The origin and subgenome dynamics of the octoploid strawberries

A. Liston[1] and T.-L. Ashman[2]

[1]Oregon State University, Corvallis, USA; [2]University of Pittsburgh, Pittsburgh, USA.

Abstract

The octoploid strawberries include the cultivated *Fragaria ×ananassa* and its two wild progenitors, *F. chiloensis* and *F. virginiana*. These two progenitors share a common origin an estimated one million years ago. The 2019 publication of the *F. ×ananassa* 'Camarosa' genome sequence has provided new opportunities to identify the diploid ancestry of the four subgenomes comprising the octoploid genome. Studies agree that two of the diploid ancestors are western North American *F. vesca* subsp. *bracteata* and East Asian *F. iinumae*. The identity of the other two subgenomes is controversial. One hypothesis presented in the genome publication implicates the extant diploids *F. nipponica* from Japan and *F. viridis* from western Eurasia. The second hypothesizes that these two subgenomes originated from an extinct species most closely related to *F. iinumae* and may represent an ancestral autotetraploid. A new phylogenomic analysis using a 10-fold increase in the number of data points strongly supports the second hypothesis. This high-resolution analysis finds that chromosomal segments from the *vesca* subgenome have replaced approximately 5% of the other 3 subgenomes. This process of homeologous exchange (HE) has occurred in regions of significantly higher gene content, suggesting an adaptive basis. These regions are expected to have less orthologous gene diversity due to this sequence replacement. Putative HE to the *vesca* subgenome is less frequent (3.9%) and occurs in smaller segments in significantly gene poor regions, indicating that it may be the result of non-adaptive processes. A small fraction (0.23%) of the *vesca* subgenome shows evidence of potential reciprocal HE with one or more of the other three subgenomes. Identifying the diploid progenitors of the octoploid will illuminate the environment and potential selective forces experienced by each subgenome prior to their merger into a single genome. This legacy has shaped the genetic variation available for breeding and crop improvement in strawberry. For this reason, it is critically important to correctly reconstruct these evolutionary events.

Keywords: *Fragaria,* octoploid subgenomes, phylogeny, homeologous exchange

INTRODUCTION

The cultivated garden strawberry, *Fragaria ×ananassa* is a hybrid between the North American *F. virginiana* and the South American *Fragaria chiloensis*. The North American *F. virginiana* was introduced to Europe by 1624, and the South American *F. chiloensis* was first brought to France in 1714 by the French engineer and explorer Amédée-François Frézier (Lee, 1966). However, Frézier only transported five female plants from Chile, unaware that the species is dioecious. This was presumably common knowledge to the Mapuche and Picunche people, who had domesticated *F. chiloensis* at least 1000 years previously (Finn et al., 2013). If Frézier had consulted with these indigenous farmers, he could have avoided his mistake. Frézier lived in Brittany at various times after his return to France, and in the decades after the introduction of *F. chiloensis*, the Plougastel region of Brittany became an important center of strawberry cultivation. At some point the Breton farmers discovered that interplanting *F. virginiana* and *F. chiloensis* resulted in the production of large strawberries in *F. chiloensis* (Lee, 1966). The seeds of these would have been the original *F. ×ananassa*. The Bretons may have kept this information to themselves for some decades, as the "pineapple strawberry" first appears in the written record in the 1750s. Thus the origin of *Fragaria ×ananassa* is between

Acta Hortic. 1309. ISHS 2021. DOI 10.17660/ActaHortic.2021.1309.17
Proc. IX International Strawberry Symposium
Eds.: B. Mezzetti et al.

270 and 300 years ago.

While the origin of *Fragaria* ×*ananassa* can be discerned from the historical record (Lee, 1966), different approaches are needed to decipher the origin of its wild ancestors, *F. virginiana* and *F. chiloensis*. Based on morphological similarities, a relationship between these two species and *F. vesca* has long been suspected. The French botanist Antoine Nicolas Duchesne was the first to propose this connection in his pioneering and comprehensive monograph of *Fragaria* (Duchesne, 1766). In fact, he illustrated this in a figure that is considered only the second published illustration of a phylogenetic network (Morrison, 2014).

It took another 150 years before additional sources of evidence became available to address this question. In some of the first cytogenetic studies to document polyploidy in plants, it was established that these two species and *F.* ×*ananassa* are octoploids, with 28 pairs of chromosomes, compared to seven in the diploid *F. vesca* (Ichijima, 1926; Longley, 1926). Subsequent cytogenetic investigations examined the chromosome behavior at meiosis in interploidy hybrids. These observations were used to develop competing hypotheses for the octoploid subgenome composition. Fedorova (1946) proposed that three distinct diploid species were involved in the ancestry of the octoploids – she associated *F. vesca* with one of the subgenomes, an unknown species with a second, and suggested that the two remaining subgenomes showed "reciprocal homology" indicative of autopolyploidy. Her model for the subgenome composition was AABBBBCC, where CC is *Fragaria vesca*. Senanayake and Bringhurst (1967) conducted similar interploidy crosses, and observed "partial homology" between the AA and CC subgenomes, suggesting a closer relationship than proposed previously. They designated these subgenomes as AA and A'A' and associated them with *F. vesca*, reversing the CC designation of Fedorova. Later, Bringhurst (1990) revised the model to AAA'A'BBB'B' to reflect the absence of multivalent pairing in the octoploids; this model has been generally accepted.

Beginning in the 1990s, the widespread adoption of DNA sequence-based comparisons to estimate phylogenetic relationships provided a new opportunity to examine the origin of the octoploid strawberries. Due to their ease of crossing, a close relationship between *F. virginiana* and *F. chiloensis* has long been suspected. Consistent with this, numerous molecular phylogenetic studies support a single origin of these octoploid progenitors of *F.* ×*ananassa* (Harrison et al., 1997; Potter et al., 2000; Rousseau-Gueutin et al., 2009; Njuguna et al., 2013; DiMeglio et al., 2014; Tennessen et al., 2014; Kamneva et al., 2017; Dillenberger et al., 2018). Likewise, all of these cited studies resolve *F. vesca* as a direct ancestor of the octoploids. Using nearly complete chloroplast genomes, Njuguna et al. (2013) further identified the western North American subspecies *F. vesca* subsp. *bracteata* as the source of the A genome (following the usage of Bringhurst, 1990). This result was replicated in later studies using the plastid (Dillenberger et al., 2018) and nuclear genomes (Tennessen et al., 2014; Edger et al., 2019).

