3-28-2016

A High-Density Linkage Map of the Ancestral Diploid Strawberry, Fragaria iinumae, Constructed with Single Nucleotide Polymorphism Markers from the IStraw90 Array and Genotyping by Sequencing

Lise L. Mahoney
University of New Hampshire, Durham, Lise.Mahoney@unh.edu

Daniel J. Sargent
San Michele all'Adige

Feseha Abebe-Akele
University of New Hampshire, Durham

David J. Wood
University of New Hampshire, Durham

Judson A. Ward
Driscoll's

Recommended Citation

Mahoney, L. L., D. J. Sargent, F. Abebe-Akele, D. J. Wood, J. A. Ward, N. V. Bassil, J. F. Hancock, K. M. Folta, and T. M. Davis. 2016. A High-Density Linkage Map of the Ancestral Diploid Strawberry, Fragaria iinumae, Constructed with Single Nucleotide Polymorphism Markers from the IStraw90 Array and Genotyping by Sequencing. Plant Genome 9. https://dx.doi.org/10.3835/plantgenome2015.08.0071

Authors

Lise L. Mahoney, Daniel J. Sargent, Feseha Abebe-Akele, David J. Wood, Judson A. Ward, Nahla V. Bassil, James F. Hancock, Kevin M. Folta, and Thomas M. Davis

A High-Density Linkage Map of the Ancestral Diploid Strawberry, *Fragaria iinumae*, Constructed with Single Nucleotide Polymorphism Markers from the IStraw90 Array and Genotyping by Sequencing

Lise L. Mahoney,* Daniel J. Sargent, F. Abebe-Akele, Dave J. Wood, Judson A. Ward, Nahla V. Bassil, James F. Hancock, Kevin M. Folta, and Thomas M. Davis

Abstract

Fragaria iinumae Makino is recognized as an ancestor of the octoploid strawberry species, which includes the cultivated strawberry, *Fragaria ×ananassa* Duchesne ex Rozier. Here we report the construction of the first high-density linkage map for *F. iinumae*. The *F. iinumae* linkage map (Fii map) is based on two high-throughput techniques of single nucleotide polymorphism (SNP) genotyping: the IStraw90 Array (hereafter "Array"), and genotyping by sequencing (GBS). The F_2 generation mapping population was derived by selfing *F. iinumae* hybrid F1D, the product of a cross between two divergent *F. iinumae* accessions collected from Hokkaido, Japan. The Fii map consists of seven linkage groups (LGs) and has an overall length of 451.7 cM as defined by 496 loci populated by 4173 markers: 3280 from the Array and 893 from GBS. Comparisons with two versions of the *Fragaria vesca* ssp. *vesca* L. 'Hawaii 4' pseudo-chromosome (PC) assemblies reveal substantial conservation of synteny and colinearity, yet identified differences that point to possible genomic divergences between *F. iinumae* and *F. vesca*, and/or to *F. vesca* genomic assembly errors. The Fii map provides a basis for anchoring a *F. iinumae* genome assembly as a prerequisite for constructing a second diploid reference genome for *Fragaria*.

THE OCTOPLOID ($2n = 8x = 56$) genomes of the cultivated strawberry (*Fragaria ×ananassa* Duchesne ex Rozier and its immediate ancestors *Fragaria chiloensis* (L.) Miller and *Fragaria virginiana* Duchesne are recognized as having at least partially allopolyploid compositions (Rousseau-Gueutin et al., 2009; DiMeglio et al., 2014). As such, the ancestries of these octoploids must trace to two or more diploid ancestors via processes of hybridization and doubling of chromosome number. Based on their allelic signatures in phylogenetic studies, ancestors of the diploids *F. vesca* L. and *F. iinumae* Makino have emerged as likely subgenome contributors

L.L. Mahoney, D.J. Wood, and T.M. Davis, Dep. of Biological Sciences, Rudman Hall, Univ. of New Hampshire, Durham, NH 03824; D.J. Sargent, Fondazione Edmund Mach, Research and Innovation Centre, San Michele all'Adige, 38010 Trento, Italy, and Driscoll's Genetics Limited, New Road, East Malling, Kent ME19 6BJ, UK; F. Abebe-Akele, Hubbard Center for Genome Studies, Univ. of New Hampshire, 444 Gregg Hall, Durham, NH 03824; J.A. Ward, Molecular Biology Dep., Driscoll's, 151 Silliman Rd, Watsonville, CA 95076; N.V. Bassil, USDA-ARS, National Clonal Germplasm Repository, 33447 Peoria Rd., Corvallis, OR 97333; J.F. Hancock, Dep. of Horticulture, 342 Plant and Soil Sciences Building, Michigan State Univ., East Lansing, MI 48824; K.M. Folta, Horticultural Sciences Dep., Univ. of Florida, Gainesville, FL 32611. Received 20 Aug. 2015. Accepted 29 Nov. 2015. *Corresponding author (lise.mahoney@unh.edu).

Abbreviations: Fii map, *Fragaria iinumae* linkage map; Fvb, *F. vesca* 'Hawaii 4' genome assembly anchored to a *F. vesca* ssp. *bracteata* linkage map; FvH, *F. vesca* 'Hawaii 4' v1.1 genome assembly anchored to the FV × FB linkage map; GBS, genotyping by sequencing; GDR, Genome Database for Rosaceae; LG, linkage group; NCGR CFRA, National Clonal Germplasm Repository, Corvallis collection of *Fragaria*; PC, pseudo-chromosome; PHR, Poly High Resolution; RB, reference-based; RZ, rearrangement zone; SNP, single nucleotide polymorphism; SSR, simple sequence repeat; UNEAK, Universal Network Enabled Analysis Kit.

to the octoploids (Folta and Davis, 2006; Rousseau-Gueutin et al., 2009; DiMeglio et al., 2014; Tennessen et al., 2014; Sargent et al., 2015). On the basis of its many favorable features (Folta and Davis, 2006), including a diverse spectrum of available germplasm, *F. vesca* was adopted as the initial genomic model species for the genus *Fragaria* (Sargent et al., 2011; Shulaev et al., 2011), with relevance to all of the Rosaceae (Shulaev et al., 2008).

Early molecular phylogenetic studies of *Fragaria* that were based on chloroplast DNA and nuclear internal transcribed spacer regions did not discern the ancestral role of *F. iinumae* in the evolution of the octoploid *Fragaria* species (Potter et al., 2000). The first molecular insight into its ancestral role came from the nuclear protein-encoding gene for alcohol dehydrogenase (*ADH*) as first reported by DiMeglio and Davis in 2003 (Folta and Davis, 2006; DiMeglio et al., 2014). Confirmatory evidence from two other nuclear, protein-encoding genes—those for granule-bound starch synthase I (*GBSSI*-2) and dehydroascorbate reductase (*DHAR*)—was provided by Rousseau-Gueutin et al. (2009). The initial establishment of the contributory role of *F. iinumae* in 2003 prompted expansion of the USDA assemblage of *F. iinumae* germplasm via a collection expedition in 2004 to Japan (Hummer et al., 2006), where *F. iinumae* is endemic (Staudt, 2005). The acquired *F. iinumae* accessions, which are archived by the USDA National Clonal Germplasm Repository (NCGR) in Corvallis, Oregon, represent a broad range of molecular diversity as assessed through interrogation with simple sequence repeat (SSR) markers (Njuguna et al., 2011), thus providing a diverse germplasm resource for genomic research in *F. iinumae*.

At present, the only published reference genomes for strawberry, are versions of the *F. vesca* 'Hawaii 4' genome. The initial anchoring of *F. vesca* 'Hawaii 4' scaffolds to produce PCs (Shulaev et al., 2011) utilized a linkage map based on an F_2 mapping population derived from a cross of *F. vesca* '815' by *F. bucharica* Losinsk '601' (FV × FB) (Sargent et al., 2011). This mapping population was genotyped with SSR markers using both the full mapping population and bin mapping (Sargent et al., 2009). Since its initial publication as v1.0 (Shulaev et al., 2011), the *F. vesca* 'Hawaii 4' reference assembly has been revised twice. First, Sargent et al. (2011) added newly developed SSR markers to the FV × FB map, enabling additional scaffold anchoring that resulted in *F. vesca* 'Hawaii 4' v1.1 (FvH). More recently, Tennessen et al. (2014) re-anchored the *F. vesca* 'Hawaii 4' assembly using new linkage maps developed in *F. vesca* ssp. *bracteata* (Tennessen et al., 2013, 2014), resulting in another *F. vesca* 'Hawaii 4' assembly (Fvb).

Assembly of a reference genome for a second ancestral diploid, *F. iinumae*, would provide a valuable new genomic resource for the *Fragaria* research community. As a step toward anchoring an *F. iinumae* reference genome, an appropriate segregating population was needed as the basis for linkage map development. As a foundation for map construction, a set of 3751

F. iinumae discovery-based markers was designed and incorporated, along with over 90,000 octoploid-based and non-discovery-based markers, into the first high-throughput strawberry SNP genotyping platform, the IStraw90 Array built on the Affymetrix Axiom SNP chip platform, as described by Bassil and Davis et al. (2015). As a precaution against possible bias in the Array-based SNP discovery process, an independent, non-discovery-based approach, genotyping by sequencing (GBS; Elshire et al., 2011; Poland et al., 2012; Lu et al., 2013), was also used for marker development and linkage map construction. GBS has been used successfully for linkage map construction in *Rubus idaeus* L. (Ward et al., 2013), *Rubus occidentalis* L. (Bushakra et al., 2015), *Prunus avium* L. (Guajardo et al., 2015), *Prunus persica* (L.) Batsch (Bielenberg et al., 2015), and *Malus × domestica* Borkh. (Gardner et al., 2014), which, like *Fragaria*, are members of the Rosaceae (Shulaev et al., 2008). In this paper we describe the development of the first genetic linkage map for *F. iinumae*, which was constructed with a combination of Array and GBS markers, and compare its structure to that of two PC assemblies of the *F. vesca* reference genome.

