

RESEARCH ARTICLE

Open Access

# Identification of a strawberry flavor gene candidate using an integrated genetic-genomic-analytical chemistry approach

Alan H Chambers<sup>1</sup>, Jeremy Pillot<sup>1</sup>, Anne Plotto<sup>2</sup>, Jinhe Bai<sup>2</sup>, Vance M Whitaker<sup>3</sup> and Kevin M Folta<sup>1,4\*</sup>

## Abstract

**Background:** There is interest in improving the flavor of commercial strawberry (*Fragaria × ananassa*) varieties. Fruit flavor is shaped by combinations of sugars, acids and volatile compounds. Many efforts seek to use genomics-based strategies to identify genes controlling flavor, and then designing durable molecular markers to follow these genes in breeding populations. In this report, fruit from two cultivars, varying for presence-absence of volatile compounds, along with segregating progeny, were analyzed using GC/MS and RNAseq. Expression data were bulked *in silico* according to presence/absence of a given volatile compound, in this case  $\gamma$ -decalactone, a compound conferring a peach flavor note to fruits.

**Results:** Computationally sorting reads in segregating progeny based on  $\gamma$ -decalactone presence eliminated transcripts not directly relevant to the volatile, revealing transcripts possibly imparting quantitative contributions. One candidate encodes an omega-6 fatty acid desaturase, an enzyme known to participate in lactone production in fungi, noted here as *FaFAD1*. This candidate was induced by ripening, was detected in certain harvests, and correlated with  $\gamma$ -decalactone presence. The *FaFAD1* gene is present in every genotype where  $\gamma$ -decalactone has been detected, and it was invariably missing in non-producers. A functional, PCR-based molecular marker was developed that cosegregates with the phenotype in F<sub>1</sub> and BC<sub>1</sub> populations, as well as in many other cultivars and wild *Fragaria* accessions.

**Conclusions:** Genetic, genomic and analytical chemistry techniques were combined to identify *FaFAD1*, a gene likely controlling a key flavor volatile in strawberry. The same data may now be re-sorted based on presence/absence of any other volatile to identify other flavor-affecting candidates, leading to rapid generation of gene-specific markers.

**Keywords:** RNA-seq,  $\gamma$ -decalactone, *Fragaria*, Flavor, Molecular marker, Strawberry

## Background

The commercial strawberry (*Fragaria × ananassa*) ( $2n = 8x = 56$ ) is a popular fresh and processed fruit with substantial value worldwide. It is recognized for its sweet flavors and appealing aromas. The volatile profiles of strawberry are relatively complicated among berries, with over 360 volatile compounds reported [1]. A reduced set of approximately 20 volatiles are commonly reported to be important components of strawberry flavor [2-4]. The principle flavor compounds include esters [5], ketones, terpenes [6], furanones [7], aldehydes [8], alcohols, and

sulfur-containing compounds. The concentrations of individual volatiles are highly dependent on species [9], environment and harvest date [2, 10-12], cultivar, post-harvest treatment and fruit developmental stage [13].

One important volatile compound is  $\gamma$ -decalactone ( $\gamma$ -D; CAS 706-14-9). This volatile is described as “fruity”, “sweet”, or “peachy” [9] and contributes to fruit aroma [14,15]. The volatile tends to be undetectable in some genotypes [2], while in others its accumulation varies greatly within and between harvest seasons [12]. This pattern suggests that a critical biosynthetic step or substrate may be missing or limited, and under strong environmental influence. The high variability may be due to differences in expression of genes encoding enzymes linked to the process. The observation that some genotypes never produce the compound when others do presents an excellent

\* Correspondence: kfolta@ufl.edu

<sup>1</sup>Horticultural Sciences Department, University of Florida, Gainesville, FL, USA

<sup>4</sup>Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, FL, USA

Full list of author information is available at the end of the article

basis to use global transcriptome profiling to identify candidate genes associated with its production or stability. Because the commercial strawberry is octoploid, F<sub>1</sub> progeny from a volatile producer and a non-producer have led to predictions about inheritance of a given volatile [2,16].

A number of researchers have used genomics approaches to identify marker-trait associations in polyploids using SNPs. Advances in marker discovery have been made in allohexaploid wheat (*Triticum* spp.) cultivars using the Illumina GoldenGate Assay [17], allohexaploid oat (*Avena sativa* L.) using Roche 454 sequencing [18], and allotetraploid oilseed rape (*Brassica napus*) using Illumina Solexa sequencing [19]. Other approaches have focused on developmental changes in the transcriptome associated with ripening to identify gene candidates. This strategy was effective for grape (*Vitis* spp.) using Illumina sequencing [20], and recently in peach (*Prunus persica* L.) using microarrays [21].

The goal of this work is to use a transcriptome-based approach to identify the genes required for  $\gamma$ -D production. The approach leverages the presence/absence nature of  $\gamma$ -D from specific genotypes, its predictable inheritance, environmental lability, and variation during the growing season. The analysis identified one transcript from a narrow set of gene candidates that is functionally related to genes implicated in biosynthesis of this compound in certain fungi [22-24] and the related compound  $\gamma$ -dodecalactone [25]. A PCR-based amplicon corresponding to the candidate sequence co-segregates with the volatile in a breeding population, corresponding backcrosses, and in select cultivars and wild accessions. We demonstrate that computational bulking of RNAseq data based on the presence or absence of a volatile can identify transcripts likely playing a direct role in volatile production.

## Results

The gene segregating with the presence of the  $\gamma$ -D volatile has been shown to segregate as a single dominant locus, making it a prime candidate for the approach outlined in Additional file 1: Figure S1. Briefly, a cross was constructed between Elyana, a  $\gamma$ -D producing cultivar, and Mara des Bois, a cultivar where  $\gamma$ -D has not been detected. Progeny were grown, and fruits from each individual plant were analyzed for volatiles and coincident gene expression. The fruits from each plant were analyzed and sequenced separately so that transcriptomes from producers and non-producers could be bulked computationally, with the hypothesis that candidate genes would be common to producers, while being expressed low levels or go undetected in non-producers. Results could be experimentally validated in the parental lines and in segregating progeny using gene expression analysis.

## $\gamma$ -D quantity is genetically and environmentally influenced

The first tests examined  $\gamma$ -D accumulation in the 'Elyana' and 'Mara des Bois' parental lines and representative progeny over a growing season, using detection by GCMS. Genotypes were assayed for  $\gamma$ -D production on three harvest dates. The data are presented in Figure 1, showing data for a single genotype representing each of five general trends. Approximately 50% of the progeny produced no  $\gamma$ -D, similar to 'Mara des Bois'. The largest portion of the  $\gamma$ -D producers followed a similar trend to 'Elyana', with higher amounts in the second harvest compared to the first and third harvests. The reciprocal trend was observed in five genotypes that showed less  $\gamma$ -D during the second harvest compared to the other two harvests. Three genotypes examined produced the highest amount in the first harvest, yet levels remained low the second and third harvests. A single genotype exhibited higher levels as the season progressed. These same volatile patterns were also observed in backcross progeny during the 2012/13 season (data not shown). While there is an approximately three-fold difference in accumulation in the 'Elyana' background over the season, no  $\gamma$ -D was ever detected in 'Mara des Bois' above background noise.

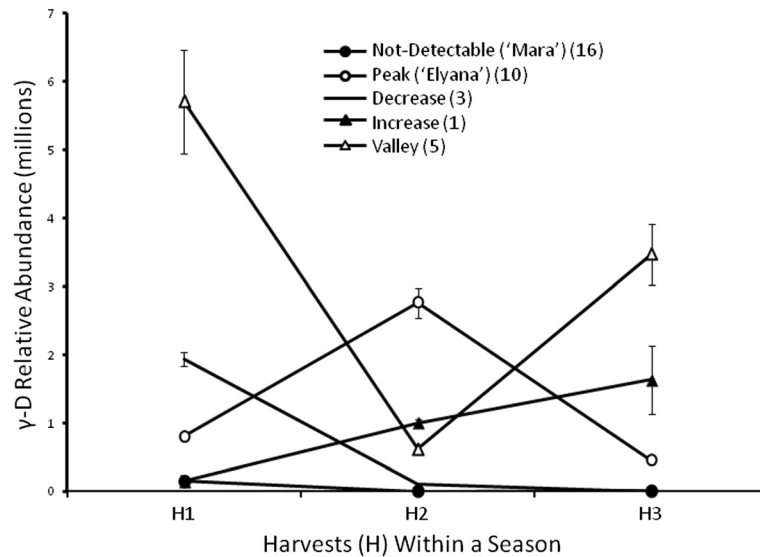
## $\gamma$ -Decalactone estimation

The levels of  $\gamma$ -D were estimated by comparing amounts detected in berries from the population using GCMS, with standards derived by adding the pure volatile to half-ripe strawberry fruit. Figure 2 shows the  $\gamma$ -D volatile phenotype for a subset of the 'Elyana' x 'Mara des Bois' progeny. The top producing 30 genotypes from the 2012/13 season ranged from 0.018 to 0.035 mM  $\gamma$ -D (data not shown).

## Transcriptome profiling

Fourteen progeny and both parents were individually analyzed by RNA-seq. The  $\gamma$ -D non-producers included in this analysis included 'Mara des Bois', 42, 89, 193, and 203. The producers were 'Elyana', 6, 24, 37, 51, 91, 93, 98, 103, 152, and 204. Many of the producers had higher  $\gamma$ -D levels than the 'Elyana'  $\gamma$ -D positive parent (Figure 2). The transcriptomes from individual lines were computationally pooled based on "producer" or "non-producer". Over 106 million MID-tagged RNAseq reads were generated from each of the parents and progeny. The average number of filtered and mapped reads per genotype was 5.5 million per genotype and ranged from 3 million to 8.5 million. Both parents had more than 7 million filtered, mapped reads.