In a landmark phylogenetic analysis of two low-copy nuclear genes, Rousseau-Gueutin et al. (2009) provided the first evidence that *F. iinumae*, a taxonomically isolated species native to Japan and Sakhalin Island, Russia, is a direct ancestor of the octoploids. This result was confirmed in subsequent studies based on nuclear genes (DiMeglio et al., 2014; Kamneva et al., 2017) and indicate that *F. iinumae* is the progenitor of the B genome.

A limitation of the above studies is that the individual nuclear genes cannot be associated with a particular subgenome in the octoploids. To address this, Tennessen et al. (2014), obtained phylogenetically informative SNPs from ca. 2600 targeted sequences used to construct genetic linkage maps in *F. virginiana* and *F. chiloensis* (Tennessen et al., 2014). When the phylogenetically informative SNP is on the same 100-bp sequence read as a linkage mapped SNP, it can be assigned to a subgenome. The result was the first evidence that three of the four octoploid subgenomes share a common ancestry with *Fragaria iinumae*. One of these three subgenomes was consistently resolved as sister to *F. iinumae*. The other two were always part of the *F. iinumae* clade (Figure 1) but were either sister to each other, or paraphyletic relative to *F. iinumae* and its sister subgenome. Based on these results, a new AvAvBiBiB1B1B2B2 model was proposed (Tennessen et al., 2014). The "v" and "i" subscripts indicate that a living diploid progenitor has been identified for the subgenome, while the "1" and "2" have not. Furthermore, the partitioning of homeologous chromosomes between B1 and B2 was provisionally based on their relative divergence from *F. iinumae*, with the more divergent chromosome assigned to B1. It is noteworthy that this is more similar to the model of Fedorova (1946) with a single *F. vesca* subgenome, as opposed to the Bringhurst (1990) model with two *F. vesca* subgenomes.

Sargent et al. (2016) also used a sequence-based genetic linkage mapping approach to assign homeologous chromosomes to subgenomes, although they did not conduct a phylogenetic analysis. They assigned two subgenomes to *F. vesca* and *F. iinumae*, respectively, based on the affinity of linkage mapped sequences. The remaining chromosomes were attributed to "one or more unknown diploid ancestors" and "arbitrarily assigned" to subgenomes designated X1 and X2. Sargent et al. (2016) noted the similarity of their model to the subgenome designations of Fedorova (1946) and Tennessen et al. (2014).

The results of Tennessen et al. (2014) were based on a very small sample size (317-902 SNPs per chromosome) with an average of 64% missing data for each phylogenetic matrix. Considering this sparse data, a genome-scale analysis is desirable to provide confidence in these results. This became feasible with the publication of the octoploid *F. ×ananassa* 'Camarosa' genome (Edger et al., 2019). Relying on the comprehensive collection maintained at the USDA National Germplasm Repository in Corvallis, Oregon (Hummer, 2008), these authors conducting phylogenetic analyses of thousands of octoploid genes together with transcriptomes from 31 accessions representing the 12 diploid species of *Fragaria*. Using a method named PhyDS (phylogenetic identification of subgenomes) they presented a novel hypothesis that each octoploid subgenome originated from a distinct diploid progenitor. The four diploid progenitors were identified as *F. nipponica*, *F. iinumae*, *F. viridis,* and *F. vesca* subsp. *bracteata* (Edger et al., 2019). The *F. nipponica* and *F. viridis* subgenomes correspond to B1 and B2 of Tennessen et al. (2014) and X1 and X2 of Sargent et al. (2016).

Taking an alternative phylogenomic approach, Liston et al. (2020) analyzed an alignment of the four *F.×ananassa* 'Camarosa' subgenomes identified by Edger et al. (2020), seven diploid *Fragaria* genomes, and the outgroup *Potentilla micrantha* to the chromosome-scale *F. vesca* 'Hawaii 4' genome. Phylogenetic analyses were conducted for each of the 7 base chromosomes, and for 2191 non-overlapping 100 kbp windows along the chromosomes (Liston et al., 2020). The results were congruent with those of Tennessen et al. (2014) with three of the four subgenomes part of a clade with a single extant diploid, *F. iinumae* (Figure 1).

In a subsequent re-evaluation of the octoploid subgenomes, the genomes of three additional diploid *Fragaria* species were assembled, and reads from 73 sequenced *F.×ananassa* accessions were mapped to these and the genome assemblies of *F. vesca* and *F. iinumae* (Feng et al., 2020). In every octoploid, approximately ⅓ of the reads mapped to *F. vesca* and *F. iinumae*, and the other third were equally split between *F. viridis, F. nilgerrensis, and F. nubicola.* They concluded that *F. viridis* (and *F. nilgerrensis* and *F. nubicola*) are not diploid progenitors of the octoploid subgenomes. Although they did not sequence *F. nipponica*, it is a close relative of *F. nubicola* (Kamneva et al., 2017), and thus is also unlikely to be a diploid progenitor.

A notable feature of the original PhyDS algorithm (Edger et al., 2019) is that when multiple octoploid genes resolved as sister to the same diploid, these were considered to represent ambiguous orthology, and excluded from the tabulation. This forces each of the four octoploid subgenomes to have a different diploid ancestor, and will not account for unsampled or extinct progenitors, as predicted for the B1 and B2 subgenomes by others (Tennessen et al., 2014; Liston et al., 2020). This limitation was addressed in PhyDS v. 2.1 (Edger et al., 2020). However, in their reanalysis of the original data, they only tabulated cases where the B1 and B2 subgenomes are sister to *F. iinumae*, excluding the Bi subgenome. They did not evaluate how many orthologous loci share the pattern of the B1, B2 and Bi subgenomes sister to *F. iinumae*, and thus did not test the hypothesis of Tennessen et al. (2014) and Liston et al. (2020).

