Tree Genetics & Genomes (2012) 8:237–247
DOI 10.1007/s11295-011-0436-9

ORIGINAL PAPER

Rosaceae conserved orthologous sequences marker polymorphism in sweet cherry germplasm and construction of a SNP-based map

Antonio Cabrera · Umesh R. Rosyara · Paolo De Franceschi · Audrey Sebolt · Suneth S. Sooriyapathirana · Elisabeth Dirlewanger · Jose Quero-Garcia · Mirko Schuster · Amy F. Iezzoni · Esther van der Knaap

Received: 1 June 2011 / Revised: 12 August 2011 / Accepted: 14 September 2011 / Published online: 18 October 2011

Abstract The Rosaceae Conserved Orthologous Set (RosCOS) provides a gene-based genome-wide set of markers that have been used in comparative analyses of peach (*Prunus persica*), apple (*Malus* × *domestica*), and strawberry (*Fragaria* spp.). In order to extend the use of these RosCOS to sweet cherry (*Prunus avium* L.), we identified markers that are polymorphic in breeding germplasm. Ninety-five percent (595/627) of previously designed RosCOS primer pairs amplified a product in six sweet cherry cultivars predicted to represent the range of genetic diversity in breeding germplasm. A total of 45% (282/627) RosCOS were polymorphic among the six cultivars, and allele number ranged from 2 to 6, with a genome-wide mean of 2.35. A subset of 92 genome-wide

Communicated by A. Dandekar

Electronic supplementary material The online version of this article (doi:10.1007/s11295-011-0436-9) contains supplementary material, which is available to authorized users.

A. Cabrera · E. van der Knaap (✉)
Department of Horticulture and Crop Science, Ohio Agricultural Research and Development Center, The Ohio State University, Wooster, OH 44691, USA
e-mail: vanderknaap.1@osu.edu

A. Cabrera
e-mail: cabrera.16@buckeyemail.osu.edu

U. R. Rosyara
Department of Crop and Soil Science, Michigan State University, East Lansing, MI 48824, USA
e-mail: rosyara@msu.edu

P. De Franceschi
Department of Fruit Tree and Woody Plant Sciences, University of Bologna, Bologna, Italy
e-mail: pdefra@agrsci.unibo.it

A. Sebolt · S. S. Sooriyapathirana · A. F. Iezzoni (✉)
Department of Horticulture, Michigan State University, East Lansing, MI 48824, USA
e-mail: iezzoni@msu.edu

A. Sebolt
e-mail: grantzau@msu.edu

S. S. Sooriyapathirana
e-mail: sunethss09@gmail.com

E. Dirlewanger · J. Quero-Garcia
INRA-UR419, Unité de Recherches sur les Espèces Fruitières, Centre de Bordeaux, 33140 Villenave d'Ornon, France

E. Dirlewanger
e-mail: dirlewan@bordeaux.inra.fr

J. Quero-Garcia
e-mail: jose.querogarcia@bordeaux.inra.fr

M. Schuster
Julius Kühn-Institut, Federal Centre for Cultivated Plants, PillnitzerPlatz 3a, 01326 Dresden, Germany
e-mail: mirko.schuster@jki.bund.de

single nucleotide polymorphisms (SNPs) corresponding to 76 RosCOS was analyzed in 36 founder accessions and progeny. The expected and observed heterozygosity suggested that 83% of the RosCOS were in Hardy–Weinberg equilibrium, implying that most RosCOS behave as neutral markers. Principal coordinate analysis (PCO) identified one wild accession and two Spanish landraces that clustered differently from the other accessions. The relatively high number of unique alleles found in the three differentially clustered selections suggested that their use as parents has potential to increase the genetic diversity in future US-bred cultivars. Of the 92 RosCOS SNPs, 81 SNPs that represented 68 genome-wide RosCOS segregated in four mapping populations. These RosCOS were mapped in four F_1 populations, thereby greatly improving the genetic linkage map of sweet cherry.

Keywords *Prunus* · RosCOS · Diversity

Introduction

Sweet cherry (*Prunus avium* L.) is a diploid ($2n=2\times=16$) species and a member of the Rosaceae family. Domesticated and wild forms, grown for their fruit and wood, respectively, have coexisted in Europe for many centuries. Wild sweet cherry is indigenous to all of mainland Europe ranging east into Russia, with the greatest prevalence between the Caspian and Black Seas (Watkins 1976). For centuries, peasants and gardeners selected and propagated the best clones by root suckers or grafting, providing a rich source of diversity (Iezzoni et al. 1991). From this wild germplasm, the domestication and cultivation of sweet cherry is proposed to have started in the region of Central Asia–Caucasia (Hedrick et al. 1915), resulting in improved fruit size and quality traits. The cultivated form was subsequently introduced into the Mediterranean region and later to Central and Northern Europe.

