New Phytologist Supporting Information

Article title: A multi-omics framework reveals strawberry flavor genes and their regulatory elements

Authors: Zhen Fan, Denise M. Tieman, Steven J. Knapp, Philipp Zerbe, Randi Famula, Christopher R. Barbey, Kevin M. Folta, Rodrigo R. Amadeu, Manbo Lee, Youngjae Oh, Seonghee Lee, Vance M. Whitaker*

Article acceptance date: 21 July 2022

Note S1

eQTL mapping

The eQTL were also mapped with the 50K Fana SNP array (Hardigan *et al.*, 2020) for 185 individuals. Genomic DNA was isolated from either fruit or leaf tissue. CEL files containing sample fluorescence data were imported into the Affymetrix Axiom Analysis Suite (v1.1.1.66) and run in “polyploid” mode to predict marker genotype clusters (Hardigan *et al.*, 2021). The SNP calls of the 185 individuals (Dataset S1) were concatenated and imputed using AlphaImpute version 1.9.8.3 (Whalen & Hickey, 2020). The outputted genotype probabilities were imported for LMM models in GEMMA to associate with gene expression data. The approach to define an eQTL was identical to RNAseq variants with a lenient *p*-value ($0.05/49,330=1.01\times10^{-6}$).

Trans-eQTL hotspots were searched using the density function in R with the default “Gaussian” kernel. The number of windows was set to 2^{15}=32,768, equivalent to a window size of 870 bp. Hotspots within 5 kb were merged and the total number of embedded trans-eQTL were counted for each hotspot. The density threshold was computed via permutation tests. Permutation tests were conducted with the following steps: all lead SNP sites were randomly shuffled a number of times equal to the total number of eQTL (repeated sampling was allowed), and the highest density was stored; the first step was repeated 1,000 times; the distribution of the highest density was drawn, and the density at $p = 0.05$ was used as the threshold.

Note S2

Genome annotation

A repeat library was constructed using EDTA version 1.9.0 with “—sensitive 1” to allow RepeatModeler to identify remaining TEs (Ou *et al.*, 2019). The TE annotation library was generated by EDTA in a separate run. TE regions of both haploid assemblies were masked by ReapeatMasker version 4.1.1 provided with the repeat library. Protein-coding genes were annotated following the MAKER-P annotation pipeline (Campbell *et al.*, 2014). In the initial run, MAKER integrated transcript and protein evidence. Transcript evidence included the *Fragaria ×ananassa* GDR RefTrans V1 (https://www.rosaceae.org/analysis/230) and non-overlapping transcripts assembled from over 20 genotypes (Barbey *et al.*, 2020). RNAseq reads were first cleaned using Trimmomatic version 0.39 and mapped to the new assemblies without the gene annotation file by STAR version 2.7.6a (Dobin *et al.*, 2013). A unified set of transcripts was assembled by PsiCLASS version 1.0.1 (Song *et al.*, 2019). The curated protein database for tracheophyta was downloaded from UniProt (https://www.uniprot.org/) and transposases filtered out. In the sequential iterative runs, ab initio gene predictors SNAP (Korf, 2004) and Augustus (Stanke & Waack, 2003) were iteratively trained and used for gene prediction. Functional annotations were assigned via the UniProt/Swiss-Prot protein database using iprscan version 5.50. The final set of annotated genes were either supported by the evidence (AED < 1.0) or encoding a Pfam domain. KEGG K-numbers for annotated genes were assigned by KofamKOALA (Aramaki *et al.*, 2020).

Evaluation of genome assembly and annotation

Basic summary stats for the assemblies such as total length, N50, and N95 were obtained using QUAST version 5.0.2 (Mikheenko *et al.*, 2018). Genome quality and phasing quality (QV, completeness, switching error and hamming error) were evaluated by Merqury version 1.1 and Yak version 0.1 (Rhie *et al.*, 2020; Cheng *et al.*, 2021). The completeness of the haploid assemblies and protein-coding gene annotations were assessed with the BUSCO eudicots database including 2,326 conserved plant genes (Seppey *et al.*, 2019). The quality of scaffolding was inspected on the Hi-C contact map and a FL 15.89-25 linkage map that was built with 1676 SNPs using 66 progeny of 'Florida Beauty' × FL 15.89-25 with the 'onemap' package in R (Margarido *et al.*, 2007). The public Hi-C data (Lee *et al.*, 2021) from *Fragaria ×ananassa* was downloaded from the NCBI SRA database (accession number: SRX10474285 & SRX10474284) to evaluate scaffolding quality. Hi-C reads were mapped to the Bea haploid assembly using HiC-Pro version 3.0 and visualized in Juicebox version 1.11.