Materials and Methods

Plant Material

Two *F. iinumae* locally numbered accessions, J17 and J4, both collected from Hokkaido, Japan, in 2004 (Hummer et al., 2006), were used in the initial J17 × J4 cross to produce progeny plant F1D, which was subsequently self-pollinated to produce an F_2 generation population of 150 individuals collectively designated F2D (Fig. 1). The parental trio (J17, J4, and F1D) and the F2D population plants were maintained in 8-inch-diameter pots in the Macfarlane Greenhouse facility at the University of New Hampshire, Durham. The members of the parental trio are available for distribution from the USDA National Clonal Germplasm Repository, Corvallis, Oregon, under PI numbers 637969 (CFRA1855), 637963 (CFRA1849), and 674769 (CFRA2214), respectively.

DNA Isolation

The DNA for genotyping on the Array and by GBS was isolated from fresh or lyophilized young, furled leaf tissue using the E-Z 96 Plant DNA kit (Omega Bio-tek) following the procedure described by Gilmore et al. (2011), including the use of Proteinase K as required by Affymetrix. The DNA for the Array was quantified with Quant-IT PicoGreen dsDNA (Life Technologies) with a Tecan Infinite 200 (Tecan Austria GmbH) microplate reader; for GBS it was quantified with a Qubit fluorometer (Invitrogen). A minimum of 750 ng (15 μL of a concentration of 50 ng/μL) and 1 μg of DNA were provided for genotyping with the Array and GBS, respectively.
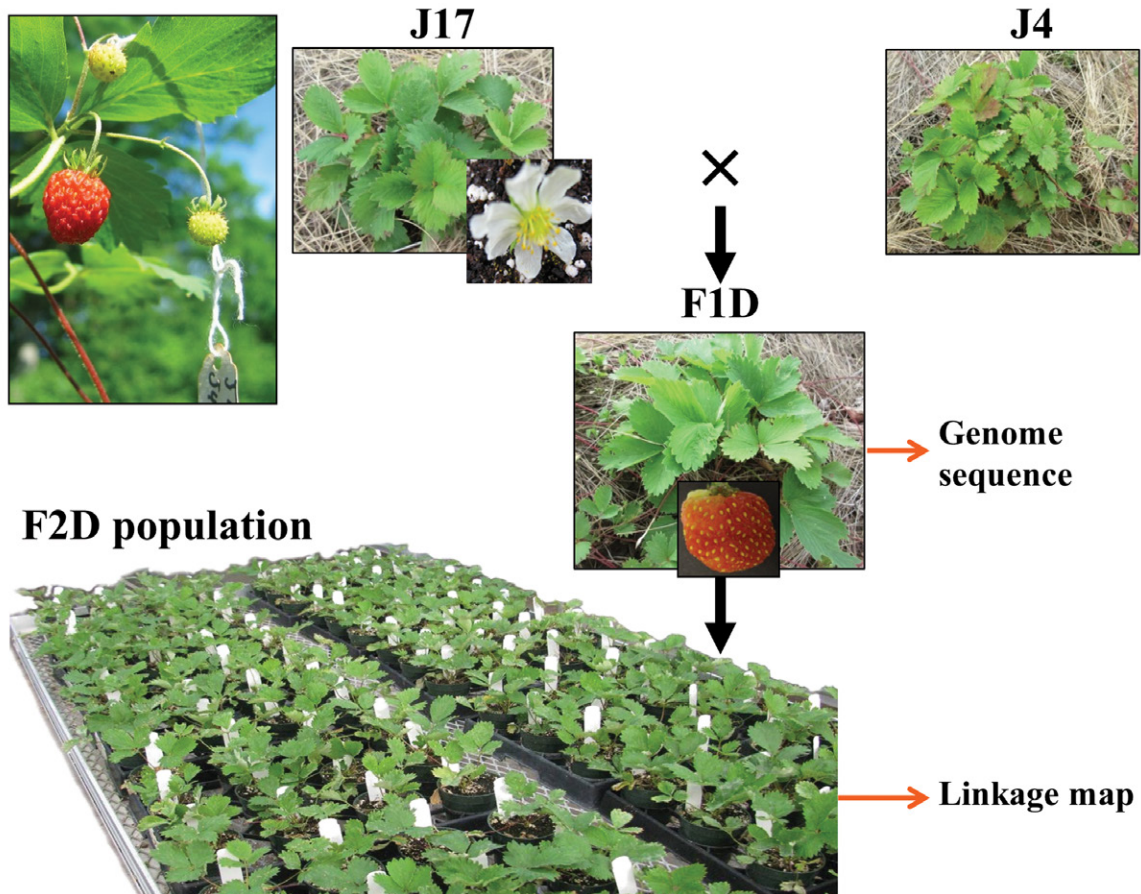
Fig. 1. *Fragaria iinumae* mapping population. This figure, showing the parental trio and F2D population, displays the distinctive appearance of *F. iinumae* plants, flowers, and fruit and indicates the research roles of each plant and population.

Table 1. Numbers of *Fragaria iinumae* parental trio and F2D samples genotyped in two IStraw90 Array runs and by genotyping by sequencing (GBS).

Genotyping Platform[†]	IStraw90 Array		GBS
	Run 1	Run 2	
Total samples	24	96	95
Trio samples			
J 17	1	1	2
J4	1	2	2
F1D	1	2	2
F2D samples	21	91	89

† A total of 112 F2D samples were assayed, representing 109 different F2D plants. Plants J17, J4, F1D, F2D17, and F2D77 were included in both runs, and plants J4, F1D, and F2D110 were each represented by two technical replicates in the second run. The parental trio members were each represented by two technical replicates in GBS.

Array Marker Data

The parental trio and 109 randomly selected F2D plants (Table 1) were genotyped using the Array following the procedures detailed by Bassil and Davis et al. (2015). The Array contains 3751 SNPs that were based on SNP discovery in F1D (F1D SNPs), 85,663 SNPs based on an octoploid strawberry discovery panel (octoploid-based SNPs), and 5648 non-discovery-based (codon-based) SNPs as described by Bassil and Davis et al. (2015).

GBS Marker Data

For GBS analysis, 95 samples, including 89 F2D plants chosen at random from the subset of 109 genotyped on the Array, and duplicate trio member samples were assayed (Table 1). A two-enzyme GBS approach as outlined by Poland et al. (2012) using the methylation sensitive enzymes *Msp*I and *Pst*I (Fellers, 2008; Deschamps et al., 2012) was performed by the Hubbard Center for Genome Studies (Univ. of New Hampshire, Durham). A pooled, multiplexed library preparation containing DNA from the 95 samples and one negative control was sequenced as 100 × 100 paired end reads on one lane of Illumina HiSeq 2500 in single-flow-cell, high-output run mode.

Genotyping and Marker Selection

The samples on the Array were genotyped with the Affymetrix Genotyping Console and SNPolisher software packages using the default settings for diploid organisms. The genotyped SNPs were classified into six cluster classes according to the Affymetrix rating system, of which the markers rated as Poly High Resolution (PHR; Affymetrix SNP genotyping highest-quality cluster designation) met the strictest quality criteria, as described in Bassil and Davis et al. (2015), and contained the expected three genotypic classes (aa:ab:bb). The SNPs from the GBS dataset were identified by means of two pipelines: the TASSEL 3.0 Universal Network Enabled

Analysis Kit (UNEAK), a non-reference-based pipeline (http://www.maizegenetics.net) and a reference-based (RB) pipeline that employed a draft sequence of the genome of the *F. iinumae* F1D parental genotype (unpublished data). This draft sequence is to be archived in the GDR (Jung et al., 2014) and is subject to ongoing revision.

Markers were considered for linkage analysis when the following conditions were met: they were classified as heterozygous in F1D; their genotypes were concordant in the parental trio; and the genotype calls for all replicates of parental trio members were consistent. In addition, for the GBS data, a minimum sequence alignment read depth of six was required at each SNP site.

Linkage Map Construction

Marker genotypes were coded for an F_2 population according to the founding parent genotypes on the basis of the standard JoinMap (Van Ooijen, 2006) conventions. Where the founding parents were both heterozygous, the progeny genotypes were coded in both coupling and repulsion phases, and the correct phase was determined during the mapping process as correct if mapped and incorrect if unmapped or polarizing.

Linkage mapping was performed with JoinMap 4.1 using the maximum likelihood algorithm and the Haldane mapping function. A minimum logarithm of odds score of 20.0 to group was specified, and default settings were in effect for the remaining parameters. The markers with identical segregation patterns were grouped into genotypic bins during the mapping process. As a convenient basis for data presentation, markers were placed in provisional orders within bins based on comparisons to three reference assemblies archived in the GDR (Jung et al., 2014): the *F. vesca* 'Hawaii 4' FvH and Fvb assemblies, and a preliminary assembly of the *F. iinumae* genome (unpublished data). The LGs were numbered in correspondence with the homologous *F. vesca* PCs (Shulaev et al., 2011).

For each locus, a Chi-square goodness-of-fit test was performed in JoinMap using a 1:2:1 expected genotypic segregation ratio. Within each LG, the locus-specific segregation data sets were subjected to homogeneity Chi-square testing performed in Microsoft Excel to determine whether these data sets could be pooled as a basis for more sensitively detecting LG-wide segregation distortion.

