Theor Appl Genet (2015) 128:1261–1275
DOI 10.1007/s00122-015-2503-3

ORIGINAL PAPER

A near-isogenic line (NIL) collection in diploid strawberry and its use in the genetic analysis of morphologic, phenotypic and nutritional characters

María Urrutia[1] · Julio Bonet[1,2] · Pere Arús[1] · Amparo Monfort[1]

Received: 15 October 2014 / Accepted: 20 March 2015 / Published online: 5 April 2015

Abstract
Key message **First near-isogenic line collection in diploid strawberry, a tool for morphologic, phenotypic and nutritional QTL analysis.**
Abstract Diploid strawberry (*Fragaria vesca*), with a small genome, has a high degree of synteny with the octoploid cultivated strawberry (*F.* × *ananassa*), so can be used as a simplified model for genetic analysis of the octoploid species. Agronomically interesting traits are usually inherited quantitatively and they need to be studied in large segregating progenies well characterized with molecular markers. Near-isogenic lines (NILs) are tools to dissect quantitative characters and identify some of their components as Mendelian traits. NILs are fixed homozygous lines that share the same genetic background from a recurrent parent with a single introgression region from a donor parent. Here, we developed the first NIL collection in *Fragaria*, with *F. vesca* cv. Reine des Vallées as the recurrent parent and *F. bucharica* as the donor parent. A collection of 39 NILs was identified using a set of single sequence repeat markers. The NILs had an average introgression of 32 cM (6 % of genome) and were phenotyped over several years in two locations. This collection segregates for agronomic characters, such as flowering, germination, fruit size and shape, and nutritional content. At least 16 QTLs for morphological and reproductive traits, such as round fruits and vegetative propagation, and seven for nutritional traits such as sugar composition and total polyphenol content, were identified. The NIL collection of *F. vesca* can significantly facilitate understanding of the genetics of many traits and provide insight into the more complex *F.* × *ananassa* genome.

Communicated by A. H. Schulman.

M. Urrutia and J. Bonet contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2503-3) contains supplementary material, which is available to authorized users.

✉ Amparo Monfort
amparo.monfort@irta.es

[1] IRTA, Center for Research in Agricultural Genomics (CSIC-IRTA-UAB-UB), Campus UAB, 08193 Bellaterra, Barcelona, Spain

[2] Present Address: Plant Response Biotech S.L. Centro de Empresas, Parque Científico y Tecnológico Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain

Introduction

Strawberry, one of the economically most important soft fruits, belongs to the *Fragaria* genus. This includes 23 species with different ploidy levels, from diploid wild strawberries such as *F. vesca* to octoploid cultivated (*F.* × *ananassa*) and decaploid (*F. iturupensis*). Diploid species ($2n = 2x = 14$) can be used as models for genetic analysis of cultivated strawberry, as synteny studies have demonstrated that the octoploid genome is composed of four genomes that are essentially collinear and syntenic with the genomes of the diploid species (Vilanova et al. 2008; Rousseau-Gueutin et al. 2008; Sargent et al. 2009). The reference linkage map of diploid strawberry was constructed from a cross between *F. vesca* ssp. *vesca* (FDP815) and *F. bucharica* with 68 single sequence repeat (SSR), one SCAR and six STS markers (Sargent et al. 2004) and three morphological traits: seasonal perpetual flowering, runner development and pale-dark green leaf color. Several versions increasing the number and source of markers have

been done (Sargent et al. 2006, 2008; Vilanova et al. 2008; Ruiz-Rojas et al. 2010) until reaching its genome anchoring (Shulaev et al. 2011).

In this paper, we describe the development of a collection of near-isogenic lines (NILs) using two accessions of two diploid species: *F. vesca*, used as recurrent parent, and *F. bucharica*, as donor parent. *F. vesca,* the most widely distributed diploid, has been reported as a principal contributor of the octoploid genomes (Rousseau-Gueutin et al. 2009). It includes the 'woodland' strawberry *F. vesca* ssp. *vesca*, a fruit of minor economic importance, found in the wild throughout the northern hemisphere and which has been described as a predominantly selfing species (Arulsekar and Bringhurst 1981). The other diploid species, *F. bucharica,* is a self-incompatible (Bošković et al. 2010) stolon-propagated species, belonging to the Asiatic group, probably of hybrid origin (Staudt 2006, 2008). These two species are phylogenetically close (Rousseau-Gueutin et al. 2008) but morphologically divergent, allowing fertile crosses between them, with a wide range of segregating phenotypes from the offspring. The selected *F. bucharica* accession was that used by Sargent et al. (2006) as one of the parents of the reference map for diploid strawberry.

NIL collections are important resources for analyzing complex characters. An NIL is identical to an original genotype, except for a single DNA introgressed fragment from a donor line. Comparison of the phenotypes of an NIL and its recurrent parent means any significant difference can be attributed to genetic factors in the introgressed fragment. In addition, phenotyping can be done with high precision, as each NIL can be duplicated as many times as necessary and grown under different environments, circumventing some of the major limitations of conventional mapping populations (Paterson et al. 1988; Eshed and Zamir 1995). A collection of NILs covering the whole genome of a species can be considered a genomic library containing overlapping introgressions from a donor genome on a recipient or recurrent genome. NIL collections can also be used to introduce new genetic variability from wild species into elite cultivars (Tanksley and McCouch 1997; Zamir 2001), to study the genetic bases of heterosis (Melchinger et al. 2007; Fernández-Silva et al. 2009), to study interactions between different QTLs and QTLs with the environment (Monforte et al. 2001), or as initial materials for map-based cloning (Frary et al. 2000; Fridman et al. 2000; Fujita et al. 2013). Several NIL collections have been developed as genomic resources in *Arabidopsis thaliana* (Koumproglou et al. 2002; Keurentjes et al. 2007; Fletcher et al. 2013) and in crops such as wheat (Pestsova et al. 2001), lettuce (Jeuken and Lindhout 2004) and melon (Eduardo et al. 2005). There are also partial collections covering specific genomic regions of *Brassica oleracea* (Ramsay et al. 2000), rice (Wan et al. 2004; Zhang et al. 2009), maize (Szalma et al. 2007), sorghum (Harris et al. 2007), tobacco (Lewis et al. 2007), barley (Marcel et al. 2007) and soybean (Kopisch-Obuch and Diers 2006; Bolon et al. 2010).

The diploid strawberry NIL collection was phenotyped for a set of characters responsible for morphological and phenotypic characteristics of different parts of the plant and for characters related with the chemical composition of the fruit, which have allowed us to find the positions of a set of Mendelian genes and QTLs responsible for their variation. We propose this new resource as a tool for the broad Rosaceae community. It will be of use for the in-depth study of the genetics of important characters in this family and to complement other resources which have been developed in *F. vesca*, such as the reference linkage map (Sargent et al. 2006), the collection of mutants (Oosumi et al. 2010) and the whole genome sequence (Shulaev et al. 2011).

Materials and methods

Plant materials and NIL development

One of the selected parental lines was *F. vesca* cv. Reine des Vallées (RV), a French, non-runnering, day-neutral, homozygous cultivated variety with long petioles, large leaflets, small flowers, and bright-red, long conic berries with a powerful aroma and low susceptibility to the common pests in strawberry crops. The other was *F. bucharica* (PI657844, also known as FDP601) (FB), one of the parents of diploid strawberry reference map. It is a short-day variety with short petioles, small leaflets and large flowers. The berries, with no commercial value, are necked, round flattened on top, and dark wine-red. The *F. vesca* parent was chosen as the recurrent genome and *F. bucharica* parent as the donor.

Pollen from an F_1 individual of a cross between RV × FDP601 was used as the donor to pollinate an *F. vesca* individual. The resulting backcross 1 (BC_1) seeds were germinated and DNA from the seedlings were extracted and genotyped using seven SSR markers (EMFvi072, EMFvi099, EMFv029, CEL-1, EMFvi018, EMFn228 and EMFv021) (Supplementary table 1). The seven markers are located in one of the extremes of the seven linkage groups of the diploid *Fragaria* reference map (Supplementary figure 1). Individuals with less than four introgressions were selected and then genotyped for seven additional SSR markers (EMFv025, EMFv003, EMFn207, EMFv007, EMFvi108, EMFv160BC and EMFv023) mapping to the other linkage group ends (see reference in Supplementary table 1). Individuals harboring <5 of these markers in heterozygosity, therefore containing at least the same number of introgressions, were selected and genotyped for a

