Disease Resistance Genetics and Genomics in Octoploid Strawberry

Christopher R. Barbey,*†‡‡ Seonghee Lee,*‡ Sujeet Verma,* Kevin A. Bird,*‡‡ Alan E. Yocca,*†† Patrick P. Edger,*¶** Steven J. Knapp,** Vance M. Whitaker,* and Kevin M. Folta*††

*Horticultural Sciences Department, †Graduate Program in Plant Molecular and Cellular Biology, University of Florida, Gainesville, FL, ‡Gulf Coast Research and Education Center, University of Florida, Wimauma, FL, ¶Department of Horticulture, **Ecology, Evolutionary Biology and Behavior, ‡‡Department of Plant Biology, Michigan State University, East Lansing, MI, and ‡§Department of Plant Sciences, University of California, Davis, CA

ABSTRACT Octoploid strawberry (Fragaria ×ananassa) is a valuable specialty crop, but profitable production and availability are threatened by many pathogens. Efforts to identify and introgress useful disease resistance genes (R-genes) in breeding programs are complicated by strawberry’s complex octoploid genome. Recently-developed resources in strawberry, including a complete octoploid reference genome and high-resolution octoploid genotyping, enable new analyses in strawberry disease resistance genetics. This study characterizes the complete R-gene collection in the genomes of commercial octoploid strawberry and two diploid ancestral relatives, and introduces several new technological and data resources for strawberry disease resistance research. These include octoploid R-gene transcription profiling, dN/dS analysis, expression quantitative trait loci (eQTL) analysis and RenSeq analysis in cultivars. Octoploid fruit eQTL were identified for 76 putative R-genes. R-genes from the ancestral diploids Fragaria vesca and Fragaria iinumae were compared, revealing differential inheritance and retention of various octoploid R-gene subtypes. The mode and magnitude of natural selection of individual F. ×ananassa R-genes was also determined via dN/dS analysis. R-gene sequencing using enriched libraries (RenSeq) has been used recently for R-gene discovery in many crops, however this technique somewhat relies upon a priori knowledge of desired sequences. An octoploid strawberry capture-probe panel, derived from the results of this study, is validated in a RenSeq experiment and is presented for community use. These results give unprecedented insight into crop disease resistance genetics, and represent an advance toward exploiting variation for strawberry cultivar improvement.

KEYWORDS Strawberry Disease Resistance R-gene eQTL Subgenome Dominance RenSeq

Cultivated strawberry (Fragaria ×ananassa) is an important specialty crop that is cultivated world-wide for its sweet and flavorful fruit. However, marketable yields and post-harvest quality are significantly affected by disease. The strawberry fruit presents a vulnerable target for microbial pathogens (Farzaneh et al. 2015), as it is soft, moist, carbohydrate rich, and subject to damage from forces as seemingly innocuous as rain (Herrington et al. 2011). Genetic disease resistance has been a long-standing breeding priority. While breeders have made progress in producing varieties with tolerance to some pathogens, growers remain dependent on exogenous crop protection strategies to reduce pathogen loads (Cordova et al. 2017).

Plant R-genes are mediators of resistance to specific pathogens via effector triggered immunity, which results in the hypersensitive response and cell death (Amil-Ruiz et al. 2011). R-genes require a high degree of regulation to maintain homeostatic transcript levels to mitigate off-target protein interactions (Hammond-Kosack and Jones 1997). For this reason, many classes of functional R-genes are expressed at low levels unless elicited by pathogens (Lai and Eulgem 2017), contributing...
to the challenges of R-gene genomic and functional annotation. About 60% of characterized plant R-genes contain nucleotide-binding (NB-ARC) and leucine-rich-repeat (LRR) domains, and are referred to NLR genes (Funk et al. 2018). Plant R-genes are frequent targets for genetic improvement via breeding and genetic engineering (Baumgartner et al. 2015; Djian-Caporalino et al. 2014), and gene editing methods may accelerate their introduction into already-dice varieties. However, progress has been hindered because relatively few R-genes conferring novel resistance have been characterized (Amil-Ruiz et al. 2011). This problem is appreciable in strawberry, where the genetic complexity of octoploid cultivars presents unique challenges for functional identification and cloning of causal variants. An analysis of diploid R-genes across the Rosaceae genus was previously conducted (Arya et al. 2014). New genetic resources for high-resolution genotyping in octoploid strawberry have resulted in the recent identification of several disease resistance loci (Mangandi et al. 2017; Nellist et al. 2019; Cockerton et al. 2018; Pincot et al. 2018; Salinas et al. 2018; Anciro et al. 2018; Roach et al. 2016; Verma et al. 2018). However, the specific genes mediating resistance in these QTL intervals typically remain unresolved, as genomic resources for octoploid strawberry have not kept pace with genetic mapping.

Cultivated strawberry shares common ancestors with the extant diploid species F. vesca, F. iinnuanae, F. nipponica, and F. viridis (Edger et al. 2019). A high-quality octoploid strawberry genome has been recently developed (Edger et al. 2019), enabling new kinds analyses and improved resolution compared with previous studies involving Fragaria NLRs (Jia et al. 2015; Zhong et al. 2018). Analysis of this F. xananassa ‘Camarosa’ genome identified the repertoire of octoploid R-gene sequences and further demonstrated a general genomic retention bias toward F. vesca-like sequences (Edger et al. 2019).

This research compares R-genes from octoploid strawberry with its diploid ancestors and provides additional analysis into the genetic control of R-gene expression and retention patterns. Additional bias toward retention of F. vesca-like R-genes was detected in octoploid strawberry, beyond the bias observed in non-R-gene coding sequences. This finding provides insight into potential practical drivers of biased gene retention. Conserved domains were compared to describe specific R-gene phylogenetic relationships. The octoploid genome was used to assemble 61 fruit transcriptomes, and used to discover subgenomic expression quantitative trait loci (eQTL) for R-genes expressed in octoploid strawberry. The results demonstrate robust capture and resequencing of octoploid and diploid R-genes using only short second-generation sequence reads and with relatively deep genomic multiplexing.

This report characterizes the complete R-gene collection in the genomes of commercial octoploid strawberry and two diploid ancestral relatives, providing the genome-level resolution necessary for fully exploiting genetic disease resistance in strawberry. This research introduces several new technology and data resources that now may be applied in study of strawberry disease resistance.

MATERIALS AND METHODS

Plant populations and genetic materials
Three pedigree-connected and segregating strawberry populations were created from crosses ‘Florida Elyana’ × ‘Mara de Bois’, ‘Florida Radiance’ × ‘Mara des Bois’, and ‘Strawberry Festival’ × ‘Winter Dawn’ (Figure S1). These cultivars and 54 progeny were selected for RNAseq and Istraw35 SNP genotyping analysis (Verma et al. 2017), and were used to identify expressed genes and R-gene eQTL. De novo assemblies of ‘Mara des Bois’ and ‘Florida Elyana’ were also used to help design RenSeq capture probes.

For RenSeq, 14 disease resistant octoploid cultivars and elite breeding lines were selected from the University of Florida breeding program, and supplemented with ‘Camarosa’ and with the ancestral diploid F. vesca. The RenSeq lines are F. vesca genotype Hawaii 4, ‘Camarosa’, Sweet Sensation ‘Florida127’, ‘Florida Elyana’, 11.28-34, 11.77-96, 11.98-41, 12.115-10, 12.121-5, 13.26-134, 13.42-5, 13.55-195, 14.100-58, 14.100-59, 14.101-154, and 14.101-225.