In polyploids, the chromosomes from different subgenomes sharing a common ancestry are called homeologs. In *Fragaria*, all diploid species have seven chromosomes, establishing this as the base number in the genus. Thus in the octoploids, there are four homeologs at each of the seven base chromosomes. Genomic rearrangement is common in newly formed polyploids, and when recombination occurs between homeologs, it is called homeologous exchange (HE). The most commonly detected outcome of this process is duplication-deletion, where the DNA sequence of one subgenome is replaced by sequence from another (Mason and Wendel, 2020). Reciprocal exchange is also a possibility, but it has not been previously documented in octoploid strawberry.

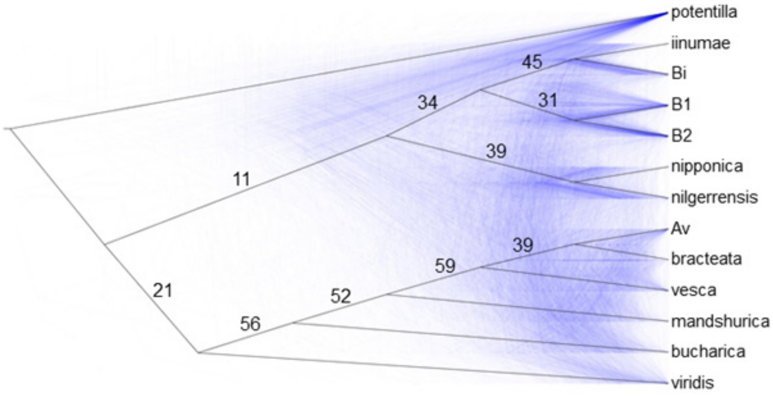

Figure 1. Densitree of 2164 distinct tree topologies for 10-kb windows from the *F. vesca* chromosome 1 alignment, with the Astral consensus of 2420 trees superimposed. Numbers indicate the percentage of trees supporting each node. The four octoploid subgenomes are labeled Av, Bi, B1 and B2.

Homeologous exchange has been observed in many recently derived allopolyploid genomes, and is also common in synthetic polyploids. The first evidence for the presence of HE in octoploid *Fragaria* was obtained by Tennessen et al. (2014), with duplication-deletion detected from the *F. vesca* subgenome to the other three. The sequencing of the *F.* ×*ananassa* 'Camarosa' genome provided further evidence for HE, and revealed that it also occurs into the *F. vesca* subgenome (Edger et al., 2020). The phylogenomic analysis of Liston et al. (2020) also showed this phenomenon.

Incomplete lineage sorting (ILS) is caused when ancestral polymorphisms persist after speciation events, and is well-known to result in phylogenetic trees that are incongruent with the evolutionary history of a group (Pease and Hahn, 2013). The process of ILS can also result in sections of a chromosome having a phylogenetic resolution that differs from the evolutionary origin of a subgenome (Edger et al., 2019; Liston et al., 2020). Phylogenomic approaches have been used to differentiate ILS from interspecific introgression (Edelman et al., 2019; Prüfer et al., 2012; Scally et al., 2012) and the same criteria can be applied in inferring ILS vs. HE. In particular, regions of ILS are expected to be shorter and have a lower gene content than inserts due to HE. This is because regions of ancestral polymorphism become shorter over time due to recombination, and they are more likely to persist in regions of the genome that are not under selection (Pease and Hahn, 2013).

A criticism of the Liston et al. (2020) analysis was that the 100 kbp windows may be too large to reveal the precise evolutionary history of the octoploid origins, especially considering the prevalence of homeologous exchange among subgenomes (Edger et al., 2020). To address this concern, we reanalyze the phylogenetic matrix (Liston et al., 2020) with 10 kbp windows. This also provides the opportunity to more precisely estimate the length and gene number in regions of putative HE. The results are entirely consistent with the results of Liston et al. (2020) and provide insights into the role of homeologous exchange and incomplete lineage sorting in the evolution of the octoploid strawberry genome.

MATERIALS AND METHODS

The analyses reported here are based on a previously published genome-scale phylogenetic matrix, and full methods are provided there (Liston et al., 2020). That matrix was created by dividing the octoploid genome into its four constituent subgenomes, following the chromosome designations of Edger et al. (2019). Each subgenome was converted into 20× coverage of synthetic 100-bp reads. These were aligned to the seven chromosomes of the *F. vesca* 'Hawaii 4' genome (Edger et al., 2018a), along with sequence reads from six diploid species (Liston et al., 2020), and synthetic 100-bp reads from *F. bucharica* (Hirakawa et al., 2014) and the outgroup *Potentilla micrantha* (Buti et al., 2018). The sampling of diploids included the four *Fragaria* species proposed by Edger et al. (2019) as progenitors of the octoploid.

The program MVFtools (Pease and Rosenzweig, 2018) was used to automate phylogenetic reconstruction of a 13-taxon matrix derived from non-overlapping 10-kbp windows from each of the seven base chromosomes. Maximum-likelihood (ML) estimates of phylogeny were obtained with RAxML v. 8.2.12 (Stamatakis, 2014) using the GTR+Γ model of sequence evolution and 100 bootstrap replicates. A taxon was excluded from the analysis of a 10-kb window if it had <10% of aligned sites. The placement of the octoploid subgenomes in each phylogenetic tree were summarized with a custom R script (Liston et al., 2020) available at https://github.com/jacobtennessen/FragariaGenome. The term "sister" was used when a single subgenome resolved as sister to a diploid species; *F. vesca* subsp. *vesca* and subsp. *bracteata* were combined as "vesca sister". If two or more octoploid subgenomes resolved as sister to a single diploid, this was called a "clade". In the summary (Table 1), sister and clade categories were combined for *F. viridis*, *F. nipponica*, and *F. nilgerrensis.*

Table 1. Phylogenetic resolution (percentage of trees) of each of the four octoploid subgenomes in 10-kbp windows (this study) and 100-kbp windows (Liston et al., 2020). The majority phylogenetic assigment of each octoploid subgenome is underlined.

