

## Construction of an Integrated High Density Simple Sequence Repeat Linkage Map in Cultivated Strawberry (*Fragaria* × *ananassa*) and its Applicability

SACHIKO N. Isobe<sup>1,\*</sup>, HIDEKI Hirakawa<sup>1</sup>, SHUSEI Sato<sup>1</sup>, FUMI Maeda<sup>2</sup>, MASAMI Ishikawa<sup>2</sup>, TOSHIKI Mori<sup>3</sup>, YUKO Yamamoto<sup>3</sup>, KENTA Shirasawa<sup>1</sup>, MITSUHIRO Kimura<sup>1</sup>, MASANOBU Fukami<sup>2</sup>, FUJIO Hashizume<sup>3</sup>, TOMOKO Tsuji<sup>3</sup>, SHIGEMI Sasamoto<sup>1</sup>, MIDORI Kato<sup>1</sup>, KEIKO Nanri<sup>1</sup>, HISANO Tsuruoka<sup>1</sup>, CHIHARU Minami<sup>1</sup>, CHIKA Takahashi<sup>1</sup>, TSUYUKO Wada<sup>1</sup>, AKIKO Ono<sup>1</sup>, KUMIKO Kawashima<sup>1</sup>, NAOMI Nakazaki<sup>1</sup>, YOSHIE Kishida<sup>1</sup>, MITSUYO Kohara<sup>1</sup>, SHINOBU Nakayama<sup>1</sup>, MANABU Yamada<sup>1</sup>, TSUNAKAZU Fujishiro<sup>1</sup>, AKIKO Watanabe<sup>1</sup>, and SATOSHI Tabata<sup>1</sup>

Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan<sup>1</sup>; Crop Breeding Institute, Chiba Prefectural Agriculture and Forestry Research Center, Mizoguchi 1-7-8, Chousei, Daizenno-cyou 808, Midori, Chiba 299-4335, Japan<sup>2</sup> and Mie Prefecture Agricultural Research Institute, Ureshino Kawakitamachi 530, Matsuzaka, Mie 515-2316, Japan<sup>3</sup>

\*To whom correspondence should be addressed. Tel. +81 438-52-3928. Fax. +81 438-52-3934.  
Email: sisobe@kazusa.or.jp

Edited by Dr Masahiro Yano  
(Received 25 September 2012; accepted 21 November 2012)

### Abstract

The cultivated strawberry (*Fragaria* × *ananassa*) is an octoploid ( $2n = 8x = 56$ ) of the Rosaceae family whose genomic architecture is still controversial. Several recent studies support the AAA'A'BBB'B' model, but its complexity has hindered genetic and genomic analysis of this important crop. To overcome this difficulty and to assist genome-wide analysis of *F.* × *ananassa*, we constructed an integrated linkage map by organizing a total of 4474 of simple sequence repeat (SSR) markers collected from published *Fragaria* sequences, including 3746 SSR markers [*Fragaria vesca* expressed sequence tag (EST)-derived SSR markers] derived from *F. vesca* ESTs, 603 markers (*F.* × *ananassa* EST-derived SSR markers) from *F.* × *ananassa* ESTs, and 125 markers (*F.* × *ananassa* transcriptome-derived SSR markers) from *F.* × *ananassa* transcripts. Along with the previously published SSR markers, these markers were mapped onto five parent-specific linkage maps derived from three mapping populations, which were then assembled into an integrated linkage map. The constructed map consists of 1856 loci in 28 linkage groups (LGs) that total 2364.1 cM in length. Macrosynteny at the chromosome level was observed between the LGs of *F.* × *ananassa* and the genome of *F. vesca*. Variety distinction on 129 *F.* × *ananassa* lines was demonstrated using 45 selected SSR markers.

**Key words:** *Fragaria* × *ananassa*; SSR marker; integrated linkage map; comparative mapping

### 1. Introduction

The cultivated strawberry (*Fragaria* × *ananassa*) is one of the most popular and globally consumed fruit crops. It is cultivated in various regions throughout the world; in 2010, 39% of the cultivated strawberry crop was produced in North and South America, followed by Europe (33%), Asia (18%), and

Africa (9%).<sup>1</sup> Because of the economic importance of this species, breeding programmes of cultivated strawberry are underway in many countries. To date, more than 3000 varieties bred in 41 countries have been registered in the International Union for the Protection and New Varieties of Plants variety database.<sup>2</sup> However, despite intensive use of *F.* × *ananassa* in the industry, the progress of genetic and genomic

research of this crop has lagged behind that of many other economically important plant species, because of its complex genome structure.

*Fragaria* × *ananassa* is an octoploid ( $2n = 8x = 56$ ) species that originated from a natural hybridization between *Fragaria virginiana* and *Fragaria chiloensis*.<sup>3</sup> The genome composition of *F. × ananassa* was initially proposed as AABBBBCC<sup>4</sup> or AAA'A'BBBB,<sup>5</sup> based on results of cytological studies. Later, Bringham<sup>6</sup> proposed the AAA'A'BBB'B' model in light of cytological and genetic evidence. Unlike the first two models, the last model suggests disomic inheritance in entire chromosomes. Although the genome composition of *F. × ananassa* has not yet been confirmed, the AAA'A'BBB'B' model has been supported by recent molecular genetic studies.<sup>7–11</sup> The genus *Fragaria* belongs to the Rosaceae family. Unlike other genera in the Rosaceae family, *Fragaria* comprises a limited number of species (approximately 22).<sup>12</sup> Several species have been nominated as candidate diploid ancestors such as *Fragaria vesca*, *Fragaria iinumae*, and *Fragaria daltoniana*.<sup>13,14</sup> Of these candidates, *F. vesca* is considered to be the most likely diploid ancestor, and it serves as a model species of *F. × ananassa*.<sup>14</sup> Therefore, genomic and genetic studies in *F. vesca* have been performed prior to those of *F. × ananassa*.

In molecular genetics studies, several types of *F. vesca* sequence-derived DNA markers, such as simple sequence repeat (SSR) and sequence-characterized amplified region (SCAR) markers, have been developed since 2003<sup>15–19</sup> (in this article, 'DNA marker' is defined as 'unique DNA sequence(s) identified by specific primers'). Linkage maps were then constructed with a population derived from an interspecific cross between *F. vesca* and *Fragaria nubicola*.<sup>20–23</sup> In genomic studies, bacterial artificial chromosome and fosmid libraries were constructed to investigate features of genome sequences of *F. vesca*.<sup>24,25</sup> Subsequently, a whole genome sequence for *F. vesca* was published in 2011, and the results have greatly contributed to advances in genomic and genetic analysis of the genus *Fragaria*.<sup>26</sup> In parallel, comparative genomics studies within the Rosaceae family have also been performed. Cabrera et al.<sup>27</sup> developed a total of 1039 Rosaceae-conserved orthologous set (RosCOS) markers. Primer pairs of the RosCOS markers were designed on intron-flanking regions of orthologous genes commonly conserved among the genera *Malus*, *Fragaria*, and *Prunus* and mapped onto a *Prunus* reference linkage map. After the genome sequence in *F. vesca* was published,<sup>26</sup> whole genome comparisons between *Fragaria*, *Prunus*, and *Malus* were performed, and major orthologous regions were identified across the genus.<sup>28</sup>

In *F. × ananassa*, the first linkage map was constructed with 235 or 285 amplified fragment length

polymorphism (AFLP) markers on 28 or 30 linkage groups (LGs) based on a 2-way pseudo-testcross strategy.<sup>29</sup> Several linkage maps were subsequently constructed with AFLP, SCAR, sequence-tagged site, random amplified polymorphic DNA, and SSR markers.<sup>8–11,30</sup> All the previously published linkage maps were generated based on single F<sub>1</sub> mapping populations, and integrated linkage maps were developed in each mapping population in all four studies.<sup>8,10,11,30</sup> Of the previously developed linkage maps, the densest map was constructed by Sargent et al.<sup>10</sup>; a total of 549 loci were mapped onto 28 LGs, with a total length of 2140.3 cM. The density of the map was greatly enhanced over previous maps, but several unsaturated LGs were still observed such as the unintegrated LG pair of FG2DA and FG2DB and LG6B that contained large gaps (36.7 cM). These results suggest the need to develop a denser linkage map in *F. × ananassa* to reveal the complex genome structure of this species.

In this study, we performed SSR marker development and constructed an integrated high density linkage map to accelerate the advancement of genomic and genetic studies in *F. × ananassa*. Three types of SSR markers were developed, namely *F. vesca* expressed sequence tag (EST)-derived SSR markers (FVES markers), *F. × ananassa* EST-derived SSR markers (FAES markers), and *F. × ananassa* transcriptome-derived SSR markers (FATS markers), using public genome sequence data. The markers were mapped onto five parent-specific linkage maps, along with previously published DNA markers, and the five parent-specific maps were then integrated into one map. The applicability of the integrated map and the markers developed in this study were also demonstrated by comparative analysis of *F. × ananassa* and *F. vesca* and by variety distinction on 129 *F. × ananassa* lines with selected markers. The markers and integrated linkage map described in this study are valuable resources for future studies that will help to elucidate the genome structure and evolutionary process in *F. × ananassa* and whole genome sequencing, genetic mapping, and molecular breeding of this species.