In the 1600s, sweet cherries were brought to North America by early settlers, first as seed and then as cultivars from Europe (Hedrick et al. 1915). Because all sweet cherry germplasm had to be imported, the genetic diversity used by North American breeders was limited (Choi and Kappel 2004). Recently, wild sweet cherry and a wider array of landrace varieties have been used to broaden the genetic base of sweet cherry breeding germplasm in both North America and Europe (A. Iezzoni, unpublished). Knowledge of the genetic diversity present in this expanded set of founders and the genetic relationships among these founders is critical for the development of genetic markers that will be useful in the breeding germplasm.

Until recently, genetic marker development lagged behind for cherry compared with other rosaceous crops such as apple, peach, and almond (*Prunus dulcis*). Currently, only medium density linkage maps have been constructed from three diploid cherry populations using amplified fragment length polymorphism, simple sequence repeat (SSR), and gene-based simple PCR markers. Intraspecific cherry mapping populations include: (1) a cross between two commercial sweet cherry cultivars, 'Regina' and 'Lapins' (Dirlewanger et al. 2004); (2) a cross between 'Emperor Francis' (EF), a founder parent for many US and northern European cultivars, and 'New York 54' (NY54), a wild mazzard cherry believed to originate from the German forest (Olmstead et al. 2008); and (3) an interspecific cross between 'Napoleon', a founder sweet cherry cultivar, and *Prunus nipponica* (Japanese alpine cherry; Clarke et al. 2009). These linkage maps provide a starting point for locating and genetically characterizing traits of interest in cherry germplasm. The utility of mapped markers for breeding applications is dependent on the allelic diversity of these markers in the breeding germplasm as only those with at least two alleles are useful for understanding genetic diversity and for tagging chromosome regions for quantitative trait locus (QTL) detection.

Single nucleotide polymorphisms (SNPs) are the most prevalent DNA polymorphisms in genomes (Syvanen 2001). Due to their prevalence and suitability for high-throughput genotyping, SNPs are well suited for whole-genome scans for QTL studies. High-throughput genotyping arrays using the GoldenGate® Assay (Illumina, Inc., San Diego, CA) that was used in this study have previously been used for SNP genotyping in soybean (Hyten et al. 2008), wheat (Akhunov et al. 2009), and maize (Yan et al. 2010). SNP assays are usually biallelic, and thus, individual SNPs are often less informative than individual SSRs. However, high-throughput multi-SNP haplotyping can be used to define DNA sequence diversity at individual loci as it allows the identification of more than two allelic variants. In addition, the greater frequency of SNPs over SSRs makes the former more useful when the polymorphism within specific genes is desired for targeted investigations.

For most major crop plants, large numbers of SNPs have been identified in public databases of DNA sequences using bioinformatic approaches (Barbazuk et al. 2007; McNally et al. 2009; Wu et al. 2010; Yang et al. 2004). Among fruit crops, alignment of expressed sequence tags (ESTs) resulted in the in silico identification of thousands of SNPs in peach (Jung et al. 2008). Genomic and EST information for cherry is minimal; therefore, the in silico identification of large numbers of SNPs in cherry breeding germplasm is not feasible with the current information in public databases. Recently, a Rosaceae Conserved Orthologous Set

(RosCOS) was developed, and 607 out of 784 were conclusively bin-mapped in the *Prunus* reference genetic map of T×E (almond 'Texas' × peach 'Earlygold') using intron-flanking primers and sequencing of the amplified products (Cabrera et al. 2009). The genome-wide coverage of these RosCOS ranges from 0.67 to 1.06 markers per centimorgan across the eight linkage groups (G1 to G8) of peach (Cabrera et al. 2009). The RosCOS has been used for whole-genome comparative analyses among some of the key species in the Rosaceae, resulting in the identification of syntenic blocks among the species within this important family (Illa et al. 2011; Shulaev et al. 2011). As the RosCOS provides an ortholog-based and genome-wide set of anchor markers for the Rosaceae, identifying SNPs for these genes in cherry breeding germplasm would provide useful markers for high-throughput genome scanning and linkage to target traits. It is expected that many RosCOS primer pairs, designed primarily from peach gene sequences, would produce amplification products in closely related cherry. Therefore, SNP discovery could be accomplished by the amplification, sequencing, and nucleotide alignment of the resulting PCR products from a set of sweet cherry selections predicted to represent the range of genetic diversity in breeding germplasm.

Our objectives were to develop a set of genome-wide RosCOS SNP markers for marker-assisted breeding in sweet cherry. We also sought to determine the genetic relationships among founder accessions and the number of unique alleles for each polymorphic RosCOS in a subset of these founders. This information would enable breeders to select breeding parents and follow their parental genetic contributions in the progeny. To achieve these goals, (1) RosCOS SNP sequencing was performed in a panel of six diverse sweet cherry selections, (2) RosCOS SNP polymorphism and diversity analysis was assessed in a subset of genome-wide RosCOS SNPs in the sweet cherry breeding germplasm, and (3) a subset of the RosCOS was placed on a consensus sweet cherry linkage map.