Annotation evaluation

Although fewer protein-coding genes were annotated in our new phased assemblies (90372 and 89218) compared to the recent reannotation (108,447) of the reference genome (Edger *et al.*, 2019), comparable completeness of BUSCO genes was achieved, of which 97.1% and 96.9% were complete including 91.8% and 92% duplicated. TE regions comprised 40.97% and 42.24% of the haploid assemblies. The increase of TE percentage compared to the reference genome (36.08%) was largely attributed to higher percentage of helitron (5.19-5.43% VS 0.09%) identified with EDTA pipeline (Ou *et al.*, 2019).

Note S3

Fruit allele-specific expression within "FL15.89-25"

To connect gene annotations in our new haploid assembles, the F12 annotated protein database was blasted to Bea annotated protein databases. Synteny blocks were classified using MCScanx version 20130328 (Wang *et al.*, 2012). A matched gene must satisfy one of the two criteria: 1. The matched gene must reside on the same chromosome with a one-to-one physical relationship. 2. If a multiple-to-one relationship was found, the difference in physical positions of the matched gene pair was limited to 1 Mb on the same chromosome. Potential presence/absence variation (PAV) genes were genes with no matched pair in the other haplotype. In addition, PAV genes were accessed by alignment of parent-specific CCS reads binned by HiCanu version 2.1. Only genes with less than 30% coverage of their whole length were considered as PAV genes. Gene ontology (GO) enrichment analysis was conducted using Bioconductor package topGO version 2.38.1. The whole gene set from both haploid assemblies were used as the universal set. The significances of GO terms were assessed by the classic Fisher's exact test and false discovery rate used for multiple test correction.

Four biological replicates of FL 15.89-25 full-red fruit samples harvested from different plants on two dates were sent for transcriptome sequencing. More than 6 Gb Illumina 150PE short-read data were generated for each sample. Gene expression was computed using the same approach as described for eQTL mapping. The concatenated haplotype assemblies were used as the reference genome. The corresponding annotations of the same gene in both haploid assemblies were identified by the abovementioned ortholog search. The allelic expression ratio was computed by dividing the F12 allelic read count by the sum of the F12 allelic read count and the Bea allelic read count. A general linear model (GLM) was fitted for each gene, specifying a binomial distribution (Crowley *et al.*, 2015) and a logit link using the GLM function in R. *P*-values from the Z-statistic, testing whether the intercept was different from 0, were adjusted with Bonferroni corrections ($\alpha = 0.05$). There was no need to correct for gene length or total number of reads since the GLM model was gene-specific.

Note S4

Volatile quantification method

Volatiles were sampled from headspace with a 2-cm tri-phase solid-phase micro-extraction (SPME) fiber, separated on an Agilent Technologies DB-5 column and analyzed on an Agilent Technologies 6890N gas chromatograph (GC) coupled with a mass spectrometry using EI ionization. Volatile identification and quantification of peak area were conducted with MassHunter Workstation software (Version 10.0; Agilent Technologies). Retention times of 12 common strawberry volatiles (butanoic acid, ethyl ester; butanoic acid, methyl ester; mesifurane; gamma-decalactone; gamma-dodecalactone; octanoic acid, methyl ester; octanoic acid ethyl ester; methyl anthranilate; 2-hexenal, (E)-; 1-hexanol; nonanal; hexanoic acid, methyl ester) were compared with authentic standards (Sigma Aldrich) for identification and calibration. The rest of compounds were tentatively identified based on matches to the NIST library version 14 (match score > 0.9). Three biological replicates (harvests) were tested for each genotype. Log-transformed mean values across biological replicates were used for genetic associations.