Synteny and Colinearity Comparisons between *F. iinumae* and *F. vesca*

Following construction of the Fii linkage map, the correspondences between the Fii LGs and the *F. vesca* (FvH and Fvb) PCs were examined to define the chromosomal patterns of synteny and colinearity between the Fii map and the *F. vesca* assemblies. The known sequence tags flanking the mapped Array and GBS markers were used as queries for BLASTN searches against the PCs of the two published genome assemblies of *F. vesca* 'Hawaii 4': FvH (Sargent et al., 2011; Shulaev et al., 2011) and Fvb (Tennessen et al., 2014), where the latter is referred to as *Fragaria vesca* Whole Genome v2.0.a1 Assembly & Annotation in the GDR database (Jung et al., 2014). In addition, comparisons were performed to the original 'Hawaii 4' scaffold assembly, v1.1, as archived in GDR (Jung et al., 2014). The comparisons of each LG of Fii to FvH and Fvb PCs were depicted visually by means of Circos diagrams (Krzywinski et al., 2009), and the Circos figures were annotated manually with Microsoft PowerPoint to include information about Fii marker correspondences to FvH PC0 (PC zero) and to the 'Hawaii 4' scaffold assembly v1.1 as well as detailed information about discrepant positional assignments.

Synteny Comparison between *F. iinumae* and an Octoploid *F. ×ananassa* Map

Following completion of the Fii map, common markers between the Fii map and an *F. ×ananassa* map also based primarily on Array markers (Sargent et al., 2015) were identified with Microsoft Excel to evaluate the extent of synteny.

Results

Mapping Population

Of the F2D individuals subjected to genotyping, 5 using the Array and 10 using GBS were excluded from further consideration due to excessive missing data, low read depth, and/or conflicting data suggestive of potential sample cross-contamination. Thus, construction of the Fii linkage map relied on Array genotyping data from 104 individuals, for which 79 also contributed GBS data.

Identification of Candidate Markers for Mapping

Array Markers

Genotyping outcomes for the three distinct SNP discovery categories of Array markers (F1D, octoploid, and codon) are summarized in Table 2. Of the 95,062 SNP sites interrogated on the Array, application of the F1D no-call and F1D homozygous filters eliminated all but 3872 candidate markers. Further candidate elimination on the basis of trio non-concordance and parental no-call filters narrowed the Array set to 3391 filtered candidates (Table 2). Overall, the F1D SNPs outperformed the octoploid-based and codon-based SNPs in terms of surviving the filtration steps (Table 2). The respective conversion rates of F1D, octoploid-based, and codon-based Array markers to filtered candidate markers were 82.1, 0.4, and 0.2%.

GBS Markers

The GBS reads were deposited in NCBI Short Read Archive (SRA312421).The GBS bioinformatics pipelines resulted in an initial 5281 candidate SNPs from the UNEAK pipeline and 2213 from the RB pipeline. After filtering of SNPs for mapping suitability on the basis of heterozygosity in F1D, trio concordance, and minimal missing data, the UNEAK and RB pipelines provided 579 and 690 filtered candidate SNPs.

Table 2. Identification and performance of the IStraw90 Array for mapping.

Description†	Total	F1D	Octoploid	Codon
Total on IStraw90 Array	95,062	3,751	85,663	5,648
Affymetrix PHR‡ class	3,509	3,209	284	16
F1D genotype filters				
F1D no-call	(235)	(1)	(203)	(31)
F1D homozygous	(90,958)	(488)	(84,881)	(5,586)
F1D heterozygous	3,872	3,262	579	31
Trio non-concordance filter	(202)	(153)	(46)	(3)
Parental no-call filter	(279)	(31)	(232)	(16)
Filtered candidate total	3,391	3,078	301	12
Filtered candidates mapped	3,280	3,050	222	8
Affymetrix PHR class	3,239	3,031	200	8
Not PHR class	41	19	22	0

† Counts are provided for the three major Array marker categories: F1D, octoploid-based, and codon-based. In the filtration pipeline, exclusion criteria were as follows: F1D no call, no genotyping call for F1D; F1D homozygous, genotyped as homozygous in F1D; non-concordant, genotype calls inconsistent among members of parental trio; parental no call, no genotyping call for one or both parents (J17 and J4).

‡ PHR, Poly High Resolution performance class.

Marker Performance in Mapping

Array Markers

Overall, 96.7% of the filtered candidate Array markers were mapped. The F1D SNPs outperformed the octoploid-based and codon-based candidates in terms of achieving map incorporation. Of the filtered candidate SNPs, the percentages mapped from the F1D, octoploid-based, and codon-based categories were 99%, (3050 out of 3078), 74% (222 out of 301), and 67% (8 out of 12), respectively (Table 2). Considering the performance of the Array SNPs by cluster class, 3239 (92.3%) of the 3509 PHR-class markers were mapped (Table 2), whereas less than 0.05% (41 of 91,553) of markers in the other (non-PHR) cluster classes were mapped (Table 2). Of the filtered candidate SNPs that were mapped, 98.8% (3239 of 3280) were classified as PHR (Table 2).

GBS Markers

The GBS dataset provided 893 mapped markers. The mapped markers included 300 markers unique to the UNEAK pipeline, 397 unique to the RB pipeline, and 196 in common between the two pipelines.

The Fii Linkage Map

Following co-segregation analysis and mapping, the Fii map (Fig. 2) contained the expected seven LGs, which ranged in length from 50.1 cM (LG2) to 92.9 cM (LG6), summing to a total map length of 451.7 cM (Table 3). Overall, 3280 Array and 893 GBS markers were incorporated into the Fii map (Table 3). Further examination of incorporated SNPs showed that 21 of the mapped GBS markers were identical in their reference genome locations to mapped Array markers. For each of these 21 SNPs, the GBS genotyping calls were identical to the respective Array genotyping calls across the 79 F2D individuals for

which both Array and GBS data were generated. Despite this apparent redundancy, the 21 duplicative GBS markers were retained in the map because their F1D sequence tags, which are based on *F. iinumae* sequence, were not necessarily identical to the corresponding Array tags, which are based on *F. vesca* sequence. Thus, the final map has a total of 4173 SNP markers archived as GDR1023 (Table 3), including the 21 redundant GBS SNPs.

The 4173 markers contained on the Fii map were distributed among 496 loci, with an average of 8 SNP markers per locus (Table 3). Each mapping bin constitutes a unique map locus, and thus the mapping bins define 496 loci. The numbers of markers and loci per LG varied across an approximately twofold range: from 439 markers in LG7 to 938 in LG6; and from 42 loci in LG4 to 101 in LG6 (Table 3). Overall, the map density of 0.91 cM/locus was less than the calculated saturated map density of 0.48 cM/locus (= $100 \times 1/[104$ individuals $\times 2$ gametes/individual]). Of the 496 loci, 473 contained at least one Array marker, 463 contained at least one F1D based Array marker, and 270 contained at least one GBS marker (Table 3). A total of 226 loci were populated exclusively by Array markers; 187 of which were populated exclusively by F1D markers, while only 23 loci were populated exclusively by GBS markers. Notably, 14% (6 of 42) of the loci on LG4 were populated only by GBS markers, while the next highest percentage was just 6% (5 of 80) on LG3 (Table 3).

Marker Distributions across the Fii Linkage Groups

When the marker densities were charted at each locus along the length of each LG, a single region of exceptionally high marker density was apparent within each LG (Fig. 3). In LGs 1, 2, 4, 5, and 7, this density peak was at, or very near, one end of the LG, while the density peak was centrally located in LG6 and sub-centrally located in LG3.