## 2. Materials and methods

### 2.1. Plant material

An integrated linkage map was constructed using three mapping populations originating from a total of five parental lines. '02–19' × 'Sachinoka' is an F<sub>1</sub> mapping population of 188 individuals derived from a one-way pseudo-testcross. The female parent '02–19' is a breeding line developed in Chiba Prefectural Agriculture and Forestry Research Center, which is resistant to powdery mildew and *Fusarium* wilt. The

male parent 'Sachinoka' is a Japanese variety bred in the National Agricultural Research Center for Kyusyu and Okinawa Region, which was derived from a cross between 'Toyonoka' and 'Aiberry'. 'Kaorino' × 'Akihime' consisted of 140 F<sub>1</sub> individuals derived from a one-way pseudo-testcross between the parental lines. 'Kaorino' and 'Akihime' were bred in Mie Prefecture Agricultural Research Institute and by Mr Kazuhiro Ogiwara in Japan, respectively. One of the ancestor lines of 'Kaorino' is 'Akihime'. 'Kaorino' exhibits crown rot resistance, whereas 'Akihime' is susceptible to the disease. The '0212921' mapping population consisted of 169 S<sub>1</sub> individuals of '0212921'. '0212921' was generated from a cross between two breeding lines developed in Mie Prefecture Agricultural Research Institute, and the male parent of '0212921' was identical to the female parent of 'Kaorino'.

### 2.2. Development of EST-derived SSR markers

SSR regions were identified from the *F. vesca* and *F. × ananassa* ESTs registered in public databases, namely NCBI's nucleotide database (<http://www.ncbi.nih.gov>). The numbers of ESTs that were investigated include 45 739 and 6117 in *F. vesca* and *F. × ananassa*, respectively. SSRs longer than 14 bases, which contained all possible combinations of dinucleotide (NN), trinucleotide (NNN) and tetranucleotide (NNNN) repeats, were identified using the FINDPATTERNS module in the GCG software package (Accelrys Inc., USA). Oligonucleotides for polymerase chain reaction (PCR) primers were designed based on the flanking regions of the identified SSRs using the Primer3 program<sup>31</sup> in such a way that the amplified products ranged from 90 to 300 bp in length. Markers corresponding to previously published SSR markers in the *Fragaria* spp and RosCOS of markers<sup>8,10,15,17–23,27,32–39</sup> were identified based on primer sequences and were excluded from the collection of FVES and FAES markers.

### 2.3. Development of transcriptome-derived SSR markers

A total of 1 188 226 *F. × ananassa* transcript sequence reads registered in NCBI's Sequence Read Archive (SRA, Accession number: SRX16008) were used for identification of the SSR regions. All reads were generated using the Roche 454 sequencing system (Roche, Basel, Switzerland). The MIRA 3.2.0 program was used for assembling non-redundant contigs.<sup>40</sup> Methods for identification of SSR regions and primer design were the same as those used for FVES and FAES marker development, except that penta- and hexanucleotide repeats were also identified and used for primer design. After designing the primer sequences, the redundancy between the FVES, FAES,

and published markers described in the above section was confirmed. The newly developed markers were designated FATS markers.

### 2.4. Polymorphic analysis of the DNA markers

DNA was extracted from the young leaves of plants using a DNeasy Plant Mini Kit (Qiagen Inc., Germany). DNA quantification and quality checks were performed using an ND1000 NanoDrop spectrophotometer (Nanodrop Technologies, DE, USA) and 0.8% agarose gels. In addition to the FVES, FAES, and FATS markers, a total of 1114 primer pairs of previously published SSR markers developed from *Fragaria* sequences and RosCOS markers were used for polymorphic analysis of the '02–19' × 'Sachinoka' mapping population (Supplementary Table S1).<sup>8,15,17–21,23,27,32–38</sup> Polymorphic analysis of the other two populations was performed without using published markers. PCR was performed in a 5- $\mu$ l reaction volume using 0.6 ng of genomic DNA in 1X PCR buffer (Bioline, UK), 3 mM MgCl<sub>2</sub>, 0.08 U of BIOTAQ DNA polymerase (Bioline, UK), 0.8 mM dNTPs, and 0.4  $\mu$ M of each primer. A modified touchdown PCR protocol was followed as described by Sato *et al.*<sup>41</sup> The PCR products were separated by 10% polyacrylamide gel electrophoresis in tris-borate-ethylenediaminetetraacetic acid (TBE) buffer according to the standard protocol or with an ABI 3730xl fluorescent fragment analyzer (Applied Biosystems, USA), according to the polymorphic fragment sizes of the PCR amplicons. In the former case, the data were analysed using the Polyans software (<http://www.polyans.kazusa.or.jp>), whereas in the latter case, polymorphisms were investigated using the GeneMapper software (Applied Biosystems, USA).

### 2.5. Linkage analysis

In this study, it was assumed that *F. × ananassa* is an allo-octoploid species and that polymorphic loci in entire chromosomes showed disomic inheritance. Therefore, linkage analysis was performed using the methodology employed with diploid and outcrossing species. The segregated data of the two pseudo-testcross mapping populations were categorized into two parent-specific data sets by comparing the sizes of polymorphic bands of the parents and progeny. The segregation data were rescored using the 'HAP1' or 'F2' population type codes employed in JoinMap analysis.<sup>42</sup> 'HAP1' codes were used for the four parental lines of the '02–19' × 'Sachinoka' and 'Kaorino' × 'Akihime', whereas 'F2' codes were employed for the parents of the S<sub>1</sub> population '0212921'. As a result, a total of five parent-specific data sets were generated. The segregation data from each parent-specific data set were then classified into multiple LGs using the colour map method<sup>43</sup> that employed a comparison

of graphical genotypes of the segregation data. During the process of colour mapping, reciprocal genotypes were converted to coupling genotypes. The robustness of the data sets for each LG was then confirmed using the Grouping Module of the JoinMap program, version 4, with a logarithm of odds (LOD) threshold of 10. The homeologous LGs within each parent-specific map were assumed based on corresponding positions on the *F. vesca* genome, which were predicted by comparative analysis (described below).

The locus orders in each parent-specific map were calculated using the Regression Mapping Module of JoinMap. '0212921'-specific data were handled as an 'F2' population type, whereas other parent-specific data sets were calculated as 'HAP1' population types. The following parameters were used for the calculation: Kosambi's mapping function,  $LOD \geq 1.0$ , REC frequency  $\leq 0.4$ , goodness of fit jump threshold for removal of loci = 5.0, number of added loci after which a ripple is performed = 1, and third round = yes. Each LG in parent-specific maps was named according to the following rule: numbers after LG (1–7) showed corresponding chromosome numbers of *F. vesca* predicted by comparative analysis (described below). The capital letters A–D distinguished four LGs in each homeologous group (HG), which corresponded to LGs in an integrated map. The A–D suffixes corresponded to the lengths of the LGs on the integrated map. When multiple LGs were integrated into a single LG of the integrated map, they were numbered using capital letters with underbars (e.g. LG1B\_1 and LG1B\_2). For unintegrated LGs, X or Y was used as a suffix.

For construction of an integrated linkage map, corresponding LGs between the parent-specific maps were assumed by identification of commonly mapped markers on each LG. Prior to integration, genotypes of dominant loci in each parent-specific map were imputed to co-dominant genotypes, according to the flanking genotypes of co-dominant loci. That is, alleles of dominant loci were converted to co-dominant alleles, according to the flanking genotypes of co-dominant loci. Parent-specific locus genotype data sets were then integrated into one dataset in each LG using the Combine Groups for Mapping Integration Module, followed by locus ordering by the Regression Mapping Module of JoinMap. The parameters used for the mapping module of an integrated map were the same as those used for parent-specific mapping. After construction of an integrated map, the mapped loci were classified into two groups, i.e. loci generated from single-locus diagnostic (SLD) and multi-loci diagnostic (MLD) markers. SLD markers were defined as markers that detected single segregation bands mapped onto single positions on an integrated map, and not amplified other

loci, such as monomorphic loci, whereas MLD markers were defined as markers that amplified more than one locus, including monomorphic loci.

## 2.6. Comparative mapping

Syntenic regions between the genomes of *F. × ananassa* and *F. vesca* were detected by identifying the conservation of the relative locations of genes and genomic regions. The source of the genome sequences of *F. vesca* was obtained from NCBI [Accession numbers: CM001053.1-CM001059.1 (LG1-7), GG775183.1- GG775301.1 (unplaced)]. The EST sequences adjacent to the mapped markers on the integrated *F. × ananassa* map were compared with the gene sequences in the reference genomes using the BLASTX program with a cutoff *E*-value  $\leq 1e-10$ . The syntenic regions defined by the top hits of the homology search were plotted using the Cicros program (<http://circos.ca>).