Identification of R-genes in strawberry spp
R-genes were predicted from the strawberry octoploid ‘Camarosa’ draft genome ‘F_ana_Camarosa_6-28-17.rn’ (Edger et al. 2019), the diploid F. vesca reassembly ‘Fragaria vesca_v2.0.a2’ (Tenneness et al. 2014), and the diploid F. iinnuanae assembly ‘FII_r1.1’ (Hirakawa et al. 2014). Domain-level analysis was performed using the CLC Genomics Workbench 11 HMM implementation to search for Pfam- v29 domains on translated gene models from all genomic and transcriptomic strawberry resources. Motif search was performed on all translated gene models, using 56 R-gene-associated motifs collected from (Van Gelder and Esmenjaud 2016; Lukasik and Takken 2009; Jupe et al. 2012). The CLC Genomics Workbench 11 (CLC Bio, Denmark) pattern discovery tool was trained on a preliminary list of strawberry R-genes, and novel motifs were re-iterated back to all protein models. The ncoils sequence analysis algorithm (Lupas et al. 1991) was used to detect coiled-coil domains, and the output was parsed into GFF3 format for protein list reannotation. BLAST2GO annotation (Conesa et al. 2005) was performed to assign putative functions to all genes and confirm sequence association with disease resistance in a cross-referenced database.

Protein models containing canonical R-gene domains (e.g., NB-ARC domain) were selected for inclusion as R-genes, as were gene models with more common domains (e.g., LRR) with supporting evidence of an R-gene-associated motif. BLAST2GO annotated disease resistance associated genes not meeting the domain and motif-level criteria were manually analyzed for potential inclusion, leading to the inclusion of many LRR-containing RLK putative R-genes.

NB-ARC phylogenetic analysis
NB-ARC domains were extracted from F. iinnuanae, F. vesca, and F. xananassa ‘Camarosa’. The CIPRES Science Gateway (Miller et al. 2010) was utilized for full-length protein sequence alignment using MUSCLE v3.7 (Edgar 2004) and Maximum likelihood analysis using RAxML v8.2.10 (Stamatakis 2014). Tree construction was performed using the PROTGAMMA rate distribution model with 100 bootstrap replicates, and rooted with human APAF-1. This process was replicated five times using different random number seeds. Trees were visualized in CLC Genomic Workbench 11 with a 50% threshold bootstrap value. Word clouds were generated per clade based on the relative domain content of the full proteins.

Cultivated strawberry shares common ancestors with the extant diploid species F. vesca, F. iinnuanae, F. nipponica, and F. viridis (Edger et al. 2019). A high-quality octoploid strawberry genome has been recently developed (Edger et al. 2019), enabling new kinds analyses and improved resolution compared with previous studies involving Fragaria NLRs (Jia et al. 2015; Zhong et al. 2018). Analysis of this F. xananassa ‘Camarosa’ genome identified the repertoire of octoploid R-gene sequences and further demonstrated a general genomic retention bias toward F. vesca-like sequences (Edger et al. 2019).

This research compares R-genes from octoploid strawberry with its diploid ancestors and provides additional analysis into the genetic control of R-gene expression and retention patterns. Additional bias toward retention of F. vesca-like R-genes was detected in octoploid strawberry, beyond the bias observed in non-R-gene coding sequences. This finding provides insight into potential practical drivers of biased gene retention. Conserved domains were compared to describe specific R-gene phylogenetic relationships. The octoploid genome was used to assemble 61 fruit transcriptomes, and used to discover subgenomic expression quantitative trait loci (eQTL) for R-genes expressed in octoploid strawberry. Data from the octoploid ‘Camarosa’ strawberry gene expression atlas (Sánchez-Sevilla et al. 2017) was also used to determine R-gene transcript accumulation throughout the strawberry plant.

Resistance gene enrichment and sequencing (RenSeq) is an advantageous method for sequencing R-genes (Andolfo et al. 2014), and is likely to be very useful for de novo resolution of causal mutations (Witek et al. 2016). This method can be used to identify casual mutations within existing disease resistance QTL. For this purpose, a novel octoploid strawberry RenSeq capture probe library was designed using the R-genes identified in this analysis. This panel was experimentally validated using the University of Florida breeding germplasm. The results demonstrate robust capture and resequencing of octoploid and diploid R-genes using only short second-generation sequence reads and with relatively deep genomic multiplexing.

This report characterizes the complete R-gene collection in the genomes of commercial octoploid strawberry and two diploid ancestral relatives, providing the genome-level resolution necessary for fully exploiting genetic disease resistance in strawberry. This research introduces several new technology and data resources that now may be applied in study of strawberry disease resistance.
dN/dS analysis

dN and dS values were computed using a set of custom scripts (https://github.com/Aeyocca/ka_lk_pipe/). Orthologous genes between the F. xananassa and F. vesca v4 (Edger et al. 2017b) genomes were identified using the compara module in JCVI utilities library (Tang et al., 2015). Filtering of the JCVI utilities output was performed using a custom Perl script to identify the best syntenic ortholog and best blast hit below e-value 1e-4. Alignment of each orthologous gene pair was performed using MUSCLE v3.8.31 (Edgar 2004), followed by PAL2NAL (v14) (Suyama et al. 2006) to convert the peptide alignment to a nucleotide alignment. Finally, dN and dS values were computed between those gene pairs using codeml from PAML Version 4.9h (Yang 2007) with parameters specified in the control file found in the GitHub repository listed above.

Tissue-specific transcriptome analysis

Raw short read RNAseq libraries from various ‘Camarosa’ tissue (Sánchez-Sevilla et al. 2017) with the study reference PRJEB12420 were download from the European Nucleotide Archive (https://www.ebi.ac.uk/ena). The complete 54 library RNAseq experiment consisted of six independent green receptacle libraries, six white receptacle libraries, six turning receptacle libraries, six red receptacle libraries, three root libraries, three leaf libraries, and six achene libraries each for all corresponding fruit stages. Raw RNAseq reads were assembled to the ‘Camarosa’ reference using the same pipeline as previously described for fruit transcriptome population analysis. Expression values from biologically-replicated libraries were averaged. Clustvis (Metsalu and Vilo 2015) was used for tissue-based RNAseq clustering and heatmap visualization using correlation distance and average linkage with scaling applied using default parameters.

Fruit transcriptome analysis

61 fruit transcriptomes were sequenced via Illumina paired-end RNAseq (Avg. 65million reads, 2x100bp), and consisted of parents and progeny from crosses of ‘Florida Elyana’ × ‘Mara de Bois’, ‘Florida Radiance’ × ‘Mara des Bois’, and ‘Strawberry Festival’ × ‘Winter Dawn’. Reads were trimmed and mapped to the F. xananassa octoploid ‘Camarosa’ annotated genome using CLC Genomic Workbench 11 (mismatch cost of 2, insertion cost of 3, deletion cost of 3, length fraction of 0.8, similarity fraction of 0.8, 1 maximum hit per read). Reads that mapped equally well to more than one locus were discarded from the analysis. RNAseq counts were calculated in Transcripts Per Million (TPM). Three-dimensional principle component analysis (PCA) was performed on all RNAseq assemblies, including two replicates of ‘Mara de Bois’ fruit harvested three years apart and sequenced independently (Figure S2). Transcript abundances were normalized via the Box-Cox transformation algorithm performed in R (Development Core Team 2014) prior to eQTL analysis. The BLAST2GO pipeline was used to annotate the full ‘Camarosa’ predicted gene complement.

Genotyping and genetic association of octoploid fruit R-genes

The Affymetrix IStraw35 Axiom SNP array (Verma et al. 2017) was used to genotype 60 individuals, including six parental lines from three independent biparental RNAseq populations (Figure S1). Sequence variants belonging to the Poly High Resolution (PHR) and No Minor Homozygote (NMH) marker classes were included for association mapping. Mono High Resolution (MHR), Off-Target Variant (OTV), Call Rate Below Threshold (CRBT), and Other marker quality classes, were discarded and not used for mapping. Individual marker calls inconsistent with Mendelian inheritance from parental lines were removed. The F. vesca physical map was used to orient marker positions as current octoploid maps do not include a majority of the available IStraw35 markers. A genome-wide analysis study (GWAS) was performed using GAPIT v2 (Tang et al. 2016) performed in R. R-gene eQTL were evaluated for significance based on the presence of multiple co-locating markers of p-value < 0.05 after false discovery rate correction for multiple comparisons. Cis vs. trans eQTL determinations were made by corroborating known ‘Camarosa’ physical gene position with the eQTL position the F. vesca map. In the example case of FaDRL28, subgenomic localization was confirmed via BLAST of the associated markers to the correct ‘Camarosa’ homeologous chromosome.