Approximately 17,000 out of ~31,000 annotated genes in the strawberry draft genome [26] were represented in the RNA-seq dataset. A cursory SNP search identified over 1.7 million SNPs total when compared to the *F. vesca* genome (SNP criteria: 95% minimum *P* not ref, 10 or



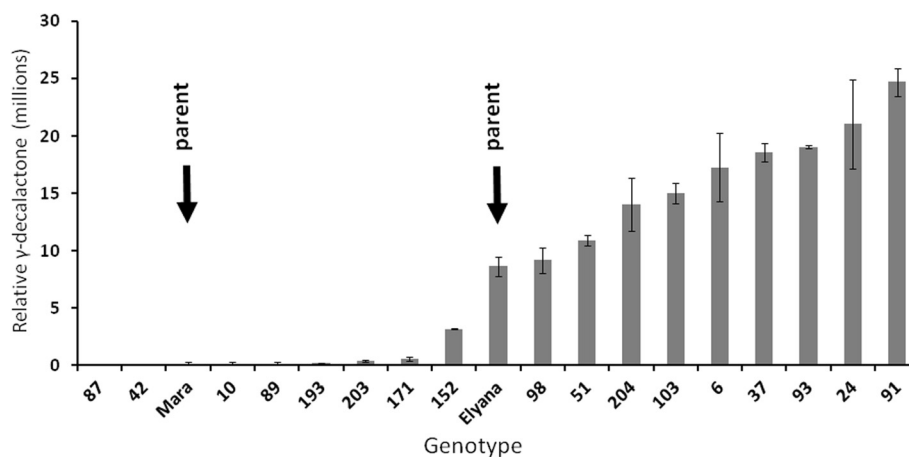
**Figure 1 Representative genotypes showing  $\gamma$ -decalactone stability over a single growing season.** 35 out of 130 'Elyana' x 'Mara des Bois' progeny produced fruit suitable for volatile analysis over three consecutive harvests. All progeny could be clustered into five classes based on  $\gamma$ -decalactone production as shown by five uniquely shaped graphs below. 'Mara des Bois' represents the class of non-producers, with 'Elyana' representing those lines that peaked mid-season. Other patterns included the inverse of 'Elyana' with a "valley" pattern, a strong "decrease" after the earliest harvest, and one progeny that showed an increase as the season progressed. Counts in each category are shown in the attached legend. Data are from two technical replicates from one example genotype per class. Error bars represent standard deviations.

greater read depth, and present in at least two of the sixteen genotypes' datasets).

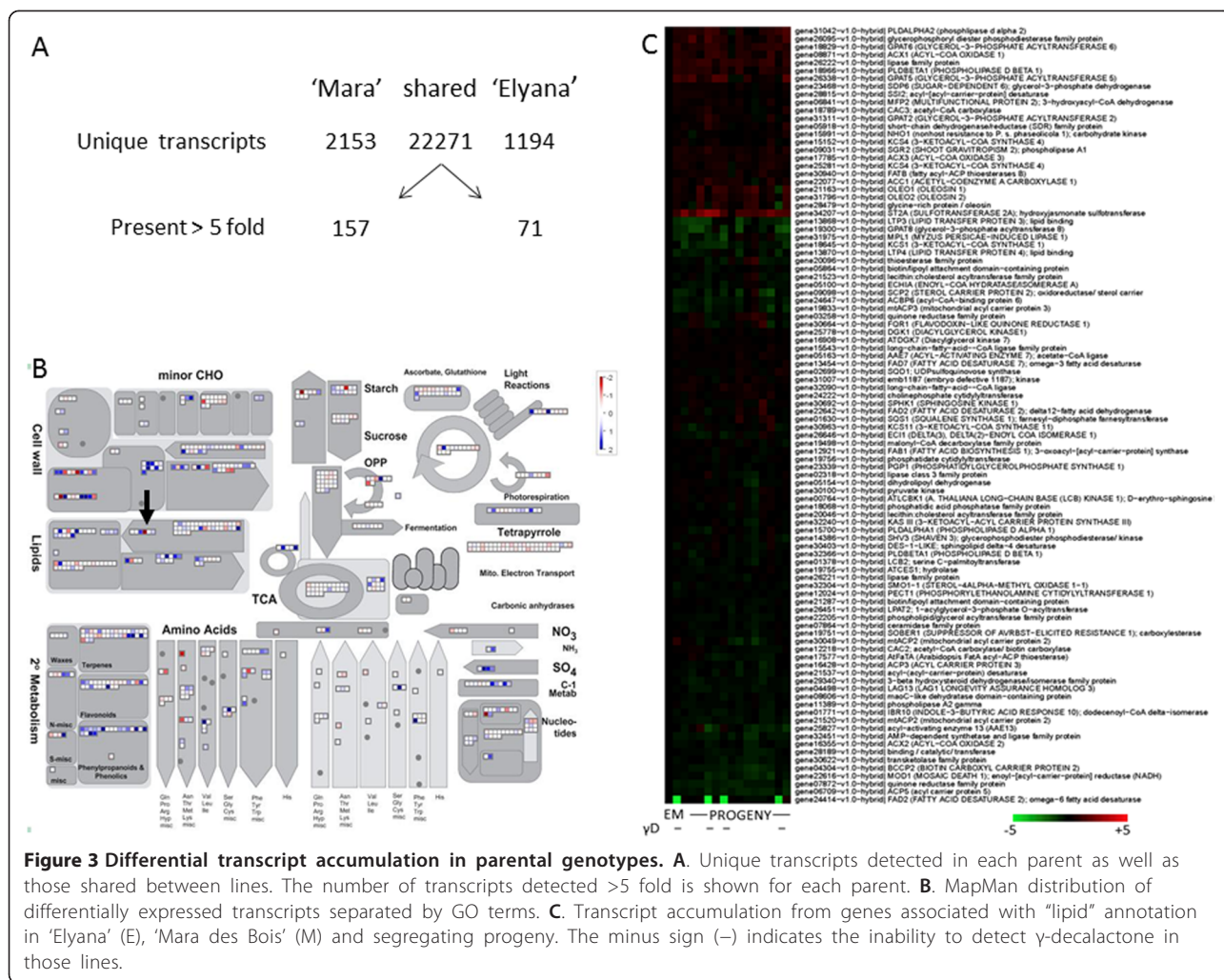
#### Gene expression trends in parental lines

Alignment of all reads against the diploid *F. vesca* genome produced transcript assemblies that provided cursory detail of gene expression difference between the two parental genotypes. Comparisons of differentially-expressed transcripts, including transcripts with low RPKM representation,

between the 'Mara des Bois' and 'Elyana' parents showed that among the 23,718 transcripts predicted from assembly of all reads, 2,153 were unique to the 'Mara des Bois' parent and 1,194 were only observed in 'Elyana' (Figure 3, Panel A). When transcripts composed of higher RPKM values were compared there were 157 that had a 5-fold or greater abundance in 'Mara des Bois' and 71 that had a >5-fold abundance in 'Elyana'. When grouped by GO function the differentially expressed genes



**Figure 2  $\gamma$ -Decalactone production in a selection of progeny from the 'Elyana' x 'Mara des Bois' cross.** Total volatiles were analyzed by GC/MS. A number of progeny produced more  $\gamma$ -decalactone than 'Elyana' during the harvest shown. A subset of the progeny, along with the 'Mara des Bois' parent, never produced  $\gamma$ -decalactone above background levels. Ripe fruit samples from some of these genotypes were split between volatile analysis and RNA-seq transcriptome analysis. Data are from two technical replicates. Error bars represent standard deviations.



**Figure 3 Differential transcript accumulation in parental genotypes.** **A.** Unique transcripts detected in each parent as well as those shared between lines. The number of transcripts detected >5 fold is shown for each parent. **B.** MapMan distribution of differentially expressed transcripts separated by GO terms. **C.** Transcript accumulation from genes associated with "lipid" annotation in 'Elyana' (E), 'Mara des Bois' (M) and segregating progeny. The minus sign (-) indicates the inability to detect  $\gamma$ -decalactone in those lines.

show no clear pattern (Figure 3B) favoring any one category.

Table 1 shows twelve transcripts that were the most abundant in the  $\gamma$ -D producing parent. The highest expressed was an omega-6-fatty acid desaturase transcript (gene24414), followed by a transcript annotated as osmotic stress/defense (gene32423). The set also includes two serine-threonine protein kinases (gene09445 and gene00774), citrate synthase (gene26778), an F-box protein (gene12328), a proline transport protein (gene21705), and several uncharacterized, hypothetical proteins.