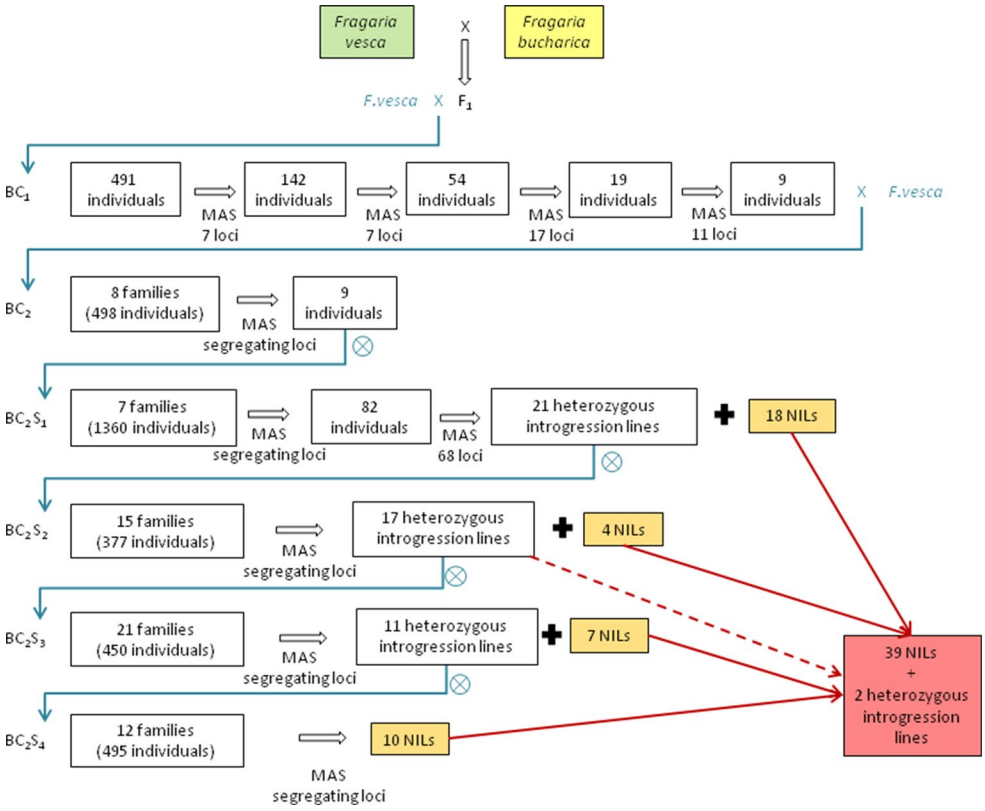

Fig. 1 Crossing scheme for the development of a collection of NILs in *F. vesca*. Each *box* contains the number of individuals chosen after a process of marker-assisted selection (MAS). *Orange boxes* include the *lines* selected as final homozygous NILs in each generation. The *red box* includes the complete set of NILs. (color figure online)

minimum of 17 and up to 28 additional markers located in intermediate regions of the seven linkage groups (Supplementary table 1), depending on the genetic composition of each line. This gave a set of plants with a low number of introgressions covering the whole genome of the donor parental line.

Selected BC_1 plants were used as female parents for a second backcross with the *F. vesca* parental line. The resulting seeds were germinated and each family evaluated for molecular markers located at the extremes of each introgression. A reduced group of plants per family was selected for further self-pollination until lines with single introgressions were obtained. Individuals resulting from a further round of self-pollination were screened with a set of 77 segregating loci (Supplementary table 1), well distributed across the genome, allowing characterization of the introgressed fragments and eliminating any plants with more than one introgression from the donor genome. The size of the introgressions and genome percentages were calculated by adding together the length between the two markers at the extreme of the introgression fragment, and half the distance of the two intervals between these markers and those flanking the introgression that were homozygous for the RV allele.

The crossing scheme for the development of the *F. vesca–F. bucharica* collection of NILs is presented in Fig. 1.

Molecular marker analysis and map construction

Tissue from young leaves was collected and disrupted by mechanical shaking using a tissue homogenizer. DNA was extracted using the method of Doyle and Doyle (1990) modified by the addition of 2 % PVP-40. For PCR, the DNA was quantified and 20 ng per reaction SSR and cleaved amplified polymorphic sequences (CAPS) markers were used to locate the introgressed regions. A set of 16 new SSR markers in the region of previously mapped RFLP markers (Vilanova et al. 2008) was developed (Supplementary table 2) and mapped to obtain additional markers in map regions with insufficient coverage of SSR markers. The region was identified by comparing the Strawberry Genome Assembly version 8 [Genome Database for Rosaceae http://www.rosaceae.org (Jung et al. 2014) with RFLP original sequences (http://www.rosaceae.org) (Vilanova et al. 2008)]. Alignments were accepted when the estimated similarity between aligned sequences (RFLP and genome) was statistically significant (*E* value $\leq$0.05). New SSR motifs were located near to the aligned sequences, with the primer pairs designed using Websat (Martins et al. 2009) and Primer3 (Rozen and Skaletsky 2000). To reduce the cost of marker detection, forward primers were modified by adding a 17 nucleotide M13-tail (5′ GTAAAACGACGGCCAGT 3′) to the 5′ extreme (Boutin-Ganache et al. 2001), allowing hybridization of any forward primer with a unique fluorescent-labeled primer in PCR.

A new marker linked to the perpetual flowering locus in linkage group 6 (LG6) was designed to amplify a specific small indel. Using the method of Iwata et al. (2011), we sequenced a 1000 bp region on LG6, scaffold 0513102 (Shulaev et al. 2011), that contained a 2 bp deletion in the *F. vesca* TFL homolog which causes a perpetual flowering phenotype (Koskela et al. 2012). Two FvTFL primers were designed to specifically amplify this deletion using Primer3 (Rozen and Skaletsky 2000). The forward primer was TCGGAACCTCTAGCTGTTGG and the FvTFL reverse primer was GAGCTCATGTCCATTGCAGA.

PCRs were in a final volume of 10 µl containing: 10 ng of DNA, 1× PCR buffer (50 mM KCl, 10 mM Tris–HCl pH 8.3, 0.001 % gelatin), 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1 U of DNA polymerase Ampli*Taq* (Perkin-Elmer, IL, USA), 0.15 µM forward primer, 0.2 µM reverse primer and 0.2 µM of M13 labeled primer, with an identical sequence to the added tail on the forward primers. The choice of dye label, 6-FAM, VIC, PET or NED (Applied Biosystems, CA, USA), depended on further multiloading capillary electrophoresis. PCR amplifications were run on a PE9600 thermal-cycler (Applied Biosystems, CA, USA) as follows: 2 min of initial denaturing at 94 °C; 10 cycles of 15 s at 94 °C, 15 s at annealing temperature and 30 s at 72 °C; 25 cycles of 15 s at 94 °C, 15 s at 50 °C and 30 s at 72 °C, followed by a final extension of 5 min at 72 °C.

Amplicons were visualized by capillary electrophoresis in an ABI3130xl Genetic Analyzer (Applied Biosystems, CA, USA), run using 2 µl of a mix containing three differently labeled PCR products, 0.3 µl of LIZ-500 ladder and 12 µl of deionized formamide. Data generated by capillary electrophoresis were analyzed using the GENEMAPPER software application (Applied Biosystems, CA, USA). For the four CAPS markers, PCR products were digested as described by Ruiz-Rojas et al. 2010, and electrophoresed in agarose gel. Segregation data of new markers were added to the segregation matrix of the FV×FB diploid *Fragaria* reference map (Sargent et al. 2006; Vilanova et al. 2008; Ruiz-Rojas et al. 2010; Zorrilla-Fontanesi et al. 2010). RFLP markers on the previously developed FV×FB map were excluded from the analysis.

Chi squared tests of goodness-of-fit, to an expected segregation ratio of 1:2:1 for co-dominant markers and 3:1 for dominant markers, were carried out for new markers using the JoinMap® software (Van Ooijen 2006). Linkage analysis was performed and the map constructed using MAPMAKER 3.0 (Lander et al. 1987), applying the Kosambi mapping function. Linkage groups were constructed and marker order determined using a minimum LOD score threshold of 3.0. Markers selected for marker-assisted selection and new markers obtained are shown in Supplementary figure 1. Graphical genotyping software GGT2.0 (van Berloo 2008) was used to present the NILs and evaluate the genome composition of the NIL collection.

Phenotyping and genetic analysis

NIL collection seeds were germinated in in vitro chambers; conditions' set was: photoperiod 12 h/12 h light/dark and temperature 24 °C. Seeds were kept in petri-plates over humid filter paper until germination occurred (appearance of the root and both cotyledons open). Four to eight plants of each genotype were transplanted to soil pots and grown under glasshouse conditions (photoperiod 12 h/12 h light/dark, day temperature between 22–24 and 17 °C during night, relative humidity 40–50 %.) for 8 weeks at the Center for Research in Agricultural Genomics (CRAG) in Bellaterra (latitude 41°29′N, longitude 2°06′E). Two-month old plants were moved to shaded greenhouses in Cabrils (latitude 41°31′N, longitude 2°22′E, altitude 146 m above sea level, coastal Mediterranean climate) and at the Center of Torre Marimon (TM) in Caldes de Montbui (latitude 41°36′N, longitude 2°10′E, altitude 203 m above sea level, pre-coastal Mediterranean climate). The shaded greenhouses were not provided with supplementary artificial light or heating, so plants were exposed to natural climatic conditions and photoperiod in north Spain (latitude 41°) from March to September (when the harvest was over). The agronomical practices were the usual for strawberry fruit production. The population was established over 3 years in Cabrils (2010–2012) and 2 years in TM (2011 and 2013). Fruits were collected in two different periods or harvests in May and July.

Qualitative morphological and phenotypic traits were evaluated visually in all seasons. All phenotypic data were provided as supplementary material. For floral traits, we evaluated the presence of pink spot on the petal base (*PSP*), extra petal number (*EPN*, above 5 petals), short and long floral stem (*SFS* and *LFS*, respectively), seasonal flowering or short-day flowering (*SDF*) as plants blooming only once a year in short day as *F. bucharica* parent, and a trait in plants flowering later than 1 year of germination (*LLF* for late–late flowering). For other plant traits, we evaluated the presence of runners (*R*), dwarfism (*DWARF*) and delayed germination (*DOG*, as more than 15 days later than in RV).