Segregation Distortion

When examined on a single-locus basis, significant segregation distortion was present in only 17 of the 496 loci composing the Fii map. Of these 17 loci, 12 were on LG4, 4 were on LG5, and 1 was on LG6. In each case, the pattern of distortion favored the allele derived from paternal parent, J4. When examined at the level of entire LGs (Table 4), homogeneity Chi-square testing (results not shown) determined that locus-specific segregation data could be pooled across all loci within each LG. Goodness-of-fit testing of LG totals to a 1:2:1 expected genotypic segregation ratio showed that for each LG, transmission of paternally derived alleles was favored ($p \leq 0.05$), although the effect was small (Table 4). Overall, of the 496 loci comprising the map, the genotypic proportions of aa:ab:bb averaged 0.24:0.48:0.28 (Table 4; Fig. 4). Similarly, the allelic proportions averaged 0.48:0.52 for a:b rather than the expected 0.50:0.50 proportions (Table 4), thus displaying a small but systematic bias toward the paternally derived alleles.
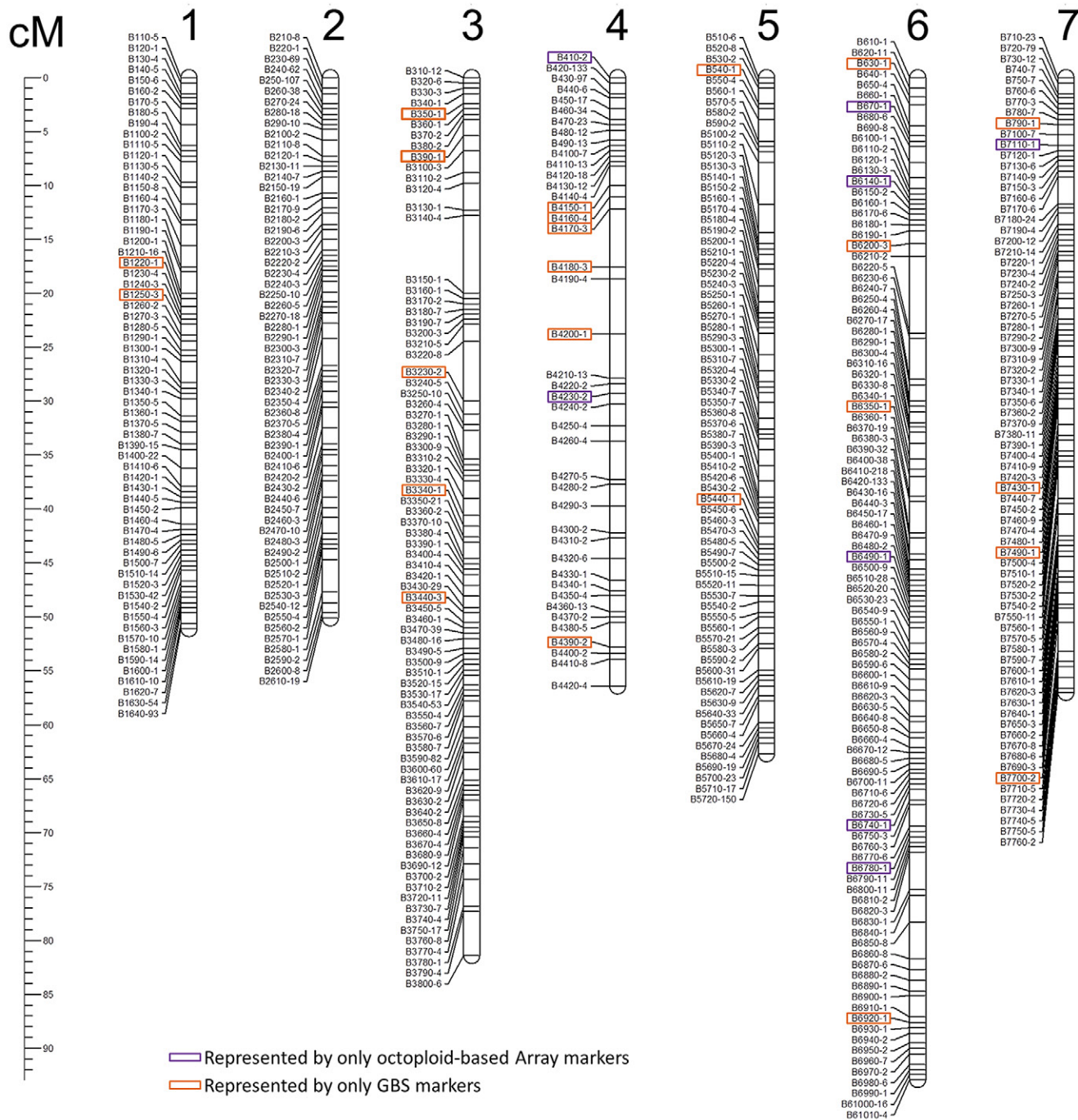
Fig. 2. *Fragaria iinumae* high-density linkage map. The nomenclature of the genotypic bins consists of three parts: (i) the letter B (for bin) followed by the appropriate linkage group (LG) number (1–7); (ii) a two- or three-digit position number, assigned sequentially in increments of 10 (beginning with 10) to indicate the linear map order; and (iii) the number of markers in the respective bin. For example: bin B110-5 is on LG1 (B1), is the first bin in LG1 (B110), and contains five markers (B110-5). Similarly, bin B220-1 belongs to LG2 (B2), is the second bin (B220), and contains one marker (B220-1).

Parental Heterozygosity

Of the 4173 mapped markers, the two parents were homozygous for the alternate alleles of 2920 markers. Of the remaining markers, 162 (3.88%) and 1071 (25.67%) markers were heterozygous in J17 and J4, respectively, and 20 markers were heterozygous in both parents.

Synteny Comparisons between *F. iinumae* and *F. vesca*

Chromosomal assignments were conserved in relation to Fii for 3702 markers in FvH and 3810 markers in Fvb (Table 5) and were mutually conserved among Fii, FvH, and Fvb for 3493 (84%) of the 4173 markers on the Fii map. The results of linkage-group-specific synteny are reported in Table 5 and graphically displayed in

Table 3. Summary of *Fragaria iinumae* linkage map statistics by linkage group (LG).

Map Statistics†	Marker Category	LG1	LG2	LG3	LG4	LG5	LG6	LG7	Total
No. SNPs	F1D only	337	434	488	343	422	687	339	3050
	Total Array‡	359	462	510	375	457	749	368	3280
	GBS§	113	133	147	121	119	189	71	893
	Total mapped	472	595	657	496	576	938	439	4173
No. loci represented	F1D	62	60	75	34	70	91	71	463
	Octoploid	14	17	15	15	21	30	20	132
	Codon	1	4	–	–	–	2	–	7
	Total Array	62	61	75	36	70	97	72	473
	GBS	38	33	51	27	34	49	38	270
No. unique loci	F1D only	24	22	28	11	28	43	31	187
	Array only	26	28	29	15	38	52	38	226
	GBS only	2	–	5	6	2	4	4	23
No. loci	Total	64	61	80	42	72	101	76	496
No. SNPs/locus	Average	7	10	8	12	8	9	6	8
Map length	cM	51.1	50.1	81.4	56.4	62.7	92.9	57.0	451.7

† The distribution of single nucleotide polymorphisms (SNPs), the numbers of loci, and the numbers of uniquely populated loci are provided by marker category, along with the total number of loci, the average number of SNPs per locus, and the map lengths in cM for each linkage group.

‡ Array, IStraw90 Array.

§ GBS, genotyping by sequencing.

Supplementary Figures S1.0–S1.7. Among the discrepancies specific to the Fii versus FvH comparison, 394 Fii markers were assigned to non-corresponding FvH PCs other than PC0, and an additional 13 GBS markers were assigned to FvH PC0. Also, 39 GBS markers that were not represented in the FvH PC assembly were assigned to scaffolds within the 'Hawaii 4' v1.1 scaffold assembly. Regarding discrepancies specific to the comparison between Fii and Fvb, 89 Fii markers were assigned to non-corresponding PCs, while 249 markers (247 GBS and 2 Array) were not represented in the Fvb assembly. The chromosomal assignments of 61 markers were conserved between FvH and Fvb but were discrepant in relation to Fii. Finally, an additional 25 Fii GBS markers were not represented in either FvH or Fvb (Table 5). A graphic, global overview of the synteny comparisons is provided by the Circos diagrams in Fig. 5, wherein the total number of discrepant connections is reduced from 394 in the FvH comparison to 89 in the Fvb comparison.

With regard to synteny comparisons by Fii LG, the percentages of markers mutually syntenic among Fii, FvH, and Fvb ranged from a low of 81% in LG6 (757 out of 938 markers) to a high of 90.7% in LG7 (398 out of 439 markers) (Table 5; Supplementary Figures S1.0–S1.7). Regarding pairwise comparisons to Fii, synteny percentages ranged from 85.6 (LG6) to 95.0% (LG7) for FvH, and 88.5 (LG4) to 94.5% (LG7) for Fvb. In pairwise comparisons to Fii, the number of syntenous marker assignments was greater in Fvb than in FvH for all LGs except for LG7, whereas the numbers of non-syntenous marker assignments was less for Fvb than for FvH in all seven LGs (Table 5). Specific synteny discrepancies of the individual Fii LGs compared with the FvH and Fvb PC assemblies are explored in detail in Supplementary Figures S1.1–S1.7.

Colinearity Comparisons between *F. iinumae* and *F. vesca*

Substantial departures from locus and marker colinearity were observed between the Fii map and FvH PCs across all seven Fii LGs (Table 6), as graphically displayed with Circos diagrams in Supplementary Figures S2.1–2.7, wherein connecting links for discrepant markers are color-coded in black. Compared with FvH, departures from colinearity were substantially reduced across all seven LGs in the Fvb PCs (Table 6), as graphically indicated by the marked reduction in the number of black connecting links in the Fvb colinearity diagram (Fig. 5). Among Fii LGs, colinearity of Fii with Fvb PCs at the locus level was greatest (100%) for LG1 and least for LG3 (89%), and at the marker level was greatest for LG7 (94%) and least for LG3 (78%).

Synteny Comparison between *F. iinumae* and an Octoploid *F. ×ananassa* Map

A total of 28 markers (Table S1) were identified to be in common between the Fii map and the recently published *F. ×ananassa* DA × MO map (Sargent et al., 2015), which contains 8407 Array markers. Synteny is uniformly conserved between the respective homologs of the Fii and DA × MO maps, yet without discernible correspondences to specific *F. ×ananassa* subgenomes (Table S1). With respect to colinearity, no rearrangements were discernible in comparisons between Fii LGs and their respective *F. ×ananassa* homologs.