**Computational bulking to limit candidate set**

The large number of differentially-expressed transcripts could be further narrowed by analyzing transcript patterns for these genes in progeny segregating for  $\gamma$ -D. Pairwise comparisons were made between genotypes with high (genotypes 'Elyana', 91, 24, 37, 103, and 006) and non-detectable (genotypes 'Mara des Bois', 203, 89, and 42)

$\gamma$ -D levels. Gene candidates were filtered to have a modest >4-fold increase in transcript support of producers over non-producers. Using this approach, a single gene candidate was identified, gene24414 on linkage group 3 (LG3:31112418..31114643, scf0513029:129621..131846), the same abundant transcript shown in Table 1 as variable between the two parents. To illustrate how the integration of segregating progeny can separate out transcripts not common to  $\gamma$ -D volatile producers, all transcripts from the "lipids" category are shown in Figure 3C from the parental genotypes and several of their progeny. In general, lipid related transcripts show limited differential accumulation between any genotypes. The clear exception is the omega-6-fatty-acid desaturase (bottom row) which is not detected in 'Mara des Bois' (M) and in three of the progeny (green squares). The  $\gamma$ -D volatile was not detected in these same genotypes. Of all candidates from Table 1, the omega-6-fatty-acid desaturase was the only transcript that correlated 100% with the ability to produce  $\gamma$ -D. The gene was



**Table 1 Transcripts abundant in gamma-decalactone-producing genotypes**

Featured ID	Seq. Description	GO biological process	GO cellular component	GO molecular function
gene24414	Omega-6 fatty acid endoplasmic reticulum isozyme 2-like	Unsaturated fatty acid biosynthesis process; oxidation-reduction process;	Endoplasmic reticulum membrane; integral to membrane;	Delta12-fatty acid dehydrogenase activity; oxidoreductase activity, acting on paired donors, with oxidation of a pair of donors resulting in the reduction of molecular oxygen to two molecules of water; omega-6 fatty acid desaturase activity;
gene32423	Osmotin-like protein osm34	Defense response to bacterium, incompatible interaction; response to salt stress; defense response to fungus, incompatible interaction;	#N/A	#N/A
gene26993	NA	#N/A	#N/A	#N/A
gene09445	Serine threonine-protein kinase rio1-like	Phosphorylation;	#N/A	ATP binding; protein serine/threonine kinase activity;
gene26778	Citrate synthase	Tricarboxylic acid cycle; cellular carbohydrate metabolic process; response to cadmium ion;	Cell wall; mitochondrial matrix; chloroplast;	Zinc ion binding; citrate (S)-synthase activity; ATP binding;
gene14440	NA	#N/A	#N/A	#N/A
gene13132	PREDICTED: uncharacterized protein in LOC101300103	#N/A	#N/A	#N/A
gene05924	atp-dependent-nad h-hydrate dehydrate-like	#N/A	#N/A	#N/A
gene12328	f-box protein at5g07610-like	#N/A	#N/A	#N/A
gene21705	Diphthamide biosynthesis protein 2-like	Proline transport;	#N/A	#N/A
gene00774	Serine threonine protein phosphatase	#N/A	#N/A	Hydrolase activity;
gene23527	Uncharacterized loc101207862	Generation of precursor metabolites and energy;	Chloroplast; membrane; mitochondrion;	#N/A

Transcripts here represent those that are correlate best with accumulation of gamma-decalactone (Pearson's  $p < 0.001$ ).

given the designation *FaFAD1*. A 1,128 bp open reading frame was cloned from 'Elyana' cDNA (Additional file 2).

#### Validation of key candidates

The steady-state transcript accumulation of *FaFAD1* (Figure 4A), gene22642 (*FaFAD2*; Figure 4B), and gene29958 (a cytochrome p450 oxidase termed *FaCYTp450*; Figure 4C) was tested. The results for *FaFAD2* and *FaCYTp450* are not consistent with the ability to produce  $\gamma$ -D, and were therefore de-prioritized as candidates. The qRT-PCR results for *FaFAD1* visually correlated more closely with the volatile phenotype than the RNA-seq RPKM values. Both methods were similar in failure to detect transcript support for *FaFAD1* in  $\gamma$ -D non-producers.

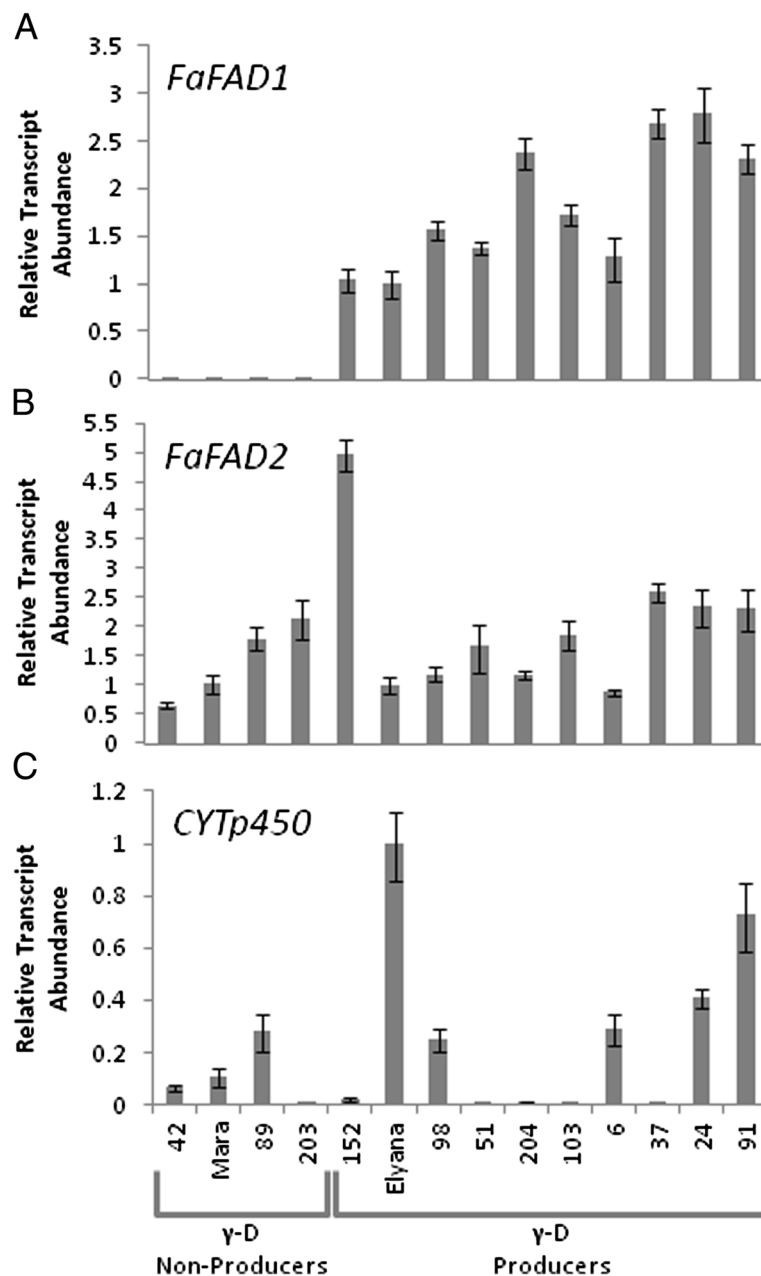
#### Candidate genes in fruit developmental series

An 'Elyana' fruit series was tested for ripening induction of *FaFAD1*, *FaFAD2*, and *FaCYTp450* as shown in Figure 5. The fold-change in transcript abundance for each gene

is shown for ripe fruit compared to blushing fruit. *FaCYTp450* showed a >11-fold increase ( $\pm 0.30$ ) in transcript abundance between ripe and blushing fruit. *FaFAD1* showed a >21-fold increase ( $\pm 0.33$ ) in transcript abundance. *FaFAD2* was not ripening induced. These results were consistent over at least three independent biological replicates.

#### Candidate genes and two environments

Environmental fluctuation of  $\gamma$ -D accumulation is shown in Figure 2. To test if *FaFAD1* matched this pattern, transcript abundance was examined in tissues obtained from two different harvests. The population average for all  $\gamma$ -D producers in environment "a" was approximately eleven-fold less than the population average for  $\gamma$ -D in environment "b".  $\gamma$ -D non-producers had levels of  $\gamma$ -D only consistent with noise during either harvest, but  $\gamma$ -D producers showed high environmental effects. Figure 6A shows the  $\gamma$ -D accumulation for four genotypes (10, 93,

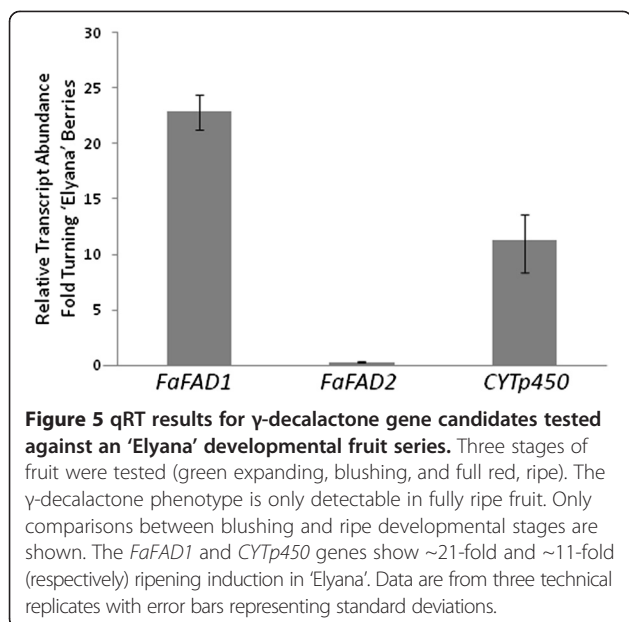


**Figure 4** qRT-PCR results from three gene candidates correlated with the  $\gamma$ -decalactone phenotype. A single gene, *FaFAD1* (gene24414), (A) encoding a putative  $\omega$ -6-fatty acid desaturase was identified as differentially expressed between high and low  $\gamma$ -decalactone genotypes. Another putative fatty acid desaturase, *FaFad2* (gene22642), was found in the *F. vesca* genome and is shown in (B). Reducing the stringency by reducing the number of progeny in each phenotypic pool resulted in another candidate (C) *CYTp450* (gene29958), a putative cytochrome p450 monooxygenase, located in proximity to *FaFAD1*. qRT-PCR results are shown for each of these genes using 'Elyana' as the comparator against a subset of progeny. Data are from three technical replicates with error bars representing standard deviations.

103, and 'Elyana') in these two environments. Figure 6B shows the qRT-PCR results for these genotypes in the two environments when transcript abundance in environment "b" was compared to environment "a". Only modest increases are shown for genotypes 93, 103, and 'Elyana' with genotype 10 (non-producer) showing no evidence for *FaFAD1* transcript accumulation.