The basic requirement for accepting these traits was that they were stable throughout the studied seasons, and that all the lines containing the introgressed region had the same phenotype. Flowering time was measured as the number of days to produce the first flower after germination, with early flowering being shorter and late flowering longer than the flowering time for RV, in all lines with the same introgression.

For berry measurements, a number of 20 fruits per line were collected during the spring (May) and summer (July) of 2011 and 2012 and measured using ImageJ (Schneider et al. 2012; http://imagej.nih.gov/ij/). Fruits of *F. vesca* cv. Reine des Vallées were collected as the reference parent,

but not enough fruits of *F. bucharica* could be collected. Photographs were taken for all the collected fruits under standardized conditions and the longest straight line of the fruits measured lengthwise (length) and crosswise (width). We then divided the length by the width to obtain a parameter (called fruit index) that measured fruit shape (round fruits being close to 1 and long fruits having higher values). All the fruits collected from an NIL plant were weighed together and the average used as the value for fruit weight. Fruit traits were analyzed for five independent harvests, in 2 years and at two locations (2011 and 2012 in May and July in Cabrils, and 2011 in May in TM).

Three biological replicates of 5 g of fruit from each genotype, in 2011–2013, were collected for nutritional analyses. The fruits were immediately frozen in liquid nitrogen, ground, divided into subsamples and stored at −80 °C until extraction. The sugars in the samples were ethanol extracted following the protocol described by Pérez et al. (1997), and then quantified by HPLC (Agilent 1200 HPLC, Agilent Technologies, CA, USA) using a Zorbax Carbohydrate Analysis column (Agilent Technologies, CA, USA). Total polyphenolic content was determined by the Folin–Ciocalteu method following the extraction and detection protocol described by Singleton et al. (1999). Within the population, the distribution parameters for the fruit sugar content (fructose, glucose and sucrose) were quantified as mg/g of fresh weight, and the total polyphenol content as mg of gallic acid equivalents per gram fresh weight.

Statistical analyses were performed using the JMP®8.0.1 statistical package (SAS Institute, NC, USA) and R 2.15.1 (Development Core Team 2008) with the R-studio interface (Rstudio 0.92.501, Rstudio, MA, USA). Distribution of all traits among the population was evaluated by a Shapiro–Wilk normality test, and the range of variation was also calculated by skewness and kurtosis of the distribution for every trait and harvest. The mean value and the standard deviation for RV and for every line of the population were calculated for each trait and harvest. Correlation of the quantitative traits in different years was calculated separately for nutritional characters. Means were compared with the recurrent parent by a Dunnett test with $\alpha \leq 0.05$ (Dunnett 1955). The QTL effects are presented in relation to the recurrent parental control. Qualitative traits and QTLs were mapped to a specific bin only when all the NILs carrying this bin had a significant effect on the phenotypic trait of study.

Results

Linkage map of FV×FB

The diploid strawberry linkage map of *F. vesca* × *F. bucharica* F_2 population (Sargent et al. 2006; Ruiz-Rojas et al. 2010) was used to monitor the introgressions for the selection of the NIL collection. This map, including new markers developed here (14 SSRs and one CAPS) and excluding RFLPs, covered 541.3, 13.2 cM more than that of Ruiz-Rojas et al. (2010), and had 219 microsatellites, one SCAR, 22 STS, 53 CAPS/dCAPS, one SNP, one indel and one morphological marker. This map is presented in Supplementary figure 1. Segregation distortions were observed on 48.8 % of the mapped markers ($P \leq 0.05$), especially on LGs 2, 4, 5 and 7.

Some of the new loci on the map reduced the extension of the original FV×FB linkage groups because segregation changes from dominant in RFLP to codominant in SSR. The RFLP markers AC-32, AC-24 and AG-53 defined the top region of LG4 in the map of Vilanova et al. (2008). The PCR markers developed here to replace these RFLPs (CFV-3135, CFV-3138 and CFV-3819, respectively) were located on the central region of LG4, in a cluster previously described by Sargent et al. (2004). In addition, the RFLPs MC-45 and Co-MET, replaced by CFV-3117 and CFV-3217, were moved to the central positions of LG7, allowing a 19.1 cM reduction of this group. There were minor rearrangements in the placement of some of the dominant marker loci presented by Sargent et al. (2006), due to a better estimation of mapping distances with the addition of the novel codominant SSR loci (e.g., EMFv164 vs. CFV-164).

Development of the NIL collection

The BC_1 seedlings (491 individuals) were analyzed for allele composition using seven SSRs, one for each linkage group (see Supplementary table 1). On average, the observed heterozygosity at these loci was $H_o = 0.57$. A set of 142 individuals (27.6 %) was selected for a second round of PCR for allele characterization of seven additional markers at the opposite end of LG. The average observed heterozygosity for these loci was $H_o = 0.52$. The marker showing the highest segregation distortion was EMFv160BC ($\chi^2 = 31.11$), located at the end of LG6 with maximum of heterozygosity. A group of 54 plants (38 %) was selected for further genome characterization. Twenty-eight additional loci were analyzed and nine BC_1 individuals, with introgressions covering the whole *F. bucharica* genome, were selected for a second backcross and segregating families generated. From the nine BC_1 individuals selected for their low number of introgressions and maximum genome coverage, only eight BC_2 families were generated. The donor genome percentage of BC_1 lines was 35.2 %, on average, ranging from 25 to 49 % in 28 introgressions, distributed as follows: eight introgressions on LG2, five on LG6, four on LG3, three on LG1, LG4 and LG7, and two on LG5. Whole introgressed linkage groups were selected for the seven linkage groups of the map

Fig. 2 Graphical genotypes of the *F. vesca* collection of 39 NILs and two heterozygous NILs. *F. bucharica* homozygous introgressions are shown in *black* and heterozygous introgressions in *pink*. The *F. vesca* genetic background is shown in *green*. The NIL names are indicated on the *right*. The *first number* indicates the LG carrying the introgression. The following two numbers, separated with a *hyphen*, indicate the marker position at the start and end point of the introgression in centiMorgans, respectively. *Dotted* NILs indicate the minimal set of lines covering the entire *F. bucharica* genome. (color figure online)

(35 % of the selected introgressions). These selected BC_1 lines had three or four introgressions per line. BC_2 families (498 individuals) were analyzed with markers that were in the extremes of, and inside for long introgressions, the introgressed fragments detected in their BC_1 progenitors. A total of 50 markers were chosen to genotype all introgressed regions (Supplementary table 1). Nine BC_2 individuals were selected as introgression donors for further generations. Five individuals had only one introgression, one had two and three had three introgressions. The average percentage of *F. bucharica* fragments (in cM) on the selected BC_2 individuals was 14.6 % (range 6.2–24.6 %) in 15 introgressions distributed as follows: three introgressions on LG3 and LG4, two on LG1, LG2, LG5 and LG6, and a single introgression on LG7. Entire introgressed LGs were obtained for LG2, LG3, LG4, LG5, LG6 and LG7 (37.5 % of the introgressions).

Selected individuals were self-pollinated and the resulting seeds germinated to generate segregating families. Between one and four selfing generations were necessary to obtain homozygous lines with only one introgression and covering together almost all the donor parent genome. Depending on the genotype of its parental line, selfed offspring (2682 in total) were analyzed with different subsets of 77 SSRs (see Supplementary table 1) to select only the homozygous target regions. This fine characterization also allowed more exact location of the introgressed regions. Lines in the BC_2 offspring with LG6 introgressed fragments blossomed more than 1 year after germination, and NIL extraction was delayed relative to the rest of the population. In regards to the CFvCT017 marker, it has not yet been possible to obtain homozygosity from the progeny in line Fb**6**:71–101 h. The introgressed region of line Fb**6:**30–39 h is only partially homozygous (see Fig. 2).

Segregation at six of the seven loci analyzed during the first stage of screening of the BC_1 population was distorted ($P < 0.05$; $n = 491$). The top LG2 region (EMFvi099) and the distal part of LG5 (EMFvi018) had the most skewed segregations ($\chi^2 = 170.96$ and $\chi^2 = 65.92$ respectively), with a higher frequency of *F. bucharica* alleles in both cases. The linkage group with the lowest *F. bucharica* allele frequency was LG1 (only 27.4 % of the plants had *F. bucharica* genotypes at LG1 loci). Biased segregation ratios were detected also in BC_2 families, especially for the markers located at the end positions of LG5 (91.8 % of *F. bucharica* genotypes). The lowest frequency of *F. bucharica* alleles was observed for markers located on LG1 (39.2 %). Significantly distorted ratios were also detected in the segregation of markers on LG2, LG3 and LG6.

The strawberry NIL collection consisted of a final set of 39 homozygous lines and two heterozygous lines (Fig. 2).