Discussion

A number of phylogenetic evaluations of the genus *Fragaria* (Rousseau-Gueutin et al., 2009; DiMeglio et al., 2014) including some recent pan-genomic studies of the genomes of the octoploid *Fragaria* (Njuguna et al., 2013;
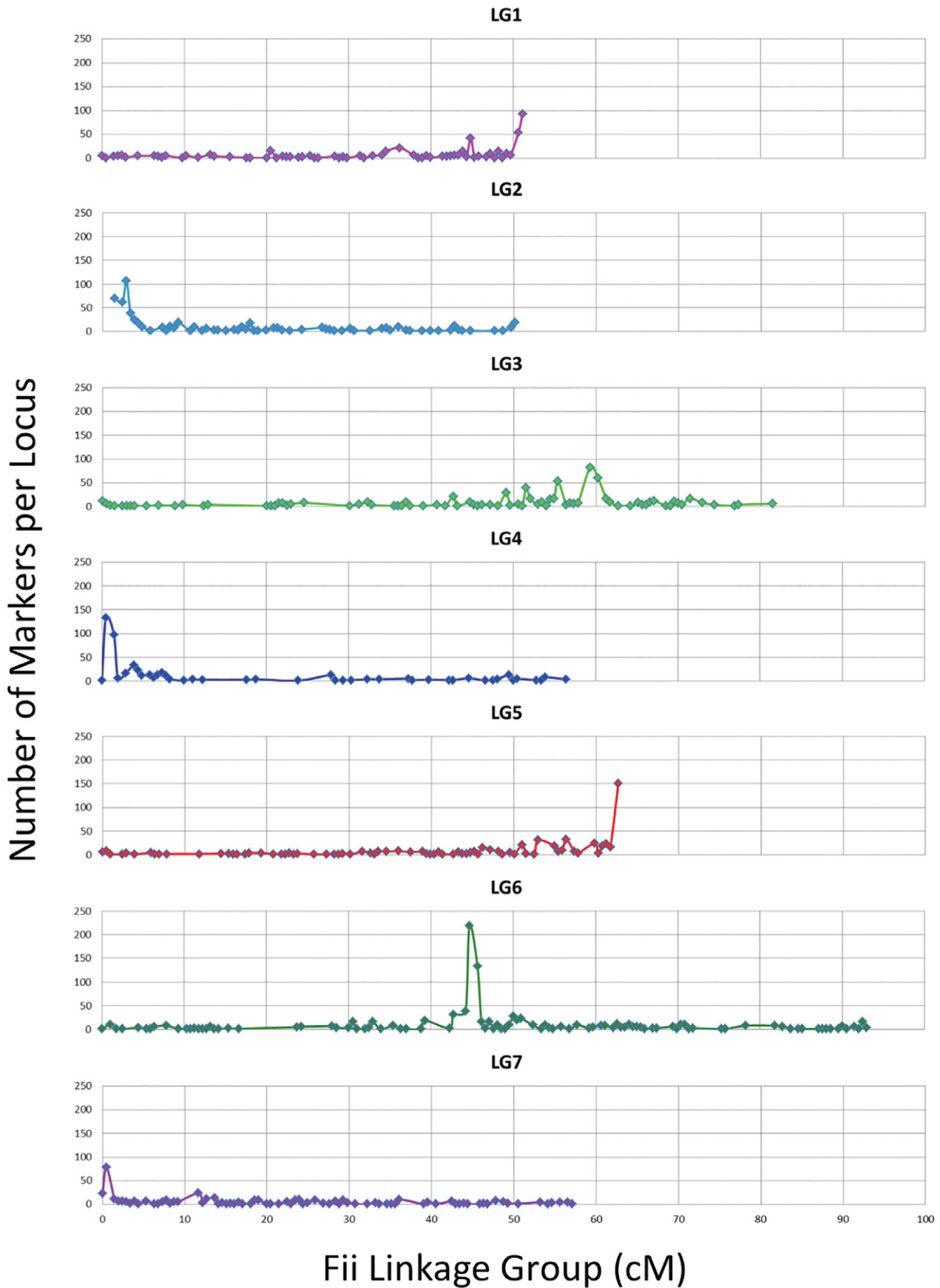
Fig. 3. Marker distributions along the seven *Fragaria iinumae* linkage groups (LGs). The *y* axis indicates the number of markers per locus (diamonds) along the length (cM) of each the LGs (*x* axis). On each LG the most pronounced peak in marker density is inferred to coincide with the centromere location.

Table 4. Segregation distortion and allele counts in the *Fragaria iinumae* linkage map.

| LG† | Missing | Genotype Counts | | | | Genotypic Proportions | | | | Allele Counts | | Allelic Proportions | | |
		aa	ab	bb	Total	aa	ab	bb	χ^2	a	b	a	b	χ^2
1	1	1,614	3,228	1,813	6,656	0.24	0.48	0.27	17.85**	6,456	6,854	0.49	0.51	11.90**
2	0	1,568	2,840	1,936	6,344	0.25	0.45	0.31	112.19**	5,976	6,712	0.47	0.53	42.69**
3	2	1,885	4,391	2,042	8,320	0.23	0.53	0.25	31.80**	8,161	8,475	0.49	0.51	5.93*
4	13	1,046	1,905	1,404	4,368	0.24	0.44	0.32	126.72**	3,997	4,713	0.46	0.54	58.86**
5	1	1,504	4,006	1,977	7,488	0.20	0.53	0.26	96.57**	7,014	7,960	0.47	0.53	59.76**
6	4	2,729	4,849	2,922	10,504	0.26	0.46	0.28	68.33**	10,307	10,693	0.49	0.51	7.10*
7	2	1,865	3,715	2,322	7,904	0.24	0.47	0.29	81.03**	7,445	8,359	0.47	0.53	52.86**
Total	23	12,211	24,934	14,416	51,584	0.24	0.48	0.28	244.08**	49,356	53,766	0.48	0.52	188.59**

* Significant at $P = 0.05$.

** Significant at $P = 0.001$.

† LG, linkage group. Within each LG, the sums of genotype counts and the calculated genotypic proportions over all loci are listed along with the outcomes of Chi-square goodness-of-fit tests to a 1:2:1 ratio, while the sums of allele counts and the calculated allelic proportions over all loci are listed along with the outcomes of Chi-square goodness-of-fit tests to a 1:1 ratio.
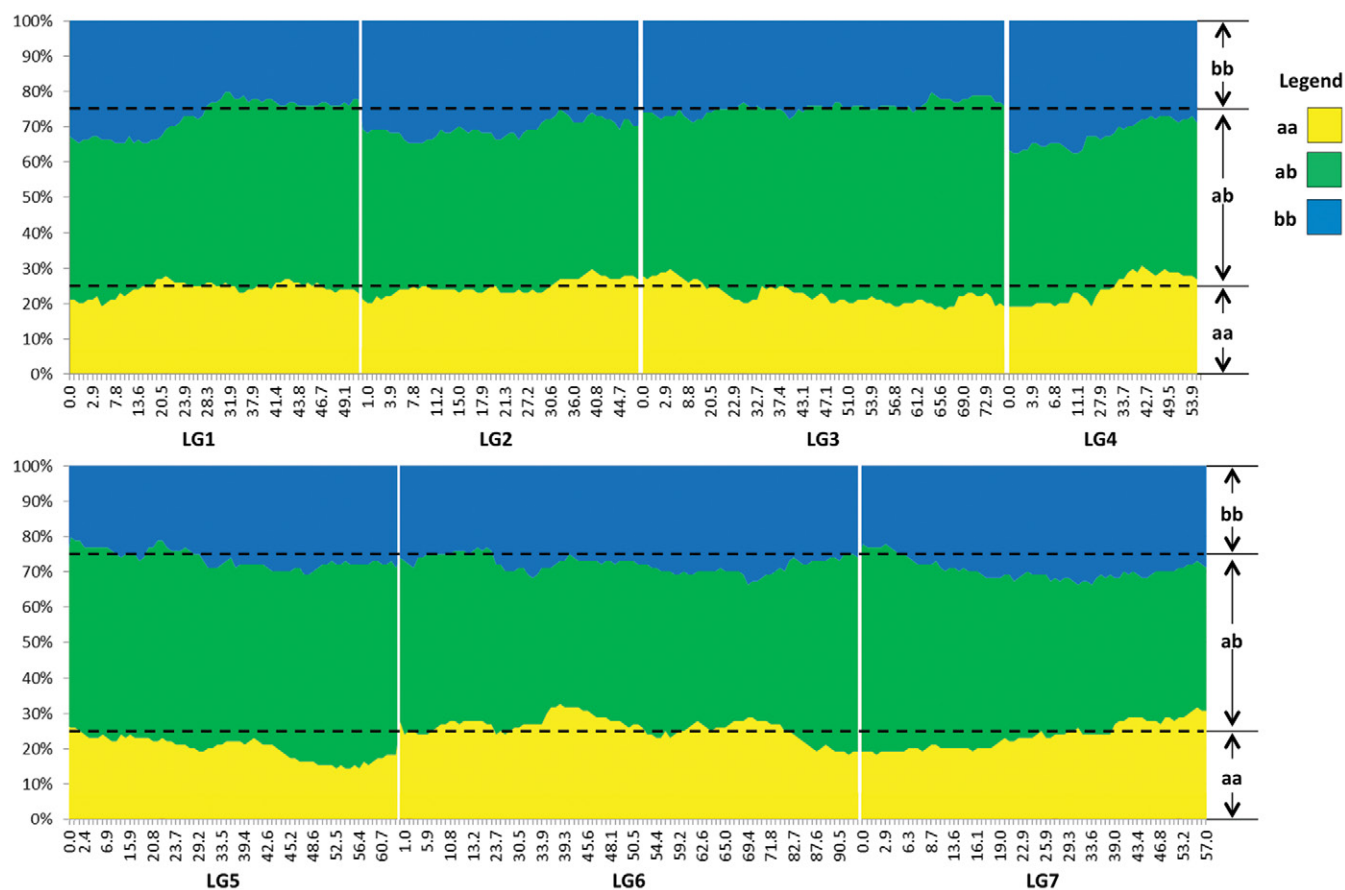

Fig. 4. Genotypic frequencies across the seven *Fragaria iinumae* linkage groups (LGs). The expected frequencies for the three genotypes are shown with dashed lines. The *y* axis represents the percentage of each genotype. The *x* axis represents cM locations along each of the LGs. The actual frequencies of the three genotypic classes are traced along each LG and are color coded.