Subgenome	*vesca* sister	*vesca* clade	*iinumae* sister	*iinumae* clade	*viridis*	*nipponica*	*Nilger-rensis*	Ambiguous	Missing
10-kbp windows									
Camarosa_vesca (Av)	51.0	26.2	0.1	1.3	1.4	1.0	0.7	0.2	18.1
Camarosa_iinumae (Bi)	0.8	3.9	54.6	19.8	2.1	1.7	3.5	0.5	13.0
Camarosa_viridis	1.2	5.3	4.9	52.4	4.6	3.9	9.7	0.7	17.3
Camarosa_nipponica	1.4	6.3	4.9	52.5	4.7	3.9	9.9	0.7	15.5
100-kbp windows									
Camarosa_vesca (Av)	60.5	31.3	0.0	0.6	0.5	0.5	0.2	0.0	6.4
Camarosa_iinumae (Bi)	0.2	2.5	80.2	14.2	0.3	0.1	0.5	0.2	1.7
Camarosa_viridis	0.7	3.1	1.6	84.0	1.2	0.8	4.7	0.2	3.7
Camarosa_nipponica	0.5	4.5	1.7	83.6	1.8	1.1	3.9	0.2	2.6

Consensus trees for each chromosome were estimated with ASTRAL v. 5.4 (Zhang et al., 2018) and DensiTree v. 2.01 (Bouckaert, 2010) was used to visualize the distribution of all tree topologies at each chromosome.

The presence of homeologous exchange (HE) or incomplete lineage sorting (ILS) was inferred when one or more adjacent windows showed a phylogenetic resolution (diploid ancestor) that differed from the predominant resolution of that subgenome. Bootstrap resampling (Efron and Tibshirani, 1994) was used to evaluate the gene density in HE vs. non-HE windows. Gene number per window was calculated with BEDTools (Quinlan, 2014) using the *F. vesca* 4.0 a2 revised genome annotation (Li et al., 2019). Gene presence per window was calculated with three different minimum criteria for gene presence: 1 bp, 10%, 20%. BEDTools was also used to calculate gene density per 100-kbp windows in the *F. iinumae* (Edger et al., 2020) and *F. vesca* 4.0 a2 genomes.

The gene count was summed for each of 10,000 bootstrap replicates equivalent to the number of windows observed with 1) putative HE from the *vesca* subgenome to the other three; 2) putative HE into the *vesca* subgenome from the other three; and 3) putative reciprocal HE between the *vesca* subgenome and one or more of the other three subgenomes. A two-tailed *P*-value was calculated as twice the fraction of replicates with sums either below or above the sum observed for these three categories of potential HE.

The assignment of individual chromosomes to subgenomes in octoploid strawberry has varied considerably across genetic linkage maps published since 2010 (reviewed in Hardigan et al., 2020). For the subgenomes associated with *F. vesca* subsp. *bracteata* (Av) and *F. iinumae* (Bi), there is complete agreement in the chromosomal assignments between Tennessen et al. (2014) and Edger et al. (2019). In contrast, the partitioning of the other two subgenomes differs for three of the seven chromosomes (Hardigan et al., 2020). When the two subgenomes are analyzed separately (Tables 1 and 2; Figure 2) we use the designations Camarosa_viridis and Camarosa_nipponica following Edger et al. (2019). Otherwise, we refer to these as subgenomes B1 and B2 when there is no need to distinguish them (Table 3 and throughout the results and discussion), and in Figure 1 where phylogenetic trees for only a single set of homeologs are shown.

RESULTS

A total of 21,932 10-kbp windows were obtained across the sequence alignment to the seven base chromosomes of *Fragaria vesca*. All four octoploid subgenomes were excluded (due to>90% missing data in the alignment) from 164 windows, resulting in 21,768 windows available for phylogenetic assignment. Between 13.0 and 18.1% of the windows had one or more excluded subgenomes due to missing data (Table 1).

Table 2. Number of 10-kbp windows (above), adjusted for consecutive windows (middle), and single window inserts (bottom) of non-predominant diploid sequence into an octoploid subgenome.

Closest diploid						
Subgenome	***vesca***	***iinumae***	***nilgerrensis***	***nipponica***	***viridis***	**Total**
Camarosa_vesca (Av)		310	159	216	304	989
		281	151	178	264	874
		256	143	157	238	794
Camarosa_iinumae (Bi)	1018					
	594					
	459					
Camarosa_nipponica	1688					
	872					
	627					
Camarosa_viridis	141					
	793					
	589					
Total	4120					
	2259					
	1675					

The majority assignment of each chromosome to an octoploid subgenome (Table 1) is identical to the results of Liston et al. (2020). Due to the 10-fold increased sampling, greater heterogeneity is revealed across the genome alignment, leading to a decrease in the size of these majority assignments and an increase in the number of windows showing the rarer assignments (Table 1). There is also a large increase in the amount of missing data per subgenome, the result of using smaller windows for phylogenetic analysis. While there is a 2.5-3.5× increase in the percentage of windows resolving *F. viridis* and *F. nipponica* with the subgenomes attributed to them by Edger et al. (2019, 2020), there is also a 2.0-2.5× increase in windows associating these subgenomes with *F. nilgerrensis* (Table 1). Thus the association of subgenomes B1 and B2 with *F. viridis* and *F. nipponica* is not supported.

High levels of phylogenetic heterogeneity are also apparent in a DensiTree plot of 2164 distinct tree topologies from 10-kbp windows on base chromosome 1 (Figure 1).

Table 3. Total gene count in 10-kb windows representing putative homeologous exchange between the Av subgenome and the Bi, B1 and B2 subgenomes, using three criteria for gene presence in a window. A two-tailed test of significance was estimated from 10,000 bootstrap replicates of summed gene count from the specified number of windows.