## 2.7. Distinguishing the varieties of 129 *F. × ananassa* lines using selected EST-SSR markers

In order to distinguish the varieties of 129 *F. × ananassa* lines, a total of 22 *F. × ananassa* varieties and breeding lines, most of which were bred in Japan, were used in pre-polymorphic analysis, including the following: 'Sachinoka', 'Fusanoka', 'Asuka Wave', 'Kaorino', 'Akihime', 'Miyoshi', 'Dover', 'Strawberry Parental Line Nou-2', 'Karenberry', 'Ohkimi', 'Sanchiigo', 'Toyonoka', 'Nyohou', 'Tochiotome', 'Nou-Hime', 'Asuka ruby', 'Sagahonoka', 'Beni Hoppe', 'Yayoihime', 'Sanukihime', '02-19', and '0212921'. PCR was performed with all the FVES and FAES markers developed in this study. The PCR protocol was same as that used for polymorphic analysis of the DNA markers (see above). The PCR products were separated by 10% polyacrylamide gel electrophoresis in TBE buffer using the standard protocol. SSR markers showing solid and polymorphic amplification were selected, and the robustness of PCR, i.e. the repeatability of results and the absence of noise peaks, was investigated for the selected markers using a fragment analyzer (ABI 3730xl, Applied Biosystems, USA) for 22 *F. × ananassa* lines. PCR was performed in a 5- $\mu$ l reaction volume using 1 ng of genomic DNA in 1X PCR buffer II (Applied Biosystems, USA), 2.5 mM MgCl<sub>2</sub>, 0.125 U of AmpliTaq Gold DNA polymerase (Applied Biosystems, USA), 0.2 mM dNTPs, and 0.4  $\mu$ M of each primer. The thermal cycling conditions were as follows: 7 min denaturation at 95°C; 30 cycles of 30 s denaturation at 95°C, 30 s annealing at 57°C, and 1 min extension at 72°C, with a final 10 min extension at 72°C. A total of 100 markers showing high repeatability were screened, and the robustness of PCR was again confirmed for 129 *F. × ananassa* lines, 119 of which were previously investigated polymorphisms identified using

CAPS markers<sup>44</sup> (Supplementary Table S2). The 45 best SSR markers were thereby selected to distinguish the varieties of 129 strawberry lines. The allelic data were converted into a binary matrix using the scores 1 or 0 for the presence or absence of the peak. The expected heterozygosity (HZ) of each identified peak was calculated using the following formula:

$$HZ_i = 1 - (P_{ip}^2 + P_{ia}^2),$$

where  $P_{ip}$  and  $P_{ia}$  are the frequency of presence and absence of the  $i$ th peak. In  $F. \times ananassa$ , SSR markers often identify multi loci due to polyploidy, and it is often difficult to investigate the exact number of loci identified in each marker in an unrelated population. Therefore, the mean HZ value of identified peaks generated by a marker was substituted for the HZ of each marker. The allelic binary data were also analysed using GGT 2.0<sup>45</sup> for the investigation of genetic similarity between the varieties using a Jaccard similarity coefficient. A unweighted pair group method (UPGMA) with arithmetic average dendrogram was constructed using MEGA version 5.05.<sup>46</sup>

### 3. Results

#### 3.1. Development of EST-SSR markers

A total of 2748 and 324 SSRs were identified in non-redundant 15 203 *F. vesca* and 1029 *F. \times ananassa* ESTs, respectively. Of the identified SSRs, 562 primers were designed based on the flanking regions of 453 *F. vesca*- and 109 *F. \times ananassa*-derived EST-SSRs and designated FVES and FAES markers. To increase the number of EST-SSR markers, additional primer pairs were designed that allowed either single-base mismatches (779 and 107 primer pairs for FVES and FAES markers, respectively) or two-base mismatches (2514 and 387 primer pairs for FVES and FAES markers, respectively) in the SSR regions. Of the markers that were generated, those corresponding to previously published markers were excluded.<sup>8,10,15,17–23,27,32–39</sup> As a result, a total of 3746 FVES and 603 FAES markers were developed. Design details of the FVES and FAES markers, including primer sequences, corresponding SSR motifs, expected product sizes, and GenBank IDs of the template EST sequences, are listed in Supplementary Tables S3 and S4 and on the web at <http://marker.kazusa.or.jp/strawberry/>.

Of the FVES and FAES markers that were developed, 2975 (79.4%) and 431 (71.5%) were trinucleotide repeats while 456 (12.2%) and 117 (19.4%) were dinucleotide repeats, and 315 (8.4%) and 55 (9.1%) were tetranucleotide repeats, respectively (Table 1).

In the FVES markers, the poly (AAG)<sub>n</sub> motif was most abundant in the trinucleotide repeats (755 SSRs, 20.1%), followed by poly (GGA)<sub>n</sub> (492 SSRs, 13.1%), poly (ATC)<sub>n</sub> (352 SSRs, 9.4%), poly (AG)<sub>n</sub> (347 SSRs, 9.3%), and poly (AGC)<sub>n</sub> (321 SSRs, 8.6%). Like the FVES markers, the poly (AAG)<sub>n</sub> motif was the most frequently observed motif in the FAES markers (132 SSRs, 21.9%). This was followed by poly (AG)<sub>n</sub> (81 SSRs, 13.4%), poly (ATC)<sub>n</sub> (68 SSRs, 11.3%), and poly (AGC)<sub>n</sub> (48 SSRs, 8.0%). Among the tetranucleotide repeats, AT-rich motifs, namely poly (AAAG)<sub>n</sub> and poly (AAAC)<sub>n</sub>, were more frequently observed than other motifs in both the FVES and FAES markers.

#### 3.2. Development of transcriptome-SSR markers

A total of 1 188 226 SRA sequences were assembled into 80 430 contigs by the MIRA 3.2.0 program.<sup>38</sup> On these contigs, 34 993 SSRs were identified. Primers that targeted the flanking regions of 129 of the 34 993 SSRs were designed and synthesized. All the 129 SSRs were located on non-redundant contigs, and a total of 125 primers, which did not overlap with the FVES, FAES, and previously published markers,<sup>8,10,15,17–23,27,32–39</sup> were designed as FATS markers. The primer sequences of the FATS markers, along with the corresponding SSR motifs, product sizes, primer sequences, and template contigs, are provided on the web at <http://marker.kazusa.or.jp/strawberry/> and in Supplementary Table S5. Of the 125 FATS markers, trinucleotide repeats were the most frequently observed (56 SSRs, 44.8%), followed by hexanucleotide repeats (37 SSRs, 29.6%), and dinucleotide repeats (25 SSRs, 20.0%, Table 1). No tetranucleotide repeats were observed, whereas seven pentanucleotide repeats were identified. Of the SSR motifs, the poly (AG)<sub>n</sub> motif was the most abundant (10 SSRs, 8.0%), followed by poly (TC)<sub>n</sub> (7 SSRs, 5.6%), and poly (AAG)<sub>n</sub> (5 SSRs, 4.0%). In the penta- and hexanucleotide repeats, each single SSR was identified for all the observed motifs except poly (GCTGT)<sub>n</sub> (two SSRs, data not shown).

#### 3.3. Construction of parent-specific linkage maps

**3.3.1. The '0212921' S<sub>1</sub> mapping population** A total of 4474 primer pairs of SSR markers, including 3746 FVES, 603 FAES, and 125 FATS, were investigated with 8 randomly chosen S<sub>1</sub> individuals of the '0212921' mapping population. Polymorphisms were observed on 605 FVES, 135 FAES, and 29 FATS markers, and segregation analysis was performed for 169 S<sub>1</sub> individuals with 769 primer pairs. A total of 881 segregation locus genotypes were generated from the 769 primer pairs, and 822 of the 881 loci were mapped onto 34 LGs