#### Molecular basis of $\gamma$ -decalactone loss-of-function

The lack of detectable transcript in the 'Mara des Bois' parent and in specific progeny may be due to at least one of two factors. First, that transcription/mRNA accumulation of the candidate *FaFAD1* is blocked. Alternatively the functional gene or allele may be missing altogether. To test these possibilities several primer pairs were designed

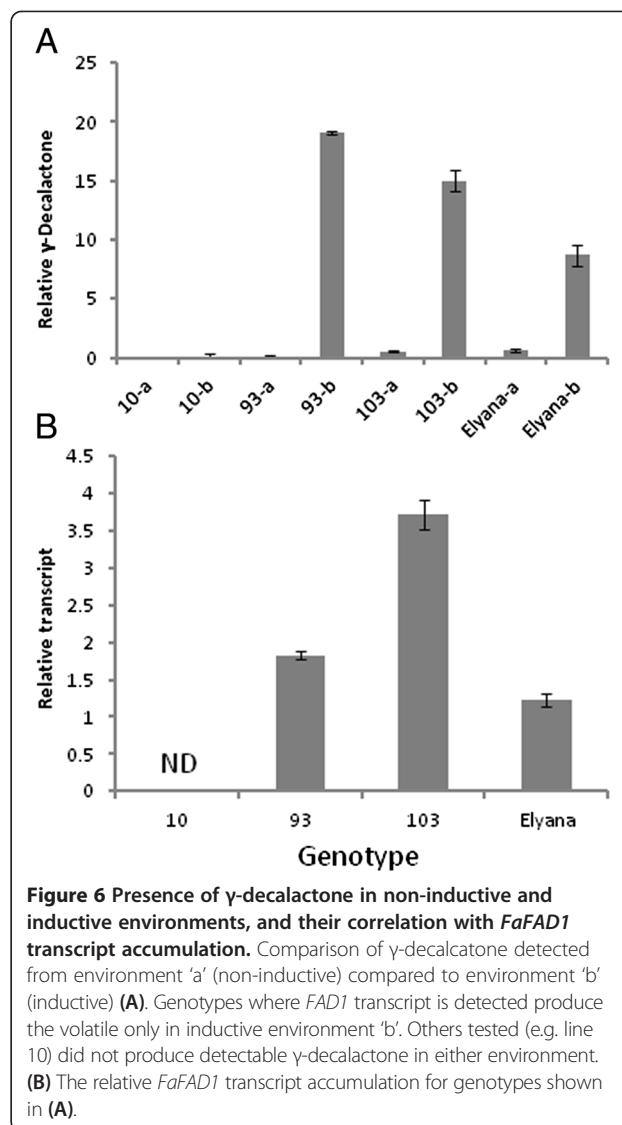


to amplify the genomic sequence upstream and internal to *FaFAD1*. A map of the genomic region and the corresponding primer pairs is provided in Figure 7. In all cases, none of the primer pairs could amplify products from genotypes unable to produce  $\gamma$ -D, while amplicons across the region were produced from every plant where  $\gamma$ -D was detected.

### $\gamma$ -decalactone marker in three populations

The ability to amplify a product specifically in  $\gamma$ -D producing genotypes provided an opportunity to design a gene-based molecular marker. A  $\gamma$ -D PCR-based assay was designed using *FaFAD1* primers and then tested against three populations. The first was a subset of 19 genotypes from the original 'Elyana' x 'Mara des Bois' F<sub>1</sub> population (Figure 8A). The presence of the 500 bp PCR amplicon (solid arrow) was detected exclusively in genotypes shown to produce  $\gamma$ -D. The positive PCR control (BFACT045) is shown by the dashed line arrow. The second and third populations tested were BC<sub>1</sub> crosses to the 'Elyana' and 'Mara des Bois' parents with progeny 98 as the male in each case. The 'Elyana' backcross contained 26 progeny, each with at least three harvests during the 2012/13 growing season. The marker co-segregated with the phenotype in all cases. Twenty-two progeny in the 'Mara des Bois' backcross had at least three volatile harvests during the 2012/13 season. Each cosegregated with the marker and phenotype (data not shown).

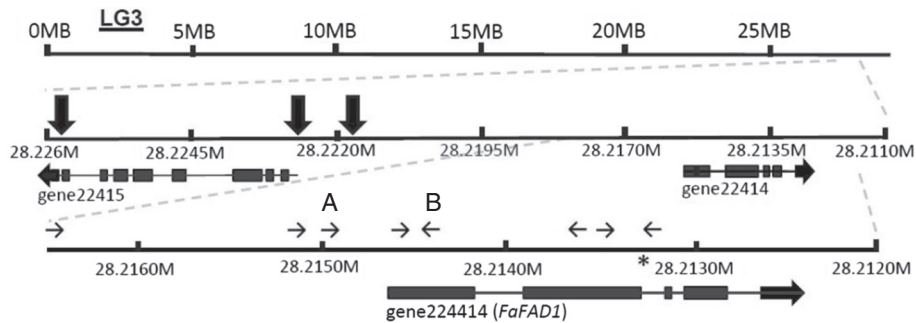
The potential molecular marker was also tested against a set of cultivars with demonstrated present or undetectable  $\gamma$ -D. 'Radiance', 'Albion', 'Winter Star', and 'Sweet Charlie' were all  $\gamma$ -D positive and positive also for the PCR product (Figure 8C). 'Deutsch Evern', 'Strawberry Festival', 'LF9', and



'Mieze Schindler' were all negative for  $\gamma$ -D and also negative also for the marker. 'Strawberry Festival' and 'LF9' were additionally interesting because 'LF9' is a seedling from self-pollination of 'Strawberry Festival' [27]. The prediction would therefore be that 'LF9' would be negative for the marker and  $\gamma$ -D phenotype. This was confirmed by volatile and marker analysis.

### SSR marker development

An SSR marker was developed to investigate cosegregation of alleles more distantly positioned relative to *FaFAD1*. Figure 9 shows the results of the SSR tested in the parents 'Elyana' and 'Mara des Bois', and 15 progeny selected from both  $\gamma$ -D producers and non-producers. Few progeny were tested because the objective of the SSR marker design was simply to demonstrate the potential for converting a gene candidate into a second type of molecular marker

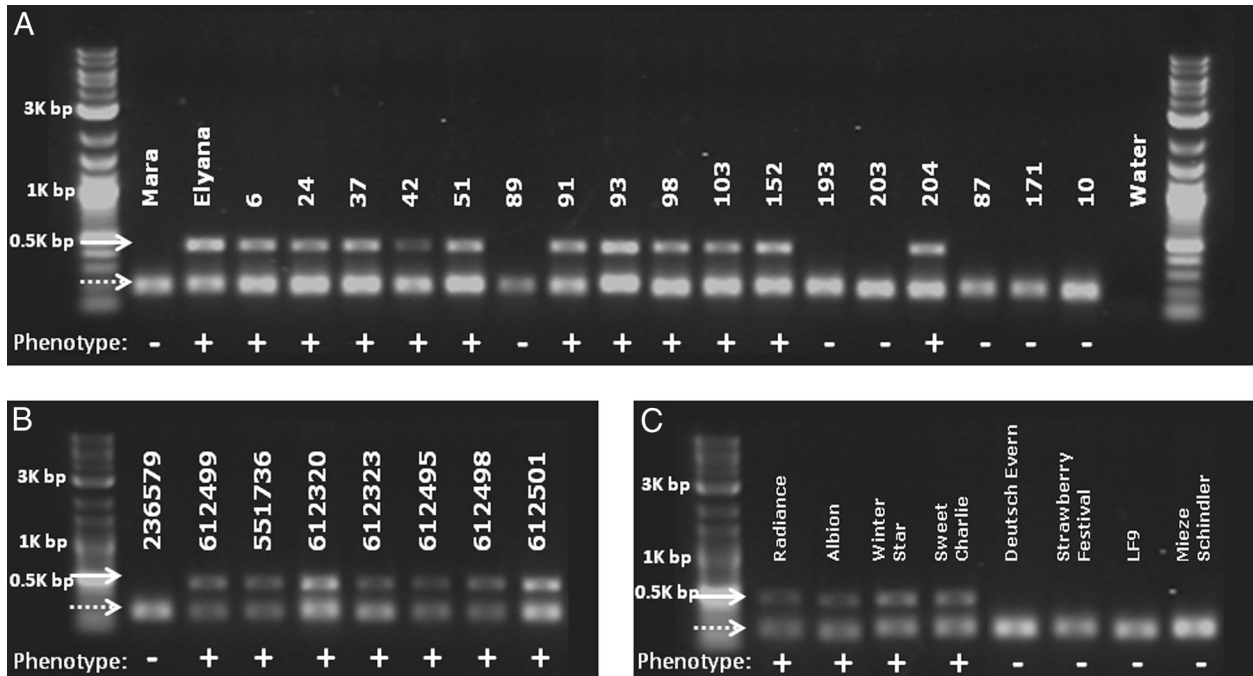


**Figure 7 FaFAD1 structure.** The *FaFAD1* gene, noted also as gene24414, is located in a region corresponding to linkage group 3 (LG3) in the *F. vesca* genome. Simple sequence repeat regions are noted by the large black arrows, and the small black arrows represent primers used to test for amplification within this region in parental lines and progeny. The letters **A** and **B** denote the primers used for marker amplification. The asterisk denotes a potential allele introducing a stop codon into the *FaFAD1* sequence.

commonly used for *Fragaria* genotyping. ‘Elyana’ exhibited four marker alleles (205, 209, 215, and 219), and ‘Mara des Bois’ only possesses the 209 allele. For clarity, only the 205 and 209 alleles are shown. Allele 209 was monomorphic in all genotypes tested, and alleles 215 and 219 were not associated with the  $\gamma$ -D phenotype. Progeny 103, 152, 204, 24, 37, 51, 6, 91, 93, and 98 were all positive for the  $\gamma$ -D phenotype and for allele 205. Progeny 171, 193, 203, 42, and 89 were negative for  $\gamma$ -D and the 205 allele.

