13.

Muranty, H., **Denancé, C.**, **Feugey, L.**, **Crépin, J.L.**, **Barbier, Y.**, **Tartarini, S.** ***et al.*** (2020) Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. *BMC Plant Biology*, **20**, 2.

Nattestad, M. & **Schatz, M.C.** (2016) Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics*, **32**, 3021–3023.

Nikiforova, S.V., **Cavalieri, D.**, **Velasco, R.** & **Goremykin, V.** (2013) Phylogenetic analysis of 47 chloroplast genomes clarifies the contribution of wild species to the domesticated apple maternal line. *Molecular Biology and Evolution*, **30**, 1751–1760.

Norelli, J.L., **Jones, A.L.** & **Aldwinckle, H.S.** (2003) Fire blight management in the twenty-first century: using new technologies that enhance host resistance in apple. *Plant Disease*, **87**, 756–765.

Ou, S., **Chen, J.** & **Jiang, N.** (2018) Assessing genome assembly quality using the LTR assembly index (LAI). *Nucleic Acids Research*, **46**, e126.

Ou, S., **Collins, T.**, **Qiu, Y.**, **Seetharam, A.S.**, **Menard, C.C.**, **Manchanda, N.** ***et al.*** (2022) Differences in activity and stability drive transposable element variation in tropical and temperate maize. *bioRxiv*. Available from: https://doi.org/10.1101/2022.10.09.511471

Ou, S. & **Jiang, N.** (2018) LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiology*, **176**, 1410–1422.

Ou, S., **Su, W.**, **Liao, Y.**, **Chougule, K.**, **Agda, J.R.A.**, **Hellinga, A.J.** ***et al.*** (2019) Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, **20**, 275.

Papp, D., **Gao, L.**, **Thapa, R.**, **Olmstead, D.** & **Khan, A.** (2020) Field apple scab susceptibility of a diverse Malus germplasm collection identifies potential sources of resistance for apple breeding. *CABI Agriculture and Bioscience*, **1**, 16.

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

Quinlan, A.R. & **Hall, I.M.** (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Ranallo-Benavidez, T.R., **Jaron, K.S.** & **Schatz, M.C.** (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, **11**, 1432.

Rhie, A., **Walenz, B.P.**, **Koren, S.** & **Phillippy, A.M.** (2020) Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, **21**, 245.

Robinson, J.P., **Harris, S.A.** & **Juniper, B.E.** (2001) Taxonomy of the genus *Malus* mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. *Plant Systematics and Evolution*, **226**, 35–58.

Rogers, S. & **Van Nocker, S.** (2013) *Gene expression atlas of development in a wild apple species*. San Diego, CA: Plant and Animal Genome Conference.

Routson, K.J., **Volk, G.M.**, **Richards, C.M.**, **Smith, S.E.**, **Nabhan, G.P.** & **De Echeverria, V.W.** (2012) Genetic variation and distribution of pacific crabapple. *Journal of the American Society for Horticultural Science*, **137**, 325–332.

Sakurai, K., **Brown, S.K.** & **Weeden, N.** (2000) Self-incompatibility alleles of apple cultivars and advanced selections. *HortScience*, **35**, 116–119.

Schrader, L.E., **Zhang, J.** & **Duplaga, W.K.** (2001) Two types of sunburn in apple caused by high fruit surface (peel) temperature. *Plant Health Progress*, **2**, 3.

Sekhon, J.S. (2011) Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *Journal of Statistical Software*, **42**, 1–52.

Sievers, F. & **Higgins, D.G.** (2018) Clustal omega for making accurate alignments of many protein sequences. *Protein Science*, **27**, 135–145.

Sim, S.B., **Corpuz, R.L.**, **Simmonds, T.J.** & **Geib, S.M.** (2022) HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics*, **23**, 157.

Simão, F.A., **Waterhouse, R.M.**, **Ioannidis, P.**, **Kriventseva, E.V.** & **Zdobnov, E.M.** (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**, 3210–3212.

Stanke, M., **Bruhn, W.**, **Becker, F.** & **Hoff, K.J.** (2019) VARUS: sampling complementary RNA reads from the sequence read archive. *BMC Bioinformatics*, **20**, 558.

Stukenbrock, E.H. & **McDonald, B.A.** (2008) The origins of plant pathogens in agro-ecosystems. *Annual Review of Phytopathology*, **46**, 75–100.

Sun, X., **Jiao, C.**, **Schwaninger, H.**, **Chao, C.T.**, **Ma, Y.**, **Duan, N.** ***et al.*** (2020) Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nature Genetics*, **52**, 1423–1432.

Tang, H., **Wang, X.**, **Bowers, J.E.**, **Ming, R.**, **Alam, M.** & **Paterson, A.H.** (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Research*, **18**, 1944–1954.

