the plant journal

SEB Society for Experimental Biology

The Plant Journal (2016) **87,** 535–547 doi: 10.1111/tpj.13215

RESOURCE

The genome of black raspberry (*Rubus occidentalis*)

Robert VanBuren[1,†], Doug Bryant[1,†], Jill M. Bushakra[2], Kelly J. Vining[3,4], Patrick P. Edger[5], Erik R. Rowley[1], Henry D. Priest[1], Todd P. Michael[6], Eric Lyons[7], Sergei A. Filichkin[4,8], Michael Dossett[9], Chad E. Finn[10], Nahla V. Bassil[2] and Todd C. Mockler[1,*]

[1]*Donald Danforth Plant Science Center, St. Louis, MO 63132, USA,*
[2]*USDA-ARS National Clonal Germplasm Repository, Corvallis, OR 97333, USA,*
[3]*Department of Horticulture, Oregon State University, Corvallis, OR 97331, USA,*
[4]*Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR 97331, USA,*
[5]*Department of Horticulture, Michigan State University, East Lansing, MI 48823, USA,*
[6]*Ibis Biosciences, Carlsbad, CA, USA,*
[7]*CyVerse, BIO5, School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA,*
[8]*Department of Botany and Plant Pathology and Center for Genome Research and Biocomputing, Oregon State University Corvallis, Corvallis, OR 97331, USA,*
[9]*B.C. Blueberry Council (in Partnership with Agriculture and Agri-Food Canada) – Agassiz Research and Development Centre, Agassiz, BC, VOM 1A0 Canada, and*
[10]*USDA-ARS Horticultural Crops Research Unit, Corvallis, OR 97330, USA*

Received 5 January 2016; revised 27 April 2016; accepted 12 May 2016; published online 20 July 2016.
*For correspondence (e-mail tmockler@danforthcenter.org)
†These authors contributed equally to this work.

SUMMARY

Black raspberry (*Rubus occidentalis*) is an important specialty fruit crop in the US Pacific Northwest that can hybridize with the globally commercialized red raspberry (*R. idaeus*). Here we report a 243 Mb draft genome of black raspberry that will serve as a useful reference for the Rosaceae and *Rubus* fruit crops (raspberry, blackberry, and their hybrids). The black raspberry genome is largely collinear to the diploid woodland strawberry (*Fragaria vesca*) with a conserved karyotype and few notable structural rearrangements. Centromeric satellite repeats are widely dispersed across the black raspberry genome, in contrast to the tight association with the centromere observed in most plants. Among the 28 005 predicted protein-coding genes, we identified 290 very recent small-scale gene duplicates enriched for sugar metabolism, fruit development, and anthocyanin related genes which may be related to key agronomic traits during black raspberry domestication. This contrasts patterns of recent duplications in the wild woodland strawberry *F. vesca*, which show no patterns of enrichment, suggesting gene duplications contributed to domestication traits. Expression profiles from a fruit ripening series and roots exposed to *Verticillium dahliae* shed insight into fruit development and disease response, respectively. The resources presented here will expedite the development of improved black and red raspberry, blackberry and other *Rubus* cultivars.

Keywords: black raspberry, tandem gene duplicates, neofunctionalization, genome assembly, improvement.

INTRODUCTION

Black raspberry (*Rubus occidentalis* L.), is a minor but important specialty fruit crop in the United States Pacific Northwest prized for its unique flavor and potential health benefits (Stoner *et al.*, 2005; Seeram, 2008). Black raspberry is diploid ($2n = 2x = 14$) and belongs to the same subgenus (*Idaeobatus*) as red raspberry (*R. idaeus* L.), with which it can be crossed (Hellman *et al.*, 1982). The global commercial raspberry production exceeds 500 000 metric tons. Black raspberry is grown on 668 hectares in Oregon, the largest US production region, with a utilized

production value of US$6.8 million (Anonymous, 2014). Black raspberry was domesticated in the 1830s (Hedrick and Howe, 1925), but formal breeding efforts did not begin until the late 1800s (Jennings, 1988). Since the early 1900s, black raspberry production in the US has seen a marked decline that many attribute to disease pressures and a lack of cultivars with sufficient resistance. Black raspberry cultivars suffer from limited genetic diversity stemming from the limited gene pool used in the elite germplasm and lack of breeding progress (Dossett *et al.*, 2008, 2012). This narrow gene pool has hindered attempts to increase yield, quality, and disease resistance. Only three new cultivars have been released over the last 20 years (Bushakra *et al.*, 2015).

Black raspberry is a member of the Rose Family (Rosaceae) that includes many economically important crops such as apple (*Malus* × *domestica* Borkh.), pear (*Pyrus* spp.), stone fruits (*Prunus* spp.), strawberry (*Fragaria* × *ananassa* Duch. ex Rozier), and roses (*Rosa* spp.). With an estimated 600–800 species worldwide (Thompson, 1995), the genus *Rubus* is one of the largest and most diverse in the Rosaceae (Alice and Campbell, 1999). The wealth of genomic resources available within the Rosaceae is rivaled only by the grasses (Poaceae) and the crucifers (Brassicaceae) (Michael and VanBuren, 2015). Draft genomes are available for apple (Velasco *et al.*, 2010), peach (Verde *et al.*, 2013), pear (Wu *et al.*, 2013; Chagné *et al.*, 2014), strawberry (Shulaev *et al.*, 2011), and Chinese plum (Zhang *et al.*, 2012), and other Rosaceae crop genomes are forthcoming. These resources allow for detailed comparisons across the Rosaceae to study evolutionary history and agronomic traits. Here we report a high-quality draft genome of black raspberry selection ORUS 4115-3 using an Illumina-only approach with genome anchoring from a high-density genetic linkage map. These resources will be useful for the raspberry, blackberry, and other Rosaceae fruit crop breeding communities.

RESULTS

Genome sequencing, assembly and annotation

The black raspberry selection ORUS 4115-3 was chosen for sequencing because of its low residual within-genome heterozygosity (Figure S1) and apparent tolerance to *Verticillium* wilt (*Verticillium dahliae* Kleb.), a soil-borne fungal disease and a leading cause of stand decline in commercial fields in Oregon (Dossett, 2011). We sequenced the black raspberry genome using eight Illumina paired-end libraries with inserts ranging in size from 165 base pairs (bp) to 4700 bp collectively representing 325 × coverage of the estimated 293 megabase (Mb) genome (Meng and Finn, 2002) (Tables S1 and S2). The final ALLPATHS (Butler *et al.*, 2008) based assembly includes 9245 contiguous sequences (contigs) in 2226 scaffolds spanning 243 Mb or 83% of the estimated genome (Table 1). The scaffold N50 length is 353 kilobase pairs (kb) with half of the assembly contained in the largest 178 scaffolds. The quality of this genome is similar to other Illumina-based Rosaceae draft genome assemblies including pear (Wu *et al.*, 2013) (scaffold N50 = 540 kb) and Chinese plum (Zhang *et al.*, 2012) (N50 = 577 kb). In addition to the nuclear genome, we assembled the chloroplast genome into a single 151 783 bp contig, and the mitochondrial genome into four scaffolds collectively spanning 449 kb (Figure S2). The black raspberry chloroplast genome has 88 protein-coding genes, four ribosomal RNA genes, and 40 tRNAs. Compared to the strawberry chloroplast genome, the black raspberry chloroplast is 3802 bp smaller, but has one more protein-coding gene, *infA*, and five more tRNAs. The *infA* gene encodes translation initiation factor 1, and has been lost from most studied rosid chloroplast genomes (Millen *et al.*, 2001), although there is evidence of its retention in the Rosaceae species *Prinsepia utilis* Royle (Wang *et al.*, 2013). Like strawberry, the atpF group II intron is absent from black raspberry.

The black raspberry scaffolds were assembled into seven pseudomolecules using a high-density genetic map derived from an F_1 population of 115 plants (Bushakra *et al.*, 2015). The genetic map has a total of 1711 markers including 1657 single nucleotide polymorphisms (SNPs) markers derived from genotyping-by-sequencing (GBS) and 54 simple sequence repeat (SSR) markers with an average distance between markers of 1.3 centiMorgans (cM) (Table S3). Markers from the genetic map were used to order and orient scaffolds. Ambiguities or inconsistencies between the two parental maps were resolved using

Table 1 Summary of assembly statistics

Metric	Value
Number of scaffolds	2226
Total size of scaffolds (bp)	239 850 264
Longest sequence (bp)	2 647 598
Shortest sequence (bp)	912
Number of sequences >10 K	756 (33.96%)
Number of sequences >100 K	591 (26.55%)
Number of sequences >1 M	26 (1.17%)
Mean sequence length (bp)	107 749
Median sequence length (bp)	26 272
Mode of sequence lengths (bp)	1261
N50 sequence length (bp)	353 335
L50 sequence number	177
N90 sequence length (bp)	72 181
L90 sequence number	753
Sequences As	30.20%
Sequences Cs	17.76%
Sequences Gs	17.76%
Sequences Ts	30.22%
Sequences Ns	4.02%
Sequences GCs	35.52%

synteny with the strawberry (Shulaev *et al.*, 2011; Tennessen *et al.*, 2014) and peach (Verde *et al.*, 2013) genomes, though ambiguities were found for less than 10% of the scaffolds. In total, 626 scaffolds were anchored to the seven black raspberry pseudochromosomes collectively spanning 203 Mb or 84.5% of the assembly (Table S4).