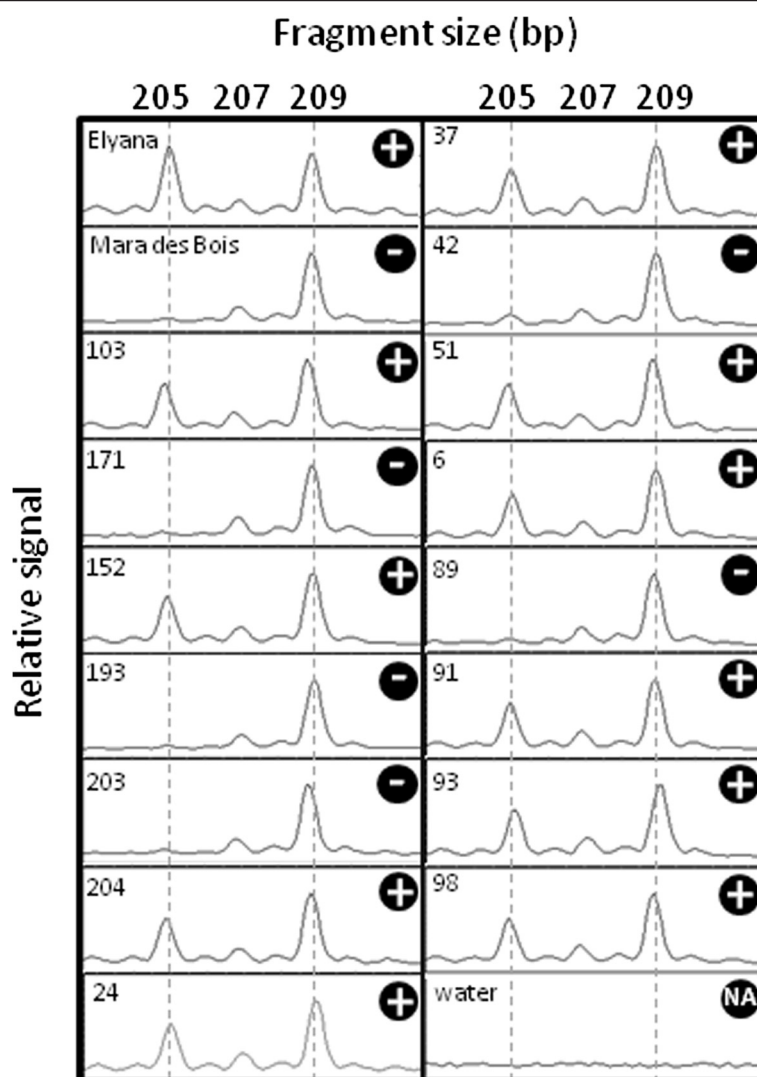
### Discussion

Fruit flavor and aroma profiles are shaped by a mixture of sugars, acids and volatile organic compounds. Individual volatile components have been demonstrated to play key roles in consumer liking, as noted in tomato [28], and strawberry [29]. Several reports have detailed the importance of key compounds to aroma production and the genetic loci or genes that control them [6,16]. One component contributing to flavor in strawberry fruits is



**Figure 8 Co-segregation of the *FaFAD1*, PCR-based marker with the  $\gamma$ -decalactone phenotype.** The *FaFAD1*-based PCR product is denoted by the single arrow and migrates at 500 bp by design. The dashed arrow denotes a positive PCR control (BFACT045) that is located in proximity of *FaFAD1*. Plus and minus signs represent the  $\gamma$ -decalactone phenotype for each genotype as determined by GC/MS. **(A)** The parental lines ‘Elyana’ x ‘Mara des Bois’ are shown with a subset of their progeny. **(B)** A series of wild octoploid materials representing one *F. chiloensis* and seven *F. virginiana* accessions (number = GRIN PI accession). **(C)** The correlation of the PCR product and  $\gamma$ -decalactone phenotype in cultivars unrelated to populations described in this work.





**Figure 9 Cosegregation of the  $\gamma$ -decalactone phenotype and an SSR-based marker 11 kb upstream of *FaFAD1*.** The parents and progeny were tested with primers corresponding to a microsatellite sequence located adjacent to the *FaFAD1* gene. While all genotypes were monomorphic for a PCR product at 209 bp, the 205 bp product was only detected in the ‘Elyana’ parent and any genotype where  $\gamma$ -decalactone was detected, as designated by the “+”.

$\gamma$ -decalactone ( $\gamma$ -D). This compound represents the recognized aromas in peaches and apricots, and is of interest to industry as a flavoring agent. The synthesis of  $\gamma$ -D is not well-understood in plants. However, several microorganisms can generate  $\gamma$ -D from fatty-acid substrates [30] and are used as bioreactors to produce this flavor compound.

To identify transcripts related to  $\gamma$ -D production, two parents varying for the volatile were crossed. ‘Elyana’ is a large, firm berry, and bred for production in Florida [31]. ‘Mara des Bois’ is smaller and soft, and not used in wide commercial cultivation. Fruits (red, ripe fruits of similar size, age, and environment) were assayed for the volatile and also for transcripts associated with receptacle/achene ripening. The RNAseq data from producers and non-producers were computationally bulked to identify sequences common

to each pool (Additional file 1: Figure S1). Strawberry fruit transcriptomes have been examined previously [32-34], and show changes in a substantial number of transcripts throughout the ripening process. The commercial strawberry is octoploid and highly heterozygous [35]. Crossing two plants produces a myriad of progeny phenotypes due to contributions from homoeologous genes. The segregation within subgenomes likely provides additional “noise” that allows for transcripts relevant to the trait to separate clearly from others, focusing the candidate set. Transcriptomes for each plant were considered independently, so the transcriptomes of  $\gamma$ -D producers and non-producers could be compared *in silico*.

The test began with detection of  $\gamma$ -D. The  $\gamma$ -D producers show fluctuations in volatile quantity throughout

the growing season (Figure 1). The genotypes with the highest concentrations were estimated to possess between 0.028 to 0.035 mM  $\gamma$ -D (data not shown). Gamma decalactone accumulation was not ever detected in the 'Mara des Bois' parent. The range of differences observed in parental fruits and progeny presented an ideal situation to assay global gene expression coinciding with volatile production.

When progeny were sorted by presence/absence of  $\gamma$ -D, a small subset of transcripts was significantly different between phenotypic groups. Pairwise comparisons of RNAseq data around a fruit phenotype resolved to an especially strong single candidate, a transcript encoding an omega-6-fatty acid desaturase (*FaFAD1*). This gene is located near a QTL previously reported to explain about 90% of the  $\gamma$ -D phenotype in strawberry [16]. Another gene recently found in peach (*PpFAD\_1B-6*, 71% identity and 82% similar to *FaFAD1* at the protein level) was correlated with  $\gamma$ -D in ripening fruit, and demonstrated  $\omega$ -6 oleate desaturase activity when overexpressed in yeast [21]. Fatty acid desaturases catalyze the formation of double bonds into fatty acyl chains. The report from Sanchez [36], and the biochemistry inferred from protein sequence, each suggest a role for this gene in lactone production, but the precise biochemical steps have not been demonstrated. Future work will examine the role of this transcript in transgenic lines.

*FaFAD1* transcript abundance correlated well with the presence of  $\gamma$ -D. The transcript accumulated with ripening, and only in certain environmental conditions (Figure 5). The ability to produce  $\gamma$ -D segregated as a single dominant locus, consistent with previous reports [16]. A PCR-based survey of materials from the population indicated that the *FaFAD1* genomic sequence could not be amplified from  $\gamma$ -D non-producers, suggesting gene deletion or radical alteration affecting PCR amplification. Eight total combinations using ten distinctly positioned primers were used to amplify regions upstream of and within the *FaFAD1* gene (Shown in Figure 7). In each case, 'Elyana' genomic template amplified the expected fragments, but 'Mara des Bois' produced no amplicons. This is consistent with the idea that a deletion is responsible for the absence of *FaFAD1*-related sequence in the 'Mara des Bois' genome, and this would explain the dominant, single gene effect for the  $\gamma$ -D phenotype. This finding allowed for development of a PCR-based molecular marker to identify genotypes with the potential to produce  $\gamma$ -D. PCR primers corresponding to the 5' sequence of *FaFAD1* were used to amplify the region shown in Figure 7 (primers A and B). The presence of the PCR product co-segregated precisely with the ability to produce the  $\gamma$ -D, in the parental cross, in progeny, and in backcross populations (Figure 8A).

The nature of the deletion is further exemplified using simple sequence repeat (SSR) markers, tools frequently

used to fingerprint strawberry germplasm [37-39]. The availability of the diploid strawberry, *F. vesca*, genome [26] makes it possible to position the microsatellite sequences within the structural context of the strawberry genome. One SSR sequence is located 11 kb adjacent to the *FaFAD1* gene, and polymorphisms would be predicted to segregate with the candidate gene. The results in Figure 8 show that the presence of allele 205 is a reliable predictor of the ability to produce  $\gamma$ -D, at least in the subset of the population evaluated. These data are important because they provide two independent, PCR-based tests can differentiate  $\gamma$ -D producers and non-producers. While not tested outside of the parental genotypes and only in several progeny, this second primer set offers another potential molecular marker that may be used to verify results from the *FaFAD1* sequence. The findings also indicate that the SSR is present in the 'Elyana' parent and not the other, suggesting that the missing *FaFAD1* gene may be part of a larger deletion.

The potential utility of this molecular marker was demonstrated when it was applied to a set of unrelated germplasm that was also varying for presence/absence of  $\gamma$ -D. In this case, the PCR product could only be amplified in genotypes demonstrated to produce  $\gamma$ -D. The evaluation was extended to wild germplasm where fruit could be obtained. Even in these distant accessions, the PCR product was only amplified in lines where  $\gamma$ -D was detectable. The putative marker not only works within a population, but likely will work across commercial breeding populations and wild octoploid accessions. Future studies will track the origin of the gene in diploid germplasm in an attempt to reconstruct its origins, as was done with alleles to trace linalool-producing variants of *FaNES1* [6,40].