Note S5

Volatile-GWAS method

Because the GWAS population was mostly constituted of bi-parental crosses, strong relatedness and stratification existed in the population. To account for population structure, we used a linear mixed model implemented in fastGWA (GCTA Version 1.93.2beta) including the top 10 principal components and a relationship matrix derived from SNPs. A high number (10) of principal components was chosen to explain the multi-family structure of the panel. All of the small off-diagonal elements (<0.05) in the relationship matrix were set to 0 (Jiang *et al.*, 2019) to counter overcorrection by population structure within families. To determine the genome-wide significance threshold, we estimated the number of effective SNPs by pruning with window size of 50 SNPs and $r^2 = 0.5$ using PLINK Version 1.90b3.39 (Purcell *et al.*, 2007). The number of effect SNPs was estimated to be 5402. The Bonferroni corrected genome-wide threshold was p = 1.85 × 10-5 (0.1/5,402). Narrow-sense heritability (h^2) and SNP-based heritability (h_{snp}^2) were computed using the approach implemented in GCTA version 1.93.2beta (Zaitlen *et al.*, 2013). This approach permitted the use of genotypes from both closely and distantly related pairs of individuals. Linkage disequilibrium (LD) analysis was conducted for significant SNPs within the GWAS clusters using the "snpStats" package version 1.42 in R.

Note S6

Miscellaneous methods

Mendelian randomization: Mendelian randomization (MR) was applied to search both master regulatory genes at trans-eQTL hotspots and candidate genes underlying volatile GWAS peaks. The T_{MR} was estimated according to (Lawlor *et al.*, 2008; Zhu *et al.*, 2016) in three steps. 1. $\hat{\beta}_{xy} = \hat{\beta}_{zy}/\hat{\beta}_{zx}$ where z denoted a shared significant genetic variant, $\hat{\beta}_{zy}$ was the effect of z on trans-regulated eGene expression y and $\hat{\beta}_{zx}$ was the effect of z on cis-regulated eGene expression x. 2. $var(\hat{\beta}_{xy}) = [var(y)(1 - R_{xy}^2)]/[nvar(x)R_{zx}^2]$ where R_{xy}^2 and R_{zy}^2 were proportions of variance in y, explained by x and z respectively. The n was equal to 196, the sample size of the eQTL study. 3. $T_{MR} = \beta_{xy}^2/var(\hat{\beta}_{xy})$, where T_{MR} follows the χ^2 distribution with one degree of freedom. The pchisq function in R was used to calculate the *p*-value. The median *p*-value across all trans-regulated genes was used to represent the significance of a potential regulatory gene. For causal gene identification, y was volatile abundance, x was eGene expression, and z denoted the significant marker shared by GWAS and eQTL. R_{xy}^2 was estimated using 59 shared genotypes between the eQTL and GWAS panels. For the master regulator search, the significance threshold was adjusted based on the number of total tests ($p = 1.04\times10^{-5}$). When used for searching candidate genes at GWAS peaks, expression correlations and predicted biological functions were considered, therefore no threshold was enforced.

Sensory evaluation: Sensory evaluations of sweetness perception and overripe flavor were collected in the previous descriptive panel study (Fan *et al.*, 2021b).

RT-PCR: Total RNA from agroinfiltrated fruits was reverse-transcribed using LunaScript® RT SuperMix Kit (New England Biolabs). The cDNA library was mixed with Forget-Me-Not™ EvaGreen® qPCR Master Mixes (Biotium). RT-PCR was performed on a LightCycler480 (Roche, Inc.). Three technical replicates were included for each sample. The qRT-PCR data was analyzed using the comparative CT method (ΔΔCT) following the manufacturer’s direction. The primers sequences can be found in Table S2.

Haplotype analysis: Phased SNP calls from Beagle output were converted to GDS files using the SeqArray package in R (Zheng *et al.*, 2012). All significant makers within a cis-eQTL for the target volatile were LD pruned ($r < 0.5$). Phased calls of the remaining markers were concatenated to form final haplotypes. Haplotype effects on phenotypes were computed using a simple linear model. Haplotype, haplotype effect and LD among makers were displayed via heatmaps using ComplexHeatmap package (Gu *et al.*, 2016).

eQTL mapping using the F12 haploid assembly: The same procedures for marker calling and eQTL mapping was used as described above, except that the markers were not filtered for MAF > 0.05.