MAKER (Cantarel *et al.*, 2008) was used to annotate the black raspberry genome. A set of 33 783 reference-guided and 71 622 *de novo*-assembled transcripts were generated using RNA sequencing (RNA-seq) reads generated from tissue of young leaf, *Verticillium*-inoculated and -uninoculated root, green fruit, red fruit, ripe fruit, and cane tissue. These transcripts were clustered into 29 460 representative sequences that were used for input into MAKER producing a preliminary set of 32 300 putative gene models. Gene models with no functional annotation based on InterPro Scan and BLASTP were removed leaving a final set of 28 005 protein-coding genes. These results are similar to strawberry (25 050 genes; Shulaev *et al.*, 2011), peach (27 852 genes; Verde *et al.*, 2013), and Chinese plum (31 390 genes; Zhang *et al.*, 2012), but less than pear (42 767 genes; Wu *et al.*, 2013) and apple (57 386 genes; Velasco *et al.*, 2010) as expected given the additional whole-genome duplication in the latter two species (Velasco *et al.*, 2010). Over 99.6% of the core eukaryotic genes (Parra *et al.*, 2007) are at least partially identified in the genome, validating the assembly quality. Gene density in the black raspberry genome is highest toward the ends of chromosomes and lowest in the pericentromeric regions and is inversely correlated with retrotransposon and DNA transposon density (Figure 1).

Repetitive elements and centromere evolution

Repetitive elements were identified using structure- and homology-based approaches and collectively spanned 136 Mb or 56% of the black raspberry genome (Table S5). This proportion of repeats is higher than any other sequenced Rosaceae species including strawberry (22.0%; Shulaev *et al.*, 2011), peach (29.6%; Verde *et al.*, 2013), apple (42.4%; Velasco *et al.*, 2010), and pear (53.1%, Wu *et al.*, 2013). These differences may in part be due to sensitivity and thoroughness of repeat identification and level of completeness of the respective genome assemblies. Long terminal repeat (LTR) retrotransposons are the most abundant type of repeats in black raspberry and the *Copia* and *Gypsy* LTR families account for 10.6 and 11.6% of the genome, respectively. LTRs are distributed non-randomly across the genome with peaks near the centromeric regions and regions of low gene density (Figure 1). Long interspersed elements (LINES) are surprisingly abundant in the black raspberry genome with 27 506 elements accounting for 4.1% of the genome. This is 10-fold higher than observed in other sequenced Rosaceae genomes and higher than that found in most plant genomes. DNA transposons comprise 8.1% of the genome, and the most abundant terminal inverted repeat (TIR) family is *Helitron* with 10 118 elements accounting for 4.1% of the genome (Table S5).

Plant centromeres are characterized by tandem arrays of satellite DNA that can exceed several Mb with the array lengths, distributions, and sequences varying greatly. Black raspberry centromeric repeats have a base monomer length of 303 bp and higher order structures of dimers (606 bp), trimers (909 bp), tetramers (1212 bp) and pentamers (1515 bp) are present (Figures S3 and S4). There is also evidence of a second distinct 95-bp centromere repeat, though this repeat is less prevalent and has lower sequence conservation (Figures S3 and S5). The within-centromere array identity for the 95-bp and 303-bp centromere repeats are 77% (minimum 52% and maximum 97%) and 73% (minimum 44% and maximum 99%) respectively. The 303-bp centromere repeat in black raspberry is longer than that described in most plants; typically centromere repeats are between 150–200 bp, although this is highly variable (Melters *et al.*, 2013). The 303- and 95-bp repeats are distinct from the 167- and 140-bp centromere repeats found in the strawberry (Shulaev *et al.*, 2011) and pear genomes (Wu *et al.*, 2013). Centromeric repeats are clustered in large arrays on chromosomes 1 and 2 that likely correspond to the centromere but surprisingly are also dispersed across entire chromosomes (Figure S6). Black raspberry has 36 225 centromeric repeats distributed across 558 genome scaffolds that collectively span over 30% of the genome. These arrays extend into relatively gene-rich regions including the entirety of scaffold 1 which is over 2.5 Mb with a normal gene density (Figure S7). A similar pattern of dispersed centromeric repeats is found in plants with holocentric chromosomes (Melters *et al.*, 2012).

Comparative genomics within the Rosaceae

The Rosaceae is an economically important family with 2800 species among 95 genera including the specialty fruit crops almond, apple, blackberry, cherry, peach, pear, plum, raspberry, and strawberry (Potter *et al.*, 2007). We compared black raspberry to the high-quality diploid strawberry (version Fvb) (Tennessen *et al.*, 2014) and peach genomes (Verde *et al.*, 2013) that are currently the best assembled genomes in the Rosaceae. Each paralogous region in black raspberry is orthologous to a single region each in peach and strawberry suggesting that there has been no lineage-specific whole-genome duplication event after the core eudicot-wide gamma hexaploidization (Bowers *et al.*, 2003) (Figure 2a–d). Black raspberry and strawberry share 15 533 pairs of collinear genes in 573 syntenic blocks while black raspberry and peach share 16 307 collinear genes in 611 syntenic blocks. The remaining genes have undergone structural rearrangements or are lineage-

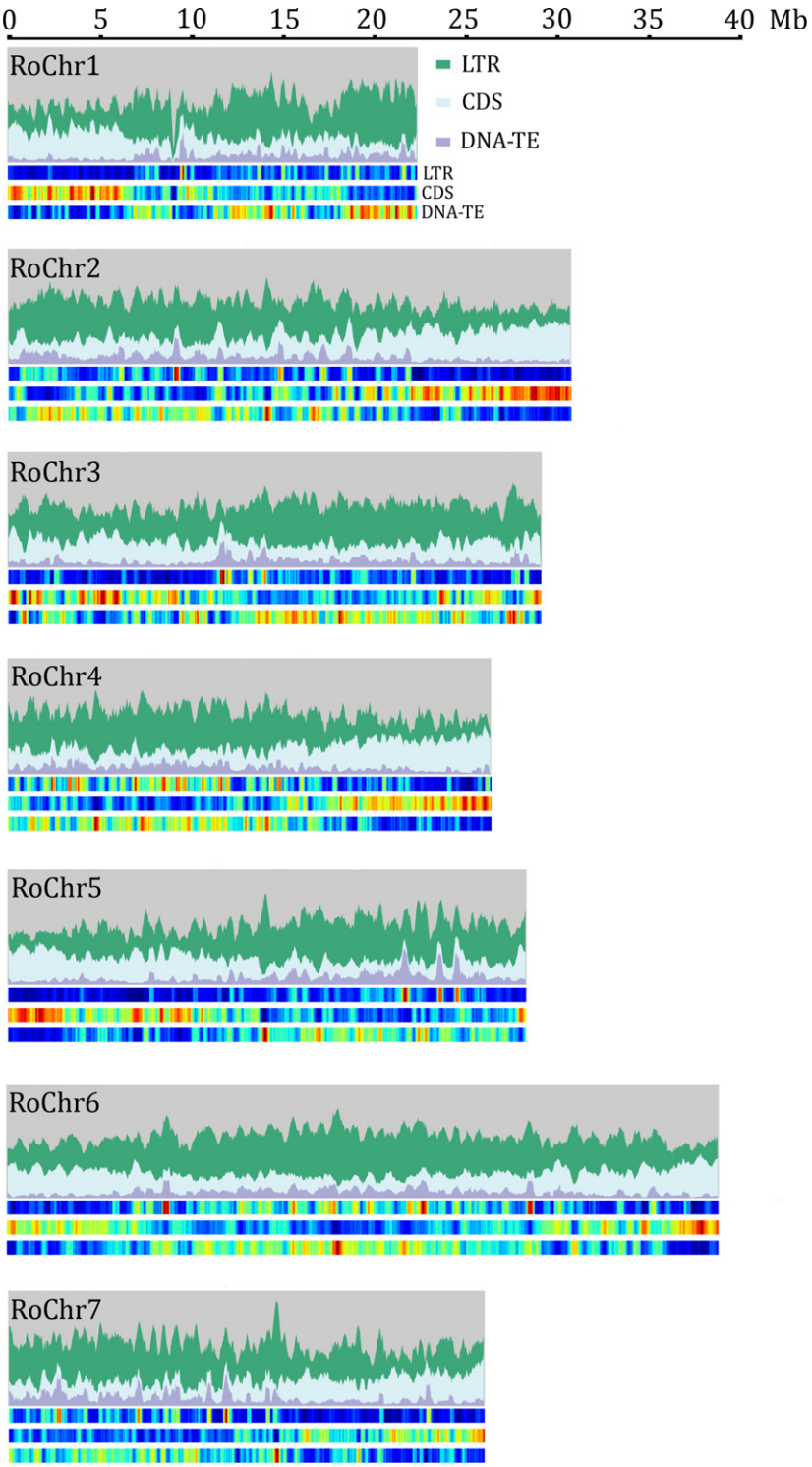

Figure 1. Chromosome landscape of the black raspberry genome.
Major DNA components are classified into coding regions (CDS; cyan), DNA transposons (purple) and retrotransposons (green). The grey (top) portion represents unclassified DNA content. Heatmaps show the distribution of elements with blue indicating lowest abundance and red indicating high abundance. The plot was generated based on a moving window of 0.5 Mb with 0.1 Mb shift along each of the pseudochromosomes.