Tennessen et al., 2014; Sargent et al., 2015) have highlighted the key role of *F. iinumae* in the evolution of the octoploid *Fragaria* and thus the importance of developing genomic resources for *F. iinumae* to assist in genomics research in the cultivated strawberry (Sargent et al., 2015). To this end, we present a robust, saturated linkage map for *F. iinumae* that will facilitate efforts to assemble a reference genome sequence for this species.

Construction of a Fii map required the development of an appropriate mapping population. The two *F. iinumae* accessions, J17 and J4, used as founder parents, were collected in 2004 from Hokkaido, Japan (Hummer et al., 2006) and correspond to the subpopulation numbers SP1855 and SP1849 specified by Njuguna et al. (2011). These two accessions were shown to belong to distinct population subgroups differentiated by SSR markers

Fii LG	Marker type	Fii Markers¶	FvH†					Fvb‡				FvH-Fvb§		
			Syn	N-syn	PC0	Scf	Total	Syn	N-syn	N-rep	Total	Syn	N-syn	N-rep
1	Array	359	326	33	0	0	359	354	5	0	359	325	4	0
	GBS	113	94	12	4	1	111	72	1	38	111	61	1	2
	Total	472	420	45	4	1	470	426	6	38	470	386	5	2
2	Array	462	439	23	0	0	462	458	3	1	462	437	2	0
	GBS	133	109	17	4	0	130	79	3	48	130	70	1	3
	Total	595	548	40	4	0	592	537	6	49	592	507	3	3
3	Array	510	460	50	0	0	510	489	21	0	510	458	19	0
	GBS	147	118	13	4	5	140	104	6	30	140	97	1	7
	Total	657	578	63	4	5	650	593	27	30	650	555	20	7
4	Array	375	359	16	0	0	375	358	17	0	375	352	8	0
	GBS	121	79	9	0	28	116	81	4	31	116	58	2	5
	Total	496	438	25	0	28	491	439	21	31	491	410	10	5
5	Array	457	402	55	0	0	457	450	8	0	458	402	7	0
	GBS	119	96	15	1	3	115	85	2	27	114	78	2	4
	Total	576	498	70	1	3	572	535	10	27	572	480	9	4
6	Array	749	648	101	0	0	749	737	12	0	749	645	9	0
	GBS	189	155	29	0	2	186	128	3	55	186	112	1	3
	Total	938	803	130	0	2	935	865	15	55	935	757	10	3
7	Array	368	349	19	0	0	368	363	4	1	368	348	4	0
	GBS	71	68	2	0	0	70	52	0	18	70	50	0	1
	Total	439	417	21	0	0	438	415	4	19	438	398	4	1
Total		4173	3702	394	13	39	4148	3810	89	249	4148	3493	61	25

† FvH, *F. vesca* 'Hawaii 4' v1.1 genome assembly anchored to the FV × FB linkage map. Columns show number of Fii markers delineated as follows: Syn (syntenic), markers assigned to corresponding FvH PC; N-syn (non-syntenic), markers assigned to non-corresponding FvH PC; PC0, markers assigned to FvH PC0; scf (scaffold), markers assigned to an *F. vesca* scaffold but not to an FvH PC.

‡ Fvb, *F. vesca* 'Hawaii 4' genome assembly anchored to a *F. vesca* ssp. *bracteata* linkage map. Columns show number of Fii markers delineated as follows: Syn, markers assigned to the corresponding Fvb PC; N-syn, markers assigned to a non-corresponding Fvb PC; and N-rep (not represented), markers not present in the Fvb assembly.

§ Number of markers qualifying as Syn, mutually syntenic among Fii LGs and PCs of FvH and Fvb; N-syn, markers not syntenic with Fii but syntenic between FvH and Fvb; and N-rep, markers not assigned to either FvH PCs and/or scaffolds, or Fvb PCs.

¶ Number of markers mapped to Fii LGs.

(Njuguna et al., 2011) and thus were expected to yield an F$_1$ hybrid that would be sufficiently heterozygous to generate an informative F$_2$ generation mapping population. As reported here, 4173 segregating Array and GBS SNP markers were genotyped and mapped in this population, thus confirming the expectation of abundant polymorphism between the founder parents.

The Fii map contains the expected seven linkage groups, corresponding to the base chromosome number (*x* = 7) of all diploid *Fragaria*. The total map length of 451.7 cM is remarkably similar to the 442.8 cM length of the revised FV × FB linkage, which was based on 76 F2 generation plants and consisted of 411 markers distributed into 275 bins (loci) (Sargent et al., 2011). The lengths of individual LGs ranged from 50.1 to 92.9 cM on the Fii map and from 48.3 to 98.1 cM on the FV × FB map, with LG6 being the longest LG on both maps. Notably, LG3 was the shortest LG (48.3 cM) on the FV × FB map (Sargent et al., 2011) but was the second longest (81.4 cM) on the Fii map. The mean map resolutions (i.e., locus densities) were one locus per 0.91 cM for the Fii map and 1.61 cM for the FV × FB map. In comparison, the Fvb

maternal and paternal linkage maps constructed by Tennessen et al. (2013) have mean resolutions of one locus per 2.43 cM and 2.73 cM, respectively. Thus, the Fii map has the highest mean resolution (locus density) of any diploid linkage map currently available for *Fragaria*.

Although individual marker segregation patterns were generally normal, there was a slight but statistically significant pattern of transmission bias favoring paternally derived alleles (Table 4). This pattern held true for each of the seven LGs. However, locus-specific segregation patterns deviated significantly from the 1:2:1 expectation for only 17 loci confined to a 12-cM region at the top of LG4, a 4-cM region near the bottom of LG5, and a single locus at position 69.4 cM on LG6 in all cases favoring the paternal allele. For comparison, locus-specific segregation distortion was substantial and widespread in two prior diploid strawberry mapping populations. On the 80-marker *F. vesca* BS × WC6 map (Davis and Yu, 1997), which consists primarily of RAPD (random amplified polymorphic DNA) marker loci, segregation distortion was observed for 38 loci located on five of the seven LGs, and without exception the transmission bias favored the
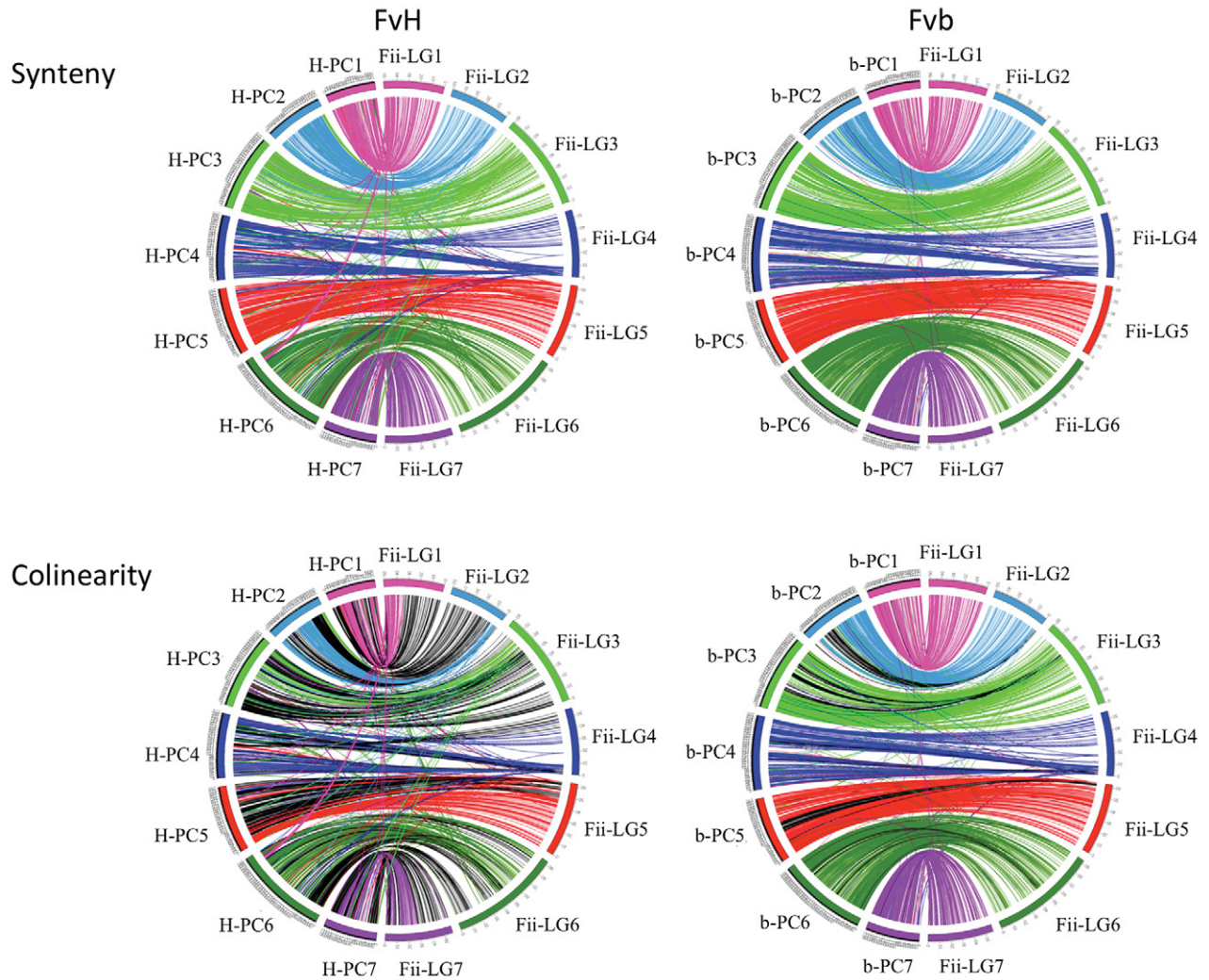
Fig. 5. Global synteny and colinearity comparisons to *Fragaria vesca* 'Hawaii 4' assemblies, FvH and Fvb. In each of the two synteny comparisons, the connecting links are color coded according to the *F. iinumae* linkage-group arc to which they are anchored. In each of the two colinearity comparisons, connecting links for discrepant markers are color coded in black. In the synteny and the colinearity comparisons, the reduction in numbers of discrepant links in the Fvb assembly is graphically evident.