Nucleotide diversity/MAF/AA frequency: Nucleotide diversity (π) was computed using vcftools version 0.1.16 with a window size of 10 kb. MAF and AA frequency were obtained using the hardy function in vcftools.

High-Resolution Melting Marker Test: The design and testing of HRM markers were according to our previous study (Oh *et al.*, 2021). In brief, the primers FaOMT01_F & FaOMT01_R were designed based on flanking sequences of the 28 bp InDel2 within *FaOMT* (Table S2). The functional marker was first tested using 38 breeding lines previously phenotyped for volatile content (Oh *et al.*, 2021). To validate the allele dosage effect, a test cross ‘FL 16.30-128’ × FL 15.89-25 was created and evaluated in year 2020. We selected 19 individuals which had the same genotype for the Chr 1C QTL to conduct the marker test. Mesifurane abundances were evaluated at three different harvests and two field replications.

Long-range alignment: Alignment between long DNA sequences (> 10kb) was conducted using Minimap2 version 2.17 in asm5 mode (Li, 2018).

Map to a reference: BWA-MEM was used for short read mapping and visualized in IGV version 2.8.9. Additional Illumina WGS sequence data were obtained for 'Mara des Bois' and 'Florida127' with a coverage of 40×.

Pairwise alignment: Pairwise alignment of *FaNES1* homologs was conducted in Geneious Prime software version 11.

Note S7

Trans-eQTL hotspots

To mine master regulatory genes at trans-eQTL hotspots, we leveraged Mendelian Randomization (MR) tests and the plausible mechanism that the master regulatory gene is often governed by a cis-eQTL in the same region as trans-eQTL for its regulated genes (Yao *et al.*, 2017; Albert *et al.*, 2018). A total of 384 master regulators were identified using this approach (Dataset S6), including five regulators controlling more than 100 genes. On Chr 5D, there were 67 hotspots, with 15 of them assigned with potential regulators including the largest hotspot (Fig. S4a). This large hotspot between 7,796,793 bp and 7,803,611 bp contained 363 trans-eQTL for eGenes across all chromosomes (Fig. S4b). FxaC_20g13890 encoding an E3 ubiquitin ligase BIG BROTHER-like gene was the sole candidate regulator for this region (*p*_MR=3.18e-12). E3 ubiquitin ligase controls hormonal pathways related to fruit development and ripening (Chen *et al.*, 2016). In *Arabidopsis,* RING-finger protein BIG BROTHER has a central regulatory role in plant growth, such that a small change in its expression can lead to substantial alternation in organ size (Disch *et al.*, 2006). We observed a decrease by 40% for FxaC_20g13890 expression with two doses of the alternative allele in our eQTL population (Fig. S4c). The majority of trans-regulated genes (n = 330) were positively influenced by FxaC_20g13890 (Fig. S4d), and molecular functions such as calmodulin binding and protein serine/threonine kinase activity, related to plant growth and cell proliferation, were enriched in the regulated gene set. Notably, most accessions in the eQTL panel had either one or two alternative alleles within the hotspot region (Fig. S4c&d), suggesting selection in favor of the alternative allele, which might be attributed to a role as a repressor of plant organ growth (Disch *et al.*, 2006). This large hotspot exemplifies the complex and orchestrated network of gene regulation in strawberry fruit and the extensive natural variations existing in breeding materials.

Note S8

Allele-specific expression

A total of 75,429 annotated genes were matched between the two haplotypes. PAV analysis revealed 1,534 genes were absent in the F12 haplotype compared to 1,861 genes absent in the Bea haplotype. Go-term Enrichment pathway analysis found that PAVs were significantly enriched for 18 GO molecular functions (fisher exact test, $p < 0.01$) including the terpene synthase activity for which 16 of 280 annotated genes were PAV (Table S3). An evaluation of SVs between two haplotypes identified a total of 22,848 deletions, 49 inversions, 23,012 insertions, 55 interspersed duplications, 72 tandem duplications and 1,381 translocations. The combined assembly of both haplotypes was used for read mapping, and mapping rates of uniquely mapped reads varied from 31.7% to 34.1%. A total of 25,600 heterozygous genes had more than 10 counts for one allele in at least two replicates.