specific. Black raspberry and strawberry share the same base chromosome number ($n = 7$), and the seven chromosomes are largely collinear with the exception of several large-scale structural rearrangements (Figure 2a,b). The high degree of collinearity with the strawberry genome validates the assembly quality of black raspberry on the scaffold and chromosome scale. The relationship between black raspberry and peach ($2n = 16$) is more complex with each chromosome in black raspberry corresponding to regions in several peach chromosomes. For instance black raspberry LG1 corresponds to regions from peach chromosomes 1, 7, and 8, and LG3 corresponds to peach

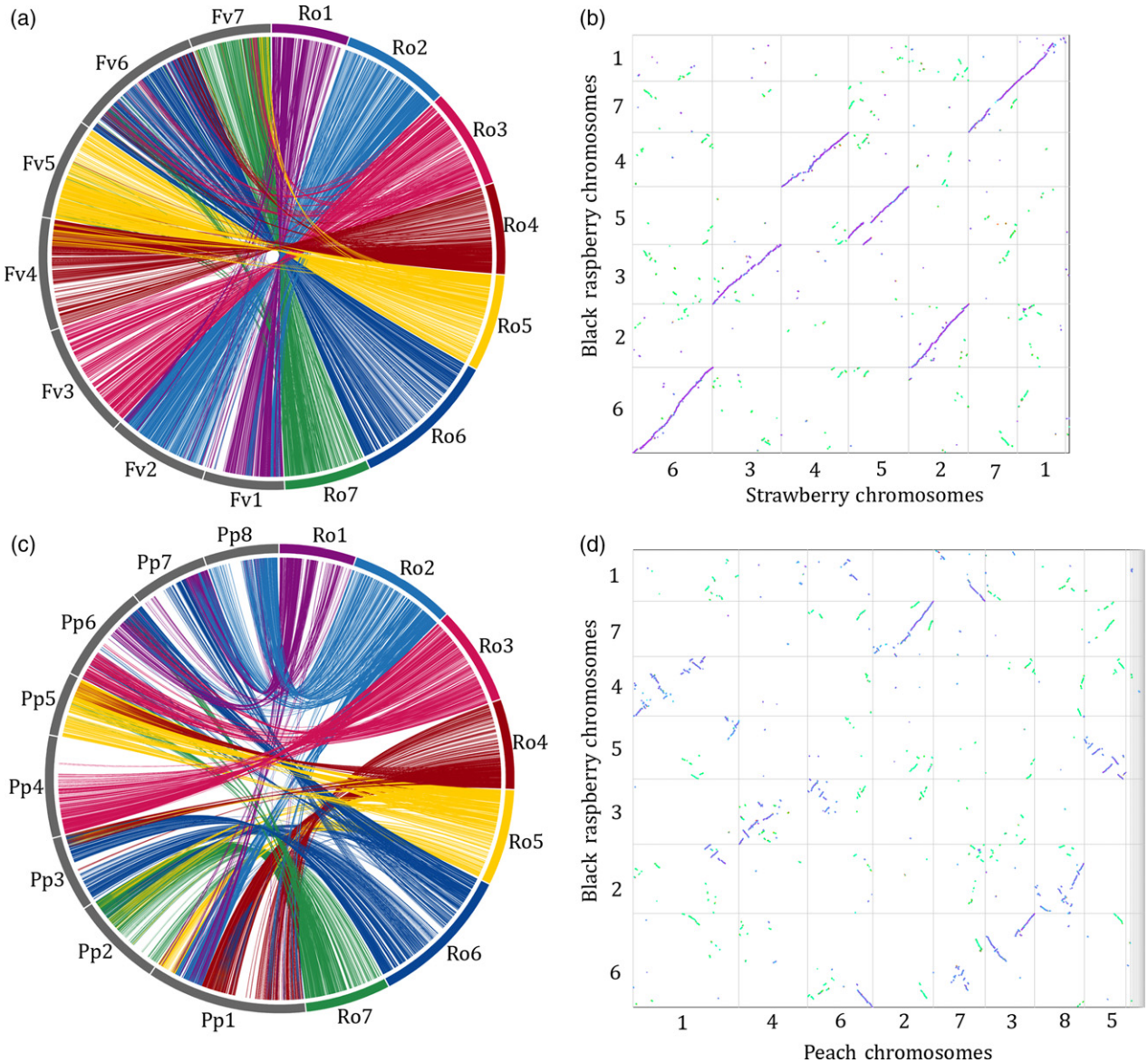

Figure 2. Comparative genomics within the Rosaceae.
(a) Circos plot (Krzywinski *et al.*, 2009) showing macro-synteny between the black raspberry and strawberry genome where each connecting line represents a syntenic gene pair.
(b) Macrosyntenic dot plot between black raspberry and strawberry (Fvb). Each dot represents a syntenic gene pair and color is based on age (K_s) where 1:1 homologs are purple and duplicated regions retained from the gamma whole-genome duplication are shown in green/blue.
(c, d) (c) Circos-based macro-synteny between black raspberry and peach and (d) macrosyntenic dot plot between black raspberry and peach.

chromosomes 4 and 6 (Figure 2c,d). This is largely consistent with previous findings in strawberry (Shulaev *et al.*, 2011) and markers from a black raspberry genetic map (Bushakra *et al.*, 2012).

We also compared the black raspberry genome sequence to published linkage maps of red raspberry (Woodhead *et al.*, 2010) and tetraploid blackberry (Castro *et al.*, 2013) by aligning the sequences of published markers on those linkage maps to the black raspberry assembly and comparing their placement. As expected, black raspberry and red raspberry were largely congruent, with 55 of the 63 matched markers occurring on the same linkage group in both species (Table S6). Differences between black raspberry and blackberry were larger and more notable. Fewer than half of the 56 matched markers occurred on the same linkage group. Most notably, none of the blackberry markers on LG7 from Castro *et al.* (2013) occurred on LG7 in black raspberry, instead most are on LG2 in black raspberry, with five markers from blackberry LG2 appearing instead on black raspberry LG7. In addition,

we noted that the orientation of the only three markers appearing in common between blackberry LG2 and black raspberry LG2 has been flipped, with Rh_MEa002cA01 appearing on the telomeric end of LG2 in blackberry (Castro *et al.*, 2013) and near the middle of LG2 in black raspberry (Bushakra *et al.*, 2015). These differences can be surveyed in more detail with the availability of additional genomic resources for blackberry.

Survey of fruit ripening expression patterns

Black raspberry fruits are high in anthocyanins, phenolics, and are a rich source of antioxidants (Wang and Lin, 2000; Moyer *et al.*, 2002; Dossett *et al.*, 2008). We surveyed gene expression in three fruit development stages (green, red, and ripe fruit) to identify genes related to fruit ripening and anthocyanin accumulation. Overall, 4446 genes were differentially expressed between green and red tissue, 9694 between green and ripe tissue, and 8376 between red and ripe tissue. Genes upregulated during fruit ripening (green versus red and red versus ripe) are enriched in gene ontology (GO) terms related to hydrolase activity, cell wall degradation, sugar transport, and anthocyanin accumulation, among others (Tables S7 and S8). We identified gene copies of each enzyme in the anthocyanin biosynthesis pathway that were significantly upregulated in ripening fruit compared to the other black raspberry tissues (Table S9). Two functional copies of the gene encoding chalcone synthase (*CHS*) and the first committed step of flavonoid biosynthesis, are highly expressed in ripe fruit with fragments per kilobase per million mapped (FPKM) of 10 019 and 18 342 compared to ranges of 105–907 in other tissues. Similar patterns of ripe fruit-specific expression were observed in genes encoding chalcone isomerase, flavanone 3-hydroxylase, flavonoid 3′-hydrozylase, leucoanthocyanidin dioxygenase, and anthocyanidin 3-*O*-glucosyltransferase (Table S9).

Gene duplicates related to fruit quality and domestication traits

Black raspberry is one of the most recently domesticated Rosaceae fruit crops with targeted breeding efforts originating in the northeastern USA in the 1880s (Jennings, 1988). Although the domestication history is short, the black raspberry genome may contain signatures of human selection for desirable fruit quality and disease-resistance traits. Gene duplications provide the raw material for evolutionary innovation and have essential roles in the evolution of new traits (Ohno, 1970). For instance, a retrotransposon-mediated gene duplication in tomato is responsible for the elongated fruit shape trait developed during domestication (Xiao *et al.*, 2008), and a tandem array of disease-resistance genes confers nematode resistance in soybean (Cook *et al.*, 2012). We surveyed small-scale duplications (SSD) which included tandem duplicates that may have transposed (Edger and Pires, 2009) to identify gene duplication events related to domestication and black raspberry-specific traits. Very recent SSDs (synonymous substitution rate (K_s) < 0.0001) may be related to human domestication traits and older duplications (K_s < 0.1) are lineage-specific and likely occurred after the divergence of black raspberry (*R. occidentalis*) from strawberry (*F. vesca*). We estimated the divergence of these two species at K_s 0.3468 (median) based on syntenic orthologs.