Torres, C.A., **Sepulveda, A.**, **Gonzalez-Talice, J.**, **Yuri, J.A.** & **Razmilic, I.** (2013) Fruit water relations and osmoregulation on apples (*Malus domestica* Borkh.) with different sun exposures and sun-injury levels on the tree. *Scientia Horticulturae*, **161**, 143–152.

Torres, C.A., **Sepúlveda, A.**, **Leon, L.** & **Yuri, J.A.** (2016) Early detection of sun injury on apples (*Malus domestica* Borkh.) through the use of crop water stress index and chlorophyll fluorescence. *Scientia Horticulturae*, **211**, 336–342.

USDA Agricultural Research Service. (2015) *Germplasm Resources Information Network (GRIN)*. Beltsville, MD: USDA Agricutlural Research Service. Available from: https://doi.org/10.15482/USDA.ADC/1212393

Van Der Zwet, T. (2006) Present worldwide distribution of fire blight and closely related diseases. *Acta Horticulturae*, **704**, 35–36.

Van der Zwet, T., **Orolaza-Halbrendt, N.** & **Zeller, W.** (2012) *Fire blight: history, biology, and management*. Amer: Phytopath Society.

Velasco, R., **Zharkikh, A.**, **Affourtit, J.**, **Dhingra, A.**, **Cestaro, A.**, **Kalyanaraman, A.** ***et al.*** (2010) The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nature Genetics*, **42**, 833–839.

Volk, G.M. (2019) Temperate tree fruits of North America: *Malus* Mill., *Prunus* L., *Diospyros* L., and *Asimina* Adans. In: Greene, S.L., Williams, K.A., Khoury, C.K., Kantar, M.B. & Marek, L.F. (Eds.) *North American crop wild relatives, Volume 2: Important species*. Cham: Springer International Publishing, pp. 353–386. Available from: https://doi.org/10.1007/978-3-319-97121-6_11

Volk, G.M., **Chao, C.T.**, **Norelli, J.**, **Brown, S.K.**, **Fazio, G.**, **Peace, C.** ***et al.*** (2015) The vulnerability of US apple (*Malus*) genetic resources. *Genetic Resources and Crop Evolution*, **62**, 765–794.

Volk, G.M., **Henk, A.D.**, **Baldo, A.**, **Fazio, G.**, **Chao, C.T.** & **Richards, C.M.** (2015) Chloroplast heterogeneity and historical admixture within the genus *Malus*. *American Journal of Botany*, **102**, 1198–1208.

Wang, Y., **Tang, H.**, **DeBarry, J.D.**, **Tan, X.**, **Li, J.**, **Wang, X.** ***et al.*** (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, **40**, e49.

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.,

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Way, R.D., **Aldwinckle, H.S.**, **Lamb, R.C.**, **Rejman, A.**, **Sansavini, S.**, **Shen, T.** ***et al.*** (1991) Apples (*Malus*). *Acta Horticulturae*, **209**, 3–46. Available from: https://doi.org/10.17660/ActaHortic.1991.290.1

Williams, A.H. (1982) Chemical evidence from the flavonoids relevant to the classification of *Malus* species. *Botanical Journal of the Linnean Society*, **84**, 31–39.

Williams, E.B. (1966) Allelic genes in Malus for resistance to Venturia inaequalis. *Proceedings of the American Society for Horticultural Science*, **88**, 52–56.

Wyllie, R. & **de Echeverria, V.** (2013) *Moolks (Pacific crabapple,* Malus fusca*) on the North Coast of British Columbia: Knowledge and Meaning in Gitga'at Culture. Masters Thesis. Victoria, CA: University of Victoria. Available from: https://dspace.library.uvic.ca/handle/1828/4596*

Zeng, Q., **Cui, Z.**, **Wang, J.**, **Childs, K.L.**, **Sundin, G.W.**, **Cooley, D.R.** ***et al.*** (2018) Comparative genomics of *Spiraeoideae*-infecting *Erwinia amylovora* strains provides novel insight to genetic diversity and identifies the genetic basis of a low-virulence strain. *Molecular Plant Pathology*, **19**, 1652–1666.

Zhang, L., **Hu, J.**, **Han, X.**, **Li, J.**, **Gao, Y.**, **Richards, C.M.** ***et al.*** (2019) A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nature Communications*, **10**, 1494.

Zhong, Z., **Feng, S.**, **Mansfeld, B.N.**, **Ke, Y.**, **Qi, W.**, **Lim, Y.-W.** ***et al.*** (2022) Haplotype-resolved DNA methylome of African cassava genome. *Plant Biotechnology Journal*, **21**, 247–249.

Zhou, Y., **Minio, A.**, **Massonnet, M.**, **Solares, E.**, **Lv, Y.**, **Beridze, T.** ***et al.*** (2019) The population genetics of structural variants in grapevine domestication. *Nature Plants*, **5**, 965–979.

© 2023 The Authors.
The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.,

1365313x, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/tpj.16433 by CochraneChina, Wiley Online Library on [14/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License