**Table 1.** Numbers of SSR motifs in the FVES, FAES, and FATS markers

Motif	FVES				FAES				FATS			
	Designed (%)		Mapped (%)		Designed (%)		Mapped (%)		Designed (%)		Mapped (%)	
Dinucleotide												
AG	347	(9.3)	120	(11.0)	81	(13.4)	34	(17.7)	10	(8.0)	4	(10.3)
AT	60	(1.6)	21	(1.9)	22	(3.6)	12	(6.3)	0	(0.0)	0	(0.0)
AC	49	(1.3)	18	(1.7)	14	(2.3)	5	(2.6)	1	(0.8)	0	(0.0)
TC	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	7	(5.6)	4	(10.3)
Others	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	7 <sup>a</sup>	(5.6)	0	(0.0)
Subtotal	456	(12.2)	159	(14.6)	117	(19.4)	51	(26.6)	25	(20.0)	8	(20.5)
Trinucleotide												
AAG	755	(20.1)	210	(19.3)	132	(21.9)	41	(21.4)	5	(4.0)	1	(2.6)
GGA	492	(13.1)	138	(12.7)	45	(7.5)	6	(3.1)	4	(3.2)	3	(7.7)
ATC	352	(9.4)	104	(9.6)	68	(11.3)	16	(8.3)	2	(1.6)	1	(2.6)
AGC	321	(8.6)	92	(8.4)	48	(8.0)	20	(10.4)	3	(2.4)	2	(5.1)
GGC	291	(7.8)	77	(7.1)	27	(4.5)	11	(5.7)	1	(0.8)	0	(0.0)
GGT	268	(7.2)	78	(7.2)	40	(6.6)	7	(3.6)	1	(0.8)	0	(0.0)
AAC	207	(5.5)	58	(5.3)	34	(5.6)	11	(5.7)	0	(0.0)	0	(0.0)
ACG	168	(4.5)	50	(4.6)	16	(2.7)	5	(2.6)	0	(0.0)	0	(0.0)
ACT	62	(1.7)	8	(0.7)	5	(0.8)	0	(0.0)	0	(0.0)	0	(0.0)
AAT	59	(1.6)	17	(1.6)	16	(2.7)	6	(3.1)	0	(0.0)	0	(0.0)
CTT	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	4	(3.2)	1	(2.6)
Others	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	36 <sup>a</sup>	(28.8)	12	(30.8)
Subtotal	2975	(79.4)	832	(76.4)	431	(71.5)	123	(64.1)	56	(44.8)	20	(51.3)
Tetranucleotide												
AAAG	100	(2.7)	37	(3.4)	15	(2.5)	4	(2.1)	0	(0.0)	0	(0.0)
AAAC	54	(1.4)	14	(1.3)	7	(1.2)	1	(0.5)	0	(0.0)	0	(0.0)
AAAT	43	(1.1)	11	(1.0)	14	(2.3)	6	(3.1)	0	(0.0)	0	(0.0)
GGGA	26	(0.7)	8	(0.7)	3	(0.5)	2	(1.0)	0	(0.0)	0	(0.0)
AAGC	21	(0.6)	6	(0.6)	2	(0.3)	1	(0.5)	0	(0.0)	0	(0.0)
AATC	17	(0.5)	4	(0.4)	5	(0.8)	1	(0.5)	0	(0.0)	0	(0.0)
AATG	15	(0.4)	7	(0.6)	1	(0.2)	1	(0.5)	0	(0.0)	0	(0.0)
Others	39	(1.0)	11	(1.0)	8	(1.3)	2	(1.0)	0	(0.0)	0	(0.0)
Subtotal	315	(8.4)	98	(9.0)	55	(9.1)	18	(9.4)	0	(0.0)	0	(0.0)
Pentanucleotide												
Subtotal	—	—	—	—	—	—	—	—	7 <sup>a</sup>	(5.6)	1	(2.6)
Hexanucleotide												
Subtotal	—	—	—	—	—	—	—	—	37 <sup>a</sup>	(29.6)	10	(25.6)
Total	3746	(100)	1089	(100)	603	(100)	192	(100)	125	(100)	39	(100)

<sup>a</sup>The numbers of types of observed 'other' SSR motifs in di- and trinucleotide repeats and all penta- and hexa- nucleotide repeats were 4, 24, 6, and 37, respectively.

(Supplementary Tables S6 and S7). The length of each LG ranged from 1.5 cM (LG4B) to 94.5 cM (LG1A), representing a total length of 1508.3 cM. The mean locus density and segregation distortion ( $P < 0.05$ ) were 1.83 cM locus<sup>-1</sup> and 22.4%, respectively, ranging from 0.22 cM locus<sup>-1</sup> (LG6D) to 4.26 cM locus<sup>-1</sup> (LG2A\_2) and from 0.0% (LG4B) to 85.7% (LG5A\_1), respectively.

3.3.2. *The '02-19' × 'Sachinoka' mapping population* Polymorphic analysis was performed with a total of 5588 primer pairs of SSR markers, including 4474 SSR markers developed in this study and 1114 previously published markers (Supplementary Table S1). A total of 1299 markers, i.e. 853 FVES, 131 FAES, 37 FATS, and 278 published markers, showed polymorphisms between the

**Table 2.** Number of mapped loci, length, locus density in an integrated map, and numbers of integrated LGs and loci of parental specific maps

LG	Interspecific map				Parental specific maps									
	Number of mapped loci	Length (cM)	Locus density (cM)	Single loci <sup>a</sup> (%)	'0212921'		'02-19'		'Sachinoka'		'Kaorino'		'Akihime'	
					LGs <sup>b</sup>	Loci <sup>c</sup>	LGs <sup>b</sup>	Loci <sup>c</sup>	LGs <sup>b</sup>	Loci <sup>c</sup>	LGs <sup>b</sup>	Loci <sup>c</sup>	LGs <sup>b</sup>	Loci <sup>c</sup>
1A	91	123.7	1.36	16 (17.5)	1	48	1	40	1	13	1	7	2	10
1B	81	80.7	1.00	25 (30.8)	1	25	1	41	2	42	2	14	0	0
1C	37	77.2	2.09	11 (29.7)	1	24	1	9	1	13	0	0	0	0
1D	24	34.2	1.43	5 (20.8)	1	24	0	0	0	0	0	0	0	0
2A	123	113.9	0.93	41 (33.3)	2	25	2	50	1	67	1	8	2	22
2B	71	95.1	1.34	15 (21.1)	1	42	2	20	1	4	1	11	1	15
2C	55	93.8	1.71	13 (23.6)	1	28	0	0	1	18	1	24	0	0
2D	64	81.1	1.27	19 (29.6)	1	37	1	28	1	7	1	8	0	0
3A	114	96.0	0.84	47 (41.2)	1	33	2	70	2	27	1	23	1	18
3B	89	82.7	0.93	19 (21.3)	1	40	2	37	1	24	0	0	1	11
3C	51	70.8	1.39	13 (25.4)	1	36	0	0	1	13	1	11	0	0
3D	38	65.8	1.73	5 (13.1)	1	30	0	0	0	0	1	9	1	3
4A	88	112.2	1.28	31 (35.2)	1	24	1	45	2	19	0	0	2	42
4B	18	91.4	5.08	2 (11.1)	1	3	1	5	1	13	0	0	0	0
4C	64	87.9	1.37	6 (9.37)	2	32	2	16	1	12	1	15	1	13
4D	45	53.0	1.18	7 (15.5)	1	27	0	0	1	13	2	14	1	5
5A	130	119.0	0.92	40 (30.7)	2	25	2	53	2	77	2	9	1	9
5B	57	79.3	1.39	13 (22.8)	1	35	0	0	0	0	1	26	2	15
5C	18	74.8	4.16	3 (16.6)	0	0	1	12	0	0	0	0	1	10
5D	23	46.0	2.00	5 (21.7)	1	10	0	0	0	0	1	15	0	0
6A	131	109.8	0.84	31 (23.6)	3	37	4	37	2	44	3	16	4	49
6B	129	87.9	0.68	32 (24.8)	2	86	1	29	2	31	2	15	2	9
6C	59	87.4	1.48	11 (18.6)	1	43	1	4	1	18	1	6	0	0
6D	49	76.4	1.56	10 (20.4)	1	11	0	0	1	15	2	24	2	23
7A	80	115.7	1.45	24 (30.0)	1	27	2	25	1	33	1	17	1	24
7B	37	72.4	1.96	11 (29.7)	1	30	0	0	0	0	1	5	1	9
7C	24	68.6	2.86	6 (25.0)	1	14	0	0	2	15	0	0	0	0
7D	66	67.3	1.02	20 (30.3)	1	21	2	35	2	14	1	3	2	14
Unintegrated	—	—	—	—	1	5	3	20	4	24	4	14	5	17
Total	1856	2364.1	1.27	481 (25.9)	34	822	32	576	34	556	32	294	33	318

<sup>a</sup>Number of mapped loci generated from SLD markers. Numbers in parentheses show percentage of all the mapped loci.

<sup>b</sup>Number of integrated LGs.

<sup>c</sup>Number of integrated loci.

parental lines. By performing segregation analysis of the 188 individuals of the mapping population, 1078 polymorphic loci were generated from 881 of the 1299 primer pairs that were tested. Of the 1078 segregation loci, 260 showed biparental polymorphisms while the other 818 were parent specific. A total of 575 and 556 loci were mapped onto '02-19' and 'Sachinoka' specific maps, respectively (Supplementary Tables S6 and S7). In the '02-19' specific map, 32 LGs were constructed on 1668.9 cM, with

lengths ranging from 2.0 cM (LG2X) to 129.2 cM (LG1A). The mean locus density and segregation distortion ( $P < 0.05$ ) were 2.90 cM locus<sup>-1</sup> and 30.2%, respectively, ranging from 0.34 cM locus<sup>-1</sup> (LG5A\_2) to 11.63 cM locus<sup>-1</sup> (LG6C) and from 0.0% (LG3B\_2, 5A\_2 and 7D\_2) to 100% (LG2X and 4B), respectively. In the 'Sachinoka' map, 34 LGs were developed, totaling 2166.6 cM in length, with a mean locus density of 3.90 cM locus<sup>-1</sup>. The length and locus density of each LG ranged from 1.6 cM (LG7C\_2) to 120.0 cM

(LG7A) and from 0.80 cM locus<sup>-1</sup> (LG7C\_2) to 16.08 cM locus<sup>-1</sup> (LG3X), respectively. Segregation distortion ( $P < 0.05$ ) of each LG ranged from 0.0% (LG7C\_2, 7D\_1 and 7X) to 85.7% (LG2X), with a mean value of 25.4%.