Within the alpha-linolenic acid metabolism pathway (Fig. S10b), several key genes showed extreme dominance of alleles from the F12 haplotype, such as two putative *peroxisomal acyl-coenzyme A oxidases 1* (*ACX*) and two putative *fatty acid beta-oxidation multifunctional protein 2* (*MFP2*), while a putative *alcohol dehydrogenase* (EC:1.1.1.1), essential to convert 3-hexenal to 3-hexenol, expressed dominantly from the Bea allele (Fig. S9b). Both *cis*-3-hexenal and *cis*-3-hexenol are among the most abundant volatiles in ripe strawberry and serve as substrates for a variety of esters (Aharoni *et al.*, 2000).

Note S9

Predict sensory characteristics using SNP markers

Our previous sensory studies have indicated medium-chain esters positively influenced sweetness perception, while sulfur esters imparted overripe flavor (Fan *et al.*, 2021b,a). We correlated the best markers underlying the respective medium-chain ester hotspot and methyl thiolacetate to sensory scores of sweetness and overripe flavor using 88 samples representing 26 accessions collected in the previous study (Fan *et al.*, 2021b). Although neither test reached significance, a discernable marker effect of AX-184745816 was observed for overripe flavor in the same direction as the effect on methyl thiolacetate abundance (Fig. S13a). Meanwhile, more samples with the alternative allele of AX-166515537 were below the predicted sweetness score based on soluble solids content (SSC%), suggesting a negative effect of the alternative allele for this medium-chain ester hotspot (Fig. S13b).

Note S10

Deletion in *FaFAD1*

To date, the effects of SVs underlying agronomic traits remains elusive in strawberry. The only natural SV that was discovered to cause a change in flavor in octoploid strawberry was the large deletion of the *FaFAD1* gene, leading to the failure to produce gamma- decalactone (Chambers *et al.*, 2014). Previously, using short-read sequencing and bacterial artificial chromosome libraries, the size of the deletion was 8,262bp and harbored the whole genic region of *FaFAD1* (Barbey *et al.*, 2021). Both of the new UF haploid assemblies carried the functional haplotype with the complete *FaFAD1* gene, whereas both 'Royal Royce' haplotypes carried the non-functional haplotype. Comparisons between haplotypes resolved a deletion of 12,871bp in the present study (Fig. S19). The *FaFAD1* gene was flanked by LTR regions. Comparisons among two *F.* ×*ananassa* haplotypes and the haplotype from the extent diploid ancestor *F. vesca* (Edger *et al.*, 2018) (*FaFAD1* is missing in the other extent diploid ancestor *F. iinumae* (Edger *et al.*, 2020)) implicated that after polyploidization, the 5' LTR has undergone contraction, whereas the 3' LTR has undergone expansion (Fig. S19). The deletion was likely mediated by illegitimate recombination, the predominant mechanism behind TE loss (Ma *et al.*, 2004; Woodhouse *et al.*, 2010).

Note S11

A regulatory network of anthocyanin biosynthesis

This comprehensive eQTL dataset could also facilitate exploration of natural variation in important biosynthetic pathways other than flavor-related pathways such as anthocyanin biosynthesis. A prominent example is anthocyanins which give strawberry fruit their characteristic color. Among 31 genes potentially involved in (Fig. S22a), 17 eQTL comprising 11 cis-eQTL and 6 trans-eQTL were found for 14 genes (Fig. S22b), with minor allele frequencies ranged from 0.046 to 0.469. Previously mapped QTL for anthocyanidins (Zorrilla-Fontanesi *et al.*, 2011) and pelargonidin-3-glucoside (Labadie *et al.*, 2020) on homoeologous group 5 and Chr 3A colocalized with four eQTL for FaANS homoeologs and one trans-eQTL for FaANR, respectively. We also identified homoeologous trans-eQTL for two FaDFRs (FxaC_6g44881 & FxaC_8g06290), co-localized with a cis-eQTL of another homoeolog (FxaC_7g41980, Fig. S22c). A significant positive correlation (t-test, $p<10e-16$) was only found between expression of FxaC_8g06290 and FxaC_6g44881 (Fig. S22d), but not with FxaC_7g41980.