We identified 290 SSDs that occurred very recently and 1021 SSDs that were lineage-specific (Figure 3a).Very recent SSDs were enriched for GO terms including metabolism, fruit development, and seed development, among others (Table S10) and were largely distributed among

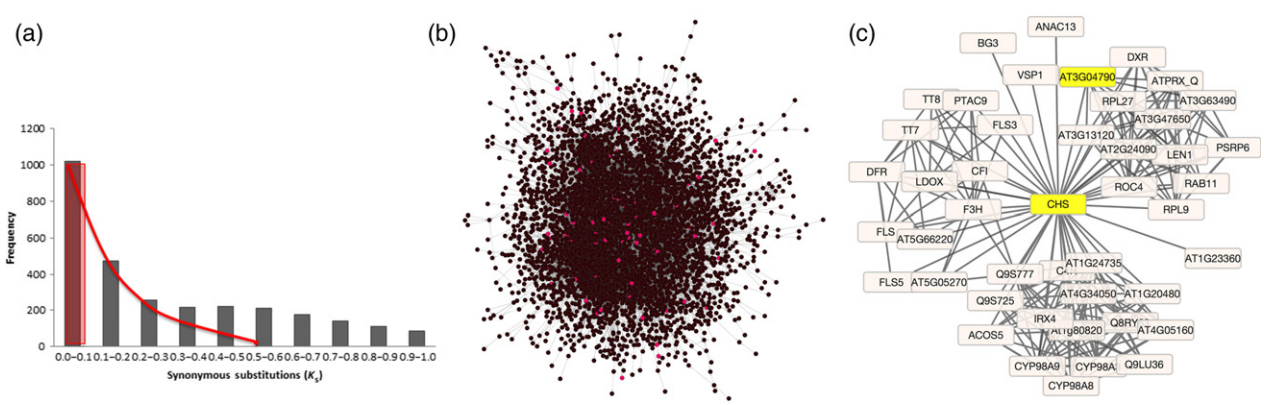

Figure 3. Recent small-scale gene duplications in the black raspberry genome.
(a) The distribution of small-scale duplicates plotted by age (K_s). Small-scale duplicates (K_s < 0. 1) are highlighted in red.
(b) The distribution of small-scale duplicates in the black raspberry metabolic network. Each node in the network represents an enzyme and edges (connections) represent shared metabolites. Network nodes containing SSDs are shown in lightly colored red and have low connectivity (i.e. number of interactions) in the global network.
(c) The gene encoding chalcone synthase (CHS) is depicted as the most central node with all direct interactors (an example from the network). Nodes with SSDs are highlighted in yellow.

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways involved in fructose and mannose metabolism and the pentose phosphate pathway (Table S11). These duplicates also included two copies of *CHS*, the first committed enzyme in flavonoid biosynthesis. Black raspberry fruits have high levels of anthocyanins, particularly cyanidin-3-rutinoside and cyanidin-3-xylosylrutinosidin, that contribute to their dark color and distinguish them in part from red raspberries (Dossett *et al.*, 2010). Older SSDs ($K_s < 0.4$) had GO enrichment for anthocyanin and pigment accumulation, in addition to flower development (Table S12). In comparison, very recent SSDs ($K_s < 0.0001$) in *F. vesca* were not enriched in any GO terms or KEGG pathways (Tables S13, S14). *F. vesca* is a wild species with no intensive breeding or domestication efforts possibly explaining the contrasting pattern observed in the domesticated black raspberry. In both species, older SSDs ($K_s < 0.1$) had GO and KEGG pathway enrichment for various primary and secondary metabolic processes (Tables S12, S15 and S16).

Black raspberry had 4295 tandem duplicated (TD) genes corresponding to 1770 ancestral gene copies, which is a slightly lower but of similar proportion (18% of genes) to other Rosaceae genomes. Most TD are single events: 1339 ancestral genes are present in two copies, 265 in three copies, 99 in four copies and 67 in five or more copies. The largest array has 17 duplicated gene copies. Duplicated genes can retain their ancestral function after duplication, partition their ancestral function between the new and old gene copies (sub-functionalization) or diverge and develop new functions (neo-functionalization) (Ohno, 1970). Using the RNA-seq-based expression profiles, we assessed the fate of duplicated genes in black raspberry. Over 56% (991) of the TD clusters were differentially expressed (DE) in all of the surveyed tissues and 89% (1575) were DE in at least one tissue. This suggests that the majority of TDs has undergone subfunctionalization or regulatory neofunctionalization. In addition, 15% (646) of TD clusters have a putatively non-functional copy (with FPKM < 1 in all tissues) and 3.8% (164) of TD clusters had no detectable expression.

We constructed a metabolic network for black raspberry where each node in the network represents an enzyme and edges (connections) represent a shared metabolite (de Oliveira Dal'Molin *et al.*, 2010). The black raspberry metabolic network has 4774 nodes with 39 005 edges and encompasses most biochemical pathways including photosynthesis, core anabolic and catabolic processes, and various signaling pathways (e.g. stress response). TDs are poorly connected in the metabolic network compared with non-duplicated nodes (Bekaert *et al.*, 2012) consistent with the gene balance hypothesis (Birchler and Veitia, 2007) (Figure 3b). A subset of the metabolic network showcasing SSDs of CHS is shown in Figure 3(c). This metabolic network is a useful resource for comparative genomics and breeding efforts within the Rosaceae.

Expression changes following *Verticillium dahliae* inoculation

In total, 144 predicted protein-coding genes in black raspberry contained motifs conserved among disease-resistance 'R' genes (Table S9). Of these, 33 had complete or partial Till-Interleukin 1 Receptor (TIR) domains, and 68 had complete or partial coiled-coil (CC) domains (Table S17). Monocots generally lack TIR domains and the proportions of CC to TIR-containing nucleotide-binding site leucine-rich repeat (NBS-LRR) genes varies among dicot species. The black raspberry genome had three times as many CC-NBS-LRR genes relative to TIR-NBS-LRR genes, a proportion similar to that found in grape (*Vitis vinifera* L.) (Jaillon *et al.*, 2007) (Table S17). Relative to Arabidopsis, the black raspberry genome is nearly twice the size, but has 62 fewer R genes. The maximum number of R genes clustering on a single black raspberry genome scaffold was 11. Along with aphid-vectored viruses, *Verticillium* wilt (VW) is a leading cause of stand decline in commercial black raspberry fields (Dossett, 2011). We surveyed gene expression changes between *V. dahliae*-inoculated and non-inoculated (wild-type) 'Jewel' roots to identify genes associated with disease response. Analysis of the RNA-seq data identified 147 genes that were differentially expressed (Tables S18 and S19). A comparison of these DE genes against the NCBI GenBank Database identified eight genes with apparent homology to known or candidate genes involved in disease resistance. Of the eight R-like genes, seven were upregulated 1.5-4.4-fold (Table S18) and one was downregulated 2.4-fold (Table S19) in inoculated plants relative to control plants. Twelve of the DE genes have homology to transcription factors. Seven of these genes were upregulated 1.5- to 3.1-fold (Table S18) and five were downregulated 1.5- to 2.4-fold (Table S19) in inoculated plants relative to control plants. In tomato, the *Ve*1 gene conferring VW resistance encodes a receptor-like protein-type cell surface receptor (Fradin *et al.*, 2011). A tBLASTx search using the tomato *Ve*1 gene as a query against the black raspberry genome sequence identified 10 putative black raspberry *Ve1* homologs. None of these genes was represented among the DE gene set.

The genes we have identified as differentially expressed (±1.5-fold) in *Verticillium*-inoculated versus control root transcriptome constitute a set of candidate genes for *Verticillium* response in black raspberry. Among the upregulated genes are homologs of *Arabidopsis* pest and pathogen resistance genes including *PDF1.4* (reviewed by Thomma *et al.*, 2002), *HSPRO2* (Murray *et al.*, 2007), and two mitogen-activated protein kinase genes *MEKK1* (Yan *et al.*, 1994) and *NPK1* (Nakashima *et al.*, 1998). A homolog of powdery mildew resistance gene *MLO3* (Freialdenhoven *et al.*, 1996) was among the downregulated genes, suggesting pathogen-specificity is invoked in black raspberries inoculated with *Verticillium.*

DISCUSSION

Specialty crops such as black raspberry are under-utilized and often suffer from disease susceptibility, variable yields, and restrictive growing conditions stemming from their low or eroded genetic diversity and limited breeding efforts (Mayes *et al.*, 2012). Despite these issues, specialty crops have great potential for increasing crop diversity and global food security and many have high-value market demand. Black raspberries are prized for their flavor and nutritional qualities but production has declined since the early 1900s, and only three new cultivars have been released over the last 20 years. Central to the success of black raspberry breeding programs is the availability of high-quality genomic resources. The chromosome-scale assembly reported here is useful for accelerating marker assisted breeding progress and comparative genomics across the Rosaceae. Black raspberry shares a high degree of macro-synteny and collinearity with the woodland strawberry genome. Despite this large-scale similarity, both species have undergone significant lineage-specific gains and losses and over one-third of the predicted genes are no longer syntenic.