Subgenome dominance in octoploid strawberry R-genes

The closest homolog for each F. xananassa ‘Camarosa’ gene in either Fragaria_vesca_v2.0.a2.cds or FII_r1.1cds was determined via BLAST analysis (e-value threshold < 0.1, word size = 25, match = 1, mismatch = 1, existence = 0, extension = 2). F. vesca-like and F. iinumae-like gene counts and TPMs were independently calculated for each octoploid chromosome. This process was performed first on all genes in the ‘Camarosa’ genome to establish the baseline gene retention and expression bias. This process was then repeated using only predicted NLR genes containing an NB-ARC domain, as

RenSeq probe design and validation

A panel of 39,501 of 120mer-length capture probes were designed based on the set of discovered strawberry R-genes from F. xananassa ‘Camarosa’, F. vesca genotype Hawaii 4, F. iinumae genomes, and de novo fruit transcriptomes from F. xananassa ‘Mara des Bois’ and ‘Florida Elyana’. A proprietary algorithm was used to select for capture probes of ideal hybridization thermodynamics and screened for potential off-target capture in the intergenic regions of ‘Camarosa’ and F. vesca (Rapid Genomics LLC, Gainesville FL). Probes were designed to not span exon-exon junction, to facilitate cross-utility for both genomic and cDNA libraries (Figure S3). A minimum baseline of 1x probe coverage was provided across the length of every predicted R-gene coding sequence, and additional probes were designed against conserved R-gene domains in order to promote capture of unknown and divergent R-genes across diverse octoploid accesions. RenSeq capture was performed on genomic libraries from fifteen octoploid disease-resistant cultivars and advanced breeding selections, and F. vesca, based on conditions set by (Jupe et al. 2014), with optimizations provided by Rapid Genomics LLC. Captured libraries were sequenced via 16x multiplexed Illumina HiSeq (2 x 100bp) and mapped to their respective annotated genomic references using CLC Genomic Workbench 11 (CLCBio, Aarhus, Denmark) (Similarity fraction = 0.9, Length fraction = 0.9, Match score = 1, Mismatch cost = 2, Insertion cost = 3, Deletion cost = 3).

Data availability

Supplementary figures, tables, and raw data are available at FigShare. Custom scripts used for performing dN/dS analysis are available at GitHub: https://github.com/Aeyocca/ka_lk_pipe/. Raw short read RNAseq data from fruit transcriptomes are available from the NCBI Short Read Archive under project SRP039356 (http://www.ncbi.nlm.nih.gov/sra/?term=SRP039356). Raw short read RNAseq data from the ‘Camarosa’ gene expression atlas (Sánchez-Sevilla et al., 2017) are available at the European Nucleotide Archive (https://www.ebi.ac.uk/ena) with the study reference PRJEB12420. Results derived from these data are compiled in Table S1. File S1 contains a GFF3 file for annotating the octoploid genome (Edger et al. 2019) with R-genes and R-gene domains. Renseq probe sequences are provided in File S2. IStraw35 marker names, map
positions, and genotype sequences used in eQTL analysis are available in File S3. Supplemental material available at FigShare: https://doi.org/10.25387/g3.8143961.

RESULTS

Octoploid and diploid R-genes

The genomes of octoploid ‘Camarosa’, diploid F. vesca, and diploid F. inumae were analyzed for R-gene signatures. The F. inumae genome was selected to represent the closely-related ‘old world’ diploid ancestors F. inumae, F. nipponica and F. viridis, which each have highly similar but fragmented genomic assemblies.

Putative R-genes were identified based on protein domain and motif analysis, which identified gene models with traditional NLR-type domains, including coiled coil (CC), Toll Interleukin Receptor-like (TIR), LRR domains) were included if there was also supporting evidence of an additional NLR-associated motif. BLAST2GO annotated disease resistance associated genes not meeting these criteria were analyzed manually, leading to the intentional inclusion of many putative Receptor-like Kinase (RLK-type) R-genes in this analysis.

Octoploid F. xananassa ‘Camarosa’ carries 1,962 putative resistance genes (1.82% of all genes) (Table S1), including 975 complete or truncated NLR genes (Table 1). NLR gene content is similar in genic proportion to the 367 complete or truncated NLR genes in F. vesca (1.09% of all genes) and 387 in F. inumae (0.5% of all genes). Traditional NLR domains comprise the majority of domain classes in all predicted resistance gene models in diploid and octoploid strawberry accessions (Figure 1). In many categories, the three genomes show somewhat dissimilar ratios of relative NLR-subtype content (Table 1). These include biases toward TIR-only proteins in F. vesca and CNI-type (CC-containing) NLR genes in F. inumae. Octoploid ‘Camarosa’ is proportionally intermediate for many NLR categories relative to F. vesca and F. inumae. A high proportion of TIR-NBS and TIR-NBS-LRR-containing genes is observed in ‘Camarosa’. However, the overall proportion of TNL-type (TNL-containing) NLR genes is similar when including TIR-only truncations. The Resistance to Powdery Mildew 8 (RPW8) domain, a disease resistance domain associated with broad-spectrum mildew resistance in Arabidopsis, appears frequently in strawberry and is present in 136 (13.9%) of octoploid NLRs (Table 1). Basic trends in NLR-subtype genomic content in ‘Camarosa’ does not more strongly resemble either F. vesca or F. inumae.

In the ‘Camarosa’ genome, 750 of 975 NLR genes contain at least one NB-ARC domain, which is the most characteristic domain of NLR-type R-genes (Table 1). The ratio of ‘Camarosa’ NB-ARC-containing genes to total predicted gene content (1:144) is higher than in F. vesca (1:171) and F. inumae (1:262), possibly indicating diversifying selection of NLR genes in octoploid F. xananassa. A substantial number of atypical domains are present on strawberry R-genes, including malectin-like carbohydrate-binding domains, RNA-binding domains, transcription factor-like WRKY and F-box domains, and several types of protein kinase domains (Figure 1).

Tandem clusters of R-genes were observed in all three of the analyzed strawberry genomes. The phenomena of R-gene expansion through tandem duplication is exemplified in the RPW8-containing R-gene class. Of the seventy-one RPW8-containing R-genes in F. vesca, all but seven reside in one of a few genomic clusters (Figure S4A-B). The major RPW8 cluster observed in F. vesca chromosome 1 is strongly retracted in ‘Camarosa’ (Figure S4C). Similar R-gene hotspots are observed throughout the diploid and octoploid strawberry reference genomes. Genome annotations for all R-genes and NLR domains in ‘Camarosa’ are provided in File S1. Annotations are also available on the JBrowse web-based genome browser at the Genome Database for the Rosaceae (www.rosaceae.org).

Phylogenetic analysis of strawberry NLRs

The conserved NB-ARC domains from ‘Camarosa’, F. vesca, and F. inumae were compared via maximum likelihood analysis to examine evolutionary trends among NLR genes. NB-ARC clusters from all three genomes phylogenetically organized mostly according to their extended R-gene domain structures, with TNLs, CNLs, and RPW8-associated NB-ARC forming clades based on this criteria (Figure 2). Minor NLR subtypes, such as WRKY-associated NLR genes, also sorted into a unique subclade based only on NB-ARC sequence. Multiple distinct clades with identical domain architectures were detected, and in a few cases these subclades are relatively distant from one another.

R-gene transcript accumulation

Raw RNAseq expression data from different tissues of ‘Camarosa’, derived from the octoploid strawberry gene expression atlas (Sánchez-Sevilla et al. 2017), were reassembled based on the homeolog-level
octoploid ‘Camarosa’ genome assembly (Edger et al. 2019). Out of 975 ‘Camarosa’ genome-predicted NLRs, 478 NLRs show evidence of RNAseq expression in any ‘Camarosa’ tissue (>1 Transcript per Million, TPM). A majority of ‘Camarosa’ NLR genes are predominantly expressed in the roots and leaves (Figure 3A-B). Compared to other NLRs, a number of NLRs are predominantly or specifically expressed in the mature receptacle (139 expressed NLRs). Many NLR type R-genes are broadly specific to only one or two tissues. Expressed NLR genes from root, leaf, green and white receptacles show fairly poor overlap. Overall NLR transcript accumulation is correlated with ripening, with strongest expression in the earlier stages and decreasing with maturity in both the receptacle and achene. Complete R-gene expression values for each tissue are provided in Table S1.