There are notable limitations to the approach used in this study. The visual correlation between the  $\gamma$ -D phenotype and qRT-PCR results is shown in Figures 4 and 5. This quantitative trend, however, is not consistent in the RNA-seq RPKM data (data not shown).  $\gamma$ -D producers, irrespective of volatile amount produced, could not be distinguished quantitatively, though producers and non-producers were still easily discernible. This suggests that while the RNA-seq approach worked well for the qualitative  $\gamma$ -D phenotype, uncovering quantitative  $\gamma$ -D effects would be very challenging. This further underscores the need to incorporate independent methods (like qRT-PCR) for examining candidate genes in the early stages of discovery.

The development of two independent molecular markers that segregate with the phenotype in a wide range of germplasm will permit improved parental selection and rapid screening of progeny possessing the ability to produce  $\gamma$ -D. A simple and robust assay can rapidly eliminate individual plants from breeding populations, saving time, fuel, space and other resources. Most importantly,

the marker allows more rapid integration of a fruity volatile into advanced selections.

## Conclusions

The results of this study demonstrate that gene candidates for strawberry fruit traits may be identified by integrating careful phenotyping and transcriptomic analysis with genetics. This approach rapidly reduced the complexity of the octoploid transcriptome down to a single candidate gene. The identified gene, *FaFAD1*, is functionally equivalent to genes involved in  $\gamma$ -D synthesis in fungi and potentially in peach [21]. Gene identification led to the development of a gene-based marker, enabling selection for  $\gamma$ -D producers at the seedling stage. Moving rapidly from candidate to marker has the potential to increase breeding efficiency and reduce the downstream costs associated with maintaining plants lacking a favorable trait. Looking forward,  $\gamma$ -D is only one of many volatiles that can be analyzed with this approach. The same exact dataset may now be re-sorted to identify candidates for other volatiles. The bulk sorting of polyploid transcriptomes is a rapid and cost effective means to identify a testable suite of genes contributing to a given trait.

## Methods

### Plant materials

Parental lines were *Fragaria x ananassa* 'Elyana' female ( $\gamma$ -D producer) and 'Mara des Bois' male ( $\gamma$ -D non-producer). F<sub>1</sub> progeny from the 'Elyana' x 'Mara des Bois' cross were clonally multiplied in a Colorado summer nursery in 2010. Two runner tips of each of approximately 130 seedlings were harvested from the nursery and grown at the Gulf Coast Research and Education Center (GCREC) in Wimauma, Florida during the 2010–11 winter season. These plants were evaluated in the field for horticultural qualities such as yield and fruit size and for volatile diversity using GC/MS. A subset of this population was selected based on flavor volatile diversity and superior plant performance. These selections were further clonally multiplied and grown again in the 2011/12 and 2012/13 seasons.

BC<sub>1</sub> populations were made by crossing one  $\gamma$ -D positive progeny (progeny 98, male) to each parent (female) in 2011. Backcross progeny were increased in the summer nursery as described above. During the 2012/13 season, the progeny were transplanted to the same commercial strawberry growing system in Florida and analyzed for volatiles and other horticultural traits. Other cultivars for marker validation were maintained at GCREC. All other genotypes used in this study were obtained from the Germplasm Resource and Information Network (GRIN) repository in Corvallis, OR, and maintained in the field at GCREC.

### Fruit volatile analysis

All fruiting progeny from the 'Elyana' x 'Mara des Bois' cross were analyzed for volatiles by GC/MS during the 2010/11 season. Harvest dates were January 20, February 11, February 25, and March 18, 2011. Backcross populations were harvested during the 2012/13 season and were harvested on January 13, January 31, and March 7, 2013. Data from the 2010/11 harvests were used to select genotypes segregating for volatiles of interest. Fruit processing for volatile analysis was conducted as follows. A representative ~25 g sample was collected from five to six fully ripe, clean, and normal-shaped berries from each genotype. The calyx from each berry was removed, and berries were blended with an equal weight of saturated NaCl solution (~35% NaCl in molecular biology grade water). The volatile 3-hexanone was added as an internal standard to a final concentration of 1 ppm prior to blending. Five ml aliquots were dispensed into 20 ml glass vials and sealed with magnetic crimp caps (Gerstel, Baltimore, MD, USA). Two technical replicates were processed for each genotype at each harvest. Samples were frozen at -20°C until analysis by Gas Chromatography/Mass Spectroscopy.

### Gas chromatography/mass spectroscopy (GC/MS)

A 2 cm tri-phase SPME fiber (50/30  $\mu$ m DVB/Carboxen/PDMS, Supelco, Bellefonte, PA, USA) was used to collect and concentrate volatiles prior to running on an Agilent 6890 GC coupled with a 5973 N MS detector (Agilent Technologies, Palo Alto, CA, USA). Before analysis, samples were held at 4°C in a Peltier cooling tray attached to a MPS2 autosampler (Gerstel). All other volatile sampling and analysis methods were as previously described [15]. The volatile 3-hexanone was used as an internal control. An authentic  $\gamma$ -D standard (Sigma Aldrich, St. Louis, MO, USA) was run under the same chromatographic conditions as berry samples for verification of volatile identify. The area of each  $\gamma$ -D peak was normalized to the peak area of the internal standard, and normalized peak areas were compared between samples.

### Estimation of $\gamma$ -decalactone in strawberry fruit

A standard curve was made in half ripe fruit of *F. x ananassa* 'Winterstar' fruit puree to estimate the amount of  $\gamma$ -D. This approach mimics volatile detection in ripe fruit. Half red fruit were processed with saturated NaCl and 3-hexanone as described in Fruit Volatile Analysis above. Pure  $\gamma$ -D (Sigma; St. Louis, MO) was added to puree aliquots from 0.005 to 0.3 mM concentrations, vortexed thoroughly, and then analyzed by GCMS as with all other samples. Baseline  $\gamma$ -D was identified in puree sampled without pure volatile added.

### Combined volatile and RNA-seq tissue

Fruit for volatile and RNA-seq analyses were harvested on December 15, 2011. Fourteen progeny and both parents ('Elyana' and 'Mara des Bois') were selected to maximize representation of segregating volatiles. Eight to ten fully-ripe fruit were collected from each genotype, cleaned, the calyx was removed, and then split longitudinally. Half of each progeny's sample was processed for volatile analysis as described above, and the other half was flash frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until RNA extraction. A blank was processed in between each GC/MS sample to minimize cross-sample contamination.

### RNA extraction

Frozen berries were crushed and then ground to a fine powder in a liquid nitrogen-cooled coffee grinder (KitchenAid Blade Coffee Grinder, St. Joseph, MI, USA). RNA was extracted using a modified method [41]. Two grams of fruit powder was used per 5 ml extraction buffer.

RNA was treated with DNase I, RNase-free (Fermentas, Waltham, MA, USA) according to the manufacturer instructions and then cleaned using the Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA, USA). RNA quality was checked on a Bioanalyzer prior to RNA-seq library construction. Each sample was individually barcoded during library construction. Two lanes with eight libraries each were run on an Illumina Genome Analyzer IIx (Illumina, San Diego, CA, USA).

### Transcriptome analysis

Reads that passed quality checks were aligned to the *F. vesca* genome using either a custom script or SeqMan NGen (DNASTAR version 2.3, Lasergene, Madison, WI, USA). *F. vesca* Genemark Hybrid version 1.1 was used for gene annotations. Gene candidates were identified in QSeq (DNASTAR) by making pairwise comparisons between high and low  $\gamma$ -D genotypes. Each pairwise comparison excluded genes less than 2 to 4-fold abundance when comparing high versus low  $\gamma$ -D genotypes. The genes remaining after these comparisons became gene candidates and were analyzed further by qRT-PCR.

A separate analysis was made using CLC Genomic Workbench software v6.5.0 (CLC Bio, Cambridge, MA, USA). The alignment parameters were adjusted to allow for expected variations in the octoploid genome relative to the diploid one (Minimum length fraction = 0.6; Minimum similarity fraction = 0.5; Maximum number of hits for a read = 5). RPKM value was used to normalize the expression data sets among individuals in the population. The data from Figure 3A overexpression data represent transcripts with a RPKM > 10 in at least one of the cultivars. A transcript was considered overexpressed when it was present at least 5 times more frequently in one genotype than the other. The functional classification of the

differentially expressed genes was performed using MapMan ontology. The heatmap used for parents comparison was designed using Mapman software (Mapman version 3.6.0RC1). A dot represents the log<sub>2</sub> of the RPKM ratio of a transcript between the 'Mara des Bois' and 'Elyana' parents.

### qRT-PCR

cDNA templates for qRT-PCR were synthesized using the Impromtu II Reverse Transcriptase kit (Promega, Madison, WI, USA) according to the manufacturer's protocol. The cDNA was diluted 1:10 prior to qRT-PCR analysis. All qRT-PCR reactions were run in 20  $\mu\text{l}$  reactions using EvaGreen qPCR Mastermix-ROX (Applied Biological Materials Inc., Richmond, BC, Canada). Each reaction contained 10  $\mu\text{l}$  2x EvaGreen mastermix, 2  $\mu\text{l}$  primer mix (2  $\mu\text{M}$  each), 1  $\mu\text{l}$  1:10 diluted cDNA, and 7  $\mu\text{l}$  DNase/RNase free water. All qRT PCR primers were designed using qRT primer design tools available online (idtdna.com), and designed to amplify fragments between 95 and 110 base pairs. Each primer-template combination was run with three technical replicates and three biological replicates. A conserved hypothetical protein (*FaCHP1* [42]) was used as a housekeeping control (5' TGCATATATC AAGCAACTTTACTACTGA 3' forward and 5' ATAGC TGAGATGGATCTTCCTGTGA 3' reverse). The qRT PCR was run on an Applied Biosystems StepOnePlus Real-Time PCR System using StepOne Software (v2.0) (Applied Biosystems, Foster City, CA, USA). The qRT-PCR data was analyzed using the comparative  $C_T$  method ( $\Delta\Delta C_T$ ) following the manufacturer's direction.