Humans have domesticated several hundred plant species for a wide range of agronomic and aesthetic purposes. Signatures of early domestication can be traced back to single events such as point mutations or transposon insertions as observed in the seed shattering gene (*qSH1*) of rice (Konishi *et al.*, 2006) or the *teosinte branched 1* gene in maize (Clark *et al.*, 2006). The process of crop improvement is more complex and is likely the product of additive effects from multiple beneficial mutations. Gene duplications are the drivers of evolutionary innovation and likely play a major role in human-guided crop improvement. Very recent gene duplications coinciding with black raspberry domestication are enriched for pathways related to agronomic traits such as fruit development, sugar accumulation, and anthocyanin biosynthesis. Older gene duplicates are involved in a more diverse array of functions unrelated to domestication traits. Very recent gene duplications in the wild woodland strawberry (*F. vesca*) show no patterns of enrichment as expected given the lack of targeted human improvements through breeding. Together this suggests gene duplications contribute to domestication and these duplications represent targets for future improvement.

EXPERIMENTAL PROCEDURES

Plant material and experimental design for *Verticillium* inoculation

In July 2012, 14 plants of black raspberry 'Jewel' were inoculated with *V. dahliae* using a modified root dip method (Bhat and Subbarao, 1999) with an additional six plants serving as non-inoculated, negative controls. Each plant was un-potted, and soil was washed from the root system with running tap water. The lowest 1 cm of the root system was cut off with sterile scissors and the trimmed root ball was soaked for 10 min in a 1:1 spore suspension (1×10^6 spores/mL) from two *V. dahliae* isolates originally collected in 2008 from naturally infected, field-grown black raspberries. Control plants were treated the same way except that the roots were soaked in distilled water instead of inoculum. The 20 plants (14 treatment and six control) were then potted into black plastic 1-gallon pots with Sunshine Professional Growing Mix (Sun Gro Horticulture, Agawam, MA, USA) and arranged in a randomized design in a glasshouse with 27°C days, 24°C nights and no supplemental light.

DNA extraction for genome sequencing

The black raspberry selection ORUS 4115-3 (National Plant Germplasm System PI 672644) was used for sequencing. ORUS 4115-3 was grown from seed collected from Rich Mountain, South Carolina, USA (Lat. 34.7528723, Long. −83.1609897). High-molecular-weight genomic DNA was extracted from young leaf tissue using the E-Z 96® Plant DNA extraction kit (Omega Bio-Tek, Norcross, GA, USA) as previously described (Gilmore *et al.*, 2011).

RNA extraction for transcriptome sequencing

Total RNA from leaves, stems, canes, green fruit, red fruit, and ripe fruit of a single field-grown plant of 'Jewel' was isolated using a modified protocol described previously (Fox *et al.*, 2009; Filichkin *et al.*, 2010). Briefly, fresh tissue was flash-frozen in liquid nitrogen and ground to a fine powder in stainless steel jars using a Mixer Mill MM 301 (Retsch, Inc., Bonn, Germany). Approximately 180 mg of frozen tissue powder was suspended in 1 mL of ice-cold PureLink® Plant RNA Reagent (Invitrogen, Carlsbad, CA, USA) and centrifuged at ~21 000 ***g*** for 3 min. After addition of 200 μL of cold 5 M NaCl, the supernatant was extracted three times with 500 μL of water-saturated chloroform. The RNA from the aqueous layer was precipitated using 0.9 volumes of 2-propanol with a 10 min incubation. After centrifugation (for 10 min at 21 000 ***g***) the RNA pellet was re-suspended in 178 μL of RNAsecure reagent (Ambion, Thermo Fisher, Waltham, MA, USA) and heated at 65°C for 10 min to inactivate ribonucleases. Residual genomic DNA was digested at 37°C for 10 min using 2 μL of Turbo-DNase and 20 μL of 10 × Turbo-DNase buffer (Ambion, Thermo Fisher). The RNA was further purified using the RNeasy mini RNA kit (Qiagen, Inc., Valencia, CA, USA) according to the manufacture's protocol.

Root tissues were harvested from *Verticillium dahliae*-treated and control plants 2 months after inoculation. Whole plants were removed from their pots, potting medium was removed, and roots rinsed in tap water. Cleaned roots were placed in 50 mL tubes and immediately submerged in liquid nitrogen. Frozen tissues were ground in stainless steel jars using Mixer Mill MM 301 (Retsch, Inc.). Black raspberry roots have a high content of polyphenolics, therefore total RNA was isolated using the protocol described above with the following modifications. Briefly, approximately 2 g of frozen root tissue powder was suspended in 10 mL of ice-cold PureLink® Plant RNA Reagent (Invitrogen) and centrifuged at ~21 000 ***g*** for 3 min. After addition of 2 mL of cold 5 M NaCl the supernatant was extracted three times with 5 mL of water-saturated chloroform. The aqueous phase containing root RNA was diluted five-fold with Buffer QRV2 from the RNA/DNA Maxi kit (Qiagen Inc., Valencia, CA, USA) and applied on the Tip 500 gravity flow column (Qiagen Inc.). After RNA binding to the Tip 500 column, the remaining RNA purification steps (i.e. washing, elution, and precipitation with 2-propanol) were carried out according to the manufacturer's protocol. After precipitation, the RNA pellet was re-suspended in 1780 μL of RNAsecure™ reagent (Thermo Fisher Scientific Inc., Waltham, MA, USA) and heated at 65°C for

10 min to inactivate ribonucleases. Residual genomic DNA was digested at 37°C for 10 min using 20 μL of Turbo-DNase and 200 μL of 10 × Turbo-DNase buffer (Thermo Fisher Scientific, Inc.). DNase-treated RNA was further purified using the RNeasy Plant Mini kit (Qiagen, Inc.) following manufacturer's instructions. RNA integrity numbers (RIN) of three biological replicates for inoculated and control tissues were calculated using a Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA).

Illumina sequencing and data filtering

Raw sequences of ORUS 4115-3 were generated using Illumina® sequencing following standard protocol with the HiSeq2000® and MiSeq® platforms (Illumina, San Diego, CA, USA) at the Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR, USA. Eight paired-end libraries were created with estimated inserts of 165 bp (two separate libraries), 212 bp, 227 bp, 347 bp, 485 bp, 4678 bp and 4703 bp generating a total of 320 × coverage (Table S1). Raw Illumina® sequence data were trimmed using Trimmomatic (V0.32) (Bolger *et al.*, 2014) to remove adaptor sequences and bases with quality scores less than 30. Reads were retained only if both reads of the pair passed quality control filtering. Table S1 shows per-library estimated coverage statistics of the quality-filtered data.

Genome assembly

ALLPATHS-LG (Butler *et al.*, 2008) was used to generate draft assemblies from the post-quality-filtered Illumina data. We also generated assemblies using Velvet (1.0) (Zerbino and Birney, 2008) but ALLPATHS-LG produced the highest quality assembly based on the number of scaffolds, contig/scaffold N50 length, percentage of Ns, and the percentage of reads used in the assembly. The percentage of reads used from each library in the final assembly can be found in Table S2 and the final assembly statistics are shown in Table S3.

High-density genetic map construction and genome anchoring

Progeny from the cross ORUS 3021-2 × ORUS 4153-1 were used to construct a high-density genetic linkage map that was used to anchor the sequence scaffolds. A total of 115 F_1 offspring were genotyped using a GBS approach (Elshire *et al.*, 2011; Bushakra *et al.*, 2015). The previously reported GBS data (Bushakra *et al.*, 2015) was reanalyzed to obtain more markers for anchoring and to improve the map quality. TASSEL (V 3.0) (Bradbury *et al.*, 2007) was used to call SNP loci using the repeat-masked genome as a reference. Missing genotypes were imputed using the FILLIN module of TASSEL and the results were manually screened to remove erroneous genotypes. GBS-based and SSR markers were mapped using JoinMap v. 4.1 (Van Ooijen, 2006) with the maximum likelihood mapping algorithm and an independence log of odds (LOD) threshold of 5. In total, 1711 markers including 1657 GBS and 54 SSR markers were used to anchor the black raspberry scaffolds (Table S4). Markers from the genetic map were used to order and orient scaffolds, and ambiguities or inconsistencies between the two parental maps were resolved using synteny with the strawberry (Shulaev *et al.*, 2011) and peach (Verde *et al.*, 2013) genomes. Two GBS markers were required for scaffold anchoring and three markers were required for proper orientation. Scaffolds containing a single mapped marker were anchored only if they had syntenic support. Less than 10% of scaffolds had ambiguous orders between the two parental genetic maps. Custom perl scripts were used to generate the seven pseudochromosomes (with 10 000 Ns between scaffolds) and updated GFF files.