Mature receptacle transcriptomes from 61 field-grown individuals of three octoploid populations reveal broadly stable R-gene expression levels (SD 0.09). R-genes comprise 1.8% of the predicted gene models in three octoploid populations reveal broadly stable R-gene expression (Table 2), and 24 R-genes are under regulation of both $cis$ and trans-eQTL status was determined. Of 76 significant R-gene eQTL transcripts, 52 R-genes are regulated via a $cis$-eQTL designation (Table 3). No solely trans-eQTL were discovered among this set of R-genes. The most significant IStraw35 SNP marker name and position for each R-gene transcript is provided with the eQTL phase, minor allele frequency, $p$-value (FDR-adjusted), heritability estimate, expression in parental lines, and BLAST2GO description.

A representative eQTL R-gene (maker-fvb5-2-snap-gene-4.75) is detailed in Figure 4. Analysis by BLAST2GO indicates a probable disease resistance gene homologous to the Arabidopsis thaliana gene At4g27220 with the Uniprot identifier DRL28_ARATH. This gene is hereafter referred to as ‘FaDRL28’. A CC-NBS-NLR structure is predicted for FaDRL28 (Figure 4A). An eQTL was detected for this gene relative to chromosome 5 on the F. vesca genome position (Figure 4B). This eQTL is associated with increased transcription of FaDRL28 from near-zero to above 25 TPM (single-marker ANOVA $p$-value 1.6E-12) (Figure 4C). This eQTL is superficially analogous to the physical position of octoploid FaDRL28 on chromosome 5, homeolog 2 (Figure 4D). The significant markers are not included in the ‘Holiday’ × ‘Korona’ octoploid genetic map, impeding a recombination-based subgenomic genetic association (van Dijk et al. 2014). However, the associated marker physical sequences match fairly uniquely to the ‘Camarosa’ chromosome 5-2 subgenomic locus, confirming a $cis$-eQTL designation (Figure 4E). One eQTL marker locates inside the FaDRL28 coding sequence.

**Evolutionary pressure on F. × ananassa R-genes**

Elevated median $dN/dS$ ratios were observed across all 1,962 predicted $F. \times ananassa$ R-genes (0.47) compared to non R-genes (0.35)
Fewer R-genes exhibited extremely low $dN/dS$ ratios, indicating that high degrees of R-gene conservation are less common. However, a similar rate of hypervariable genes ($dN > dS$) was observed between R-genes and non-R-genes. Median $dN/dS$ values for RPW8-type R-genes (0.62) are significantly higher than for general R-genes (0.47) as confirmed by one-way ANOVA [$F (1, 1911) = 10.6$, $P < 0.0012$] (Table S1).

R-gene $dN/dS$ values were compared against transcript accumulation across various strawberry tissues and receptacle stages. R-genes with low transcript accumulation across all tissues were correlated with higher $dN/dS$ ratios (Pearson’s $r = -0.69$, $P < 0.0001$) (Figure S5). In other words, R-genes with poor evidence of expression also have higher ratios of non-synonymous mutation capable of altering amino acid sequences and affecting protein function.

**Subgenome dominance in octoploid strawberry**

Polyploidization is associated with rapid genome remodeling events to establish a new homeostasis, including selective gene loss and methylation. While R-gene expansiveness is often considered evolutionarily favorable, genes that are stoichiometrically or dosage sensitive are more commonly retained in duplicate after polyploidization (Edger and Pires 2009; Birchler and Veitia 2012; Edger et al. 2017a). The ‘Camarosa’ octoploid genome, in comparison with the genomes from its diploid...
F. vesca-like and F. iinumae-like ancestors, has provided an ideal platform to study the general biological phenomena of post-hybridization genome remodeling and subgenome dominance (Edger et al. 2019).

To gauge R-gene post-hybridization retention specifically, a gene-focused baseline assessment of subgenome dominance in the 'Camarosa' octoploid genome was necessary. Putative gene ancestry was predicted

Figure 3 RNAseq-based Detection of Octoploid NLR Transcripts. A. Tissue-based heatmap of transcript accumulation (TPM) of NLRs in 'Camarosa'. B. Total and proportional number of expressed NLRs (TPM >1) (left y-axis) with averaged 'Camarosa' transcript abundance (right y-axis), per tissue and in all tissues (global). C. Generalized R-gene mature-fruit expression across three segregating populations (n = 61).
| R-gene Name | eQTL phase | iStraw35 AX-MAF | p-value (FDR) | $h^2$ estimate | Mara-Elya-Radi-Festival | Winter dawn | Description |
|-------------|------------|-----------------|---------------|----------------|------------------------|------------|-------------|
| augustus_masked-fvb1-2-processed-gene-27.2 | cis | 89817565 0.43 | 0.0418 | 30.4% | 36.6 | 113.3 | 32.2 | 21.3 | 34.3 | uncharacterized protein LOC101293711 |
| augustus_masked-fvb1-4-processed-gene-7.11 | cis | 166502627 0.16 | 0.0150 | 43.4% | 2.3 | 0.6 | 4.0 | 3.8 | 1.6 | PLT5_ARATH Sugar-proton symporter PLT5 |
| augustus_masked-fvb2-4-processed-gene-105.5 | cis | 166503861 0.29 | 0.0114 | 80.9% | 0.0 | 0.0 | 2.8 | 1.8 | 1.3 | TMVRN_NICGU TMV resistance N |
| augustus_masked-fvb3-1-processed-gene-107.3 | cis | 166511589 0.50 | 0.0294 | 41.8% | 0.0 | 0.5 | 0.7 | 0.8 | 0.3 | DRL30_ARATH Probable disease resistance AT5g04720 |
| augustus_masked-fvb3-3-processed-gene-283.7 | cis | 89786873 0.38 | 0.0009 | 30.4% | 0.8 | 0.4 | 0.0 | 2.6 | 0.4 | DRL1_ARATH Probable disease resistance At1g12280 |
| augustus_masked-fvb3-3-processed-gene-38.8 | cis | 166513199 0.23 | 0.0032 | 81.7% | 6.0 | 3.0 | 5.5 | 0.8 | 4.9 | Y4294_ARATH LRR receptor-like serine threonine-kinase |
| augustus_masked-fvb3-4-processed-gene-19.4 | cis | 166504873 0.28 | 0.0014 | 100% | 0.1 | 4.8 | 5.6 | 8.0 | 0.1 | DR100_ARATH DNA damage-repair toleration DRT100 |
| augustus_masked-fvb4-1-processed-gene-166.2 | cis | 166505902 0.43 | 0.0083 | 91.5% | 0.6 | 0.4 | 2.5 | 1.4 | 1.8 | MKKA_DICDI Mitogen-activated kinase |
| augustus_masked-fvb4-2-processed-gene-107.10 | cis | 166527457 0.18 | 0.0202 | 100% | 0.0 | 0.0 | 1.2 | 1.0 | 0.1 | RGA1_SOLBU disease resistance RGA1 RGA3-bb |
| augustus_masked-fvb4-2-processed-gene-258.8 | cis | 166505336 0.35 | 0.0015 | 25.1% | 4.6 | 0.0 | 1.4 | 1.3 | 0.1 | LRX2_ARATH Leucine-rich repeat extensin 2 |
| augustus_masked-fvb5-1-processed-gene-238.7 | cis | 123365994 0.09 | 0.0012 | 74.1% | 0.8 | 0.8 | 0.1 | 1.2 | 1.0 | HSL1_ARATH Receptor kinase HSL1 HAESA-LIKE1 |
| augustus_masked-fvb5-1-processed-gene-71.8 | cis | 166524323 0.48 | 0.0080 | 35.5% | 1.7 | 0.0 | 10.3 | 4.4 | 3.0 | TIR_ARATH Toll interleukin-1 receptor |
| augustus_masked-fvb5-3-processed-gene-135.4 | cis | 166523635 0.26 | 0.0124 | 63.7% | 0.5 | 0.2 | 0.1 | 0.0 | 0.0 | MKKA_DICDI Mitogen-activated kinase |
| augustus_masked-fvb5-4-processed-gene-18.1 | cis | 166518037 0.35 | 0.0110 | 26.6% | 0.8 | 1.4 | 0.7 | 2.5 | 0.8 | RGA1_SOBU disease resistance RGA1 RGA3-bb |
| maker-fvb1-4-augustus-gene-30.48 | cis | 123365069 0.44 | 0.0007 | 64.9% | 1.4 | 0.1 | 2.5 | 2.1 | 1.8 | LRX2_ARATH Leucine-rich repeat extensin 2 |
| maker-fvb2-1-augustus-gene-182.42 | cis | 89877559 0.46 | 0.0225 | 80.9% | 2.1 | 3.8 | 2.2 | 3.2 | 0.8 | MKKA_DICDI Mitogen-activated kinase |
| maker-fvb2-1-snap-gene-111.27 | cis | 166503168 0.29 | 0.0002 | 100% | 0.5 | 0.0 | 0.0 | 2.0 | 1.2 | RGA1_SOBU disease resistance RGA1 RGA3-bb |