References

Aharoni A, Keizer LC, Bouwmeester HJ, Sun Z, Alvarez-Huerta M, Verhoeven HA, Blaas J, van Houwelingen AM, Vos RC de, van der Voet H, *et al.* 2000. Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *The Plant Cell* **12**: 647–662.

Albert FW, Bloom JS, Siegel J, Day L, Kruglyak L. **2018**. Genetics of trans-regulatory variation in gene expression. *eLife* **7**: e35471.

Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. **2020**. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**: 2251–2252.

Barbey CR, Hogshead MH, Harrison B, Schwartz AE, Verma S, Oh Y, Lee S, Folta KM, Whitaker VM. **2021**. Genetic Analysis of Methyl Anthranilate, Mesifurane, Linalool, and Other Flavor Compounds in Cultivated Strawberry (Fragaria × ananassa). *Frontiers in Plant Science* **12**: 718.

Barbey C, Hogshead M, Schwartz AE, Mourad N, Verma S, Lee S, Whitaker VM, Folta KM. **2020**. The Genetics of Differential Gene Expression Related to Fruit Traits in Strawberry (Fragaria ×ananassa). *Frontiers in Genetics* **10**: 1317.

Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, *et al.* **2014**. MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. *Plant Physiology* **164**: 513–524.

Chambers AH, Pillet J, Plotto A, Bai J, Whitaker VM, Folta KM. **2014**. Identification of a strawberry flavor gene candidate using an integrated genetic-genomic-analytical chemistry approach. *BMC genomics* **15**: 217.

Chen J, Mao L, Lu W, Ying T, Luo Z. **2016**. Transcriptome profiling of postharvest strawberry fruit in response to exogenous auxin and abscisic acid. *Planta* **243**: 183–197.

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

Crowley JJ, Zhabotynsky V, Sun W, Huang S, Pakatci IK, Kim Y, Wang JR, Morgan AP, Calaway JD, Aylor DL, *et al.* **2015**. Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nature Genetics* **47**: 353–360.

Disch S, Anastasiou E, Sharma VK, Laux T, Fletcher JC, Lenhard M. **2006**. The E3 Ubiquitin Ligase BIG BROTHER Controls Arabidopsis Organ Size in a Dosage-Dependent Manner. *Current Biology* **16**: 272–279.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. **2013**. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**: 15–21.

Edger PP, McKain MR, Yocca AE, Knapp SJ, Qiao Q, Zhang T. **2020**. Reply to: Revisiting the origin of octoploid strawberry. *Nature genetics* **52**: 5–7.

Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR, Smith RD, Teresi SJ, Nelson ADL, Wai CM, *et al.* **2019**. Origin and evolution of the octoploid strawberry genome. *Nature Genetics* **51**: 541–547.

Edger PP, VanBuren R, Colle M, Poorten TJ, Wai CM, Niederhuth CE, Alger EI, Ou S, Acharya CB, Wang J, *et al.* **2018**. Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (Fragaria vesca) with chromosome-scale contiguity. *GigaScience* **7**: gix124.

Fan Z, Hasing T, Johnson TS, Garner DM, Barbey CR, Colquhoun TA, Sims CA, Resende MFR, Whitaker VM. **2021a**. Strawberry sweetness and consumer preference are enhanced by specific volatile compounds. *Horticulture Research* **8**: 1–15.

Fan Z, Plotto A, Bai J, Whitaker VM. **2021b**. Volatiles Influencing Sensory Attributes and Bayesian Modeling of the Soluble Solids–Sweetness Relationship in Strawberry. *Frontiers in Plant Science* **0**.

Gu Z, Eils R, Schlesner M. **2016**. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**: 2847–2849.

Hardigan MA, Feldmann MJ, Lorant A, Bird KA, Famula R, Acharya C, Cole G, Edger PP, Knapp SJ. **2020**. Genome Synteny Has Been Conserved Among the Octoploid Progenitors of Cultivated Strawberry Over Millions of Years of Evolution. *Frontiers in Plant Science* **10**: 1789.

Hardigan MA, Lorant A, Pincot DDA, Feldmann MJ, Famula RA, Acharya CB, Lee S, Verma S, Whitaker VM, Bassil N, *et al.* **2021**. Unraveling the Complex Hybrid Ancestry and Domestication History of Cultivated Strawberry. *Molecular Biology and Evolution* **38**: 2285–2305.