Table 1 NIL collection characteristics

	n	cM	%
Nb. NILs	39		
Nb. heterozygous NILs	2		
Nb. NILs covering whole LG	5		
Total genome covered in intogressions		522.0	96.4
Genome covered in homozygosity		479.3	88.5
Genome covered in heterozygosity		42.7	7.9
Average NILs per LG	6		
Average introgression size[a]		32.9	6.1
Smallest introgression[a]		3.2	0.6
Largest introgression[a]		89.3	16.5
Nb. of BINs	37		
Average BIN size		14.2	2.6
Smallest BIN		3.2	0.6
Largest BIN		65.7	12.1

[a] Not considering whole LG introgressions

Among these, 18 correspond to BC_2S_1 lines, four to BC_2S_2 lines, seven to BC_2S_3 and ten correspond to BC_2S_4 ("S_n" is the *n*th selfing generation) lines. Assuming no double crossovers in the introgressed fragments, each NIL contained a single homozygous introgression, with an average introgression size of 32.9 cM (6 % of the donor genome) (Table 1). Each linkage group was represented by an average of six NILs with overlapping introgressions. Only a 19.3 cM region on LG1, defined by markers UDF002 and CFV-4021, was not covered by *F. bucharica* introgressions. Furthermore, lines covering a region of LG6 (Fb**6**:30–39 h and Fb**6**:71–101 h), which showed large period to flowering (more than 1 year), retained heterozygous introgressions.

Taking the collection of NILs of each linkage group into consideration, the ensemble of recombination breakpoints of the *F. bucharica* genome introgressed fragments defines a set of identifiable regions, or bins, defined by two consecutive breakpoints with no recombination between them. This population had 37 bins (4–7 per linkage group), with an average size of 14.2 cM per bin and ranging between 3.2 and 65.7 cM. Each bin represents on average a small part, 2.6 %, of the genome (Table 1). A minimal subset of nine of these lines covering the whole genome (except for the missing fragment of LG1 and for the heterozygous LG6 fragment) may be used to locate any segregating locus to its chromosome when resources are limited.

Analysis of morphological and phenotypic traits with the NIL collection

In Table 2, we present the results for all the harvests analyzed. The initial and final QTL positions indicate always the smallest region that could be assigned to a QTL, taking into account all the harvests where it was found to be significant. All QTLs found were significant in overlapping NILs; the NIL carrying the shorter introgression (expressed in cM) is indicated. The effects of the QTLs are expressed as the percent mean variation between the lines carrying the QTL and the recurrent parent, RV. A single bin was determined for all qualitative morphological and phenotypic characters (Table 2A).

The presence of runners has been extensively described (Sargent et al. 2004), and in our population could be clearly assigned to the region between cM 39 and 47 in LG2. *F. bucharica* donor parent is causal for runner trait presence. In this region, we also detected one or more genes involved with long floral stems, so that plants carrying this introgression from *F. bucharica* produced flowers that emerged from the vegetative mass of the plant. This effect was recessive, not visible in parental and F_1 lines. The opposite was seen in lines carrying *F. bucharica* alleles between cM 29 and 39 in LG5: these lines produced short floral stems so that the flowers and fruits remained within the leaf canopy of the plant, as with the *F. bucharica* donor parent.

F. bucharica has a very characteristic flower that usually has more than five overlapping petals, typically 8–10 (Fig. 3). This particular phenotype can be seen in plants with introgressions in LG3 between cM 54 and 94. We also observed flowers with a pink blush in the inner part of their petals, a trait not shown by any of the parents but a little visible in hybrid F_1 plants. This dominant trait was located in LG1 between cM 50 and 61.

We also identified one or more genes in the region between cM 15 and 29 of LG3 causing extreme dwarfism in plants (Fig. 4) and leading to death before flowering. Plants carrying homozygous introgressions in this region did not grow well under our greenhouse conditions, so we have never been able to reproduce them. This is a recessive trait, not present in parents.

Flowering time, seed germination, fruit weight and fruit shape were measured as quantitative traits. Their distribution among the NIL collection is summarized in Table 3A. The flowering time was studied in two consecutive years. We were able to identify two map regions that determined flowering time (Table 2B), one in LG6, between cM 0–11 (86 days after germination on average) as early flowering, and the other in LG5, cM 11–20 (144 days after germination on average) as late flowering. The seasonal flowering behavior typical of *F. bucharica* that blossom one time a year, opposite to the continuous flowering characteristic of RV, also segregated in the NIL population. We mapped this trait in LG6 at cM 30–38, in agreement with that published by Koskela et al. (2012) in plants with few flowering success. Another specific region for everbearing was mapped in LG5, cM 35–39; these plants blossom from March to June, as donor parent, and not everbearing in autumn as *F. vesca* RV. In addition, we considered

Table 2 Position of Mendelian traits (A) and QTLs for quantitative agronomic traits (B)

Trait	Name	Effect	% Effect	LG	Initial position (cM)	Final position (cM)	Stability	Shortest NIL
(A)								
Runners	*R*[a]	Presence of runners	–	2	39 (Fvi11)	47 (EMFn134)	All	Fb**2**:39–47
Floral stem length	*LFS*	Longer floral stems	–	2	39 (Fvi11)	47 (EMFn134)	All	Fb**2**:39–47
	SFS	Shorter floral stems	–	5	29 (FvH4093)	39 (CFVCT024)	All	Fb**5**:20–76
Pink petals	*PSP*	Pink spot on petal base	–	1	50 (UFF02F02)	61 (CFV164)	All	Fb**1**:50–61
Petals number	*EPN*	>5 petals/flower	–	3	54 (CFVCT022)	94 (CFVCT012)	All	Fb**3**:54–94
Seasonal flowering	*SDF*	Short-day flowering	–	5	35 (UDF006)	39 (CFVCT024)	All	Fb**5**:20–76
Flowering	*LLF*	Flowering >1 year later	–	6	30 (EMFn117)	38 (EMFn017)	All	Fb**6**:30–39
Plant size	*DWARF*	Dwarf plants	–	3	15 (VT398)	29 (UDF017)	All	Fb**3**:8–94
Germination	*DOG*	Delay of germination	–	5	39 (CFVCT024)	41 (UDF009)	3	Fb**5**:39–76
Trait	QTL name	Effect	% Effect	LG	Initial position (cM)	Final position (cM)	Stability	Shortest NIL
(B)								
Fruit shape	FRS_2	Rounded fruit	−24	2	0 (EMFvi099)	30 (BFACT002)	5/5	Fb**2**:0–30
	FRS_4	Rounded fruit	−35	4	58 (CEL1)	78 (ChFaM23)	5/5	Fb**4**:58–78
	FES_5	Elongated fruit	25	5	0 (CFV3072)	11 (CFV3132)	4/5	Fb**5**:0–11
Weight	FLW_1	Smaller fruit	−70	1	50 (UFF02F02)	61 (CFV164)	3/3	Fb**1**:50–61
	FLW_4	Smaller fruit	−79	4	58 (CEL1)	78 (ChFaM23)	2/3	Fb**4**:58–78
Flowering	EF[b]	Early flowering	−30	6	0 (ARSFL007)	11 (EMFn228)	2/2	Fb**6**:0–5
	LF	Late flowering	36	5	11 (CFV3132)	20 (CEL2)	2/2	Fb**5**:11–76

Trait, name of QTL, effect percentage means average of trait value on NIL compared to RV trait value, bin position and markers flanking the bin, stability as number of harvests observed over total measured and the shortest NIL showing trait effect

[a] Locus R (mapped at LG2: 45.2 cM)

[b] Compared to population mean

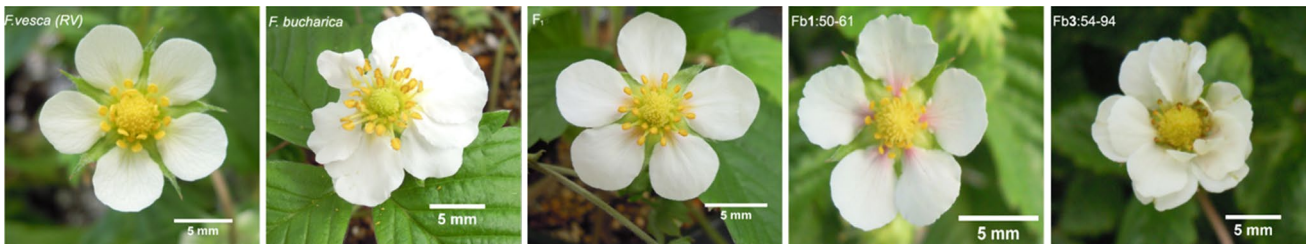

Fig. 3 Flower phenotypes. From *left* to *right*, flower morphology of, *F. bucharica*, the hybrid and NILs Fb**1**:50–61 and Fb**3**:54–94

the germination time for every line in 2013 (Table 3A). In a first germination trial, germination was considerably delayed (around 40 days, compared with 14 days for RV) in a particular line carrying an introgression in LG5 (line Fb**5:**39–76). We repeated the experiment twice, using only a few lines with introgressions in LG5 (Fb**5:**20–76, Fb**5:**39–76, Fb**5:**41–76, Fb**5:**50–76 and Fb**5:**0–76) and RV as the control. The same results were obtained in both additional trials (data not shown), and the delayed germination time was mapped to a short bin at LG5 cM 39–41.