Minimum criteria for gene presence in a window				
		1 bp	**10%**	**20%**
3256 vesca windows in Bi, B1, B2	Gene count	6845	4914	3077
	Rank	<0	<0	22
	p-value	<0.0002	<0.0002	0.0044
310 iinumae windows in Av	Gene count	552	368	218
	Rank	9996	>10000	>10000
	p-value	0.0008	<0.0002	<0.0002
49 reciprocal windows	Gene count	97	64	43
	Rank	4297	1719	3948
	p-value	0.8594	0.3438	0.7895

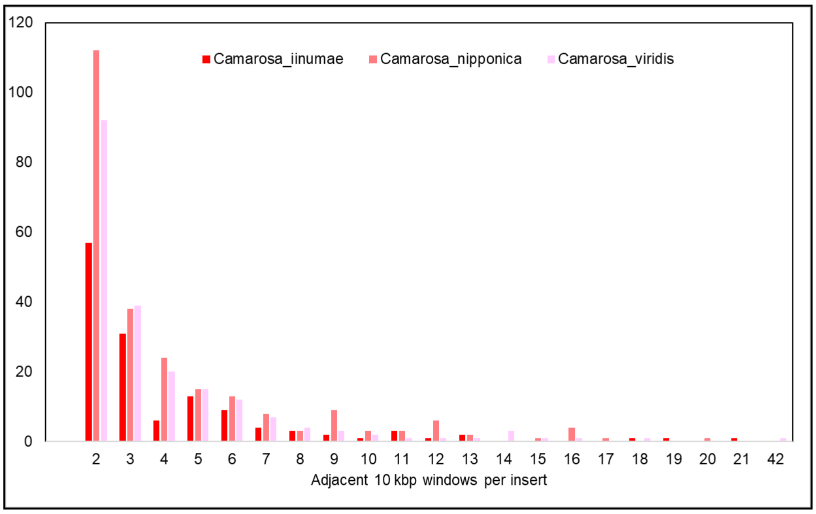

Figure 2. Size distribution of *F. vesca* inserts into the octoploid subgenomes sharing a most recent common diploid ancestor with *F. iinumae* (Bi, B1, B2). Single window inserts not shown, see Table 2 for these values.

However, the ASTRAL integration of 2420 trees resolves the same topology as a concatenation of all sequence from chromosome 1 (Liston et al., 2020). The same topology is obtained in the ASTRAL result for base chromosomes 2-6. Only chromosome 7 differs, with the clade of *F. nipponica* and *F. nilgerrensis* sister to the *vesca* clade (*F. vesca, F. mandshurica,* and *F. bucharica*) and not the clade of *F. iinumae* plus three octoploid subgenomes. At each chromosome, the subgenomes B1 and B2 are resolved as sister to each other, and *F. iinumae* is resolved as the extant diploid species sharing a most recent common ancestor (Figure 1). The phylogenetic position of each subgenome is supported by at least 31% of the 2420 trees obtained for chromosome 1 (Figure 1); similar percentages are obtained for other chromosomes. The clade of *F. viridis* and the octoploid subgenome associated with it by Edger et al. (2019) appears in 1.12% of the 2164 alternative trees, and the clade of *F. nipponica* and its Edger et al. (2019) subgenome in 0.62% of the trees.

A total of 4120 windows in the Bi, B1 and B2 subgenomes resolve as either *vesca* sister or *vesca* clade (Table 2). Of these windows, 729 are found in two of the three subgenomes, and 135 are shared by all three, resulting in 3256 unique windows. These are interpreted as representing homeologous exchange from the Av subgenome to the other three, and comprise approximately 6.3% of these three subgenomes. Counting consecutive windows as a single exchange results in evidence for 2259 HE events (Table 2).

The Av subgenome has 310 windows, representing 281 events, that resolve as *iinumae* sister or *iinumae* clade. These comprise approximately 1.4% of the Av subgenome. Three additional diploids that are not resolved as progenitors of a subgenome (Figure 1) show similar numbers of windows in the Av subgenome (Table 2).

While the majority of phylogenetic trees that suggest either HE or ILS are restricted to a single 10-kbp window (Table 2), the percentage of these small inserts into the Av subgenome is considerably higher than inserts from the Av subgenome into the other three. Inserts found in two or more adjacent windows must be longer, and these are also much more common in the the Bi, B1 and B2 subgenomes (Figure 2) compared to inserts into the Av subgenome (Figure 3).

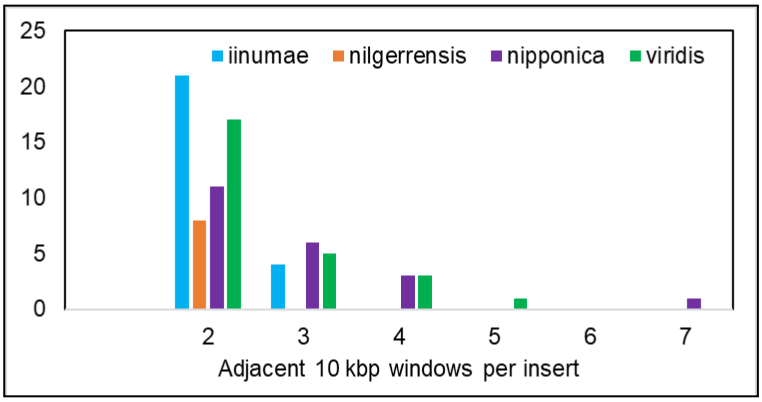

Figure 3. Size distribution of inserts associated with diploid species into the Av octoploid subgenome. Single window inserts not shown, see Table 2 for these values.

The 3256 *vesca* windows in the Bi, B1, and B2 subgenomes are significantly enriched for genes compared to 10,000 random replicates of 3256 windows (Table 3). Conversely, the 310 *iinumae* windows in the Av subgenome have a significantly lower gene count than 10,000 random replicates. The 49 windows representing putative reciprocal exchange are neither significantly enriched nor depleted of genes, which is not unexpected considering these windows are a subset of the two other comparisons. While p-values vary depending on how gene presence is counted, the significance level remains unchanged for each comparison (Table 3).

The *Fragaria iinumae* genome has 23540 genes and an average density of 9.823 genes 100 kbp^{-1}, while the *F. vesca* genome has 33953 genes and an average density of 15.496 genes 100 kbp^{-1}.

A total of 61 windows may represent reciprocal HE between the Av subgenome and one or more of Bi, B1 and B2 subgenomes. Only two pairs of windows are consecutive. Ten windows show a reciprocal pattern in three subgenomes, and one represents potential HE among all four octoploid subgenomes, resulting in 49 unique windows relative to the *F. vesca* reference genome.