(continued)
| R-gene Name                                      | eQTL phase | iStraw35 AX-MAF | p-value (FDR) | h² estimate | Mara | Elya-na | Radi-ance | Festi-val | Winter dawn | Description                                                                 |
|-------------------------------------------------|------------|-----------------|---------------|-------------|------|---------|-----------|-----------|-------------|-----------------------------------------------------------------------------|
| maker-fvb7-1-snap-gene-223.45                   | cis        | 0.45            | 0.0077        | 70.3%       | 2.4  | 0.1     | 0.1       | 0.0       | 1.7         | LRX2_ARATH Leucine-rich repeat extensin 2 2 LRR                            |
| maker-fvb7-4-snap-gene-48.49                    | cis        | 0.34            | 0.0357        | 75.1%       | 0.3  | 0.4     | 0.1       | 0.0       | 0.2         | HSL1_ARATH Receptor kinase HSL1-HAESA-LIKE1                                |
| maker-fvb7-4-snap-gene-59.59                    | cis        | 0.50            | 0.0064        | 4.7%        | 0.9  | 0.4     | 1.8       | 0.1       | 1.6         | TIR_ARATH Toll interleukin-1 receptor                                       |
| maker-fvb7-4-snap-gene-59.63                    | cis        | 0.50            | 0.0092        | 44.8%       | 0.1  | 0.0     | 0.6       | 0.0       | 0.4         | MKKA_DICDI Mitogen-activated kinase                                         |
| maker-fvb7-4-snap-gene-69.51                    | cis        | 0.17            | 0.0002        | 82.6%       | 2.3  | 1.8     | 2.1       | 3.6       | 2.2         | RGA1_SOLBU disease resistance RGA1 RGA3-bb                                 |
| snap_masked-fvb1-2-processed-gene-79.33         | cis        | 0.08            | 0.0068        | 98.3%       | 0.0  | 0.0     | 0.0       | 0.0       | 2.0         | LRX2_ARATH Leucine-rich repeat extensin 2 2 LRR                            |
| snap_masked-fvb2-1-processed-gene-107.14        | cis        | 0.22            | 0.0011        | 75.1%       | 2.1  | 2.6     | 8.5       | 4.0       | 3.4         | HSL1_ARATH Receptor kinase HSL1-HAESA-LIKE1                                |
| snap_masked-fvb3-2-processed-gene-11.25         | cis        | 0.37            | 0.0084        | 11.4%       | 0.2  | 0.2     | 1.1       | 1.5       | 0.7         | TIR_ARATH Toll interleukin-1 receptor                                       |
| snap_masked-fvb3-3-processed-gene-288.15        | cis        | 0.45            | 0.0095        | 98.9%       | 0.9  | 0.5     | 2.0       | 1.4       | 1.3         | MKKA_DICDI Mitogen-activated kinase kinase kinase A RGA1                   |
| snap_masked-fvb6-1-processed-gene-37.31         | cis        | 0.34            | 0.0411        | 24.2%       | 13.3 | 19.2    | 0.9       | 0.5       | 0.0         | RGA1_SOLBU disease resistance RGA1                                          |
| snap_masked-fvb7-2-processed-gene-254.35        | cis        | 0.33            | 0.0011        | 36.5%       | 0.0  | 0.0     | 0.0       | 0.0       | 0.0         | LRX2_ARATH Leucine-rich repeat extensin 2 2 LRR                            |
| maker-fvb7-1-snap-gene-273.51                   | cis        | 0.43            | 0.0319        | 35.7%       | 0.0  | 0.0     | 0.0       | 0.4       | 0.0         | Y3475_ARATH LRR receptor-like serine threonine- kinase                      |
| maker-fvb3-4-augustus-gene-265.40               | cis        | 0.13            | 0.0007        | 54.7%       | 4.4  | 5.3     | 2.6       | 18.8      | 5.2         | GLO5_ARATH Peroxisomal(S)-2-hydroxy-acid oxidase GLO5                       |
| maker-fvb5-1-snap-gene-191.37                   | cis        | 0.13            | 0.0000        | 36.5%       | 0.0  | 0.0     | 0.0       | 0.0       | 0.0         | RPM1_ARATH Disease resistance RPM1                                         |
| maker-fvb5-2-augustus-gene-59.20                | cis        | 0.46            | 0.0036        | 46.0%       | 0.0  | 0.3     | 4.0       | 12.0      | 0.8         | probable LRR receptor-like serine threonine- kinase At5g48740             |
| maker-fvb5-2-augustus-gene-61.13                | cis        | 0.46            | 0.0014        | 71.7%       | 0.0  | 0.3     | 2.2       | 1.6       | 0.8         | P2B10_ARATH F-box PP2-B10 PHLOEM PROTEIN 2-LIKE                              |
| maker-fvb5-2-augustus-gene-63.17                | cis        | 0.46            | 0.0072        | 16.5%       | 0.0  | 0.0     | 3.5       | 3.7       | 1.1         | P2B11_ARATH F-box PP2-B11 PHLOEM PROTEIN 2-LIKE                              |
| maker-fvb5-2-snap-gene-4.75                     | cis        | 0.32            | 0.0000        | 29.9%       | 2.1  | 9.3     | 37.2      | 43.0      | 1.4         | DRL28_ARATH Probable disease resistance At4g27220                          |
| maker-fvb5-2-snap-gene-61.17                    | cis        | 0.46            | 0.0291        | 78.4%       | 0.0  | 0.2     | 1.3       | 1.4       | 0.3         | P2B11_ARATH F-box PP2-B11 PHLOEM PROTEIN 2-LIKE                              |
| maker-fvb5-3-augustus-gene-135.25               | cis        | 0.37            | 0.0040        | 11.5%       | 0.0  | 0.0     | 1.6       | 0.3       | 1.3         | RPM1_ARATH Disease resistance RPM1                                         |
| maker-fvb4-3-snap-gene-155.68                   | cis        | 0.21            | 0.0016        | 100%        | 0.1  | 0.2     | 0.6       | 0.1       | 0.5         | TMVRN_NICGU TMV resistance N                                               |
| maker-fvb5-3-snap-gene-221.67                   | cis        | 0.11            | 0.0045        | 94.4%       | 0.5  | 0.2     | 0.1       | 1.6       | 0.1         | TMVRN_NICGU TMV resistance N                                               |
| maker-fvb5-3-snap-gene-254.50                   | cis        | 0.21            | 0.0155        | 51.8%       | 0.1  | 0.1     | 0.4       | 0.5       | 0.1         | TMVRN_NICGU TMV resistance N                                               |
| maker-fvb5-4-snap-gene-125.42                   | cis        | 0.18            | 0.0078        | 93.2%       | 0.6  | 0.0     | 2.0       | 2.3       | 1.4         | RGA3_SOLBU RGA3 Blight resistance B149                                      |
| eQTL | Phase | Profile | P-value | MAF | h² | Description |
|------|-------|---------|---------|-----|----|-------------|
| cis  | 0.17  | 0.0000  | 76.3%   | 0.0 | 0  | Gv958ra56   |
| cis  | 0.47  | 0.0167  | 44.9%   | 0.0 | 0  | Gv958ra56   |
| cis  | 0.04  | 0.0139  | 95.1%   | 0.0 | 0  | Gv958ra56   |
| cis  | 0.28  | 0.0084  | 38.4%   | 0.0 | 0  | Gv958ra56   |
| cis  | 0.12  | 0.0340  | 42.6%   | 0.0 | 0  | Gv958ra56   |