Table 6. Colinearity between *Fragaria iinumae* (Fii) linkage groups (LGs) and corresponding *F. vesca* 'Hawaii 4' v1.1 genome assembly anchored to the FV × FB linkage map (FvH) and *F. vesca* 'Hawaii 4' genome assembly anchored to a *F. vesca* ssp. *bracteata* linkage map (Fvb) pseudo-chromosome (PC) assemblies.†

| | | Loci | | | Markers | | | | | | |
| | | FvH | Fvb | | FvH | | | Fvb | | |
Fii LG	Fii	Colin	Colin	Fii	Syn	Colin	N-colin	Syn	Colin	N-colin
1	64	27	64	472	420	238	182	426	416	10
2	61	8	59	595	548	236	312	537	453	84
3	80	14	71	657	578	132	446	593	512	81
4	42	21	37	496	437	268	169	439	414	25
5	72	50	69	576	498	230	268	535	462	73
6	101	41	98	938	803	478	325	865	821	44
7	76	35	74	439	417	185	232	415	412	3
Total	496	196	472	4173	3701	1767	1934	3810	3490	320

† Under the "Loci" heading, the columns (left to right) provide the number of loci on the indicated Fii LG, and the numbers of Fii loci that are colinear (Colin) within the corresponding FvH and Fvb PCs. Under the heading "Markers," the columns provide the numbers of markers on the indicated Fii LG that are syntenous (Syn) to and colinear (Colin) or non-colinear (N-colin) within the corresponding FvH and Fvb PCs.

maternally derived allele. On the FV × FB SSR-based map (Sargent et al., 2006), 59% of the markers displayed segregation distortion, and every LG had at least one distorted marker. It is important to emphasize that in the prior diploid mapping studies, in which locus-specific segregation distortion was prevalent and often extreme, the founding parental crosses were between widely divergent parents: 'BS' and 'WC6' belong to separate *F. vesca* subspecies (*vesca* and *americana*, respectively), whereas FV and FB belong to different species (*F. vesca* and *F. bucharica*, respectively). Wide crosses were employed in these earlier studies in part to assure the availability of sufficient numbers of segregating markers, given the marker platforms employed. In the present study, which relied on SNP markers, the *F. iinumae* accessions used as founder parents were sufficiently genetically distinct to yield an abundance of markers. Although a systematic segregation bias favoring paternally derived alleles was evident in the *F. iinumae* F2D population, the effect was slight and on a locus-specific basis rose to the level of statistical significance for only 3.4% of the 496 loci.

Inclusion on the Array of a set of 3751 SNPs identified on the basis of their apparent heterozygosity in hybrid F1D provided a valuable source of SNPs for use in map construction while also providing a measure of Array utility to complement the assessments of its performance on octoploid germplasm by Bassil and Davis et al. (2015) and Sargent et al. (2015). Of the 3751 F1D based SNPs on the Array, 488 were genotyped as homozygous in F1D and did not segregate in the F2D population, while 3050 of the remaining 3262 F1D based Array SNPs, or 93.5%, provided segregation data suitable for mapping. Thus, the Array performed very well with respect to SNPs designed for a specific diploid population. In contrast, of the 85,663 Array SNPs based on discovery in octoploid germplasm, only 222 (0.26%), were incorporable into the Fii map. Of the 5648 speculative, codon-based Array SNPs, which had been designed using a non-discovery-based process, only 8 (0.14%), were incorporable into the Fii map. Thus, in the absence of the *F. iinumae*–specific F1D SNPs, the Array provided useful segregation data for only 230 SNPs, a number far short of that required to construct a high-resolution linkage map. It will be of considerable interest to see how the Array performs in other diploid species as well as in other *F. iinumae* populations.

Overall, 3.45% of the 95,062 Array markers were mapped (Table 2). Considering the performance of Array by cluster class, 92.3% (3239 out of 3509) PHR-class markers were mapped (Table 2), whereas less than 1% (41 of 95,062) of the markers assigned to the other cluster classes was mapped. The high conversion rate of the PHR-class markers confirms the expected superiority of this marker class for use in mapping. Thus, at the expense of only 41 mapped Array markers, SNP candidate selection could have been conveniently achieved by application of only three criteria: heterozygosity in F1D, trio concordance, and marker membership in the PHR class.

As a complementary, non-discovery-based platform, GBS provided useful segregation data for 893 markers, distributed among 270 loci. Due to co-segregation between many GBS and Array markers, the GBS data added only 23 unique loci (bins) to the map (Table 3). Nevertheless, 5 marker bins containing a total of 12 GBS markers were clustered in a poorly marked region of LG4, which otherwise contained only a single Array marker bin (Fig. 2). These 5 uniquely GBS-marker loci contributed importantly to the continuity of LG4. The absence of Array-based markers in the respective LG4 region may have been the consequence of a deficiency in the assembly of the FvH *F. vesca* reference genome (Shulaev et al., 2011), precluding marker discovery in that region. Similarly, Tennessen et al. (2013, 2014) encountered difficulty in constructing LG4 in their *F. vesca* ssp. *bracteata* linkage mapping project, which employed a sequence-capture approach that had also relied in its design on the original *F. vesca* 'Hawaii 4' reference assembly. Thus our findings emphasize the critical importance of reference genome completeness and accuracy in relation to reference-based processes of marker development and, by extrapolation, the importance of a genomic resource for *F. iinumae* to facilitate studies of the genome of the cultivated strawberry.

Fii Map Comparison to FvH and to Fvb Assemblies and Associated Linkage Maps

The structure of the Fii linkage map provided a basis for three types of comparisons with the *F. vesca* genome, involving (i) inferred centromere position, (ii) syntenic relationships (i.e., LG assignments), and (iii) colinearity (i.e., locus order within LGs). On the FV × FB map of Sargent et al. (2011), centromere positions were inferred on a qualitative basis and assigned on each chromosome to its major focal point of marker clustering, which was taken to be an indicator of recombination suppression. In the present study, a quantitative approach was taken, with much the same rationale and outcome. In both the Fii and FV × FB maps, centromeres were positioned near the ends of all LGs except on LG3 and LG6.

On the Fii linkage map, markers were assigned to LGs entirely on the basis of recombination frequency data, whereas the comparator PC assemblies are the products of two processes: genomic sequence assembly and anchoring of assembled contigs and scaffolds to linkage maps. Our synteny comparisons relied on the use of marker sequence tags as BLASTN queries of the two PC assemblies, and thus correspondences are subject to the possibility of artifactual discrepancies in cases where the true orthologous site is atypically divergent or is missing from the respective PC assembly due to either deficient assembly or to site loss from the comparator genome itself.

Several important conclusions can be drawn from the synteny and colinearity comparisons of the Fii map to the FvH and Fvb assemblies. First, the marked reduction in discrepancies in the comparisons to the Fvb assembly as compared with the FvH assembly indicates

that the Fvb reassembly (Tennessen et al., 2014) has significantly improved the accuracy of the *F. vesca* 'Hawaii 4' reference genome. This improvement is particularly apparent with respect to the colinearity comparisons, indicating that errors in the ordering and orientation of scaffolds within PCs have been markedly reduced in the Fvb assembly. Yet it must be noted that in several instances, new discrepancies may have been introduced into the Fvb assembly. Additionally, 247 GBS markers that were assigned to LGs in the FvH PC assembly were not represented on Fvb PCs (Table 5), even though all but two of the markers are present in the FvH v1.1 scaffolds.

Second, comparison of the Fii and Fvb maps documents substantial similarity between the *F. iinumae* and *F. vesca* genomes with respect to synteny. Of the 3899 Fii markers that could be assigned to specific Fvb PCs, only 89 (2.3%) were assigned to a non-corresponding PC. Of these 89 discrepancies, which involve 70 Array and 19 GBS markers, almost all are markers that are located in putative centromeric regions on the Fii map but that are widely dispersed on the Fvb PCs (see Supplementary Figures S1.0–S1.7). Thus, no substantial translocational differences were revealed between the *F. iinumae* and *F. vesca* genomes. However, for 61 markers (Table 5), the respective chromosomal assignments agreed in the FvH and Fvb PC assemblies but differed in the Fii map. As noted above, these discrepancies may be artifactual consequences of the employed homology comparisons; however, at least some may be evidence of actual structural differences between the *F. iinumae* and *F. vesca* genomes.

Third, the comparison of colinearity between the Fii map and Fvb assembly documents substantial similarity yet reveals a number of discrepancies, which are identified as rearrangement zones (RZs) in Supplementary Table S2 and Figures S2.1–2.7. Four distinct RZs are evident on LG3 (Supplementary Figure S2.3), and one or more such zones exist on every LG except for LG7 (Supplementary Figures S2.1–2.7). The presence of these RZs points to areas that warrant close scrutiny as sites of possible error remaining in the Fvb assembly. However, it is likely that at least some of these constitute true inversional and other intrachromosomal rearrangements differentiating the *F. iinumae* and *F. vesca* genomes. Completion of an *F. iinumae* reference assembly, now in progress, will provide a basis for further exploring such potential structural divergences.