Candidate genes from the RNA-seq results were validated by qRT-PCR against templates from multiple genotypes, developmental stages, and/or environments. Candidates were initially validated using cDNA templates from all 16 genotypes included in the RNA-seq dataset. Further, candidates were tested for induction during ripening in 'Elyana' (only ripe fruits have detectable  $\gamma$ -D), and two environments with low or high  $\gamma$ -D production. The qRT-PCR sequences for *FaFAD1* (gene24414) were forward 5' GTGCCCTTACTGATAACAAACG 3' and reverse 5' TCGCAACCAATCCCACTC 3', for *CYTp450* (gene 29958) forward 5' ACCCAAAGGTCTATCACATGAC 3' and reverse 5' TGAGCTTCAGTTCCTAACCAC 3', and for *FaFAD2* (gene22642) forward 5' AACTGGTGTC TGGGTCATTG 3' and reverse 5' GAAAGGAGTGAAG GATCAGGC 3'.

### Cloning full length transcript for *FaFAD1*

The full length transcript for gene24414 was cloned using primer sequences guided by the transcriptome data. Primers were designed to include attb sites for Gateway (Invitrogen) cloning into pDONR222. Forward primer 5' AAAAAGCAGGCTGCATGGGAGCCGATACCAAGTT



CGAAGAG 3' and a poly T reverse primer 5' AGAAAGCTGGGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 3' were initially used to obtain the full length transcript including 3' UTR. A second reverse primer was designed from the cloned sequence up to and including the stop codon from the open reading frame (reverse 5' AGAAAGCTGGGTGTTAGTCCGGTACCAAA AACACCTTTGGT 3'). A second round of PCR reconstructed full length attb sites using forward 5' GGGGACAAGTTTGTACAAAAAAGCAGGCT 3' and reverse 5' GGGGACCACTTTGTACAAGAAAGCTGGGT 3' primers. Standard PCR conditions were used with GoTaq polymerase (Promega) according to the manufacturer's recommendations. The full length sequence, predicted protein translation, and physical map coordinates are provided in Additional file 2.

#### Designing a functional molecular marker

A molecular marker for  $\gamma$ -D production was designed using the *F. vesca* genome sequence [26]. Primers were designed to amplify a 500 bp fragment from the 5' end of gene24414 into the 5' UTR: (forward) 5' CGGGA TTAATGGTTTTGTTGTTGACCGACC 3' and (reverse) 5' GTAGAAGAGAGACCAAGACGAG 3'. BFACT045 primers were previously shown to be linked to  $\gamma$ -D production and these primers were used as PCR controls forward 5' CGACAAATGTAGTTGCTAGTCTTCTCA 3' and reverse 5' GAGGCAGAAGTGTTTTTTCGTG 3' [43]. The BFACT045 primers did not produce alleles that segregated with the  $\gamma$ -D phenotype in our populations as tested by capillary electrophoresis (data not shown). All PCR reactions (12.5  $\mu$ l total volume) were run using GoTaq Hot Start polymerase (Promega) according to the manufacturer's instructions. All thermocycler conditions were as follows: 94°C 4 min, followed by 25 cycles of 94°C 30 s, 56°C 30 s, 72°C 30 s, and with a final extension for 10 min at 72°.

The putative marker for gene24414 was tested against a subset of 17 progeny from the 'Elyana' x 'Mara des Bois' cross including parents, and two backcross populations to test segregation with the  $\gamma$ -D phenotype. Progeny were included in this analysis if they had three or more volatile samplings during the season to ensure accurate phenotyping of  $\gamma$ -D non-producers. Backcross population 'Elyana' x progeny 98 produced 26 progeny suitable for analysis, and backcross population 'Mara des Bois' x progeny 98 had 22 progeny suitable for analysis. Other octoploids suitable for consideration as breeding parents and/or for this study were also tested for marker cosegregation with the  $\gamma$ -D phenotype. These genotypes included 'Radiance', 'Deutsch Evern', 'Festival', 'LF9' (a self-pollination of 'Festival'), 'Albion', 'Winterstar'<sup>™</sup> ('FL 05-107'), 'Mieze Schindler', 'Sweet Charlie', and 'Winter Dawn'.

A selection of wild, octoploid genotypes were also tested for the presence of the gene24414 marker, but were only included in this study if a minimum of three volatile samplings were performed. These genotypes are listed here by their Germplasm Resources Information Network Plant Inventory (PI) number and include accessions 236579, 612323, 612495, 612498, and 612499.

#### Statistical marker analysis

Chi-square analysis was performed on marker and volatile data for all genotypes with at least three separate harvests during a season. Due to a strong environmental component, three harvests were necessary to provide the strongest evidence for accurately identifying a  $\gamma$ -D non-producer.

#### SSR marker design

A 50 kb region of the genome surrounding gene24414 was downloaded using the Strawberry Genome Browser at the Genome Database for Rosaceae (rosaceae.org). The web version of BatchPrimer3 [44] was used to search for Simple Sequence Repeats (SSRs) near gene24414. Primers were designed to flank a dinucleotide repeat approximately 11 kb from gene24414. The primer pair (forward 5' TGTAACGACGACGCCAGTGAAGAAG ATGACACTAGGGACGAGGAAG 3' and reverse 5' GTTCTATGTGAGAACATGGGAAGAAACATGAC 3' with fluorescently labeled 5' 6FAM-TGTAACGACGCCAGT 3') exhibited variation between 'Elyana' and 'Mara des Bois' and segregation in the fifteen progeny tested. These primers were used for allele detection during capillary electrophoresis as previously described [44].

#### Availability of supporting data

The RNAseq reads have been deposited into the NCBI Short Read Archive and are accessible under project SRP039356 (<http://www.ncbi.nlm.nih.gov/sra/?term=SRP039356>). Data represent all reads from parental lines ('Mara des Bois' and 'Elyana') as well as reads from individual F1 plants.

#### Additional files

**Additional file 1: Figure S1.** An overview of the process used to identify candidate genes through analysis of bulk segregation of transcripts corresponding to a flavor volatile.

**Additional file 2: The DNA and protein sequences of FAD1.**

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

AHC collected and prepared fruit samples for GCMS, isolate RNA for libraries, analyzed transcript reads, analyzed GCMS results, performed qPCR validations and structural genomic analyses. JP performed computational analyses on RNAseq data. AP performed GCMS analysis, provided interpretations and assisted in preparation of relevant areas of the manuscript. JB provided technical expertise with GCMS. VMW performed initial genetic crosses, raised progeny, collected fruit and assisted in manuscript preparation. KMF



conceived the concept of the study, supervised and coordinated experiments, and prepared the manuscript. All authors have read and approved the manuscript.

#### Acknowledgements

The authors thank Dr. Jeffrey Brecht (University of Florida) for the loan of a Chemstation software license, and Paul V. Stodghill (Robert W. Holley Center for Agriculture and Health, Agriculture Research Service, USDA, Ithaca, NY) for their assistance with the project. The authors also thank Shane Alan Evans and Christy Evans for their assistance in processing fruit samples. This work was performed with funding from the University of Florida Plant Molecular Breeding Initiative, with support from the Florida Strawberry Research and Education Foundation.

#### Author details

<sup>1</sup>Horticultural Sciences Department, University of Florida, Gainesville, FL, USA. <sup>2</sup>Horticultural Research Laboratory, Agriculture Research Service, USDA, Ft. Pierce, FL, USA. <sup>3</sup>Gulf Coast Research and Education Center, University of Florida, Gainesville, FL, USA. <sup>4</sup>Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, FL, USA.