Repetitive element annotation

We used the REPET (Flutre *et al.*, 2011) packages TEdenovo and TEannot to identify uncharacterized repeat sequences in the black raspberry genome. The TEdenovo pipeline compares the genome with itself to identify and classify repeated genomic elements. All-By-All alignments were conducted with NCBI-BLAST + using default TEdenovo parameters. LTRharvest (Ellinghaus *et al.*, 2008) was used for structural detection. Consensus building was performed using default parameters. During consensus detect features, RepeatScout (Price *et al.*, 2005) was invoked and Pfam26.0 HMM profiles (Finn *et al.*, 2013) and repbase (v18.08) nucleotide and amino acids databanks were used. Consensus classification, filtering, and clustering were performed using default parameters.

Output from the TEdenovo pipeline was used as input to the TEannot pipeline. This pipeline mines the genome sequence using repeated sequences identified in the previous TEdenovo pipeline to produce classified, non-redundant consensus sequences. A set of perfect matching sequences from the TEdenovo-output transposable elements (TE) library was selected by running a subset of the TEannot pipeline, producing a working reference TE library. This TE library was then used in a full run of the TEannot pipeline. For alignment of the reference TE library, NCBI-BLAST+ was used, and BLASTER, repeat masker, and censor steps were run on the reference TE library and on randomized segments. Filtering and subsequent merging were performed using default parameters. For comparisons, repbase18.08 nucleic acids and amino acids databanks were used. Annotations were exported to GFF3 format. Overall, 136 249 014 bp (56% overall genomic sequence) were annotated as repetitive elements.

Prediction of disease-resistance genes

A set of 20 previously-identified disease-resistance gene motifs (Bailey *et al.*, 2015) was used to scan the amino acid sequences of the 28 005 black raspberry predicted protein-coding genes using Motif Alignment and Scanning Tool (MAST) v. 4.10.0. MAST output was compiled using the NLR-Parser tool (Steuernagel *et al.*, 2015).

RNA-seq library preparation and sequencing

RNA was extracted using the protocol outlined previously (Filichkin *et al.*, 2010), and RNA integrity was monitored using a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). Only RNA with RNA integrity number > 6 was used for preparation of Illumina RNA-seq libraries. Three biological replicates were used for each tissue. The RNA-seq libraries were prepared and bar coded using TruSeq RNA Sample Prep Kits (Illumina, San Diego, CA, USA) according to the manufacturer's protocol. Paired-end 101 bp sequencing runs were performed on the Illumina HiSeq2000 instrument.

Gene prediction and annotation

RNA-seq reads were assessed for quality using Trimmomatic (V0.32) for filtering as described above. Each individual quality-controlled library was assembled along with its replicates against the black raspberry reference genome using TopHat (v2.0.8b) (Trapnell *et al.*, 2012). Assemblies from all of the tissues were then merged and the transcripts were extracted. Putative transcript sequences were evaluated for coding potential and filtered by length. This process yielded 33 783 putative transcripts. Trinity (r2012-10-05) was used to assemble the RNA-seq data *de novo* using default parameters to obtain a set of putative transcripts. The assembled transcripts were evaluated for coding potential and filtered by length leaving a total of 71 622 high-quality transcripts.

MAKER (v2.28) (Cantarel *et al.*, 2008) was run on the 33 783 reference-guided and 71 622 *de novo*-assembled transcripts to produce a set of reference gene models. This combined set of putative transcript sequences was clustered by homology. For each cluster of sequences found to be at least 95% similar, the longest sequence was retained. This process yielded 29 460 representative sequences that were input to MAKER as expressed sequence tag (EST) evidence. These representative sequences were translated to amino acid sequences, and were combined with protein sequences from *Arabidopsis* (Swarbreck *et al.*, 2008), *Brachypodium* (Vogel *et al.*, 2010), rice (Project IRGS, 2005), poplar (Tuskan *et al.*, 2006), Sorghum (Paterson *et al.*, 2009), and maize (Schnable *et al.*, 2009) (total: 254 380 sequences). These sequences were clustered as described above, and the resulting 238 978 sequences were used as multi-organismal protein homology evidence. The black raspberry genome was masked based on the repetitive elements annotation, and the set of representative TE was input to MAKER as a custom repeat library. SNAPhmm, Augustus, and GeneMarkHMM were invoked by MAKER, and were initially trained using rice, rice, and maize data, respectively. SNAPhmm was subsequently iteratively trained using the output of a series of six MAKER runs. Only genes for which the encoded protein was predicted to contain a complete open reading frame (ORF) were retained. Genes encoding proteins with fewer than 33 amino acids in length were removed. Genes with overlapping coding regions were pruned based on MAKER-assigned annotation edit distanced (AED) score. In total this process identified 32 300 putative protein-coding genes that were then functionally annotated.

InterProScan version 5-RC6 (http://www.ebi.ac.uk/interpro/) was used to functionally annotate putative protein-coding sequences using default parameters. Additionally, NCBI BLASTP (blast.ncbi.nlm.nih.gov) was used to functionally annotate the gene models against the Swiss-Prot protein database (http://www.uniprot.org/). Sequence matches with e-value $\leq 10^{-5}$ were retained. Protein sequences with MAKER AED = 1 and that also had no functional annotation assignment by at least one of InterProScan or BLASTP versus Swiss-Prot were removed. This process yielded the 28 005 protein-coding genes that comprised our final set of black raspberry gene models.

RNA-seq expression analysis

For each replicate of each tissue type and condition, data were aligned against the black raspberry reference genome using TopHat2 (Trapnell *et al.*, 2012) (v2.0.10) and Bowtie2 (Langmead and Salzberg, 2012) (v2.1.0.0). Cuffdiff (v2.0.0) (Trapnell *et al.*, 2012) was used to quantify per-gene expression and differential expression between sets of conditions. For condition-specific gene expression analyses, genes expressed at a level of at least 5 FPKM in the highest expressed condition, and at no greater than 1, 2, 3, 4, and 5% of that highest expressed level in the next highest expressed condition.

Network construction

Pairwise genomic alignments of the black raspberry genome were made with *Arabidopsis* (AG Initiative, 2000), *F. vesca (Shulaev et al., 2011)*, and *V. vinifera* (Jaillon *et al.*, 2007) using QUOTA-ALIGN (Tang *et al.*, 2011) then filtered to identify orthologous gene clusters and tandemly duplicated gene copies. Pairs within a physical distance of less than ten genes apart were classified as tandem duplicates. Smaller scale gene duplicates were identified in black raspberry and *F. vesca* by calculating substitutions per synonymous sites (K_s) between all genes in the genome and constructing K_s age distributions (Maere *et al.*, 2005). Duplicate pairs were identified with BLAST (Altschul *et al.*, 1997), aligned with GeneWise and MUSCLE (Birney *et al.*, 2004; Edgar, 2004), and K_s values calculated with the PAML package (Yang, 1997). The black raspberry ortholog list was then filtered to include genes with functional data in the STRING v9.1 global *Arabidopsis* protein interaction network (Franceschini *et al.*, 2013). Gene expression patterns and duplicated genes were mapped onto this network using Cytoscape v3.1.1 (Saito *et al.*, 2012) to identify clusters of co-expressed and interacting duplicate genes. Enrichment analyses were calculated using tools and curated descriptions of gene functions from the GO Consortium databases (Ashburner *et al.*, 2000) and KEGG (Kanehisa and Goto, 2000). Network statistics were calculated using Network Analyzer (Doncheva *et al.*, 2012) including average number of neighbors (i.e. protein interactions).

Chloroplast genome assembly

The black raspberry chloroplast genome was assembled using the reference-guided 'alignreads' pipeline (Straub *et al.*, 2011). The *F. vesca* (Shulaev *et al.*, 2011) v1.1 chloroplast sequence, including a single copy of the inverted repeat, was used as a reference. The resulting black raspberry and strawberry consensus chloroplast sequence was then input as a reference for a second alignreads run. The consensus sequences for both runs were realigned to the strawberry reference using MEGA6 (Tamura *et al.*, 2011). Differences between the consensus sequences were manually inspected to identify the correct version. Six gaps ranging from 100 to 1100 bp remained in the consensus sequence. Primer3 was used to design primers to flank the gaps. The PCR protocol began with an initial denaturation at 98°C for 30 sec followed by 40 cycles of 98°C for 8 sec, 55°C for 30 sec, 72°C for 1:20, followed by a final extension of 72°C for 10 min. Sanger sequencing (CGRB, Corvallis, OR, USA) was then employed on the amplified fragments. A draft annotation of the chloroplast genome was produced using DOGMA (Wyman *et al.*, 2004) and visualized with OGDraw v. 1.2 (Lohse *et al.*, 2007).

Comparative genomics

Syntenic dot plots were generated with SynMap (Lyons *et al.*, 2008) using LAST for whole-genome comparisons. A minimum of five syntenic genes was used to seed a syntenic block allowing for a maximum distance of 20 genes between adjacent syntenic genes. Syntenic gene pairs were colored based on their K_s values in order to differentiate orthologous and out-paralog syntenic blocks. Circularized syntenic plots were constructed using Circos (Krzywinski *et al.*, 2009) with each connecting line representing orthologous gene sequences.

Accession codes and genome links

The genome assembly and annotation are available for download on the Genome Database for Rosaceae (https://www.rosaceae.org/).