DISCUSSION

The results presented here provide further evidence that only two extant diploids, *Fragaria vesca* subsp. *bracteata* and *F. iiinumae* are direct ancestors of the octoploid strawberry species that gave rise to the cultivated strawberry (Tennessen et al., 2014; Liston et al., 2020; Feng et al., 2020; Session and Rokhsar, 2020). The previous identification of *F. nipponica* and *F. viridis* as progenitors of the octoploid strawberry (Edger et al., 2019, 2020) was apparently the result of enforcing models of evolutionary relationships that exclude the most likely phylogenetic hypothesis (Figure 1). By ignoring the predominant signal, less common topologies were given greater significance than warranted.

Fragaria iinumae is a morphologically and ecologically isolated species in the genus (Staudt, 2005), and it has no close phylogenetic relatives among extant diploids (Njuguna et al., 2013; Kamneva et al., 2017). While the Bi subgenome is closely related to *F. iinumae*, the relationship of B1 and B2 to *F. iinumae* is more distant, comparable to the relationships among the morphologically similar species *F. bucharica*, *F. mandshurica*, and *F vesca*. While we have previously speculated that these subgenomes could originate with an undiscovered relative of *F. iinumae*, we think it is more likely that the progenitor or progenitors of the B1 and B2 subgenomes are extinct. Extinct diploid ancestors have been inferred for other polyploids, and this phenomenon may not be uncommon (Edger et al., 2018b).

The sister group relationship between the B1 and B2 subgenomes is suggestive of autopolyploidy, and this is consistent with early cytogenetic observations (Fedorova, 1946). However, we do not know if the octoploid originated from the hybridization of two tetraploids, or between a diploid and a hexploid. The latter was proposed by Edger et al. (2019) and they hypothesized that *F. moschata* was this ancestral hexaploid which hybridized with *F. vesca* subsp. *bracteata*. Sequence-based genetic linkage mapping found no support for *F. iinumae, F. nipponica* nor *F. viridis* ancestry in *F. moschata* (Liston et al., 2020).

If we cannot associate two octoploid subgenomes with *F. nipponica* and *F. viridis,* a remaining question is how to partition two sets of homeologous chromosomes between the B1 and B2 subgenomes. Genetic linkage mapping and our phylogenomic approach do not address this. Session and Rokhsar (2020) have recently used short sequences that are associated with repetitive elements and unique to a subgenome to provide a novel division of the B1 and B2 chromosomes in the octoploid strawberry genome.

In addition to confirming the results obtained with 100-kbp windows (Liston et al., 2020), the 10-fold increase in genomic sampling provided more details on genomic regions that deviate from the predominant phylogenetic results. Consistent with previous analyses (Tennessen et al., 2014; Edger et al., 2019) we interpret the presence of *F. vesca* sequence in the Bi, B1 and B2 as evidence for homeologous exchange from the Av subgenome. These HE regions have a significantly larger number of genes (Table 3). While 74% of these *F. vesca* inserts were restricted to a single 10-kbp window (Table 2), the other 26% were found in two or more consecutive windows (Figure 2), indicative of a size larger than 10 kbp. Without finer scale analyses the precise size of these remains unknown. Using a conservative estimate that the insert covers at least half of a 10-kbp window, the largest HE events per subgenome range from 190 kbp (20 windows) to 410 kbp (42 windows).

In contrast, incomplete lineage sorting (ILS) is a more likely explanation for presence of sequences associated with *F. iinumae* in the Av subgenome. These regions are much shorter than the *F. vesca* sequence tracts in the B subgenomes (Figure 3), and they have significantly fewer genes (Table 3). It is also noteworthy that similar results are observed for sequences in the Av genome that are associated with *F. nilgerrensis, F. nipponica* and *F. viridis* (Table 3; Figure 3). Since we have no evidence that these diploid species are progenitors of octoploid subgenomes, the most likely explanation is ILS. These patterns are also consistent with regions of ILS from gorilla and bonobo in in the human genome (Prüfer et al., 2012; Scally et al., 2012) and more distantly related species in *Heliconius* butterfly genomes (Edelman et al., 2019). A caveat is that the *F. iinumae* genome has 44% fewer genes and much lower gene density than *F. vesca*, so we would also expect fewer genes in true HE tracts from *F. iinumae*. However, the above analyses are based solely on the *F. vesca* genome annotation, so any bias resulting from this would cause us to overestimate the gene density in these putative regions of ILS. Comparing our results to ones based on the *F. ×ananassa* genome annotation is a topic for future analyses.

The phylogenomic approach utilized here and in Liston et al. (2020) relies on the mapping of 100-bp reads to a single reference genome. This has the potential to create incorrect sequence homology assessment across divergent species (Edger et al., 2020). We agree with this concern, but counter that this type of error should result in arbitrary perturbations of the phylogenetic signal, and inconsistent results among the 7 base chromosomes. The fact that this was not observed suggests that incorrect read mapping is unlikely to be influencing the overall phylogenetic results. However, some of the inconsistencies among 10-kbp windows could certainly be the result of homology error. Whole genome alignments are becoming increasingly feasible (Armstrong et al., 2020) and represent a promising approach for subsequent studies.

An example of the increased resolution of HE events can be seen at the dehydroascorbate reductase (DHAR) locus. This locus was used to estimate the phylogeny of *Fragaria* (Rousseau-Gueutin et al., 2009). In the octoploid samples, 3 alleles resolved with *F. vesca*, suggesting contributions from at least 2 subgenomes. Tennessen et al. (2014) predicted that the locus would fall within an HE region, but it was 6 kbp outside of the nearest event they observed, which was estimated to be as large as 386 kbp. In our study, a DHAR locus is found within an HE tract of 40-50 kbp on Bi subgenome chromosome 7, fulfilling the prediction of Tennessen et al. (2014), and greatly refining the size of the insertion.