**3.3.3. The 'Kaorino' × 'Akihime' mapping population** A total of 4474 primer pairs of SSR markers developed in this study were used for polymorphic analysis with the parental lines. Polymorphisms were observed in 438 FVES, 95 FAES, and 33 FATS markers, and segregation analysis was performed with the 140 mapping plants, for a total of 566 primer pairs. A total of 612 segregation locus genotypes were generated from 537 of the 566 primer pairs that were tested. Of the 612 segregation loci, 140 showed biparental polymorphisms while the other 472 were parent specific. A total of 294 and 318 loci were mapped onto 'Kaorino' and 'Akihime' specific maps, respectively (Supplementary Tables S6 and S7). The 'Kaorino' map consisted of 32 LGs, representing a total length of 1103.4 cM. The mean locus density and segregation distortion ( $P < 0.05$ ) were 3.75 cM locus<sup>-1</sup> and 24.8%, respectively, ranging from 0.45 cM locus<sup>-1</sup> (LG1B\_2) to 7.61 cM locus<sup>-1</sup> (LG1A) and from 0.0% (LG1B\_2, 1X, 2A, 2D, 4X, 5A\_1, 5X, 6A\_2, 6A\_3 and 7D) to 100% (LG4D\_2), respectively. In the 'Akihime' map, 33 LGs were developed, with a total length of 951.4 cM and a mean locus density of 2.99 cM locus<sup>-1</sup>. The length and locus density of each LG ranged from 2.2 cM (LG5X) to 77.7 cM (LG7A) and from 0.55 cM locus<sup>-1</sup> (LG5X) to 12.72 cM locus<sup>-1</sup> (LG6A\_4), respectively. Segregation distortion ( $P < 0.05$ ) of each LG ranged from 0.0% (LG3X, 4A\_2, 4C, 5A, 5C, 5X, 6A\_1, 6B\_1, 6B\_2, and 7D\_2) to 100% (LG1A\_1), with a 14.8% mean value.

#### 3.4. Construction of an integrated linkage map

Prior to integration, the subsets of parent-specific LGs to be integrated were determined by the numbers of commonly mapped markers across parent-specific maps, that is, pairs of LGs showing the largest number of commonly mapped markers were assembled into the same subset. When one LG showed the same numbers of commonly mapped markers in different pairs, the LG was excluded from the integration. Locus genotype data of each subset of LG were integrated using the Combine Groups for Mapping Integration Module in JoinMap, and the locus orders of an integrated map were then calculated. When grouped LG subsets were inadequate, the segregation loci of incorrect LGs were excluded during the process of locus ordering. When a LG subset was misassembled with incorrect pairs of

parent-specific LGs, the loci of the mis-integrated LG overlapped with loci onto other integrated LGs which correctly paired LGs were integrated. Therefore, the correct assembly of a LG subset was determined by confirming the number of overlapping loci on each integrated LG using the Mapping module in JoinMap.

The integrated linkage map consisted of 1856 loci on 28 LGs, totalling 2364.1 cM in length (Table 2, Supplementary Table S7 and Fig. S1). The length of each LG varied from 34.2 (LG1D) to 123.7 cM (LG1A). The mean locus density was 1.27 cM locus<sup>-1</sup>, ranging from 0.68 cM locus<sup>-1</sup> (LG6B) to 5.08 cM locus<sup>-1</sup> (LG4B). The largest gap was 27.3 cM, between FVES1598\_7a and FVES1351\_7a on LG7A, followed by a 23.4 cM gap between FAES0326\_4b and FAES3462\_4b on LG4B, and a 21.6 cM gap between FVES0576\_4c and FAES0045 on LG4C. The number of integrated LGs ranged from 1 (LG1D) to 16 (LG6A) parent-specific LGs (Table 2). The number of unintegrated parent-specific LGs, which totalled 17, ranged from 1 ('0212921') to 5 ('Akihime') in each parental map. The ratios of mapped FVES, FAES, and FATS loci were 77.5, 15, and 3.4%, respectively (Supplementary Fig. S2). The ratio of mapped FVES markers in each LG ranged from 62.5 to 93.3%.

Of the mapped loci, 481 (25.9%) were generated from SLD markers, whereas the other 1375 were from MLD markers. The ratio of mapped loci generated from SLD markers in each LG varied from 9.4% (LG4C) to 41.2% (LG3A). These loci were mapped randomly onto many of the LGs while several clusters were observed on parts of LGs (LG1A, 1C, 2C, 4C, 4D, 6C, 6D, 7B, and 7C, Supplementary Fig. S1). Loci generated from MLD markers were classified into four types; multi loci mapped onto homoeologous LGs (Multi\_H in Supplementary Table S7), non-homeologous LGs (Multi\_NH), the same LGs (Multi\_S), and single position (Multi\_NM). For the Multi\_NM loci, all the observed multiple bands were monomorphic except for mapped ones. Of the four types of multi loci, Multi\_H was most frequently observed (684 loci), followed by Multi\_NM (523), Multi\_S (206), and Multi\_NH (77). Of the 684 Multi\_H loci, 16 and 99 were observed, corresponding to loci on non-homeologous LGs (Multi\_H&NH) and the same LGs (Multi\_H&S), respectively. The Multi\_H loci were randomly distributed along the entire integrated linkage map, whereas Multi\_S loci were not observed on several LGs, i.e. LG2B, 2D, 3C, 4D, 5A, 5B, 7B, and 7C (Supplementary Fig. S1). In each HG, the homeologous regions differed depending on the paired LGs (Supplementary Fig. S3). For example, most homeologous regions of LG2A and LG2C were observed on 20–40 cM while those of LG2B and LG2D were



identified on 60–80 cM. In the case of HG1, homeologous regions were not observed between LG1C and LG1D. Such biases of homeologous pairs were identified across the genome.

### 3.5. Comparison with the genomes of a wild relative, *F. vesca*

Of the 3746 ESTs that corresponded to the designed FVES markers, 3743 showed significant similarities to the genome sequences of *F. vesca*, while 3 ESTs, corresponding to FVES1248, FVES2637, and FVES2807, did not (Supplementary Table S3). Of the 3743 ESTs, 3608 showed similarities to genome sequences placed on 7 chromosomes of *F. vesca*, and the other 135 were mapped onto unplaced genomic scaffolds. In *F. × ananassa* sequence-derived markers, significant similarities to the *F. vesca* genome sequences were observed in all 603 ESTs and 124 of the 125 transcript contigs from which FAES and FATS markers were designed, respectively (Supplementary Table S4 and S5). Of these, 593 and 120 sequences corresponding to FAES and FATS markers, respectively, were mapped onto *F. vesca* genome sequences placed on the 7 chromosomes, whereas the other 14 ESTs or transcript contigs were mapped onto unplaced scaffolds. Similarity searches were also performed for the previously published markers that were located on the integrated map. Of the 91 mapped previously published markers, 68 ESTs or genome sequence are available on the NCBI dbGSS database (<http://www.ncbi.nlm.nih.gov/projects/dbGSS>, Supplementary Table S1) from which the markers were designed. Of the 68 sequences, 62 showed significant similarities to the *F. vesca* genome sequences placed onto the 7 chromosomes.