ACKNOWLEDGEMENTS

This work is supported in part by funding from the National Science foundation (DBI-1401572 to RV), Donald Danforth Plant Science Center to TCM and the USDA-National Institute of Food and Agriculture (NIFA) Specialty Crop Research Initiative (SCRI) USDA-ARS CRIS 2072-21000-044-00D, 2072-21000-047-00D and 2072-21220-002-00D to CEF and NVB. PPE is supported by USDA-NIFA and Michigan State University AgBioResearch.

AUTHOR CONTRIBUTIONS

R.V., D.B., N.V.B., T.C.M. designed and conceived research; D.B., J.M.B., M.D., E.R.R., identified biological material, and extracted DNA and RNA; D.B., and T.C.M. annotated genome features; R.V., D.B., K.J.V., P.P.E., E.R.R., H.D.P., S.A.F., E.L., T.P.M., M.D., C.E.F., N.V.B., T.C.M., analyzed data; R.V., D.B., J.M.B., T.C.M. wrote the paper. All authors read and approved the final manuscript.

SUPPORTING INFORMATION

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

Figure S1. Kmer distribution of unassembled black raspberry Illumina reads.

Figure S2. Map of the black raspberry chloroplast genome.

Figure S3. Identification of centromeric satellite DNA repeats.

Figure S4. Alignment of 303 bp monomeric centromere satellite repeats.

Figure S5. Alignment of 95 bp monomeric centromere satellite repeats.

Figure S6. Centromere repeat distributions on black raspberry chromosomes 1 and 2.

Figure S7. Centromere repeat distribution on scaffold 1.

Table S1. Sequencing statistics of Illumina paired end libraries.

Table S2. Allpaths-LG data usage and measured coverage statistics.

Table S3. Summary of black raspberry markers used for scaffold anchoring.

Table S4. Summary of anchored scaffolds in each pseudochromosome.

Table S5. Summary of repetitive elements in the black raspberry genome.

Table S6. Comparison of linkage group placement for markers from red raspberry and blackberry markers against the black raspberry assembly using the revised naming system for raspberry of Bushakra *et al.* (2012).

Table S7. Enriched GO terms in genes upregulated in red fruit tissue compared to green tissue.

Table S8. Enriched GO terms in genes upregulated in ripe fruit tissue compared to red tissue.

Table S9. Expression patterns of genes involved in flavonoid biosynthesis.

Table S10. GO enrichment of recent SSD with $K_s < 0.0001$ in black raspberry.

Table S11. KEGG pathway enrichment of recent SSD with $K_s < 0.0001$ in black raspberry.

Table S12. GO enrichment of SSD with $K_s < 0.4$ in black raspberry.

Table S13. GO enrichment of SSD with $K_s < 0.0001$ in *F. vesca.*

Table S14. KEGG pathway enrichment of recent SSD with $K_s < 0.0001$ in *F. vesca.*

Table S15. GO enrichment of SSD with $K_s < 0.1$ in *F. vesca.*

Table S16. KEGG pathway enrichment of recent SSD with $K_s < 0.0001$ in *F. vesca.*

Table S17. Nucleotide binding site-leucine-rich repeat (NBS-LRR) disease resistance gene homologs identified in the black raspberry genome.

Table S18. RNA sequence results showing the black raspberry genes that are upregulated when comparing untreated (wild-type) versus *Verticillium*-infected 'Jewel' roots.

Table S19. RNA sequence results showing black raspberry genes that are downregulated when comparing untreated (wild-type) versus *Verticillium*-infected 'Jewel' roots.

REFERENCES

AG Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature*, **408**, 796.

Alice, L.A. and Campbell, C.S. (1999) Phylogeny of Rubus (Rosaceae) based on nuclear ribosomal DNA internal transcribed spacer region sequences. *Am. J. Bot.* **86**, 81–97.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.

Anonymous (2014) National Statistics for Raspberries. National Agricultural Statistics Service

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S. and Eppig, J.T. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29.

Bailey, T.L., Johnson, J., Grant, C.E. and Noble, W.S. (2015) The MEME Suite. Nucleic acids research: gkv416

Bekaert, M., Edger, P.P., Hudson, C.M., Pires, J.C. and Conant, G.C. (2012) Metabolic and evolutionary costs of herbivory defense: systems biology of glucosinolate synthesis. *New Phytol.* **196**, 596–605.

Bhat, R.G. and Subbarao, K.V. (1999) Host range specificity in *Verticillium dahliae*. *Phytopathology*, **89**, 1218–1225.

Birchler, J.A. and Veitia, R.A. (2007) The gene balance hypothesis: from classical genetics to modern genomics. *The Plant Cell Online*, **19**, 395–402.

Birney, E., Clamp, M. and Durbin, R. (2004) GeneWise and genomewise. *Genome Res.* **14**, 988–995.

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

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.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.

Bushakra, J., Stephens, M., Atmadjaja, A., Lewers, K., Symonds, V., Udall, J., Chagné, D., Buck, E. and Gardiner, S. (2012) Construction of black (*Rubus occidentalis*) and red (R. idaeus) raspberry linkage maps and their comparison to the genomes of strawberry, apple, and peach. *Theor. Appl. Genet.* **125**, 311–327.

Bushakra, J.M., Bryant, D.W., Dossett, M., Vining, K.J., VanBuren, R., Gilmore, B.S., Lee, J., Mockler, T.C., Finn, C.E. and Bassil, N.V. (2015) A genetic linkage map of black raspberry (*Rubus occidentalis*) and the mapping of Ag 4 conferring resistance to the aphid Amphorophora agathonica. *Theor. Appl. Genet.* **128**, 1631–1646.

Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I.A., Belmonte, M.K., Lander, E.S., Nusbaum, C. and Jaffe, D.B. (2008) ALLPATHS: de novo assembly of whole-genome shotgun microreads. *Genome Res.* **18**, 810–820.

Cantarel, B.L., Korf, I., Robb, S.M., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A.S. and Yandell, M. (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196.

Castro, P., Stafne, E., Clark, J. and Lewers, K. (2013) Genetic map of the primocane-fruiting and thornless traits of tetraploid blackberry. *Theor. Appl. Genet.* **126**, 2521–2532.

Chagné, D., Crowhurst, R.N., Pindo, M., Thrimawithana, A., Deng, C., Ireland, H., Fiers, M., Dzierzon, H., Cestaro, A. and Fontana, P. (2014) The draft genome sequence of European pear (Pyrus communis L'.Bartlett'). *PLoS ONE*, **9**, e92644.

Clark, R.M., Wagler, T.N., Quijada, P. and Doebley, J. (2006) A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* **38**, 594–597.

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

Doncheva, N.T., Assenov, Y., Domingues, F.S. and Albrecht, M. (2012) Topological analysis and interactive visualization of biological networks and protein structures. *Nat. Protoc.* **7**, 670–685.

Dossett, M. (2011) Evaluation of genetic diversity in wild populations of black raspberry (*Rubus occidentalis* L.).

Dossett, M., Lee, J. and Finn, C.E. (2008) Inheritance of phenological, vegetative, and fruit chemistry traits in black raspberry. *J. Am. Soc. Hortic. Sci.* **133**, 408–417.

Dossett, M., Lee, J. and Finn, C.E. (2010) Variation in anthocyanins and total phenolics of black raspberry populations. *J. Funct. Foods*, **2**, 292–297.

Dossett, M., Bassil, N.V., Lewers, K.S. and Finn, C.E. (2012) Genetic diversity in wild and cultivated black raspberry (*Rubus occidentalis* L.) evaluated by simple sequence repeat markers. *Genet. Resour. Crop Evol.* **59**, 1849–1865.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797.

Edger, P.P. and Pires, J.C. (2009) Gene and genome duplications: the impact of dosage-sensitivity on the fate of nuclear genes. *Chromosome Res.* **17**, 699–717.

Ellinghaus, D., Kurtz, S. and Willhoeft, U. (2008) LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics*, **9**, 18.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.

Filichkin, S.A., Priest, H.D., Givan, S.A., Shen, R., Bryant, D.W., Fox, S.E., Wong, W.-K. and Mockler, T.C. (2010) Genome-wide mapping of alternative splicing in Arabidopsis thaliana. *Genome Res.* **20**, 45–58.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L. and Mistry, J. (2013) Pfam: the protein families database. Nucleic acids research: gkt1223

Flutre, T., Duprat, E., Feuillet, C. and Quesneville, H. (2011) Considering transposable element diversification in de novo annotation approaches. *PLoS ONE*, **6**, e16526.

Fox, S., Filichkin, S. and Mockler, T.C. (2009) Applications of ultra-high-throughput sequencing. In *Methods Mol. Biol.* **553**, 79–108.

Fradin, E.F., Abd-El-Haliem, A., Masini, L., van den Berg, G.C., Joosten, M.H. and Thomma, B.P. (2011) Interfamily transfer of tomato Ve1 mediates Verticillium resistance in Arabidopsis. *Plant Physiol.* **156**, 2255–2265.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P. and von Mering, C. (2013) STRING v9. 1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–D815.