Table 2, continued

Genetic association results for 61 transcriptomes are shown, detailing cis genetic factors controlling differentially expressed R-genes. The most significant marker name, minor allele frequency, FDR-adjusted p-value, and p-value for association with the expression of each R-gene are shown. The most significant marker name, minor allele frequency, FDR-adjusted p-value, and p-value for association with the expression of each R-gene are shown.

NLR-gene subgenome dominance

Significant gene retention bias toward NLR genes that are more F. vesca-like is observed in ‘Camarosa’ gene models (Figure 7). Of the 750 predicted NLR-gene models containing an NB-ARC domain, 69.3% more closely resemble a F. vesca gene rather than an F. iinumae gene (Figure 7A). This is somewhat higher than the baseline retention bias toward F. vesca-like genes in octoploid (63.8%) from this analysis. In every chromosome group, the F. vesca-like homeologous chromosomes (yellow highlight) retained the greatest number of NB-ARC domain-containing NLRs. Overall, 538 NB-ARC domain-containing-NLRs demonstrate the highest sequence identity with an F. vesca gene, 210 show highest sequence identity with an F. iinumae gene, and 2 (an RPW8-only gene, and an LRR_8-only gene) are without significant matches to either diploid genome. While F. vesca-like genes contribute the most to total NLR expression across 61 mature fruit transcriptomes (70.5% of transcripts), this is proportional to F. vesca-like NLR genome content (71.3%) and is similar in magnitude to general F. vesca expression bias (73.7%) (Figure 7B). In other words, F. vesca-like NLR genes are retained in the octoploid genome somewhat above the baseline bias, but do not experience the additional expression magnitude bias that is a generic feature of F. vesca-like transcripts. A greater breadth of genome-predicted NLR genes had evidence of expression across mature fruit transcriptomes (246 genes, n = 61) (Figure 7B) than in the ‘Camarosa’ mature fruit transcriptome alone (139 genes, n = 5) (Figure 3B) however averaged transcriptional magnitude was similar (Figure 3C).

RenSeq for strawberry resistance genes

A panel of sequence capture probes was designed based on putative R-gene sequences discovered in the genomes of F. ×ananassa ‘Camarosa’, F. vesca genotype Hawaii 4, F. iinumae, and de novo fruit transcriptomes from
### Table 3 Cis and trans eQTL pertaining to fruit-expressed genes in *F. ×ananassa*

| R-gene Name                  | eQTL phase | eQTL | lStraw35 AX-MAF | p-value (FDR) | h² estimate | Mara  | Elya-na | Radi-ance | Festi-val | Winter dawn | Description                                                                 |
|------------------------------|------------|------|----------------|---------------|-------------|-------|---------|-----------|-----------|-------------|-----------------------------------------------------------------------------|
| maker-fvb4-2-snap-gene-5.61  | cis        | 166505436 | 0.38           | 0.0214        | 34.3%       | 1.5   | 0.9     | 1.6       | 3.0       | 2.3         | PSKR1_ARATH LRR receptor kinase 1 serine-threonine kinase                  |
| maker-fvb5-1-augustus-gene-139.47 | cis | 123361503 | 0.38            | 0.0214        | 77.5%       | 0.3   | 0.4     | 5.8       | 1.5       | 1.0         | PIRL5_ORYSJ Plant intracellular Ras-group-related LRR 5                     |
| maker-fvb5-1-augustus-gene-263.34 | trans | 123357041 | 0.18            | 0.0415        | 14.6%       | 8.6   | 4.0     | 3.3       | 3.7       | 3.2         | R13L1_ARATH disease resistance RPP13 1                                     |
| maker-fvb5-2-snap-gene-213.46 | cis        | 89817904  | 0.10            | 0.006         | 79.8%       | 1.7   | 1.6     | 3.3       | 0.0       | 1.4         | DRL21_ARATH disease resistance At3g14460                                   |
| maker-fvb5-2-snap-gene-46.64  | trans      | 166508680 | 0.48            | 0.0192        | 67.1%       | 3.4   | 0.0     | 1.3       | 1.0       | 1.4         | GSO1_ARATH LRR receptor-like serine threonine- kinase GSO1                  |
| maker-fvb5-3-snap-gene-244.48 | trans      | 166522785 | 0.48            | 0.0192        | 39.5%       | 0.4   | 0.5     | 4.7       | 5.6       | 2.8         | Y4265_ARATH Probable LRR receptor-like serine threonine- kinase             |
| maker-fvb5-4-snap-gene-110.38 | trans      | 166524268 | 0.48            | 0.0046        | 100%        | 1.0   | 1.3     | 0.4       | 1.6       | 0.3         | PIRL5_ORYSJ Plant intracellular Ras-group-related LRR 5                     |
| maker-fvb6-1-augustus-gene-27.58 | trans | 123363787 | 0.38            | 0.0028        | 91.2%       | 0.3   | 0.0     | 0.2       | 0.5       | 0.0         | DRL42_ARATH Probable disease resistance At5g66900                          |
| maker-fvb6-3-snap-gene-412.71 | trans      | 123361427 | 0.27            | 0.0457        | 74.4%       | 1.8   | 1.1     | 1.4       | 2.1       | 0.9         | TIR_ARATH Toll interleukin-1 receptor                                        |
| augustus_masked-fvb6-2-processed-gene-268.10 | trans | 166525307 | 0.27            | 0.0457        | 91.2%       | 0.8   | 0.8     | 0.3       | 1.7       | 1.2         | HSL1_ARATH Receptor kinase HSL1 HAESA-LIKE1                                |
| augustus_masked-fvb6-3-processed-gene-176.9 | trans | 123356218 | 0.48            | 0.0021        | 65.6%       | 0.4   | 0.0     | 7.9       | 20.0      | 0.1         | ADT3_ARATH ADP,ATP carrier mitochondrial ADP ATP                            |
| augustus_masked-fvb7-2-processed-gene-63.0 | trans | 12335691  | 0.33            | 0.0005        | 27.7%       | 0.5   | 0.1     | 1.3       | 0.1       | 0.0         | TMVRN_NICGU TMV resistance N                                               |
| augustus_masked-fvb7-3-processed-gene-243.6 | trans | 166517344 | 0.38            | 0.0204        | 66.5%       | 47.5  | 14.1    | 0.1       | 8.1       | 0.7         | DR100_ARATH DNA damage-repair toleration DRT100                            |
| maker-fvb1-2-snap-gene-63.40  | trans      | 166510935 | 0.17            | 0.0015        | 86.2%       | 0.5   | 0.1     | 1.3       | 0.1       | 0.0         | TMVRN_NICGU TMV resistance N                                               |
| maker-fvb1-2-snap-gene-96.40  | trans      | 166525497 | 0.16            | 0.0015        | 67.5%       | 0.0   | 0.0     | 1.0       | 0.0       | 0.4         | DRL42_ARATH Probable disease resistance                                     |
| maker-fvb1-4-snap-gene-66.72  | trans      | 166517617 | 0.18            | 0.0020        | 68.4%       | 3.9   | 0.2     | 12.0      | 12.0      | 7.4         | U496_ARATH UPF0496 At2g18630                                               |
| maker-fvb1-4-snap-gene-76.45  | trans      | 166516240 | 0.18            | 0.0093        | 100%        | 0.7   | 0.2     | 0.0       | 0.0       | 0.0         | DRL43_ARATH Probable disease resistance                                     |