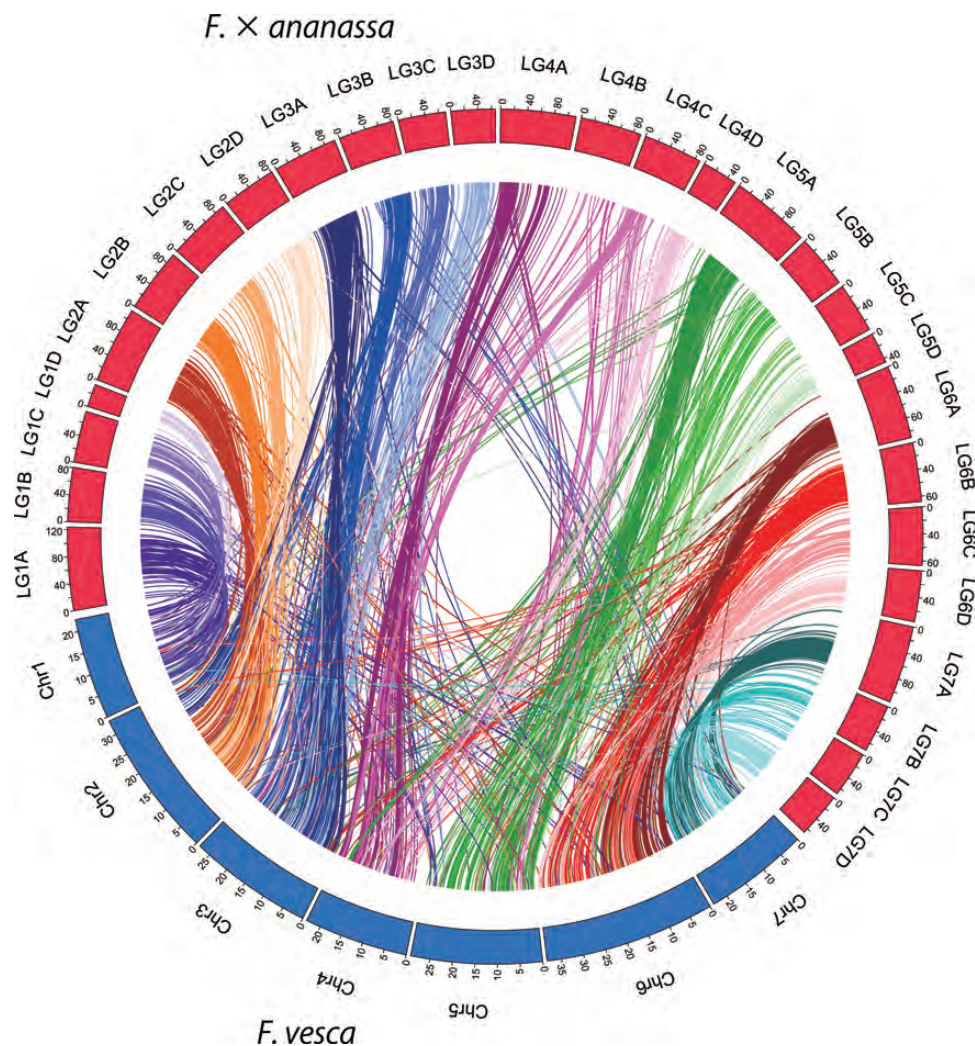
A total of 1354 markers that generated 1783 mapped loci onto the integrated linkage map showed significant similarity with *F. vesca* genome sequences placed onto the 7 chromosomes. By considering the sequences with highest similarity scores to be putative orthologs, the map locations of the SSR markers and the corresponding *F. vesca* genome sequences were compared. As shown in Fig. 1, the alignment of homologous sequence pairs along each LG revealed an obvious syntenic relationship between corresponding HGs in *F. × ananassa* (Fa-HG) and chromosomes (Chr) in *F. vesca* (Fv-Chr). Most of the regions did not show synteny between non-corresponding Fa-HGs and Fv-Chr, except between Fa-HG3 and 25–30 Mb on Fv-Chr2. In some LGs in *F. × ananassa*, syntenic regions were observed on whole corresponding chromosomes in *F. vesca*, whereas segmental syntenic blocks were identified between pairs of the other Fa-LGs and Fv-Chr (Supplementary Fig. S3-1). For

example, Fa-LG1A displayed syntenies against entire regions of Fv-Chr1, whereas segmental syntenic blocks were observed between 0 and 11 Mb in Fv-Chr1 and Fa-LG1B or Fa-LG1D, as well as between 8 and 23 Mb in Fv-Chr1 and Fa-LG1C.

### 3.6. Distinguishing the varieties of 129 *F. × ananassa* lines using selected EST-SSR markers

To select the best marker set to determine the varieties of the *F. × ananassa* lines, polymorphic analysis of the 22 cultivated strawberry lines was performed using primer pairs targeting 3746 FVES and 603 FAES markers. In the primer pairs of FVES markers, 2949 resulted in solid amplification, whereas of the remaining primer pairs, 431, 263, and 103 resulted in no amplification, multiple bands, and rare bands, respectively (Supplementary Table S3). In the FAES markers, solid amplification was observed for 460 primer pairs, whereas of the remaining primer pairs, 88, 52, and 3 resulted in no amplification, multiple, and rare bands, respectively (Supplementary Table S4). A total of 751 primer pairs, including 650 FVES and 101 FAES, showed solid and polymorphic amplification and were used for the subsequent investigation of PCR, i.e. the stability of PCR and detection of noise peaks, using a fragment analyzer for all 22 lines. A total of 100 of the 751 primer pairs showing high stability and few noise peaks were selected and used for polymorphic analysis with 129 strawberry lines to reconfirm the robustness of PCR in a more diversified collection. From this experiment, the 45 best markers, which showed high stability in PCR and few noise peaks, were selected. PCR was then performed twice more with the 129 strawberry lines to exclude genotyping errors (Supplementary Table S3).

The primer pairs of 45 selected SSR markers, including 4 FAES and 41 FVES markers, generated a total of 158 peaks in the 129 strawberry lines that were tested (<http://vim.kazusa.or.jp/Strawberry/>). The number of identified peaks per a primer pair ranged from two to seven, with a mean value of 3.51 (Supplementary Fig. S4). The HZ value of each identified peak was calculated based on a score of 1 or 0 for the presence and absence of the peak and ranged from 0 to 0.5. The mean HZ value of each marker ranged from 0.11 to 0.43, with an average value of 0.25. Similarity coefficients were used to examine the genetic relationships between the 129 strawberry lines. All possible genotypes showed similarity coefficients ranging from 0.00 to 0.41 (Supplementary Table S2). No genetic differences were identified with the 45 markers between 'Nyoho' and 'Shinnyoho' or between 'Himatsuri' and 'Toyonoka'. The highest similarity coefficient, 0.41, was found



**Figure 1.** Graphical view of syntentic relationship between *F. × ananassa* and *F. vesca*. Red and blue bars show LGs of *F. × ananassa* and chromosomes of *F. vesca*, respectively. Syntenic regions between the two species are connected by coloured lines.

between the UK variety ‘Serenata’ and the Japanese variety ‘Summer berry’.

#### 4. Discussion

In this study, a total of 4474 SSR markers, including 3746 FVES, 603 FAES, and 125 FATS, were designed from public EST and transcript sequences of *F. vesca* and *F. × ananassa*. Of the 4474 SSR markers, 672 resulted in rare or no amplification, and the remaining 3802 SSR markers resulted in amplification. To our knowledge, 441 SSR markers have previously been published for the genus *Fragaria*.<sup>8,10,15,17–23,27,32–39</sup> The number of SSR markers that were developed in this study is approximately 8.6 times that of previously published markers. Recent advances in genome sequencing technology have enabled the large-scale development of single nucleotide polymorphism (SNP) markers in

many of plant species. However, SNP discovery and genotyping are still difficult for polyploidy species because of the difficulty in distinguishing between homeologous allelic SNPs. Therefore, SSR markers are currently the most rapid and conclusive marker system for most polyploid species, including *F. × ananassa*. The numerous SSR markers developed in this study are an important resource for genetic and genomic studies of this species.

In this study, linkage analysis was performed using a methodology previously employed in diploid and outcrossing species with the assumption that *F. × ananassa* is an allo-octoploid species with an AAA’A’BBB’B’ genome structure. The large number of mapped loci that showed disomic inheritance supported this assumption.

In the ‘02–19’ × ‘Sachinoka’ and ‘Kaorino’ × ‘Akihime’ populations, polymorphic marker screening was only performed with parental lines that caused failure of screening polymorphic markers showing a

<ab × ab> segregation pattern. Meanwhile, eight randomly selected S<sub>1</sub> individuals were used in polymorphic marker screening of the '0212921' population, and all possible polymorphic markers were selected. Therefore, we considered that the large differences of the number of mapped loci in the parent-specific maps were due to the steps in polymorphic marker screening and the genetic distances of haplotypes within parents. All the previously published maps in *F. × ananassa* were derived from F<sub>1</sub> mapping populations, and the '0212921' specific map is the first parent-specific map derived from an S<sub>1</sub> population. The density of the parent-specific map '0212921' suggests that an S<sub>1</sub> population is available for linkage map construction in *F. × ananassa*.

According to a previous study<sup>10</sup> and the integrated map developed in this study, the length of a saturated linkage map in *F. × ananassa* would exceed 2000 cM. The lengths of all the parent-specific maps were shorter than 2000 cM, except the 'Sachinoka' map that was 2166.6 cM. In addition, no parent-specific map had 28 LGs, i.e. the number of chromosomes in the haploid genome of *F. × ananassa*. Therefore, it was concluded that all parental specific maps were unsaturated.

The integrated linkage map comprises 1856 loci on 28 LGs. The total length of the map, 2364 cM, is slightly longer than the previously reported densest map, 2140 cM, by Sargent *et al.*<sup>10</sup> The mean locus density of our integrated map was 1.27 cM locus<sup>-1</sup>, which was three times denser than that of the map of Sargent *et al.* Moreover, the integrated map presented in this study is the first linkage map derived from multiple mapping populations of *F. × ananassa*, suggesting that the map reflects wider genetic diversity in the species. Several LGs of the integrated map seemed to be saturated, whereas some of others were not, such as LG1D, LG4B, LG5C, LG5D, LG7B, and LG7C. In addition, the parent-specific LGs were not evenly merged into the LGs in the integrated map. For example, LG6A in the integrated map consists of 16 parent-specific LGs, whereas LG1D was derived from a single parent-specific LG of the '0212921'-specific map. One of the causes for the uneven positions of the mapped loci and the number of merged parent-specific LGs may be the bias of heterozygous (polymorphic) genomic regions within each parental line. Moreover, the source of EST-SSR markers might affect the evenness of the mapping. Of the mapped loci on the integrated map, 82% were derived from *F. vesca* sequences, whereas 14 and 4% markers were generated from *F. × ananassa* and other species, respectively. Previous molecular and cytogenetic studies suggested that *F. vesca* and *F. iinumae* or *F. daltoniana* were candidate ancestral species of *F. × ananassa*.<sup>13,14</sup> This theory suggests that partial genome regions in *F. × ananassa*

were derived from *F. iinumae* or *F. daltoniana* and may show non-homology with the genomes of *F. vesca*. In addition, the ratio of mapped FVES loci ranged from 62.5 to 93.3% (Supplementary Fig. S2). The large contribution of *F. vesca* sequence-derived SSR markers on the integrated map may affect the biased position of the mapped loci.