CONCLUSIONS

Accurate knowledge of the ancestry of a polyploid can have practical applications. Each diploid progenitor brings a unique evolutionary legacy to the polyploid. This legacy includes potential adaptations to different climates and unique pathogens experienced prior to the formation of the polyploid. For example, if one diploid was better adapted to saline conditions, crop improvement efforts in this area could be directed to that subgenome. The correct identification of diploid ancestors can also inform efforts to conserve and utilize these wild relatives (Liston et al. 2014; Wei et al., 2020).

ACKNOWLEDGEMENTS

The authors thank Jacob Tennessen for assistance with analyses. Funding was provided by the US National Science Foundation grants DEB1912203 to A.L. and DEB 1912180 to T.-L.A.

Literature cited

Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I.T., Novak, A.M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., et al. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature *587* (*7833*), 246–251 https://doi.org/10.1038/s41586-020-2871-y. PubMed

Bouckaert, R.R. (2010). DensiTree: making sense of sets of phylogenetic trees. Bioinformatics *26* (*10*), 1372–1373 https://doi.org/10.1093/bioinformatics/btq110. PubMed

Bringhurst, R.S. (1990). Cytogenetics and evolution in American *Fragaria.* HortScience *25* (*8*), 879–881 https://doi.org/10.21273/HORTSCI.25.8.879.

Buti, M., Moretto, M., Barghini, E., Mascagni, F., Natali, L., Brilli, M., Lomsadze, A., Sonego, P., Giongo, L., Alonge, M., et al. (2018). The genome sequence and transcriptome of *Potentilla micrantha* and their comparison to *Fragaria vesca* (the woodland strawberry). Gigascience *7* (*4*), 1–14 https://doi.org/10.1093/gigascience/giy010. PubMed

Dillenberger, M.S., Wei, N., Tennessen, J.A., Ashman, T.-L., and Liston, A. (2018). Plastid genomes reveal recurrent formation of allopolyploid *Fragaria.* Am J Bot *105* (*5*), 862–874 https://doi.org/10.1002/ajb2.1085. PubMed

DiMeglio, L.M., Staudt, G., Yu, H., and Davis, T.M. (2014). A phylogenetic analysis of the genus *Fragaria* (strawberry) using intron-containing sequence from the ADH-1 gene. PLoS One *9* (*7*), e102237 https://doi.org/10.1371/journal.pone.0102237. PubMed

Duchesne, A.-N. (1766). Histoire Naturelle des Fraisiers (Paris: chez Didot le jeune).

Edelman, N.B., Frandsen, P.B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R.B., García-Accinelli, G., Van Belleghem, S.M., Patterson, N., Neafsey, D.E., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. Science *366* (*6465*), 594–599 https://doi.org/10.1126/science.aaw2090. PubMed

Edger, P.P., McKain, M.R., Bird, K.A., and VanBuren, R. (2018a). Subgenome assignment in allopolyploids: challenges and future directions. Curr Opin Plant Biol *42*, 76–80 https://doi.org/10.1016/j.pbi.2018.03.006. PubMed

Edger, P.P., VanBuren, R., Colle, M., Poorten, T.J., Wai, C.M., Niederhuth, C.E., Alger, E.I., Ou, S., Acharya, C.B., Wang, J., et al. (2018b). Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (*Fragaria vesca*) with chromosome-scale contiguity. Gigascience *7* (*2*), 1–7 https://doi.org/10.1093/gigascience/gix124. PubMed

Edger, P.P., Poorten, T.J., VanBuren, R., Hardigan, M.A., Colle, M., McKain, M.R., Smith, R.D., Teresi, S.J., Nelson, A.D.L., Wai, C.M., et al. (2019). Origin and evolution of the octoploid strawberry genome. Nat Genet *51* (*3*), 541–547 https://doi.org/10.1038/s41588-019-0356-4. PubMed

Edger, P.P., McKain, M.R., Yocca, A.E., Knapp, S.J., Qiao, Q., and Zhang, T. (2020). Reply to: revisiting the origin of octoploid strawberry. Nat Genet *52* (*1*), 5–7 https://doi.org/10.1038/s41588-019-0544-2. PubMed

Efron, B., and Tibshirani, R.J. (1994). An Introduction to the Bootstrap (CRC Press).

Fedorova, N. (1946). Crossability and phylogenetic relations in the main European species of *Fragaria.* Compilation of the National Academy of Sciences USSR *52*, 545–547.

Feng, C., Wang, J., Harris, A.J., Folta, K.M., Zhao, M., and Kang, M. (2020). Tracing the diploid ancestry of the cultivated octoploid strawberry. Mol. Biol. Evol. PubMed

Finn, C.E., Retamales, J.B., Lobos, G.A., and Hancock, J.F. (2013). The Chilean strawberry (*Fragaria chiloensis*): over 1000 years of domestication. HortScience *48* (*4*), 418–421 https://doi.org/10.21273/HORTSCI.48.4.418.

Hardigan, M.A., Feldmann, M.J., Lorant, A., Bird, K.A., Famula, R., Acharya, C., Cole, G., Edger, P.P., and Knapp, S.J. (2020). Genome synteny has been conserved among the octoploid progenitors of cultivated strawberry over millions of years of evolution. Front Plant Sci *10*, 1789 https://doi.org/10.3389/fpls.2019.01789. PubMed

Harrison, R., Luby, J., Furnier, G., and Hancock, J. (1997). Morphological and molecular variation among populations of octoploid *Fragaria virginiana* and *F. chiloensis* (*Rosaceae*) from North America. Am J Bot *84* (*5*), 612–620 https://doi.org/10.2307/2445897. PubMed

Hirakawa, H., Shirasawa, K., Kosugi, S., Tashiro, K., Nakayama, S., Yamada, M., Kohara, M., Watanabe, A., Kishida, Y., Fujishiro, T., et al. (2014). Dissection of the octoploid strawberry genome by deep sequencing of the genomes of *Fragaria* species. DNA Res *21* (*2*), 169–181 https://doi.org/10.1093/dnares/dst049. PubMed

Hummer, K.E. (2008). Global Conservation Strategy for *Fragaria* (Strawberry). Scripta Hortic. *6*.