(continued)
## Table 3, continued

| Gene Name                      | Phase | Winter  | p-value (FDR) | MAF   | Winter RMA | NCI GU TMV resistance | Description                        |
|--------------------------------|-------|---------|---------------|-------|------------|------------------------|------------------------------------|
| maker-fvb7-2-augustus-gene-136.47 | cis   | 0.43    | 0.0116        | 79.9% | 0.43       | 0.0116                 | 79.9%                              |
| trans                          |       | 0.46    | 0.0371        | 45.7% | 0.46       | 0.0371                 | 79.9%                              |
| maker-fvb7-2-augustus-gene-147.47 | cis   | 0.45    | 0.0003        | 11.1% | 0.2        | 0.0003                 | 45.2%                              |
| trans                          |       | 0.45    | 0.0011        | 20.4% | 0.3        | 0.0011                 | 45.2%                              |
| maker-fvb7-2-augustus-gene-163.44 | cis   | 0.39    | 0.0011        | 52.3% | 0.7        | 0.0011                 | 52.3%                              |
| trans                          |       | 0.42    | 0.0011        | 52.3% | 0.7        | 0.0011                 | 52.3%                              |
| maker-fvb7-2-augustus-gene-65.21 | cis   | 0.21    | 0.0031        | 69.2% | 0.3        | 0.0031                 | 69.2%                              |
| trans                          |       | 0.21    | 0.0031        | 69.2% | 0.3        | 0.0031                 | 69.2%                              |
| maker-fvb7-2-augustus-gene-161.50 | cis   | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| trans                          |       | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| maker-fvb7-2-augustus-gene-66.21 | cis   | 0.42    | 0.0031        | 69.2% | 0.3        | 0.0031                 | 69.2%                              |
| trans                          |       | 0.42    | 0.0031        | 69.2% | 0.3        | 0.0031                 | 69.2%                              |
| maker-fvb7-2-augustus-gene-161.50 | cis   | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| trans                          |       | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| maker-fvb7-2-augustus-gene-35.21 | cis   | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| trans                          |       | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| maker-fvb7-2-augustus-gene-161.50 | cis   | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| trans                          |       | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| maker-fvb7-2-augustus-gene-35.21 | cis   | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| trans                          |       | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| maker-fvb7-2-augustus-gene-161.50 | cis   | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| trans                          |       | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| maker-fvb7-2-augustus-gene-35.21 | cis   | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| trans                          |       | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| maker-fvb7-2-augustus-gene-161.50 | cis   | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |
| trans                          |       | 0.42    | 0.0055        | 69.2% | 0.3        | 0.0055                 | 69.2%                              |

**DISCUSSION**

These results provide a characterization of the R-gene complement of cultivated octoploid strawberry and the relationship to the extant diploid relatives, *F. vesca* and *F. inumae*. Commercial strawberry is hypothesized to contain a single *F. vesca*-like subgenome, and three highly-similar ‘old world’ subgenomes which are likely derived from *F. inumae*, *F. viridis*, and *F. nipponica* (Edger et al. 2019). Polyploidization is associated with massive genome remodeling events including gene loss (Edger et al. 2018; Edger et al. 2017a). Polyploidization involves massive genome remodeling events including gene loss (Edger et al. 2018; Edger et al. 2017a). Linkage-map comparisons in octoploid and diploid strawberry have uncovered extensive unidirectional homeologous exchanges which have broadly converted the three ‘old world’ *F. inumae*-like subgenomes to be more *F. vesca*-like (Tennesen et al. 2014). This finding explains the difficulties of clear ancestral delineation of strawberry homeologs (Vining et al. 2017). Recent analysis of the octoploid genome reveals that biased homologous exchanges have converted other subgenomes to be more like the dominant *F. vesca*-like subgenome (Edger et al. 2019). The present gene-level homology and expression analysis shows the majority of *F. vesca*-like dominance is derived from *F. vesca*-like genes residing on alternate subgenomes. For NLR genes in particular, the bias toward *F. vesca*-like genomic retention was more pronounced. Unlike general octoploid genes, expression of *F. vesca*-like and *F. inumae*-like NLRs is proportional to their genomic representation. This finding provides potential insight into the practical drivers of subgenome conversion. Consolidation of redundant genes and maintenance of stoichiometrically sensitive genes has been hypothesized as a driver for gene retention bias (Edger and Pires 2009; Birchler and Veitia 2012; Edger et al. 2017a). NLRs are involved in consequential and sensitive protein-level interactions, including signaling functions requiring homo- and hetero-dimerizations (El Kasmi and Nishimura 2016). Avoidance of dysfunctional NLR molecular interactions may have contributed to the observed biases in NLR retention and expression, post-polyploidization.

Multiple distinct NLR clades with identical domain architectures were detected, likely distinguishing intra-subgenome homologs from different subgenomes. These likely reflect broad ancestral sequence divergences prior to hybridization. Comparison of R-genes in octoploid and diploid strawberry reveals enrichment of different subtypes. The ‘Camarosa’ genome shows a large increase in complete TNL-type R-genes and a concomitant decrease in truncated TIR-only genes, relative to its diploid ancestral relatives. The *F. inumae* genome shows a...
considerably larger amount of CNL-types. TNLs have been nearly eliminated from most monocot genomes in bias toward CNL-types (Nepal et al. 2017). The reasons for emerging divisions in TNL/CNL content in plant genomes remains unclear. In hybrid F. xananassa, it is possible that relatively high number of complete TNL genes is a result of higher rate of retention post-polyploidization in this category. Many of the non-classical domains discovered in 'Camarosa' R-genes have also been found and characterized in the R-genes of other species. These include an LRR/Malectin-like RLK protein, which mediates powdery mildew resistance in barley and wheat (Rajaraman et al. 2016). Atypical R-gene domains physically associated with NB-ARCs have been implicated in a variety of active disease resistance functions, including signal transduction and defense gene activation, and serving as decoy endogenous sequences to bait pathogen effectors into direct interaction and detection (Khan et al. 2016).

A large proportion of strawberry NLR genes from octoploid and diploid genomes are associated with RPW8 domains. It has been suggested that the RPW8 domain emerged with the earliest land plants and subsequently merged with NLR genes, however their prevalence across plant genomes varies widely (Zhong and Cheng 2016). The RPW8 domain appears to have been completely lost in monocots, and is rare in many other species. R-gene genomic studies frequently neglect to assess the presence of NLR-associated RPW8 domains. Two NBS-RPW8 proteins conferring mildew-resistance have been described in the Arabidopsis thaliana genome (Xiao et al. 2001) that retained their function when expressed in grape (Hu et al. 2018). The AtRPW8.2 gene was recently shown to induce the expression of defense-related genes when expressed in strawberry leaves (Cui et al. 2017). This R-gene subtype has apparently expanded in strawberry, possibly due to unusually high mildew disease pressure exerted on strawberry species and intense selection for resistance. However, R-gene domain content is not reliably predictive of resistance specificity, and close R-gene paralogs are known to confer resistance to pathogens in entirely different kingdoms (Wen et al. 2015). Interestingly, the strawberry RPW8 domain is frequently found in association with NB-ARC-containing genes but never with TIRs. RPW8-containing genes are similar to general R-genes in terms of their frequency of expression, eQTL discovery rate, and putative orthology with F. vesca (Table S1). However, a significant and large difference was noted in terms of dN/dS ratio, indicating a higher degree of protein-level variability for this subclass. The purpose of RPW8 gene diversification and expansion in strawberry remains an interesting open question.