Although a total of 107 loci derived from previously published markers were mapped onto the integrated linkage map, we did not employ the corresponding names used in previously published maps. This was due to the insufficient number of commonly mapped markers across linkage maps, along with the large number of mapped multi loci. In this study, 'A' to 'D' suffixes were added to the LG names, which corresponded to the lengths of the LGs; A was used for the longest LG, whereas D was added to the shortest. Therefore, there are no biological relationships among the LGs designated with the same capital letter suffix. We request the replacing of the capital letters as to represent the genome structure in *F. × ananassa*, for example AAA'A'BBB'B', after corresponding LGs to each genome, will be identified in future.

Loci generated from SLD markers should play an important role in the identification of specific LGs and chromosomes in *F. × ananassa*. In this study, we mapped both SLD and MLD markers, and 25.9% (481) of the mapped loci were derived from SLD markers. Indeed, the karyotypes of octoploid *Fragaria* species were already reported in *F. chiloensis* and *F. virginiana*,<sup>4,7</sup> as well as in *F. × ananassa*.<sup>48</sup> Using genomic region mapped single loci as probes for cytological analysis such as fluorescence *in situ* hybridization, corresponding karyotypes of *F. × ananassa* could be identified on each LG in the integrated map. In addition, the genome structure of *F. × ananassa* could be resolved by comparative mapping with candidate ancestral *Fragaria* species using markers detecting single loci.

While 25.9% of the mapped loci were generated from SLD markers, the remaining 74.1% (1375) were derived from MLD markers, comprising 569 Multi\_H, 16 Multi\_H&NH, 99 Multi\_H&S, 61 Multi\_NH, 107 Multi\_S, and 523 Multi\_NM. The ratio of all Multi\_H (including Multi\_H&NS and H&S) to the sum of mapped loci (except Multi\_NM) was 0.51 [(569 + 16 + 99)/(1856 - 523)]; that is approximately half of the mapped loci on the integrated map were Multi\_H. The Multi\_H loci were mapped along the entire genomes. However, corresponding homeologous positions differed depending on the paired LGs. The uneven homeologous regions between each pair of LGs might represent a feature of the genome composition of *F. × ananassa*.

Comparative mapping studies between *F. × ananassa* and *F. vesca* have been reported.<sup>8-11</sup> According

to previous results, most of the genomic regions are conserved between the two species, with the rare exception of chromosome rearrangement on HG1, HG3, and HG6. Results from our comparative mapping study generally agreed with those of previous studies, i.e. the identification of clear macrosynteny over both entire genomes. However, our results suggest that partial rearrangement occurred more frequently between homeologous genomes in *F. × ananassa* and *F. vesca* than previously reported. In addition, genomic rearrangement across HGs was first observed in this study, i.e. between Fv-Chr2 and Fa-HG3 (Fig. 1). Of the LGs belonging to HG3, LG3A, 3B and 3C showed the genomic rearrangement, whereas it was not observed on LG3D.

The varieties of 129 *F. × ananassa* lines were distinguished to demonstrate the practicable utility of the markers developed in this study. Ninety-one percent (118 of 129) of the tested lines were developed in Japan. Because most Japanese lines were developed from a limited number of ancestral lines, such as 'Haward17' and 'General Chanzy', genetic diversity is considered to be generally narrow in Japanese germplasm.<sup>49</sup> Despite the expected narrow genetic diversity, the 45 SSR markers employed in this study identified most of the tested lines except for 2 pairs, 'Nyoho' and 'Shin-Nyoho' and 'Toyonoka' and 'Himatsuri'. 'Shin-Nyoho' and 'Himatsuri' were developed from mutation lines of 'Toyonika' and 'Nyoho', respectively. Therefore, it was assumed that the genetic diversities were quite narrow in each pair of undistinguished varieties. Meanwhile, it is notable that the 'Anter', 'Pihyaradondon', and 'Nyoho' lines and the 'Akita Berry' and 'Morioka 16' lines could be distinguished from each other, although the former two and one of the varieties are mutant lines of 'Nyoho' and 'Morioka 16', respectively.

Variety identification is one of the major practical uses of DNA markers. Such identification should protect breeders' rights and inhibit the contamination of clonal seedlings during propagation. Because unknown samples are often used in variety identification, the accurate detection of targeted peaks is essential. In addition, it is well known that PCR results can vary depending on experimental conditions such as the type of thermal cycler and Taq polymerase employed during PCR analysis. In a previous study, Kuniyama<sup>44</sup> investigated the stability of polymorphic analysis of CAPS markers by comparing the results obtained in 14 laboratories to verify the adequacy of the markers used for variety identification tests in *F. × ananassa*. Govan et al.<sup>50</sup> also screened 32 SSRs that produced reliable PCR results for the identification of 60 varieties. In the present study, the 45 SSR markers were carefully screened by confirming the results of 3 independent PCRs. The 36 out of the 45

markers were mapped onto a total of 20 LGs of the integrated map. It suggested that the polymorphisms of selected markers roughly reflect genetic diversity in entire genome. Therefore, the markers were found to be reliable for variety identification.

In this study, we developed large numbers of SSR markers and constructed the densest linkage map for *F. × ananassa*. By performing comparative mapping and variety distinction, the resources that we developed were shown to be useful for genetic and genomic analysis as well as practical applications. Our results should contribute to the acceleration of advances in the study of *F. × ananassa*, along with the genus *Fragaria*.

**Acknowledgements:** We thank Dr Satoru Matsumoto at the National Agriculture and Food Research Organization, Institute of Fruit Tree Science, Dr Toshi Nishimoto at Nara Prefectural Agriculture Research Center, and Toshihiro Tanaka at Wakayama Research Center of Agriculture, Forestry and Fisheries, for providing *F. × ananassa* DNA or plant material.

**Supplementary Data:** Supplementary Data are available at [www.dnaresearch.oxfordjournals.org](http://www.dnaresearch.oxfordjournals.org).

## Funding

This work was supported by the Kazusa DNA Research Institute Foundation, research and development projects for application in promoting new policies of agriculture, forestry and fisheries (21010) funded by the Ministry of Agriculture, Forestry and Fisheries.

## References

1. FAOSTAT. <http://faostat3.fao.org>.
2. UPOV. <http://www.upov.int>.
3. Darrow, G. 1966, *The Strawberry. History, Breeding and Physiology*. Holt, Rinehart and Winston, New York, USA.
4. Federova, N.J. 1946, Cross ability and phylogenetic relationships in the main European species of *Fragaria*, *Natl. Acad. Sci. USSR*, **52**, 545–7.
5. Senanayake, Y.D.A. and Bringham, R.S. 1967, Origin of the *Fragaria* polyploids. I. Cytological evidence, *Am. J. Bot.*, **54**, 221–8.
6. Bringham, R.S. 1990, Cytogenetics and evolution in American *Fragaria*, *Hortic. Sci.*, **25**, 879–81.
7. Kuniyama, M., Fukino, N. and Matsumoto, S. 2005, CAPS markers improved by cluster-specific amplification for identification of octoploid strawberry (*Fragaria × ananassa* Duch.) cultivars, and their disomic inheritance, *Theor. Appl. Genet.*, **110**, 1410–1418.
8. Rousseau-Gueutin, M., Lerceteanu-Kohler, E., Barrot, L., et al. 2008, Comparative genetic mapping between octoploid and diploid *Fragaria* species reveals a high