Conclusions

The Fii linkage map provides an important new resource for *Fragaria* genomics. Although not explored here, the F2D mapping population is segregating with respect to traits of economic interest, providing a basis for identification of marker-trait associations in *F. iinumae*. The fact that the 3280 Array–based markers are all located in coding sequences adds to the potential for identifying gene-trait associations in *F. iinumae*. Due to its high resolution and substantial number of 4173 sequence-characterized

markers, the Fii map will constitute a key resource for the anchoring of an *F. iinumae* genome assembly. Comparisons with the *F. iinumae* genome will help to reveal remaining errors in the *F. vesca* reference genome assembly and will reveal true biological differences between these two divergent genomes. The synteny between the markers common to the Fii and the *F. ×ananassa* DA × MO maps, and the improvement of marker representation on LG4, support the expectation that the Fii map and anticipated *F. iinumae* genome assembly will provide an important new resource for reference-based marker discovery in *F. ×ananassa*. Finally, the establishment of genomic resources for *F. iinumae* will support continuing efforts to dissect the subgenomic structure and evolutionary history of the octoploid, cultivated strawberry and its wild octoploid ancestors.

Supplemental Information Available

The supplemental materials include two tables and two series of figures. Table S1 includes a list of markers common between the Fii map and an *F. ×ananassa* map, and Fii linkage map information for the 4,173 mapped markers are included in Table S2. Comparisons of Fii LGs to individual PCs of FvH and Fvb are included for synteny in Figures S1.0 –S1.7 and for colinearity in Figure S2.1 – S2.7.

References

Bassil, N.V., T.M. Davis, H. Zhang, S. Ficklin, M. Mittman, T. Webster, et al. 2015. Development and preliminary evaluation of a 90K Axiom® SNP Array for the allo-octoploid cultivated strawberry *Fragaria ×ananassa*. BMC Genomics 16(155):30. doi:10.1186/s12864-015-1310-1

Bielenberg, D.G., B. Rauh, S. Fan, K. Gasic, A.G. Abbott, G.L. Reighard, W.R. Okie, and C.E. Wells. 2015. Genotyping by sequencing for SNP-based linkage map construction and QTL analysis of chilling requirement and bloom date in peach [*Prunus persica* (L.) Batsch]. PLoS ONE 10(10): e0139406. doi:10.1371/journal.pone.0139406.

Bushakra, J.M., D.W. Bryant, M. Dossett, K.J. Vining, R. VanBuren, B.S. Gilmore, J. Lee, T.C. Mockler, C.E. Finn, and N.V. Bassil. 2015. A genetic linkage map of black raspberry (*Rubus occidentalis*) and the mapping of Ag_4 conferring resistance to the aphid *Amphorophora agathonica*. Theor. Appl. Genet. 128: 1631–1646. doi:10.1007/s00122-015-2541-x

Davis, T., and H. Yu. 1997. A linkage map of the diploid strawberry, *Fragaria vesca*. J. Hered. 88:215–221. doi:10.1093/oxfordjournals.jhered.a023091

Deschamps, S., V. Llaca, and G.D. May. 2012. Genotyping-by-sequencing in plants. Biology 1:460–483. doi:10.3390/biology1030460

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

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6(5): e19379. doi:10.1371/journal.pone.0019379

Fellers, J.P. 2008. Genome filtering using methylation-sensitive restriction enzymes with six base pair recognition sites. Plant Gen. 1:146–152. doi:10.3835/plantgenome2008.05.0245

Folta, K.M., and T.M. Davis 2006. Strawberry genes and genomics. Crit. Rev. Plant Sci. 25:399–415. doi: 10.1080/07352680600824831

Gardner, K.M., P. Brown, T.F. Cooke, S. Cann, F. Costa, C. Bustamante, R. Velasco, M. Troggio, and S. Myles 2014. Fast and cost-effective genetic mapping in apple using next-generation sequencing. G3 Genes Genomes Genet. 4:1681–1687.

Gilmore, B., N. Bassil, and K. Hummer. 2011. DNA extraction protocols from dormant buds of twelve woody plant genera. J. Am. Pomol. Soc. 65:201–207.

Guajardo, V., S. Solís, B. Sagredo, F. Gainza, C. Muñoz, K. Gasic, and P. Hinrichsen. 2015. Construction of high density sweet cherry (*Prunus avium* L.) linkage maps using microsatellite markers and SNPs detected by genotyping-by-sequencing (GBS). PLoS ONE 10(5): e0127750. doi:10.1371/journal.pone.0127750

Hummer, K.E., T.M. Davis, H. Iketani, and H. Imanishi. 2006. American-Japanese expedition to Hokkaido to collect berry crops in 2004. HortScience 41:993.

Jung, S., S.P. Ficklin, T. Lee, C.-H. Cheng, A. Blenda, P. Zheng, J. Yu, A. Bombarely, I. Cho, S. Ru, K. Evans, C. Peace, A.G. Abbott, L.A. Mueller, M.A. Olmstead, and D. Main. 2014. The genome database for Rosaceae (GDR): Year 10 update. Nucleic Acids Res. 42:D1237–D1244. doi:10.1093/nar/gkt1012

Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, and M.A. Marra. 2009. Circos: An information aesthetic for comparative genomics. Genome Res. 19:1639–1645. doi:10.1101/gr.092759.109

Lu, F., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney, M.D. Casler, E.S. Buckler, and D.E. Costich. 2013. Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. PLoS Genet. 9(1): e1003215. doi:10.1371/journal.pgen.1003215

Njuguna, W., K.E. Hummer, C.M. Richards, T.M. Davis, and N.V. Bassil. 2011. Genetic diversity of diploid Japanese strawberry species based on microsatellite markers. Genet. Resour. Crop Evol. 58:1187–1198. doi:10.1007/s10722-010-9652-7

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

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7(2): e32253. doi:10.1371/journal.pone.0032253

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

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

Sargent, D.J., J. Clarke, D.W. Simpson, K.R. Tobutt, P. Arus, A. Monfort, S. Vilanova, B. Denoyes, M. Rousseau-Gueutin, K.M. Folta, N.V. Bassil, and N.H. Battey. 2006. An enhanced microsatellite map of diploid *Fragaria*. Theor. Appl. Genet. 112:1349–1359. doi:10.1007/s00122-006-0237-y

Sargent, D.J., F. Fernandez-Fernandez, J.J. Ruiz-Roja, B.G. Sutherland, A. Passey, A.B. Whitehouse, and D.W. Simpson. 2009. A genetic linkage map of the cultivated strawberry (*Fragaria ×ananassa*) and its comparison to the diploid *Fragaria* reference map. Mol. Breed. 24(3):293–303. doi:10.1007/s11032-009-9292-9

Sargent, D.J., P. Kuchta, E.L. Girona, H. Zhang, T.M. Davis, J.-M. Celton, A. Marchese, M. Korbin, K.M. Folta, V. Shulaev, and D.W. Simpson. 2011. Simple sequence repeat marker development and mapping targeted to previously unmapped regions of the strawberry genome sequence. Plant Gen. 4:165–177. doi:10.3835/plantgenome2011.05.0014

Sargent, D.J., Y. Yang, N. Surbanovski, L. Bianco, M. Buti, R. Velasco, L. Giongo, and T.M. Davis. 2015. HaploSNP affinities and linkage map positions illuminate subgenome composition in the octoploid, cultivated strawberry (*Fragaria ×ananassa*). Plant Sci. 242:140–150. doi:10.1016/j.plantsci.2015.07.004

Shulaev, V., S.S. Korban, B. Sosinski, A.G. Abbott, H.S. Aldwinckle, K.M. Folta, A. Iezzoni, D. Main, P. Arús, A.M. Dandekar, K. Lewers, S.K. Brown, T.M. Davis, S.E. Gardiner, D. Potter, and R.E. Veilleux. 2008. Multiple models for Rosaceae genomics. Plant Physiol. 147:985–1003. doi:10.1104/pp.107.115618

Shulaev, V., D.J. Sargent, R.N. Crowhurst, T.C. Mockler, O. Folkerts, A.L. Delcher, et al. 2011. The genome of woodland strawberry (*Fragaria vesca*). Nat. Genet. 43:109–116. doi:10.1038/ng.740

Staudt, G. 2005. Notes on Asiatic *Fragaria* species: IV. *Fragaria iinumae*. Bot. Jahrb. 126:163–175. doi:10.1127/0006-8152/2005/0126-0163

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

Tennessen, J.A., R. Govindarajulu, A. Liston, and T.-L. Ashman. 2013. Targeted sequence capture provides insight into genome structure and genetics of male sterility in a gynodioecious diploid strawberry, *Fragaria vesca* ssp. *bracteata* (Rosaceae). G3 Genes Genomes Genetics 3:1341–1351. doi:10.1534/g3.113.006288

Van Ooijen, J.W. 2006. JoinMAP 4.1, Software for the calculation of genetic linkage maps in experimental populations. Wageningen, Netherlands: Kyazma B.V.

Ward, J.A., J. Bhangoo, F. Fernández-Fernández, P. Moore, J. Swanson, R. Viola, R. Velasco, N. Bassil, C.A. Weber, and D.J. Sargent. 2013. Saturated linkage map construction in *Rubus idaeus* using genotyping-by-sequencing and genome-independent imputation. BMC Genomics 14(2). doi:10.1186/1471-2164-14-2