Figure 4 Example cis-eQTL of a Fruit-Expressed Strawberry R-gene. A. Domain analysis of the ‘Camarosa’ putative resistance gene FaDRL28, a CC-NBS-LRR R-gene. Gray lines delineate exon-exon borders in the predicted mature transcript. B. Octoploid fruit expression of FaDRL28 associates with a single locus on chromosome 5. C. A single dose of an “A” allele increases mean transcription of FaDRL28 to above 25 TPM (“AB” genotype) from a mean approximately 1 TPM (“BB” genotype) (single-marker ANOVA p-value 1.6E-12). D. The Fvb5-2 subgenomic location of FaDRL28 in the octoploid ‘Camarosa’ genome is indicated (purple vertical line). E. Seven equally-significant eQTL markers (p-value 2.73E-05, post-FDR adjustment) show close subgenomic co-localization with FaDRL28, including one marker within the FaDRL28 coding sequence (AX-123366087).
Octoploid NLR transcript accumulation is low throughout the strawberry plant, but is particularly low in the mature receptacle. This is an unexpected result due to the many pathogens targeting this susceptible organ. It is possible that only certain R-genes are highly upregulated in the response to pathogen attack. Another possibility is that resistance based on the hypersensitive response may be less effective at mature stages, where cell wall disruption has already initiated with ripening and the intercellular environment is conducive to pathogen growth. Transcriptional response to *Botrytis cinerea* infection in the mature octoploid receptacle led to differential expression of over 1,500 genes, including secondary metabolism and pathogenesis-related (PR) genes, but only 15 NLR genes (Xiong et al. 2018). In the present study,

Figure 5 Evolutionary Pressures on *F. × ananassa* R-genes. The median dN/dS ratio for R-genes (0.47) is higher than for non R-genes (0.35). Density curves for *F. × ananassa* R-genes (blue) and non R-genes (red) are calculated based on comparison to the closest ancestral diploid homolog from *F. vesca*.

Figure 6 General Retention and Expression Bias in Octoploid Strawberry. 'Camarosa' gene models from every chromosome are categorized as either more *F. vesca*-like, more *F. iinumae*-like, or neither. Red-green color scale indicates low-to-high gene content, respectively. Yellow highlight indicates the most *F. vesca*-like homeologous chromosome. A. Gene content per homeologous chromosome, by putative ancestral gene similarity. B. Relative transcript accumulation of all genes in the fruit, by putative ancestral similarity.
elevated NLR transcription in the green, white, and turning stages suggest NLR-based resistance may be more prevalent at these earlier developmental stages. The highest levels of NLR expression were seen in the roots and leaves, indicating this mode of resistance may be more common in these tissues. Root-dominant expression of NLRs is common in many but not all plant species (Munch et al. 2018). Strawberry NLR expression overlaps poorly between tissues, supporting the concept that NLRs are optimized for each tissue (Munch et al. 2018). Based on these patterns of expression, resistance to soil-borne pathogens via NLR-genes may be more common in strawberry. It would be interesting to examine the patterns of tissue-specific expression of R-genes against different strawberry pathogens, particularly the common soil-borne pathogens causing strawberry verticillium wilt (Verticillium dahlia), charcol rot (Macrophomina phaseolina), and Fusarium wilt (Fusarium oxysporum f.sp. fragariae) (Zurn et al. 2018). Across 61 mature fruit transcriptomes, a greater number of putatively expressed NLR genes were found compared to 'Camarosa' mature fruit alone. This signifies possible genetic/environmental variabilities which can be measured via eQTL analysis.

The genetics of differential fruit expression of all R-genes in strawberry cultivars was examined via eQTL analysis. In many cases, the identified genetic markers described presence/absence of R-gene expression. The identified eQTLs were often due to a cis variant at a single detectable locus, very close to the physical position of the gene itself. This is suggestive of a mutation in a cis-regulatory element, such as the gene promoter or 5’-UTR, or a genomic presence/absence structural variation. Such presence/absence variation affects nearly 20% of genes in the Brassica oleracea pangenome and is a major contributor of agronomic trait diversity (Gólich et al. 2016). As these strawberry R-gene eQTL are derived from crosses of cultivars with differing ranges of pathogen susceptibility, these eQTL genes represent strong candidates for functional disease resistance and potential genetic improvement.

These disclosed R-gene eQTL marker sequences may be cross-referenced with existing disease-resistance QTL to potentially identify causal R-genes. As categories of R-genes are expressed at very low levels unless induced by pathogens (Lai and Eulgem 2017), the genotype × pathogen interaction may have lowered confidence values or introduced possible type II errors in eQTL detection. However, the reproducibility of cis-eQTL tends to be particularly high in related populations (Peirce et al. 2006). Additional replicates and infected/non-infected challenge conditions will likely reveal additional eQTL associations and greatly improve the confidence of heritability estimates, and may be used to validate pathogen-induced R-gene candidates.

Fig. 7 NLR-gene Retention and Expression Bias in Octoploid Strawberry. 'Camarosa' NB-ARC domain-containing NLR models from every chromosome are categorized as either more F. vesca-like, more F. ilinumae-like, or neither. Red-green color scale indicates low-to-high gene content, respectively. Yellow highlight indicates the most F. vesca-like homoeologous chromosome. A. NLR-gene content per homoeologous chromosome, by putative ancestral gene similarity. B. Relative transcript accumulation of NLR-genes in the fruit, by putative ancestral similarity.
pressure to maintain these sequences, offering evidence of maintained function despite low expression. The results of this combinatorial analysis can be used to help identify novel sources of R-gene-based resistance which may be otherwise difficult to detect. It should be noted that this analysis is performed in the context of a single cultivar, which has undergone several centuries of artificial selection. It is possible that wild octoploid species may reveal different and more natural patterns of disease-resistance selection. More sequenced accessions from geographically diverse wild and cultivated germplasm are needed. Further analysis on the octoploid pangenome will reveal more detailed selection patterns, and more importantly, reveal recent selection sweep events which may have occurred in certain R-gene groups.

Many R-genes were discovered clustered in the genomes of both octoploid and diploid strawberry, highlighting the challenges of resolving individual R-genes via association mapping and positional cloning. The difficulty of isolating functional R-genes from strawberry disease resistance QTL was the principle motivator of this analysis. A thorough identification of R-genes in the octoploid genome is necessary for future genomics and genetics analysis in strawberry disease-resistance breeding programs. Additionally, this information is prerequisite for creating a RenSeq probe panel, to facilitate targeted R-gene sequencing in breeding programs.

A novel strawberry RenSeq capture-probe library was developed based on the R-gene sequences identified from genomic and transcriptomic resources. This 39,501-probe panel was experimentally validated using octoploid and diploid genomes and resulted in an average ~20× R-gene resequencing depth per genomic library, using only multiplexed short reads. RenSeq assembly in 'Camarosa' and the F. vesca genotype Hawaii 4 resulted in significant coverage of R-genes. Despite having the highest total resequencing depth, the capture efficiency in F. vesca (R-gene reads over total reads) was somewhat lower. It is possible that this represents the saturation of capture probes in a smaller genome. It is also possible that probes designed from non-F. vesca-like octoploid sequences bound spuriously in the F. vesca genome. Similar rates of perfect sequence matching along the entire read in 'Camarosa' and F. vesca (66.24% and 69.68%, respectively) indicates that theoretical octoploid reference sequence errors are not likely promoting RenSeq assembly error in 'Camarosa'. However, 14.4% of mapped 'Camarosa' R-gene reads have an equally valid alternative R-gene mapping locus, compared with just 4.83% in F. vesca. This difference indicates that homeologous sequence redundancy is an appreciable issue for mapping short-reads in polyploids, even with an isogenic (but not haplotype-specific) mapping reference. Longer sequencing read-lengths, spanning less well-conserved non-coding sequences, will assist in de novo resolution of similar loci in octoploid strawberry. Combining RenSeq with long-read sequencing technologies will allow for improved de novo assembly of R-gene loci, and will greatly facilitate causal mutation detection within disease resistance QTL in octoploid strawberry.

ACKNOWLEDGMENTS
We acknowledge Aristotle Koukoulidis for ncoils prediction and annotation, Matthew Robinson for assistance with data normalization.