Fruit size was phenotyped as a quantitative trait by measuring the length and width of fruits, and their ratio (the fruit index) that corresponds to shape of fruit. We observed that the population means were lower than the mean values for RV except for the first harvest (Table 3A). The mean fruit index values for both RV and the population were >1, indicating that the fruits were generally elongated (except for RV in the first harvest). However, the population fruit index ranges showed that the phenotype among the NILs varied from rounded (0.5) to very elongated fruits (2.5) (Table 3A; Fig. 5). We mapped two QTLs for rounded fruits, as donor parent, stable through all five harvests and present in overlapping NILs (Table 2). The first one was in LG2, from cM 0 to 30 and the second one in LG4, cM

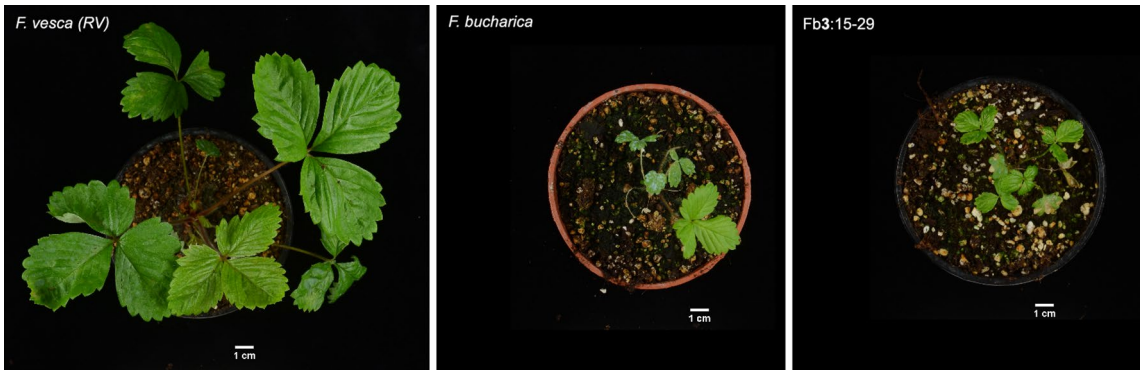

Fig. 4 Size phenotypes. Images show the characteristic plant size 12 weeks after germination for both parents, RV and *F. bucharica*, and NIL Fb**3**:8–94

58–78, accounting for 24 and 35 % of the total variation, respectively (Table 2B). An additional QTL for elongated fruits mapped to LG5 cM 0–11, accounting for 25 % of the variation and stable in four out of five seasons (Table 2B).

Also two QTLs explaining more than 70 % of fruit weight reduction were detected, one in LG4, cM 58–78, on the same region of fruit shape QTL, and another was observed in LG1, cM 50–61 where fruits are smaller but elongated.

QTLs of nutritional content in the NIL collection

Results for genetic analysis of fruit nutritional characters are summarized in Table 3B. Measured parameters showed a normal distribution over most of the harvests.

The sugar composition of the fruits was mainly based on fructose and glucose, with fructose more abundant in almost all the analyses. Fructose, glucose and total sugar content showed a very high correlation within each harvest (always above 0.9). These correlations existed also between harvests and between locations with the exception of Cabrils 2012 and TM 2013 harvests (data not shown). Sucrose, a minor sugar in fruit, where its accumulation is very variable, had little significant correlation with other sugars for each harvest, and no significant correlation over the different years. All the analyzed harvests showed that RV accumulates more sugars than the NIL collection mean. However, the ranges of the population were very extreme and some lines surpassed the mean sugar values of RV (Table 4B). Values for the F_1 individual were collected in 2013. These fruits had lower levels of fructose and glucose than RV.

As a result of the analyses, a QTL resulting in a decrease of fructose, glucose and total sugar content was mapped in LG2, cM 45–63 (Table 4), accounting for 35, 45 and 35 % of the total character variation for these characters, respectively. This QTL was detected in all 3 years. For fructose, we found one additional QTL decreasing also its content (in LG3 cM 54–94), and for glucose we found also one additional QTL decreasing content (in LG5, cM 41–50). Sugar content was not measured in *F. bucharica* but results in F_1 fruits point to donor parent as causal of decrease.

There was a significant correlation of the total polyphenolic content among all the harvests, and the population and RV values were very similar within every harvest (data not shown). There was, however, no correlation between the sugar and the polyphenol content: the 2012 harvest had the highest concentration of fruit sugars and the lowest total polyphenol content. Two regions were associated with the variation in total polyphenolic content. One region was in LG2 cM 0–30, where the alleles from *F. bucharica* increased it by 26 %, and another in LG5 cM 50–76, where the introgressed region caused a 35 % decrease in total polyphenolic content (Table 4). Both were detected in two of the 3 years.

Discussion

Fragaria vesca possesses several features that make it attractive as a model for the octoploid cultivated strawberry, the most evident being that the diploid nature of its genome allows the genetic complexity of the cultivated strawberry to be circumvented. Furthermore, wild strawberries are potential donors of genes of interest for strawberry fruit quality improvement, including disease resistances, and quality characters such as those involved in volatile compound expression and concentration. To enhance our capability to analyze complex traits in strawberry, we developed an NIL collection in an *F. vesca* background, using an Asiatic species *F. bucharica,* as introgression donor and *F. vesca* cv. Reine des Vallées, a French non-runnering variety commonly cultivated in Spain, as the recurrent parent. Considerable research has

Table 3 Agronomic trait distribution (A) and nutritional trait distribution (B)

Trait	Year	Location	*F. vesca* "Reine des Vallées"		NIL population				
			Mean	SD	Mean	SD	Range	Skewness	Kurtosis
(A)									
Fruit length (mm)	2011-May	Cabrils	11.96	1.56	13.85	2.48	7.3–22.0	0.22	3.84
	2011-May	TM	15.72	1.56	15.41	2.66	6.9–30.5	0.01	3.04
	2011-July	Cabrils	23.17	0.17	17.48	3.62	10.9–27.4	0.21	1.85
	2012-May	Cabrils	19.35	2.73	17.63	3.27	7.31–25.09	−0.22	2.20
	2012-July	Cabrils	18.64	2.10	15.72	2.10	8.09–25.07	−0.18	2.61
Fruit width (mm)	2011-May	Cabrils	14.08	1.63	12.06	1.86	5.5–17.8	−2.04	7.87
	2011-May	TM	13.52	1.56	12.42	2.29	4.6–21.4	−0.60	3.01
	2011-July	Cabrils	14.12	1.83	10.02	1.63	6.1–14.9	−0.03	2.60
	2012-May	Cabrils	12.71	1.52	12.58	1.34	7.03–18.28	−0.40	2.99
	2012-July	Cabrils	11.32	1.20	10.89	1.25	5.59–17.19	−0.90	3.15
Index (length/width)	2011-May	Cabrils	0.84	0.03	1.17	0.22	0.58–1.93	0.38	2.77
	2011-May	TM	1.18	0.20	1.28	0.25	0.57–2.42	0.51	3.48
	2011-July	Cabrils	1.65	0.18	1.77	0.30	0.94–2.6	−0.64	2.17
	2012-May	Cabrils	1.54	0.28	1.42	0.27	0.55–2.45	−0.22	2.39
	2012-July	Cabrils	1.66	0.28	1.46	0.20	0.59–2.54	−0.15	2.17
Weight (g)	2011-July	Cabrils	1.66	0.36	0.72	0.28	0.26–1.58	0.53	2.17
	2011-July	TM	2.36	0.57	0.68	0.22	0.26–1.47	0.58	2.75
	2012-July	Cabrils	0.93	0.18	0.66	0.19	0.19–1.09	−0.35	3.64
Flowering (days)	2012	Cabrils	86.00	2.87	125.20	37.58	78–264	2.22	8.54
	2013	TM	118.00	14.62	117.46	28.00	74–189	1.03	3.90
Germination (days)	2013	In vitro	14.88	4.37	17.05	6.08	9.95–42.00	0.02	10.08

Trait	Year	Location	*F. vesca* "Reine des Vallées"		Hybrid (F_1)		NILs population				
			Mean	SD	Mean	SD	Mean	SD	Range	Skewness	Kurtosis
(B)											
Fructose content (mg/g fw)	2011	Pool[a]	27.78	12.48			20.82	5.04	8.33–33.15	0.08	3.00
	2012	Cabrils	25.57	0.79			24.80	5.57	13.05–39.94	0.24	2.90
	2013	TM	26.56	5.46	17.74	2.98	24.18	7.78	8.96–51.42	1.18	4.93
Glucose content (mg/g fw)	2011	Pool[a]	24.64	11.75			17.46	5.19	1.96–30.85	−0.30	3.46
	2012	Cabrils	26.82	3.28			23.02	6.18	11.59–41.27	0.51	3.00
	2013	TM	18.15	4.22	9.90	2.10	15.71	5.64	0.875–33.20	0.63	3.93
Sucrose (mg/g fw)	2011	Pool[a]	16.52	4.84			12.32	5.48	1.67–26.76	0.52	2.91
	2012	Cabrils	12.04	2.21			10.40	2.67	2.92–17.16	−0.13	2.81
	2013	TM	18.42	2.67	21.55	6.10	14.18	6.47	0–31.79	0.08	2.77
Total sugar content (mg/g fw)	2011	Pool[a]	68.94	26.72			50.62	12.32	11.95–77.23	−0.52	3.65
	2012	Cabrils	64.44	5.41			58.23	11.99	29.66–92.28	0.13	3.21
	2013	TM	63.14	9.79	49.20	8.14	54.09	16.63	20–107.26	0.54	3.38
Total polyphenol content (Gae/g fw)	2011	Pool[a]	4.12	1.02			4.49	1.14	2.26–8.35	0.77	3.71
	2012	Cabrils	2.09	0.16			2.23	0.73	0.95–3.96	0.51	2.45
	2013	TM	4.27	0.63	2.63	0.30	4.05	0.82	2.28–6.72	0.27	3.15

[a] Cabrils and TM

been published on *F. vesca* and *F. bucharica* genomes and maps have been updated (Sargent et al. 2004, 2006; Vilanova et al. 2008; Ruiz-Rojas et al. 2010). Here, we provide an additional set of SSR markers, for a region previously covered by RFLPs, which improve the current reference map.