Ichijima, K. (1926). Cytological and genetic studies on *Fragaria.* Genetics *11* (*6*), 590–604 https://doi.org/10.1093/genetics/11.6.590. PubMed

Kamneva, O.K., Syring, J., Liston, A., and Rosenberg, N.A. (2017). Evaluating allopolyploid origins in strawberries (*Fragaria*) using haplotypes generated from target capture sequencing. BMC Evol Biol *17* (*1*), 180 https://doi.org/10.1186/s12862-017-1019-7. PubMed

Lee, V. (1966). The Strawberry - History, Breeding and Physiology, Chapter 3–5 (New York: Holt, Rinehart & Winston), p.15–72.

Li, Y., Pi, M., Gao, Q., Liu, Z., and Kang, C. (2019). Updated annotation of the wild strawberry *Fragaria vesca* V4 genome. Hortic. Res. *6*, 1–9.

Liston, A., Cronn, R., and Ashman, T.-L. (2014). *Fragaria*: a genus with deep historical roots and ripe for evolutionary and ecological insights. Am J Bot *101* (*10*), 1686–1699 https://doi.org/10.3732/ajb.1400140. PubMed

Liston, A., Wei, N., Tennessen, J.A., Li, J., Dong, M., and Ashman, T.-L. (2020). Revisiting the origin of octoploid strawberry. Nat Genet *52* (*1*), 2–4 https://doi.org/10.1038/s41588-019-0543-3. PubMed

Longley, A. (1926). Chromosomes and their significance in strawberry classification. J. Agric. Res. *32*, 559.

Mason, A.S., and Wendel, J.F. (2020). Homoeologous exchanges, segmental allopolyploidy, and polyploid genome evolution. Front Genet *11*, 1014 https://doi.org/10.3389/fgene.2020.01014. PubMed

Morrison, D.A. (2014). Is the tree of life the best metaphor, model, or heuristic for phylogenetics? Syst Biol *63* (*4*), 628–638 https://doi.org/10.1093/sysbio/syu026. PubMed

Njuguna, W., Liston, A., Cronn, R., Ashman, T.-L., and Bassil, N. (2013). Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. Mol Phylogenet Evol *66* (*1*), 17–29 https://doi.org/10.1016/j.ympev.2012.08.026. PubMed

Pease, J.B., and Hahn, M.W. (2013). More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. Evolution *67* (*8*), 2376–2384 https://doi.org/10.1111/evo.12118. PubMed

Pease, J.B., and Rosenzweig, B.K. (2018). Encoding data using biological principles: the multisample variant format for phylogenomics and population genomics. IEEE/ACM Trans Comput Biol Bioinform *15* (*4*), 1231–1238 https://doi.org/10.1109/TCBB.2015.2509997. PubMed

Potter, D., Luby, J.J., and Harrison, R.E. (2000). Phylogenetic relationships among species of *Fragaria* (*Rosaceae*) inferred from non-coding nuclear and chloroplast DNA sequences. Syst. Bot. *25* (*2*), 337–348 https://doi.org/10.2307/2666646.

Prüfer, K., Munch, K., Hellmann, I., Akagi, K., Miller, J.R., Walenz, B., Koren, S., Sutton, G., Kodira, C., Winer, R., et al. (2012). The bonobo genome compared with the chimpanzee and human genomes. Nature *486* (*7404*), 527–531 https://doi.org/10.1038/nature11128. PubMed

Quinlan, A.R. (2014). BEDTools: the Swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics *47* (*1*), 1–34, 34 https://doi.org/10.1002/0471250953.bi1112s47. PubMed

Rousseau-Gueutin, M., Gaston, A., Aïnouche, A., Aïnouche, M.L., Olbricht, K., Staudt, G., Richard, L., and Denoyes-Rothan, B. (2009). Tracking the evolutionary history of polyploidy in *Fragaria* L. (strawberry): new insights from phylogenetic analyses of low-copy nuclear genes. Mol Phylogenet Evol *51* (*3*), 515–530 https://doi.org/10.1016/j.ympev.2008.12.024. PubMed

Sargent, D.J., Yang, Y., Šurbanovski, N., Bianco, L., Buti, M., Velasco, R., Giongo, L., and Davis, T.M. (2016). HaploSNP affinities and linkage map positions illuminate subgenome composition in the octoploid, cultivated strawberry (*Fragaria×ananassa*). Plant Sci *242*, 140–150 https://doi.org/10.1016/j.plantsci.2015.07.004. PubMed

Scally, A., Dutheil, J.Y., Hillier, L.W., Jordan, G.E., Goodhead, I., Herrero, J., Hobolth, A., Lappalainen, T., Mailund, T., Marques-Bonet, T., et al. (2012). Insights into hominid evolution from the gorilla genome sequence. Nature *483* (*7388*), 169–175 https://doi.org/10.1038/nature10842. PubMed

Senanayake, Y., and Bringhurst, R. (1967). Origin of *Fragaria* polyploids. I. Cytological analysis. Am. J. Bot. *54* (*2*), 221–228 https://doi.org/10.1002/j.1537-2197.1967.tb06912.x.

Session, A., and Rokhsar, D. (2020). Discovering subgenomes of octoploid strawberry with repetitive sequences. BioRxiv 2020.11.04.330431.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30* (*9*), 1312–1313 https://doi.org/10.1093/bioinformatics/btu033. PubMed

Staudt, G. (2005). Notes on Asiatic *Fragaria* species: IV. *Fragaria iinumae.* Botanische Jahrbücher *126* (*2*), 163–175 https://doi.org/10.1127/0006-8152/2005/0126-0163.

Tennessen, J.A., Govindarajulu, R., Ashman, T.-L., and Liston, A. (2014). Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. Genome Biol Evol *6* (*12*), 3295–3313 https://doi.org/10.1093/gbe/evu261. PubMed

Wei, N., Du, Z., Liston, A., and Ashman, T.-L. (2020). Genome duplication effects on functional traits and fitness are genetic context and species dependent: studies of synthetic polyploid *Fragaria.* Am J Bot *107* (*2*), 262–272 https://doi.org/10.1002/ajb2.1377. PubMed

Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics *19* (*S6*, *Suppl 6*), 153 https://doi.org/10.1186/s12859-018-2129-y. PubMed