Jiang L, Zheng Z, Qi T, Kemper KE, Wray NR, Visscher PM, Yang J. **2019**. A resource-efficient tool for mixed model association analysis of large-scale data. *Nature Genetics* **51**: 1749–1755.

Korf I. **2004**. Gene finding in novel genomes. *BMC bioinformatics* **5**: 59.

Labadie M, Vallin G, Petit A, Ring L, Hoffmann T, Gaston A, Potier A, Munoz-Blanco J, Caballero JL, Schwab W. **2020**. The genetic architecture of fruit colour in strawberry (Fragaria× ananassa) uncovers the predominant contribution of the F. vesca subgenome to anthocyanins and reveals underlying genetic variations. *bioRxiv*.

Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Smith GD. **2008**. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine* **27**: 1133–1163.

Lee H-E, Manivannan A, Lee SY, Han K, Yeum J-G, Jo J, Kim J, Rho IR, Lee Y-R, Lee ES, *et al.* **2021**. Chromosome Level Assembly of Homozygous Inbred Line 'Wongyo 3115' Facilitates the Construction of a High-Density Linkage Map and Identification of QTLs Associated With Fruit Firmness in Octoploid Strawberry (Fragaria × ananassa). *Frontiers in Plant Science* **12**: 696229.

Li H. **2018**. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

Ma J, Devos KM, Bennetzen JL. **2004**. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome research* **14**: 860–869.

Margarido GRA, Souza AP, a. F. Garcia A. **2007**. OneMap: software for genetic mapping in outcrossing species. *Hereditas* **144**: 78–79.

Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. **2018**. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**: i142–i150.

Oh Y, Barbey CR, Chandra S, Bai J, Fan Z, Plotto A, Pillet J, Folta KM, Whitaker VM, Lee S. **2021**. Genomic Characterization of the Fruity Aroma Gene, FaFAD1, Reveals a Gene Dosage Effect on γ-Decalactone Production in Strawberry (Fragaria × ananassa). *Frontiers in Plant Science* **12**: 639345.

Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, *et al.* 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* **20**: 275.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, *et al.* 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**: 559–575.

Rhie A, Walenz BP, Koren S, Phillippy AM. **2020**. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology* **21**: 245.

Seppey M, Manni M, Zdobnov EM. **2019**. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods in Molecular Biology (Clifton, N.J.)* **1962**: 227–245.

Song L, Sabunciyan S, Yang G, Florea L. **2019**. A multi-sample approach increases the accuracy of transcript assembly. *Nature Communications* **10**: 1–7.

Stanke M, Waack S. **2003**. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**: ii215–ii225.

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, *et al.* 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* **40**: e49.

Whalen A, Hickey JM. **2020**. AlphaImpute2: Fast and accurate pedigree and population based imputation for hundreds of thousands of individuals in livestock populations. *BioRxiv*: 299677.

Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. **2010**. Following Tetraploidy in Maize, a Short Deletion Mechanism Removed Genes Preferentially from One of the Two Homeologs. *PLOS Biology* **8**: e1000409.

Yao C, Joehanes R, Johnson AD, Huan T, Liu C, Freedman JE, Munson PJ, Hill DE, Vidal M, Levy D. **2017**. Dynamic Role of trans Regulation of Gene Expression in Relation to Complex Traits. *The American Journal of Human Genetics* **100**: 571–580.

Zaitlen N, Kraft P, Patterson N, Pasaniuc B, Bhatia G, Pollack S, Price AL. **2013**. Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits. *PLOS Genetics* **9**: e1003520.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. **2012**. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**: 3326–3328.

Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, *et al.* 2016. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics* **48**: 481–487.

Zorrilla-Fontanesi Y, Cabeza A, Domínguez P, Medina JJ, Valpuesta V, Denoyes-Rothan B, Sánchez-Sevilla JF, Amaya I. **2011**. Quantitative trait loci and underlying candidate genes controlling agronomical and fruit quality traits in octoploid strawberry (Fragaria× ananassa). *Theoretical and Applied Genetics* **123**: 755–778.