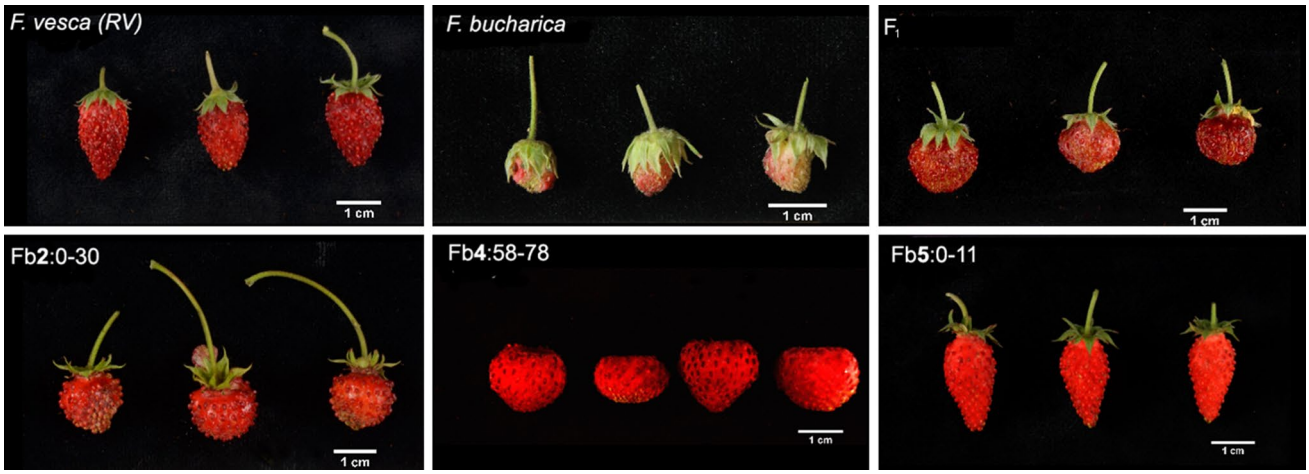

Fig. 5 Fruit phenotypes. *Left to right* and *top to bottom*, characteristic fruit morphology of the parents, Reine des Vallées and *F. bucharica*, their F_1 and NILs Fb**2:**0–30, Fb**4:**58–78 and Fb**5:**0–11

Table 4 QTLs for nutritional traits

Nutritional trait	QTL	% Effect[a]	QTL stability	LG	Initial position (cM)	Final position (cM)	Shortest NIL
Fructose content	Fru_2	−35	3/3	2	45 (EMFn134)	63 (EMFxa379796)	Fb**2**:0–63
	Fru_3	−40	2/3	3	54 (CFVCT022)	94 (CFVCT012)	Fb**3**:54−94
Glucose content	Glu_2	−45	3/3	2	45 (EMFn134)	63 (EMFxa379796)	Fb**2**:0–63
	Glu_5	−28	2/3	5	41 (UDF009)	50 (CFV3821)	Fb**5**:41–76
Total sugar content	TotSug_2	−35	3/3	2	45 (EMFn134)	63 (EMFxa379796)	Fb**2**:0–63
Total polyphenol content	Tphe_2	26	2/3	2	0 (EMFvi099)	30 (BFACT002)	Fb**2**:0–30
	Tphe_5	−35	2/3	5	50 (CFV3821)	76 (VT010)	Fb**5**:50–76

Trait, name of QTL, effect as mean percentage compared to RV, stability as number of harvests observed, bin position (LG, markers flanking bin and cM) and shortest NIL showing trait effect

[a] % Effect with respect to parental RV

The 39 NILs of the collection covered 96.4 % of the *F. bucharica* genome. Each line had a single introgressed chromosome fragment from the *F. bucharica* parent genome, covering 6 % of its genome on average. The resolution capacity, 14.2 cM per bin on average, with regions defined by only 3.2 cM, makes this collection a powerful tool for QTL analysis.

This collection of introgression lines was obtained in only two generations of backcrossing and one to four of selfing. A larger number of backcross generations have been used to extract other NIL collections. Eduardo et al. (2005) used a group of eight double haploid lines to produce a melon NIL collection in six BC generations, Jeuken and Lindhout (2004) began from an F_1 and needed to get to the fifth generation backcross to obtain lettuce NILs, and Eshed and Zamir (1994) needed up to 12 BC generations, using a hybrid between tomato and *Solanum pennellii*. Our strategy was to select from a large BC_1 progeny, which has the advantage of minimizing the risk of unnoticed small introgressions that may be problematic for the use of the NIL collection, and reduces the number of generations needed for its completion. In our case, this was further facilitated by the low number of chromosomes of *Fragaria* ($x = 7$) compared to melon and tomato (both with $x = 12$), or lettuce ($x = 9$). In *Arabidopsis* ($x = 5$), the NIL collection was obtained after backcrossing an RIL collection, and several selfing stages (Keurentjes et al. 2007). We made an initial selection with only 14 molecular markers located on the ends of chromosomes, discarding over 70 % of the BC_1 individuals to select a final working population of nine plants, covering the entire genome with a maximum of four introgressions per plant. Selection at the BC_1 level also facilitated extraction of five NILs containing a whole chromosome of *F. bucharica*. Backcrossing heterozygous plants for these introgressions would produce a series of recombinant individuals with diverse coverage of these chromosomes, useful for fine mapping of genes/QTLs located on them. Few major genes have been mapped in strawberry:

two of them, gen controlling runner formation mapped in LG2 and gen controlling continuous flowering mapped in LG6 (Sargent et al. 2004), segregated in our collection and were assigned to the expected position.

A study with different species of *Rosa* flowers including *Rosa* × *hybrida* determined that the expression pattern of the rose AGAMOUS ortholog gene (*RhAG*) is responsible for the number of petals. Higher expression of *RhAG* has been associated with low petal number cultivars while a more restricted *RhAG* expression towards the flower center has been associated with high petal number cultivars (Dubois et al. 2010). The flower developmental stages and AG expression between *F. vesca* and rose have been described as similar. The *F. vesca* AG homolog (*FvAG*) was identified as the Hybrid Gene Model #24852 (Hollender et al. 2012). This gene is located in *F. vesca* LG3 3939971–3944769, within the boundaries of the exotic introgression in Fb**3**:54–94. Further studies are needed to validate if a different expression pattern of *FvAG F. bucharica* in relation to *F. vesca* could be responsible for the *EPN* phenotype.

F. vesca is a perennial plant and there are accessions with opposite photoperiodic responses. Short-day (SD) *F. vesca* accessions require short days and low temperatures to initiate flowering; in contrast, everbearing (or semperflorens) *F. vesca* accessions initiate flowering on long days and high temperatures. Control of these contrasting phenotypes was recently attributed to a single dominant gene seasonal flowering locus (SFL) that was mapped and characterized as the *F. vesca* TFL1 homolog (FvTFL1), a photoperiodically controlled flowering repressor (Iwata et al. 2011; Koskela et al. 2012). There is functional evidence that FvTFL1 functional alleles cause the SDF flowering habit and that a non-functional allele (2 bp deletion generating a truncated protein) is associated with the everbearing flowering habit in *F. vesca*. The *F. vesca* RV parental is an everbearing species that carries the truncated allele of FvTFL1 while the *F. bucharica* parental has an SD flowering habit and carries the functional allele so we observe segregation of this character in our NIL collection. FvTFL1 is located in LG6 12836105–12837237 bp between EMFn117 (30.1 cM) and EMFn017 (38.8 cM) markers. We have observed that all NILs with exotic introgressions between these two markers have an SD habit that can be assumed by the functional FvTFL1 alleles but in addition, these NILs did not flower until the 2nd or the 3rd year after germination and their blooming was reduced to 2 or 3 inflorescences. This could be caused by other independent loci, included in the same introgression, affecting the juvenile period of the plants.

In many plant species, domestication has resulted in a modified architecture of the reproductive organs, mainly fruit and seed, and, generally, variability has been amplified by human selection. This is the case of fruit morphology, where shape and size traits have often been subjected to intense selection. The genetic control of these characters has been studied in melon (Fernández-Silva et al. 2010) and in tomato (Frary et al. 2000; Liu et al. 2002), where two QTLs with major effects on fruit weight (fw2.2) and shape (ovate) have been positionally cloned. Studies on the genetic basis of these traits in the Rosaceae family are few. Zhang et al. (2006) have shown that the evolution of fruit size in *Pyrus pyrifolia* during domestication is due to changes in the ability of cells to divide after pollination, and Zhang et al. (2010) found several QTLs for size and mesocarp cell number, in cherry, that suggested that both characters were positively correlated. In this work, we located three QTLs involved in strawberry fruit shape, two (LG2 and LG4) showing rounded fruits, and one (LG5) showing elongated fruits. The positions of these QTLs in the strawberry map are not syntenic with the *Prunus* chromosome 2 (Illa et al. 2011) where the QTL for fruit size in cherry occurs, suggesting that they correspond to other genes involved in determining this character. Fine mapping of these QTLs in future studies would be facilitated by the NIL collection presented here, which could be an incentive for the study of quantitative traits of cultivated strawberry fruit morphology and its comparison with other species of the Rosaceae family.