Received: 14 November 2013 Accepted: 17 March 2014

Published: 17 April 2014

#### References

1. Latrasse A: **Fruits**. In *Volatile compounds in foods and beverages*. Edited by Maarse H. New York, NY: Marcel-Decker; 1991:329–387.
2. Olbricht K, Grafe C, Weiss K, Ulrich D: **Inheritance of aroma compounds in a model population of *Fragaria x ananassa* Duch.** *Plant Breeding* 2008, **127**(1):87–93.
3. Schieberle P, Hofmann T: **Evaluation of the character impact odorants in fresh strawberry juice by quantitative measurements and sensory studies on model mixtures.** *J Agric Food Chem* 1997, **45**(1):227–232.
4. Carrasco B, Hancock J, Beaudry R, Retamales J: **Chemical composition and inheritance patterns of aroma in *Fragaria x ananassa* and *Fragaria virginiana* progenies.** *Hort Sci* 2005, **40**(6):1649–1650.
5. Bood KG, Zabetakis I: **The biosynthesis of strawberry flavor (II): biosynthetic and molecular biology studies.** *J Food Sci* 2002, **67**(1):2–8.
6. Aharoni A, Giri AP, Verstappen FW, Berteau CM, Sevenier R, Sun Z, Jongsma MA, Schwab W, Bouwmeester HJ: **Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species.** *Plant Cell* 2004, **16**(11):3110–3131.
7. Larsen M, Poll L: **Odour thresholds of some important aroma compounds in strawberries.** *Z Lebensm Unters Forsch* 1992, **195**(2):120–123.
8. Pérez AG, Sanz C, Ollas R, Ollas JM: **Lipoxygenase and hydroperoxide lyase activities in ripening strawberry fruits.** *J Agric Food Chem* 1998, **47**(1):249–253.
9. Ulrich D, Komes D, Olbricht K, Hoberg E: **Diversity of aroma patterns in wild and cultivated *Fragaria* accessions.** *Genet Resour Crop Evol* 2007, **54**(6):1185–1196.
10. Watson R, Wright CJ, McBurney T, Taylor AJ, Linforth RST: **Influence of harvest date and light integral on the development of strawberry flavour compounds.** *J Exp Bot* 2002, **53**(377):2121–2129.
11. Pelayo-Zaldivar C, Ebeler SE, Kader AA: **Cultivar and harvest date effects on flavor and other quality attributes of California strawberries.** *J Food Quality* 2005, **28**(2005):78–97.
12. Olbricht K, Ulrich D, Weiss K, Grafe C: **Variation in the amounts of selected volatiles in a model population of *Fragaria x ananassa* Duch. As influenced by harvest year.** *J Agric Food Chem* 2011, **59**(3):944–952.
13. Forney C, Kalt W, Jordan M: **The composition of strawberry aroma is influenced by cultivar, maturity, and storage.** *Hort Sci* 2000, **35**(6):1022–1026.
14. Larsen M, Poll L, Olsen CE: **Evaluation of the aroma composition of some strawberry (*Fragaria x ananassa* Duch) cultivars by use of odour threshold values.** *Z Lebensm Unters Forsch* 1992, **195**(6):536–539.
15. Jouquand C, Chandler C, Plotto A, Goodner K: **A sensory and chemical analysis of fresh strawberries over harvest dates and seasons reveals factors that affect eating quality.** *J Am Soc Hortic Sci* 2008, **133**(6):859–867.
16. Zorrilla-Fontanesi Y, Rambla JLL, Cabeza A, Medina JJ, Sánchez-Sevilla JF, Valpuesta V, Botella MA, Granell A, Amaya I: **Genetic analysis of strawberry fruit aroma and identification of O-methyltransferase FaOMT as the locus controlling natural variation in mesiferane content.** *Plant Physiol* 2012, **159**(2):851–870.
17. Akhunov E, Nicolet C, Dvorak J: **Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay.** *Theor Appl Genet* 2009, **119**(3):507–517.
18. Oliver RE, Lazo GR, Lutz JD, Rubenfield MJ, Tinker NA, Anderson JM, Morehead NHW, Adhikary D, Jellen EN, Maughan PJ: **Model SNP development for complex genomes based on hexaploid oat using high-throughput 454 sequencing technology.** *BMC Genomics* 2011, **12**(1):77.
19. Trick M, Long Y, Meng J, Bancroft I: **Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using Solexa transcriptome sequencing.** *Plant Biotechnol J* 2009, **7**(4):334–346.
20. Zenoni S, Ferrarini A, Giacometti E, Xumerle L, Fasoli M, Malerba G, Bellin D, Pezzotti M, Delledonne M: **Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq.** *Plant Physiol* 2010, **152**(4):1787–1795.
21. Sánchez G, Venegas-Calerón M, Salas JJ, Monforte A, Badenes ML, Granell A: **An integrative “omics” approach identifies new candidate genes to impact aroma volatiles in peach fruit.** *BMC Genomics* 2013, **14**(1):343.
22. Schrader J, Etschmann M, Sell D, Hilmer J-M, Rabenhorst J: **Applied biocatalysis for the synthesis of natural flavour compounds—current industrial processes and future prospects.** *Biotechnol Lett* 2004, **26**(6):463–472.
23. Hagedorn S, Kaphammer B: **Microbial biocatalysis in the generation of flavor and fragrance chemicals.** *Annu Rev Microbiol* 1994, **48**(1):773–800.
24. Romero-Guido C, Belo I, Ta TMN, Cao-Hoang L, Alchihab M, Gomes N, Thonart P, Teixeira JA, Destain J, Waché Y: **Biochemistry of lactone formation in yeast and fungi and its utilisation for the production of flavour and fragrance compounds.** *Appl Microbiol Biotechnol* 2011, **89**(3):535–547.
25. Schöttler M, Boland W: **Biosynthesis of dodecano-4-lactone in ripening fruits: Crucial role of an epoxide-hydrolase in enantioselective generation of aroma components of the nectarine (*Prunus persica* var. *nucipersica*) and the strawberry (*Fragaria ananassa*).** *Helv Chim Acta* 1996, **79**(5):1488–1496.
26. Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, Burns P, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Harkins T, Kodira C, Desany B, Crasta OR, Jensen RV, Allan AC, Michael TP, Setubal JC, Celton JM, Rees DJ, Williams KP, Holt SH, Ruiz Rojas JJ, Chatterjee M, et al: **The genome of woodland strawberry (*Fragaria vesca*).** *Nat Genet* 2011, **43**(2):109–116.
27. Folta KM, Dhingra A, Howard L, Stewart PJ, Chandler CK: **Characterization of LF9, an octoploid strawberry genotype selected for rapid regeneration and transformation.** *Planta* 2006, **224**(5):1058–1067.
28. Tieman D, Bliss P, McIntyre L, Blandon-Ubeda A, Bies D, Odabasi A, Rodriguez G, van der Knaap E, Taylor M, Goulet C, Mageroy MH, Snyder DJ, Colquhoun TA, Moskowitz H, Clark DG, Sims C, Bartoshuk L, Klee HJ: **The chemical interactions underlying tomato flavor preferences.** *Curr Biol* 2012, **22**(11):1035–1039.
29. Schwieterman ML, Colquhoun TA, Bartoshuk L, Jaworski EA, Gilbert JL, Tieman DM, Odabasi AZ, Moskowitz HR, Folta KM, Klee HJ, Sims CA, Whitaker VM, Clark DG: **Strawberry flavor: diverse chemical recipes, a seasonal influence and their effect on sensory perception.** *PLoS One* 2014, **9**(2):e88446.
30. Blin-Perrin C, Molle D, Dufosse L, Le-Quere J, Viel C, Mauvais G, Feron G: **Metabolism of ricinoleic acid into gamma-decalactone: beta-oxidation and long chain acyl intermediates of ricinoleic acid in the genus *Sporidiobolus* sp.** *FEMS Microbiol Lett* 2000, **188**(1):69–74.
31. Chandler CK, Santos BM, Peres NA, Jouquand C, Plotto A: **Florida Elyana’ strawberry.** *Hort Sci* 2009, **44**(6):1775–1776.
32. Aharoni A, Keizer LC, Bouwmeester HJ, Sun Z, Alvarez-Huerta M, Verhoeven HA, Blaas J, van Houwelingen AM, de Vos RC, van der Voet H, Jansen RC, Guis M, Mol J, Davis RW, Schena M, van Tunen AJ, O’Connell AP: **Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays.** *Plant Cell* 2000, **12**(5):647–662.
33. Bombarely A, Merchante C, Csukasi F, Cruz-Rus E, Caballero JL, Medina-Escobar N, Blanco-Portales R, Botella MA, Muñoz-Blanco J, Sánchez-Sevilla JF, Valpuesta V: **Generation and analysis of ESTs from strawberry (*Fragaria x ananassa*) fruits and evaluation of their utility in genetic and molecular studies.** *BMC Genomics* 2010, **11**:503.

34. Kang C, Darwish O, Geretz A, Shahan R, Alkharouf N, Liu Z: **Genome-scale transcriptomic insights into early-stage fruit development in woodland strawberry *Fragaria vesca***. *Plant Cell Online* 2013, **25**(6):1960–1978.
35. Ashley MV, Wilk JA, Styan SM, Craft KJ, Jones KL, Feldheim KA, Lewers KS, Ashman TL: **High variability and disomic segregation of microsatellites in the octoploid *Fragaria virginiana* Mill. (Rosaceae)**. *Theor Appl Genet* 2003, **107**(7):1201–1207.
36. Sánchez G, Besada C, Badenes ML, Monforte AJ, Granell A: **A non-targeted approach unravels the volatile network in peach fruit**. *PLoS One* 2012, **7**(6):e38992.
37. Monfort A, Vilanova S, Davis T, Arús P: **A new set of polymorphic simple sequence repeat (SSR) markers from a wild strawberry (*Fragaria vesca*) are transferable to other diploid *Fragaria* species and to *Fragaria* × *ananassa***. *Mol Ecol Notes* 2006, **6**(1):197–200.
38. Brunings A, Moyer C, Peres N, Folta KM: **Implementation of simple sequence repeat markers to genotype Florida strawberry varieties**. *Euphytica* 2010, **173**(1):63–75.
39. Chambers A, Carle S, Njuguna W, Chamala S, Bassil N, Whitaker VM, Barbazuk WB, Folta KM: **A genome-enabled, high-throughput, and multiplexed fingerprinting platform for strawberry (*Fragaria* L.)**. *Mol Breed* 2013, **31**(3):1–15.
40. Chambers A, Whitaker VM, Gibbs B, Plotto A, Folta KM: **Detection of the linalool-producing NES1 variant across diverse strawberry (*Fragaria* spp.) accessions**. *Plant Breed* 2012, **131**(3):437–443.
41. Chang S, Puryear J, Cairney J: **A simple and efficient method for isolating RNA from pine trees**. *Plant Mol Biol Report* 1993, **11**(2):113–116.
42. Clancy MA, Rosli HG, Chamala S, Barbazuk WB, Civello PM, Folta KM: **Validation of reference transcripts in strawberry (*Fragaria* spp.)**. *Mol Genet Genomics* 2013, **288**(12):671–681.
43. Rousseau-Gueutin M, Lerceteau-Köhler E, Barrot L, Sargent DJ, Monfort A, Simpson D, Arus P, Guérin G, Denoyes-Rothan B: **Comparative genetic mapping between octoploid and diploid *Fragaria* species reveals a high level of colinearity between their genomes and the essentially disomic behavior of the cultivated octoploid strawberry**. *Genetics* 2008, **179**(4):2045–2060.
44. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers**. *Methods Mol Biol* 2000, **132**(3):365–386.

doi:10.1186/1471-2164-15-217

**Cite this article as:** Chambers *et al.*: Identification of a strawberry flavor gene candidate using an integrated genetic-genomic-analytical chemistry approach. *BMC Genomics* 2014 **15**:217.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