- level of colinearity between their genomes and the essentially disomic behavior of the cultivated octoploid strawberry, *Genetics*, **179**, 2045–60.
9. Sargent, D.J., Fernández-Fernández, F., Ruiz-Roja, J.J., et al. 2009, A genetic linkage map of the cultivated strawberry (*Fragaria times ananassa*) and its comparison to the diploid *Fragaria* reference map, *Mol. Breed.*, **24**, 293–303.
  10. Sargent, D.J., Passey, T., Šurbanovski, N., et al. 2012, A microsatellite linkage map for the cultivated strawberry (*Fragaria times ananassa*) suggests extensive regions of homozygosity in the genome that may have resulted from breeding and selection, *Theor. Appl. Genet.*, **124**, 1229–40.
  11. Lerceteau-Köhler, E., Moing, A., Guérin, G., et al. 2012, Genetic dissection of fruit quality traits in the octoploid cultivated strawberry highlights the role of homoeo-QTL in their control, *Theor. Appl. Genet.*, **124**, 1059–77.
  12. Staudt, G. 1989, The species of *Fragaria*, their taxonomy and geographic distribution, *Acta Hort.*, **265**, 23–33.
  13. Hancock, J.F. 1999, *Strawberries*. CABI Publishing, New York, USA.
  14. Rousseau-Gueutin, M., Gaston, A., Aïnouche, A., et al. 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**, 515–30.
  15. James, C.M., Wilson, F., Hadonou, A.M. and Tobutt, K.R. 2003, Isolation and characterization of polymorphic microsatellites in diploid strawberry (*Fragaria vesca* L.) for mapping, diversity studies and clone identification, *Mol. Ecol. Notes*, **3**, 171–3.
  16. Albani, M.C., Battey, N.H. and Wilkinson, M.J. 2004, The development of ISSR-derived SCAR markers around the SEASONAL FLOWERING LOCUS (SFL) in *Fragaria vesca*, *Theor. Appl. Genet.*, **109**, 571–9.
  17. Cipriani, G. and Testolin, R. 2004, Isolation and characterization of microsatellite loci in *Fragaria*, *Mol. Ecol. Notes*, **4**, 366–8.
  18. Hadonou, A.M., Sargent, D.J., Wilson, F., James, C.M. and Simpson, D.W. 2004, Development of microsatellite markers in *Fragaria*, their use in genetic diversity analysis, and their potential for genetic linkage mapping, *Genome*, **47**, 429–38.
  19. Monfort, A., Vilanova, S., Davis, T.M. and Arús, P. 2006, A new set of polymorphic simple sequence repeat (SSR) markers from a wild strawberry (*Fragaria vesca*) are transferable to other diploid *Fragaria* species and to *Fragaria times ananassa*, *Mol. Ecol. Notes*, **6**, 197–200.
  20. Sargent, D.J., Davis, T.M., Tobutt, K.R., Wilkinson, M.J., Battey, N.H. and Simpson, D.W. 2004, A genetic linkage map of microsatellite, gene-specific and morphological markers in diploid *Fragaria*, *Theor. Appl. Genet.*, **7**, 1385–91.
  21. Sargent, D.J., Clarke, J., Simpson, D.W., et al. 2006, An enhanced microsatellite map of diploid *Fragaria*, *Theor. Appl. Genet.*, **112**, 1349–59.
  22. Sargent, D.J., Rys, A., Nier, S., Simpson, D.W. and Tobutt, K.R. 2007, The development and mapping of functional markers in *Fragaria* and their transferability and potential for mapping in other genera, *Theor. Appl. Genet.*, **114**, 373–84.
  23. Sargent, D.J., Cipriani, G., Vilanova, S., et al. 2008, The development of a bin mapping population and the selective mapping of 103 markers in the diploid *Fragaria* reference map, *Genome*, **51**, 120–7.
  24. Bonet, J., Girona, E.L., Sargent, D.J., et al. 2009, The development and characterisation of a bacterial artificial chromosome library for *Fragaria vesca*, *BMC Res. Notes*, **2**, 188.
  25. Pontaroli, A.C., Rogers, R.L., Zhang, Q., et al. 2009, Gene content and distribution in the nuclear genome of *Fragaria vesca*, *Plant Genome*, **2**, 93–101.
  26. Shulaev, V., Sargent, D.J., Crowhurst, R.N., et al. 2011, The genome of woodland strawberry (*Fragaria vesca*), *Nat. Genet.*, **43**, 109–16.
  27. Cabrera, A., Kozik, A., Howad, W., Arus, P., Iezzoni, A.F. and Knaap, E. 2009, Development and bin mapping of a Rosaceae conserved ortholog set (COS) of markers, *BMC Genomics*, **10**, 562.
  28. Jung, S., Cestaro, A., Troggio, M., et al. 2012, Whole genome comparisons of *Fragaria*, *Prunus* and *Malus* reveal different modes of evolution between Rosaceae subfamilies, *BMC Genomics*, **13**, 129.
  29. Lerceteau-Köhler, E., Guérin, G., Laigret, F. and Denoyes-Rothan, B. 2003, Characterization of mixed disomic and polysomic inheritance in the octoploid strawberry (*Fragaria times ananassa*) using AFLP mapping, *Theor. Appl. Genet.*, **107**, 619–28.
  30. Zorrilla-Fontanesi, Y., Cabeza, A., Domínguez, P., et al. 2011, Quantitative trait loci and underlying candidate genes controlling agronomical and fruit quality traits in octoploid strawberry (*Fragaria times ananassa*), *Theor. Appl. Genet.*, **123**, 755–78.
  31. Rozen, S. and Skaletsky, H. 2000, Primer3 on the WWW for general users and for biologist programmers, In: Krawetz, S. and Misener, S. (eds.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, USA, pp. 365–86.
  32. Ashley, M.V., Wilk, J.A., Styan, S.M.N., et al. 2003, High variability and disomic segregation of microsatellites in the octoploid *Fragaria virginiana* Mill. (Rosaceae), *Theor. Appl. Genet.*, **107**, 1201–7.
  33. Sargent, D.J., Hadonou, A.M. and Simpson, D.W. 2003, Development and characterization of polymorphic microsatellite markers from *Fragaria viridis*, a wild diploid strawberry, *Mol. Ecol. Notes*, **3**, 550–2.
  34. Lewers, K.S., Styan, S.M.N. and Hokanson, S.C. 2005, Strawberry GenBank derived and genomic simple sequence repeat (SSR) markers and their utility with strawberry, blackberry, and red and black raspberry, *J. Am. Soc. Hortic. Sci.*, **130**, 102–15.
  35. Bassil, N.V., Gunn, M., Folta, K. and Lewers, K. 2006, Microsatellite markers for *Fragaria* from 'Strawberry Festival' expressed sequence tags, *Mol. Ecol. Notes*, **6**, 473–76.
  36. Gil-Ariza, D.J., Amaya, I., Botella, M.A., et al. 2006, EST-derived polymorphic microsatellites from cultivated strawberry (*Fragaria times ananassa*) are useful for diversity studies and varietal identification among *Fragaria* species, *Mol. Ecol. Notes*, **6**, 1195–7.

37. Keniry, A., Hopkins, C.J., Jewell, E., et al. 2006, Identification and characterization of simple sequence repeat (SSR) markers from *Fragaria times ananassa* expressed sequences, *Mol. Ecol. Notes*, **6**, 319–22.
38. Spigler, R.B., Lewers, K.S., Main, D.S. and Ashman, T.L. 2008, Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome, *Heredity*, **101**, 507–17.
39. Sargent, D.J., Kuchta, P., Lopez Girona, E., et al. 2011, Simple sequence repeat marker development and mapping targeted to previously unmapped regions of the strawberry genome sequence, *Plant Genome*, **4**, 1–13.
40. Chevreux, B., Pfisterer, T., Drescher, B., et al. 2004, Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs, *Genome Res.*, **14**, 1147–59.
41. Sato, S., Isobe, S., Asamizu, E., et al. 2005, Comprehensive structural analysis of the genome of red clover (*Trifolium pratense* L.), *DNA Res.*, **12**, 301–64.
42. Van Ooijen, J.W. 2006, *JoinMAP® 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*, Kyazma B.V., Wageningen, Netherlands.
43. Kiss, G.B., Kereszt, A., Kiss, P. and Endre, G. 1998, Colormapping; a non-mathematical procedure for genetic mapping, *Acta Biol. Hung.*, **19**, 125–42.
44. Kuniyama, M. 2011, Studies using DNA markers in *F. times ananassa*: genetic analysis, genome structure, and cultivar identification, *J. Jpn. Soc. Hortic. Sci.*, **80**, 231–43.
45. van Berloo, R. 2008, GGT 2.0: versatile Software for visualization and analysis of genetic data, *J. Hered.*, **99**, 232–236.
46. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. 2011, MEGA5: Molecular Evolutionary Genetics Analysis using likelihood, distance, and parsimony methods, *Mol. Biol. Evol.*, **28**, 2731–9.
47. Nathewet, P., Yanagi, T., Iwatsubo, Y., Sone, K., Tamura, T. and Okuda, N. 2010, Karyotype analysis in octoploid and decaploid wild strawberries in *Fragaria* (Rosaceae), *Cytologia*, **75**, 277–88.
48. Nathewet, P., Iwatsubo, Y., Sone, K., Hummer, K. and Yanagi, T. 2008, Chromosome numbers and morphology in Japanese strawberry cultivars, *Hortic. Res.*, **7**, 195.
49. Kimura, M. 2004, *History of Japanese Strawberry Varieties and their Adaptability, Yasai-Engei Dai-Hyakka* (in Japanese), ISBN 4-540-03284-4, Nou-Bunkyou, Japan, pp. 116–7.
50. Govan, C.L., Simpson, D.W., Johnson, A.W., Tobutt, K.R. and Sargent, D.J. 2008, A reliable multiplexed microsatellite set for genotyping *Fragaria* and its use in a survey of 60 *F. × ananassa* cultivars, *Mol. Breeding*, **22**, 649–61.