Given that the recurrent parent RV is cultivated commercially, any NIL containing an interesting new character could be rapidly evaluated as a possible new cultivar. For example, NIL Fb**2:**39–47 has the "*r*" locus that allows the plant to develop stolons. If other agronomic and organoleptic characteristics of the range remain unchanged, this line may represent a direct improvement of RV, which can be propagated vegetatively. The same could be true for the QTLs of increased sugar or polyphenolic content found in this paper. For lines with selected phenotypes which have other undesirable traits associated, subNILs, with shorter introgressed fragments, could be obtained that would eventually eliminate unwanted regions.

Consumers now call for healthy products with enhanced nutritional and organoleptic characteristics. Strawberry is an important source of natural products such as vitamins, flavonoids and other (poly-)phenolic compounds with antioxidant capacity. Sugars and organic acids of strawberries have been investigated as indicators of fruit development and ripening and as components of fruit flavor (Pérez et al. 1997). The anticancer effects of specific phytochemical constituents of strawberries, as well as whole strawberry extracts, have been demonstrated (Seeram 2008). The variation of nutritional parameters in different genotypes indicates a genetic control of these parameters, suggesting that the strawberry can potentially be improved for nutrition-enhanced traits, and some examples of genome regions carrying genes associated with sugar content phenotypic

variation have been found in this paper: a QTL in LG2 determined a decrease (around 35 %) of total sugars, fructose and glucose, a QTL for a decrease in fructose or glucose content was detected in LG3 and LG5, respectively. The polyphenol content in NIL **2**:0–30 was increased in 26 %, and another region (LG5) decreased total polyphenol content.

NILs have allowed cloning of genes involved in fruit sugar content (Fridman et al. 2000), and mapping of QTLs for aroma (Tadmor et al. 2002) and color (Liu et al. 2003) of tomato fruit. Developing melon NILs has been crucial in clarifying the genetic basis of some climacteric accessions (Vegas et al. 2013). Analysis of the metabolome and the dissection of its inheritance in tomato (Schauer et al. 2008) were thanks to a collection of introgression lines of the cultivated tomato with other wild species. Therefore, it is likely that, through the collection of introgression lines developed here, complex traits such as nutritional quality, flavor or shelf-life of the fruit can also be studied. These characters are crucial targets for commercial plant breeding in rosaceous fruit such as strawberry, apple, peach, pear, apricot and raspberry, as they are increasingly in demand by consumers. Being these fruit crops a significant part of the human diet, new cultivars based on the generated knowledge can contribute to the improvement of the health and well-being of our society.

Author contribution statement JB obtained the first generation of NILs, phenotyped plants for the first season and was involved in edition of the manuscript. MU made crosses to increase the number of NILs, phenotyped plants for the three seasons and participated in edition of the manuscript. PA designed the crossing and selection scheme and contributed to writing the manuscript. AM lead the project, participated in all steps of selection and phenotyping and in writing the manuscript.

Acknowledgments The authors thank J. Ribes and A. Ortigosa for technical assistance. This work was funded by Grants RTA2007-00063 and AGL2010-21414 from the Spanish Ministry of Science. MU was supported by a fellowship FPI from the Spanish Ministry of Education and JB was supported by a fellowship FI from the Generalitat de Catalunya.

Conflict of interest The authors declare that they have no conflict of interest.

References

Arulsekar S, Bringhurst RS (1981) Genetic model for the enzyme marker PGI in diploid California *Fragaria vesca*. Its variability and use in elucidating the mating system. J Hered 72:117–120

Bolon YT, Bindu J, Cannon SB, Grahan MA, Diers BW, Farmer AD, May GD, Muehlbauer GJ, Specht JE, Tu ZJ, Weeks N, Xu WW, Schoemaker RC, Vance CP (2010) Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. BMC Plant Biol 10:41–62

Bošković RI, Sargent DJ, Tobutt KR (2010) Genetic evidence that two independent S-loci control RNase-based self-incompatibility in diploid strawberry. J Exp Bot 61:755–763

Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF (2001) M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. Biotechniques 31:24–28

Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0, http://www.R-project.org

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Dubois A, Raymond O, Maene M, Baudino S, Langlade NB et al (2010) Tinkering with the C-function: a molecular frame for the selection of double flowers in cultivated roses. PLoS One 5:e9288

Dunnett CW (1955) A multiple comparison procedure for comparing several treatments with a control. J Am Stat Assoc 50:1096–1121

Eduardo I, Arús P, Monforte AJ (2005) Development of a genomic library of near isogenic lines (NILs) in melon (*Cucumis melo* L.) from the exotic accession PI 161375. Theor Appl Genet 112:139–148

Eshed Y, Zamir D (1994) Introgressions from *Lycopersicon pennellii* can improve the soluble-solids yield of tomato hybrids. Theor Appl Genet 88:891–897

Eshed Y, Zamir D (1995) An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141:1147–1162

Fernández-Silva I, Moreno E, Eduardo I, Arús P, Álvarez JM, Monforte AJ (2009) On the genetic control of heterosis for fruit shape in melon (*Cucumis melo* L.). J Hered 100:229–235

Fernández-Silva I, Moreno E, Essafi A, Fergany M, García-Mas J, Martín-Hernández AM, Álvarez JM, Monforte AJ (2010) Shaping melons: agronomic and genetic characterization of QTLs that modify melon fruit morphology. Theor Appl Genet 121:931–940

Fletcher RS, Mullen JL, Yoder S, Bauerle WL, Reuning G, Sen S, Meyer E, Juenger TE, Mckay JK (2013) Development of a next-generation NIL library in *Arabidopsis thaliana* for dissecting complex traits. BMC Genomics 14:655–668

Frary A, Nesbitt TC, Frary A, Grandillo S, van der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science 289:85–88

Fridman E, Pleban T, Zamir D (2000) A recombination hotspot delimits a wild-species quantitative trait locus for tomato sugar content to 484 bp within an invertase gene. Proc Natl Acad Sci USA 97:4718–4723

Fujita D, Trijatmiko KR, Tagle AG, Sapasap MV, Koide Y, Sasaki K, Tsakirpaloglou N, Gannaban RB, Nishimura T, Yanagihara S (2013) NAL1 allele from a rice landrace greatly increases yield in modern indica cultivars. Proc Natl Acad Sci USA 110:20431–20436

Harris K, Subudhi PK, Borrell A, Jordan D, Rosenow D, Nguyen H, Klein P, Mullet J (2007) Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. J Exp Bot 58:327–338

Hollender CA, Geretz AC, Slovin JP, Liu Z (2012) Flower and early fruit development in a diploid strawberry, *Fragaria vesca*. Planta 235:1123–1139

Illa E, Sargent DJ, Lopez Girona E, Bushakra J, Cestaro A, Crowhurst R, Pindo M, Cabrera A, van der Knaap E, Iezzoni A, Gardiner S, Velasco R, Arús P, Chagné D, Troggio M (2011) Comparative analysis of rosaceous genomes and the reconstruction

of a putative ancestral genome for the family. BMC Evol Biol 11:9–22

Iwata H, Gaston A, Remay A, Thouroude T, Jeauffre J, Kawamura K, Oyant LHS, Araki T, Denoyes B, Foucher F (2011) The TFL1 homologue KSN is a regulator of continuous flowering in rose and strawberry. Plant J 69:116–125

Jeuken MJW, Lindhout P (2004) The development of lettuce backcross inbred lines (BILs) for exploitation of the *Lactuca saligna* (wild lettuce) germplasm. Theor Appl Genet 109:394–401

Jung S, Ficklin S, Lee T, Cheng CH, Blenda A, Zheng P, Yu J, Bombarely A, Cho I, Ru S, Evans K, Peace C, Abbott AG, Mueller LA, Olmstead MA, Main D (2014) The Genome Database for Rosaceae (GDR): year 10 update. Nucl Acids Res 42(1):1237–1244

Keurentjes JJB, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, Effgen S, Vreugdenhil D, Koornneef M (2007) Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. Genetics 175:891–905

Kopisch-Obuch FJ, Diers BW (2006) Segregation at the SCN resistance locus rhg1 in soybean is distorted by an association between the resistance allele and reduced field emergence. Theor Appl Genet 112:199–207

Koskela EA, Mouhu K, Albani MC, Kurokura T, Rantanen M, Sargent DJ, Battey NH, Coupland G, Elomaa P, Hytonen T (2012) Mutation in TERMINAL FLOWER1 reverses the photoperiodic requirement for flowering in the wild strawberry *Fragaria vesca*. Plant Physiol 159:1043–1054

Koumproglou R, Wilkes TM, Townson P, Wang XY, Beynon J, Pooni HS, Newbury HJ, Kearsey MJ (2002) STAIRS: a new genetic resource for functional genomic studies of *Arabidopsis*. Plant J 31:355–364