Freialdenhoven, A., Peterhansel, C., Kurth, J., Kreuzaler, F. and Schulze-Lefert, P. (1996) Identification of genes required for the function of non-race-specific *mlo* resistance to powdery mildew in barley. *Plant Cell*, **8**, 5–14.

Gilmore, B.S., Bassil, N.V. and Hummer, K.E. (2011) DNA extraction protocols from dormant buds of twelve woody plant genera. *J. Am. Pomo. Soc.* **65**, 201–207.

Hedrick, U.P. and Howe, G.H. (1925) Small fruits of New York.

Hellman, E., Skirvin, R. and Otterbacher, A. (1982) Unilateral incompatibility between red and black raspberries [Fruit breeding]. *J. Am. Soc. Hortic. Sci.* **107**, 781–784.

Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N. and Jubin, C. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467

Jennings, D. (1988) *Raspberries and Blackberries: Their Breeding, Diseases and Growth*. Cambridge, MA: Academic press.

Kanehisa, M. and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30.

Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T. and Yano, M. (2006) An SNP caused loss of seed shattering during rice domestication. *Science*, **312**, 1392–1396.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645.

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

Lohse, M., Drechsel, O. and Bock, R. (2007) OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **52**, 267–274.

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., Wang, X., Bowers, J., Paterson, A. and Lisch, D. (2008) Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol.* **148**, 1772–1781.

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M. and Van de Peer, Y. (2005) Modeling gene and genome duplications in eukaryotes. *Proc. Natl Acad. Sci. USA*, **102**, 5454–5459.

Mayes, S., Massawe, F., Alderson, P., Roberts, J., Azam-Ali, S. and Hermann, M. (2012) The potential for underutilized crops to improve security of food production. *J. Exp. Bot.* **63**, 1075–1079.

Melters, D.P., Paliulis, L.V., Korf, I.F. and Chan, S.W. (2012) Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Res.* **20**, 579–593.

Melters, D.P., Bradnam, K.R., Young, H.A., Telis, N., May, M.R., Ruby, J.G., Sebra, R., Peluso, P., Eid, J. and Rank, D. (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* **14**, R10.

Meng, R. and Finn, C. (2002) Determining ploidy level and nuclear DNA content in Rubus by flow cytometry. *J. Am. Soc. Hortic. Sci.* **127**, 767–775.

Michael, T.P. and VanBuren, R. (2015) Progress, challenges and the future of crop genomes. *Curr. Opin. Plant Biol.* **24**, 71–81.

Millen, R.S., Olmstead, R.G., Adams, K.L., Palmer, J.D., Lao, N.T., Heggie, L., Kavanagh, T.A., Hibberd, J.M., Gray, J.C. and Morden, C.W. (2001) Many parallel losses of infA from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell*, **13**, 645–658.

Moyer, R.A., Hummer, K.E., Finn, C.E., Frei, B. and Wrolstad, R.E. (2002) Anthocyanins, phenolics, and antioxidant capacity in diverse small fruits: Vaccinium, Rubus, and Ribes. *J. Agric. Food Chem.* **50**, 519–525.

Murray, S.L., Ingle, R.A., Petersen, L.N. and Denby, K.J. (2007) Basal resistance against pseudomonas syringae in arabidopsis involves WRKY53 and a protein with homology to a nematode resistance protein. *Mol. Plant Microbe Interact.* **20**, 1431–1438.

Nakashima, M., Hirano, K., Nakashima, S., Banno, H., Nishihama, R. and Machida, Y. (1998) The expression pattern of the gene for NPK1 protein kinase related to mitogen-activated protein kinase kinase kinase (MAPKKK) in a tobacco plant: correlation with cell proliferation. *Plant Cell Physiol.* **39**, 690–700.

Ohno, S. (1970) *Evolution by Gene Duplication*. London: George Alien & Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag.

de Oliveira Dal'Molin, C.G., Quek, L.-E., Palfreyman, R.W., Brumbley, S.M. and Nielsen, L.K. (2010) AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant Physiol.* **152**, 579–589.

Parra, G., Bradnam, K. and Korf, I. (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23**, 1061–1067.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T. and Poliakov, A. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature*, **457**, 551–556.

Potter, D., Eriksson, T., Evans, R.C., Oh, S., Smedmark, J., Morgan, D.R., Kerr, M., Robertson, K.R., Arsenault, M. and Dickinson, T.A. (2007) Phylogeny and classification of Rosaceae. *Plant Syst. Evol.* **266**, 5–43.

Price, A.L., Jones, N.C. and Pevzner, P.A. (2005) De novo identification of repeat families in large genomes. *Bioinformatics*, **21**, i351–i358.

Project IRGS (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

Saito, R., Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., Lotia, S., Pico, A.R., Bader, G.D. and Ideker, T. (2012) A travel guide to Cytoscape plugins. *Nat. Methods*, **9**, 1069–1076.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L. and Graves, T.A. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.

Seeram, N.P. (2008) Berry fruits: compositional elements, biochemical activities, and the impact of their intake on human health, performance, and disease. *J. Agric. Food Chem.* **56**, 627–629.

Shulaev, V., Sargent, D.J., Crowhurst, R.N., Mockler, T.C., Folkerts, O., Delcher, A.L., Jaiswal, P., Mockaitis, K., Liston, A. and Mane, S.P. (2011) The genome of woodland strawberry (Fragaria vesca). *Nat. Genet.* **43**, 109–116.

Steuernagel, B., Jupe, F., Witek, K., Jones, J.D. and Wulff, B.B.(2015) NLR-parser: rapid annotation of plant NLR complements. *Bioinformatics*, **31**, 1665–1667.

Stoner, G.D., Sardo, C., Apseloff, G., Mullet, D., Wargo, W., Pound, V., Singh, A., Sanders, J., Aziz, R. and Casto, B. (2005) Pharmacokinetics of anthocyanins and ellagic acid in healthy volunteers fed freeze-dried black raspberries daily for 7 Days. *J. Clin. Pharmacol.* **45**, 1153–1164.

Straub, S.C., Fishbein, M., Livshultz, T., Foster, Z., Parks, M., Weitemier, K., Cronn, R.C. and Liston, A. (2011) Building a model: developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. *BMC Genom.* **12**, 211.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R. and Ploetz, L. (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* **36**, D1009–D1014.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.

Tang, H., Lyons, E., Pedersen, B., Schnable, J.C., Paterson, A.H. and Freeling, M. (2011) Screening synteny blocks in pairwise genome comparisons through integer programming. *BMC Bioinformatics*, **12**, 102.

Tennessen, J.A., Govindarajulu, R., Ashman, T.-L. and Liston, A. (2014) Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. *Genome Biol. Evol.* **6**, 3295–3313.

Thomma, B.P.H.J., Cammue, B.P.A. and Thevissen, K. (2002) Plant defensins. *Planta*, **216**, 193–202.

Thompson, M.M. (1995) Chromosome numbers of *Rubus* species at the National Clonal Germplasm Repository. *HortScience*, **30**, 1447–1452.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578.

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S. and Salamov, A. (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science*, **313**, 1596–1604.

Van Ooijen, J. (2006) *JoinMap® 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Wageningen 33: Kyazma BV.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana, P., Bhatnagar, S.K., Troggio, M. and Pruss, D. (2010) The genome of the domesticated apple (Malus [times] domestica Borkh.). *Nat. Genet.* **42**, 833–839.

Verde, I., Abbott, A.G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M.T., Grimwood, J. and Cattonaro, F. (2013) The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat. Genet.* **45**, 487–494.

Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D., Bevan, M.W., Barry, K., Lucas, S., Harmon-Smith, M. and Lail, K. (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature*, **463**, 763–768.

Wang, S.Y. and Lin, H.-S. (2000) Antioxidant activity in fruits and leaves of blackberry, raspberry, and strawberry varies with cultivar and developmental stage. *J. Agric. Food Chem.* **48**, 140–146.

Wang, S., Shi, C. and Gao, L.-Z. (2013) Plastid genome sequence of a wild woody oil species, Prinsepia utilis, provides insights into evolutionary and mutational patterns of Rosaceae chloroplast genomes. *PLoS ONE*, **8**, e73946.

Woodhead, M., Weir, A., Smith, K., McCallum, S., MacKenzie, K. and Graham, J. (2010) Functional markers for red raspberry. *J. Am. Soc. Hortic. Sci.* **135**, 418–427.

Wu, J., Wang, Z., Shi, Z., Zhang, S., Ming, R., Zhu, S., Khan, M.A., Tao, S., Korban, S.S. and Wang, H. (2013) The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* **23**, 396–408.

Wyman, S.K., Jansen, R.K. and Boore, J.L. (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics*, **20**, 3252–3255.

Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J. and van der Knaap, E. (2008) A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science*, **319**, 1527–1530.

Yan, M., Dai, T., Deak, J.C., Kyriakis, J.M., Zon, L.I., Woodgett, J.R. and Templeton, D.J. (1994) Activation of stress-activated protein kinase by MEKK1 phosphorylation of its activator SEK1. *Nature*, **372**, 798–800.

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**, 555–556.

Zerbino, D.R. and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829.

Zhang, Q., Chen, W., Sun, L. *et al.* (2012) The genome of Prunus mume. *Nat. Commun.* **3**, 1318.