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Lewis RS, Milla SR, Kernodle SP (2007) Analysis of an introgressed *Nicotiana tomentosa* genomic region affecting leaf number and correlated traits in *Nicotiana tabacum*. Theor Appl Genet 114:841–854

Liu J, Van Eck J, Cong B, Tanksley SD (2002) A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proc Natl Acad Sci USA 99:13302–13306

Liu YS, Gur A, Ronen G, Causse M, Damidaux R, Buret M, Hirschberg J, Zamir D (2003) There is more to tomato fruit colour than candidate carotenoid genes. Plant Biotechnol J 1:195–207

Marcel TC, Aghnoum R, Durand J, Varshney RK, Niks RE (2007) Dissection of the barley 2L1.0 region carrying the 'Laevigatum' quantitative resistance gene to leaf rust using near isogenic lines (NIL) and sybNIL. Mol Plant Microbe Interact 20:1604–1615

Martins WS, Lucas DCS, Neves KFS, Bertioli DJ (2009) WebSat—a web software for microsatellite marker development. Bioinformation 3:282–283

Melchinger AE, Piepho HP, Utz HF, Muminović J, Wegenast T, Törjék O, Altmann T, Kusterer B (2007) Genetic basis of heterosis for growth-related traits in *Arabidopsis* investigated by testcross progenies of near-isogenic lines reveals a significant role of epistasis. Genetics 177:1827–1837

Monforte AJ, Friedman E, Zamir D, Tanksley SD (2001) Comparison of a set of allelic QTL-NILs for chromosome 4 of tomato: deductions about natural variation and implications for germplasm utilization. Theor Appl Genet 102:572–590

Oosumi T, Ruiz-Rojas JJ, Veilleux RE, Dickerman A, Shulaev V (2010) Implementing reverse genetics in Rosaceae: analysis of T-DNA flanking sequences of insertional mutant lines in the diploid strawberry, *Fragaria vesca*. Physiol Plant 140:1–9

Paterson AH, Lander ES, Hewitt JD, Peterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–736

Pérez AG, Olías R, Espada J, Olías JM, Sanz C (1997) Rapid determination of sugars, nonvolatile acids, and ascorbic acid in strawberry and other fruits. J Agric Food Chem 45:3545–3549

Pestsova EG, Börner A, Röder MS (2001) Development of a set of *Triticum aestivum–Aegilops tauschii* introgression lines. Hereditas 135:139–143

Ramsay L, Macaulay M, degli Ivanissevich S, MacLean K, Cardle L, Fuller J, Edwards KJ, Tuvesson S, Morgante M, Massari M, Maestri E, Marmiroli N, Sjakste T, Ganal M, Powell W, Waugh W (2000) A simple sequence repeat-based linkage map of barley. Genetics 156:1997–2005

Rousseau-Gueutin M, Lerceteau-Köhler E, Barrot L, Sargent DJ, Monfort A, Simpson DW, Arús P, Guérin G, Denoyes-Rothan B (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 179:2045–2060

Rousseau-Gueutin M, Gaston A, Aïnouche A, Aïnouche ML, Olbricht K, Staudt d G, Richard L, Denoyes-Rothan B (2009) Tracking the evolutionary history of polyploidy in *Fragaria* L. (strawberry): 3 new insights from phylogenetic analyses of low-copy nuclear genes. Mol Phyl Evol 51:515–530

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386

Ruiz-Rojas JJ, Sargent DJ, Shulaev V, Dickerman AW, Pattison J, Holt SH, Ciordia A, Veilleux RE (2010) SNP discovery and genetic mapping of T-DNA insertional mutants in *Fragaria vesca* L. Theor Appl Genet 121:449–463

Sargent DJ, Davis TM, Tobutt KR, Wilkinson MJ, Battey NH, Simpson DW (2004) A genetic linkage map of microsatellite, gene-specific and morphological markers in diploid *Fragaria*. Theor Appl Genet 109:1385–1391

Sargent DJ, Clarke J, Simpson DW, Tobutt KR, Arús P, Monfort A, Vilanova S, Denoyes-Rothan B, Rousseau M, Folta KM, Bassil NV, Battey NH (2006) An enhanced microsatellite map of diploid *Fragaria*. Theor Appl Genet 112:1347–1359

Sargent DJ, Cipriani G, Vilanova S, Gil-Ariza D, Arús P, Simpson DW, Tobutt KR, Monfort A (2008) The development of a bin mapping population and the selective mapping of 103 markers in the diploid Fragaria reference map. Genome 51:120–127

Sargent DJ, Fernández-Fernández F, Ruiz-Rojas JJ, Southerland BG, Passey A, Whitehouse AB, Simpson DW (2009) A genetic linkage map of the cultivated strawberry (*Fragaria* × *ananassa*) and its comparison to the diploid *Fragaria* reference map. Mol Breed 24:293–303

Schauer N, Semel Y, Balbo I, Steinfath M, Repsilber D, Selbig J, Pleban T, Zamir D, Fernie AR (2008) Mode of inheritance of primary metabolic traits in tomato. Plant Cell 20:509–523

Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671–675

Seeram NP (2008) Berry fruits: compositional elements, biochemical activities, and the impact of their intake on human health, performance, and disease. J Agric Food Chem 56:627–629

Shulaev V, Sargent DJ, Crowhurst RN, Mockler RN, Veilleux RE, Folkerts O, Delcher AL, Jaiswal P, Liston A, Mane SP, Burns P, Mockaitis K, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Jensen RV, Allan AC, Michael TP, Setubal JC, Celton JM, Rees DJG, Williams K, Holt SH, Dickerman A, Ruiz Rojas JJ, Chatterjee M, Liu B, Silva H, Meisel L, Filichkin AAS, Velasco R, Troggio M, Viola R, Borodovsky M, Ashman TL, Aharoni A, Bennetzen J, Dharmawardhana P, Elser J, Raja R, Priest

HD, Bryant DW, Fox SE, Givan SA, Naithani S, Christoffels A, Salama DY, Carter J, Lopez Girona E, Zdepski A, Wang W, Kerstetter RA, Salzberg SL, Schwab W, Korban SS, Davik J, Monfort A, Denoyes-Rothan B, Arus P, Mittler R, Flinn B, Folta KM (2011) The genome of woodland strawberry (*Fragaria vesca*). Nat Genet 43:109–116

Singleton VL, Orthofer R, Lamuela-Raventós RM (1999) Analysis of total phenols and other oxidation substrates and antioxidants by means of Folin–Ciocalteu reagent. Methods Enzymol 299:152–178

Staudt G (2006) Himalayan species of *Fragaria* (Rosaceae). Bot Jahrbucher fur Syst 126:483–508

Staudt G (2008) Strawberry biogeography, genetics and systematics. Acta Hortic 842:71–83

Szalma SJ, Hostert BM, LeDeaux JR, Stuber CW, Holland JB (2007) QTL mapping with near-isogenic lines in maize. Theor Appl Genet 114:1211–1228

Tadmor Y, Fridman E, Gur A, Larkov O, Lastochkin E, Ravid U, Zamir D, Lewinsohn E (2002) Identification of malodorous, a wild species allele affecting tomato aroma that was selected against during domestication. J Agric Food Chem 50:2005–2009

Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063–1066

van Berloo R (2008) GGT 2.0: versatile software for visualization and analysis of genetic data. J Hered 99:232–236

Van Ooijen JW (2006) Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen

Vegas J, Garcia-Mas J, Monforte AJ (2013) Interaction between QTLs induces an advance in ethylene biosynthesis during melon fruit ripening. Theor Appl Genet 126:1531–1544

Vilanova S, Arús P, Sargent DJ, Monfort A (2008) Synteny conservation between two distantly-related Rosaceae genomes: *Prunus* (the stone fruits) and *Fragaria* (the strawberry). BMC Plant Biol 8:67–79

Wan XY, Wan JM, Su CC, Want CM, Shen WB, Li JM, Wang HL, Jiang L, Liu SJ, Chen LM, Yasui H, Yoshimura A (2004) QTL detection for eating quality of cooked rice in a population of chromosome segment substitution lines. Theor Appl Genet 110:71–79

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2:983–989

Zhang C, Tanabe K, Wang S, Tamura F, Yoshida A, Matsumoto K (2006) The impact of cell division and cell enlargement on the evolution of fruit size in *Pyrus pyrifolia*. Ann Bot 98:537–543

Zhang Y, Luo L, Liu T, Xu C, Xing Y (2009) Four rice QTL controlling number of spikelets per panicle expressed the characteristics of single Mendelian gene in near isogenic backgrounds. Theor Appl Genet 118:1035–1044

Zhang GR, Sebolt AM, Sooriyapathirana SS, Wang DC, Bink MCAM, Olmstead JW, Iezzoni AF (2010) Fruit size QTL analysis of an F_1 population derived from a cross between a domesticated sweet cherry cultivar and a wild forest sweet cherry. Tree Genet Genomes 6:25–36

Zorrilla-Fontanesi Y, Cabeza A, Torres A, Botella MA, Valpuesta V, Monfort A, Sánchez-Sevilla JF, Amaya I (2010) Development and bin mapping of strawberry genic-SSRs in diploid *Fragaria* and their transferability across the Rosoideae subfamily. Mol Breed 27:137–156