Materials and methods

Plant materials

The sweet cherry selections evaluated consisted of 36 individuals (Table 1) that were used as parents in the Pacific Northwest breeding program located at the Washington State University (WSU) Irrigated Agricultural Research and Extension Center (IAREC) in Prosser, WA, USA. Trees for these 36 individuals are located at WSU-IAREC, the Michigan State University Clarksville Horticultural Research Station, Clarksville, MI, USA, and the North West Horticultural Research Station, Traverse City, MI, USA.

Table 1 Thirty-six sweet cherry selections genotyped for SSR and SNP markers, their parents, and countries of origin

Cultivar	Parent 1 × Parent 2	Country of origin
19–21 B[a]	U × U	Ukraine
Ambrunes[a]	U × U	Spain
Benton	Stella × Beaulieu	USA
Bing[a]	Black Republican × U	USA
Brooks	Rainier × Early Burlat	USA
Chelan	Stella × Beaulieu	USA
Chinook	Bing × Gil Peck	USA
Cowiche[b]	PC7147-4 × PC7146-11[c]	USA
Cristobalina[a]	U × U	Spain
Emperor Francis[a]	U × U	Northern Europe
Eugenia[a]	U × U	Northern Europe
Glacier	Stella × Early Burlat	USA
Katalin[a]	Germersdorfer × Podjebrad	Hungary
Kiona[b]	Glacier × Cashmere	USA
Krupnoplodnaya[a]	Drogana Zholtaya × Valeriy Chkalov[d]	Ukraine
Lambert[a]	Napoleon × Black Heart	USA
Lapins	Van × Stella	USA
Napoleon[a]	U × U	Germany
Newstar	Van × Stella	Canada
New York 54[a]	U × U	Germany
PC7147-009	Stella × U	USA
PMR-1[a]	U × U	USA
Rainier	Bing × Van	USA
Regina[a]	Schneiders × Rube	Germany
Sam	Windsor × U	Canada
Schmidt[a]	U × U	Germany
Schneiders[a]	U × U	Germany
Selah	P8-79[e] × Stella	USA
Stella	Lambert × JI2420[f]	Canada
Summit	Van × Sam	Canada
Sweetheart	Van × Newstar	Canada
Tieton	Stella × Early Burlat	USA
Ulster	Schmidt × Lambert	USA
Van[a]	Empress Eugenie × U	Canada
Vic	Bing × Schmidt	Canada
Windsor[a]	U × U	Northern Europe

U Unknown

[a] Founder

[b] The selection numbers for Cowiche and Kiona were PC7903-2 and PC800702, respectively

[c] Parentage of PC7146-11 is Stella × Beaulieu

[d] Pedigree provided by L. Taranenko, personal communication

[e] Parentage of P8-79 is Bing × Rainier

[f] Parentage of JI2420 is Emperor Francis × Napoleon

Table 2 Populations, their locations, and number of progeny individuals genotyped for SSR and RosCOS SNP markers

Population	No. of progeny genotyped		
	Location	SSR	SNP
New York 54 × Emperor Francis	Michigan, USA	190	113
Regina × Lapins	Bordeaux, France	124	100
Namati[a] × Summit	Dresden, Germany	0	77
Namati × Krupnoplodnaya	Dresden, Germany	0	80
Total		314	370

[a]The parentage of Namati is Bopparder Kracher × open-pollinated

Four populations were used in linkage mapping (Table 2). These included two previously reported mapping populations, 'NY54' × 'EF' (Olmstead et al. 2008), 'Regina' × 'Lapins' (Dirlewanger et al. 2004), and two additional mapping populations, 'Namati' × 'Summit' and 'Namati' × 'Krupnoplodnaya'. See Table 2 for the number of individuals in each population.

DNA extraction

Immature and actively growing leaves were sampled from all individuals in early spring, placed immediately on dry ice for transport to the laboratory, and stored at −80°C. Frozen leaf samples were freeze-dried for 48 to 72 h and stored at −20°C until DNA extraction. DNA was extracted using the cetyltrimethylammonium method described by Stockinger et al. (1996). DNA extractions from the 'Regina' × 'Lapins', 'Namati' × 'Summit', 'Namati' × 'Krupnoplodnaya' populations were performed from winter buds with the bud scales removed using the DNeasy plant Mini kit according to the manufacturer's protocol (Qiagen, Inc.).

Selection of the RosCOS SNP detection panel

Six sweet cherry cultivars that represented different and diverse parental lineages in sweet cherry breeding were selected for RosCOS SNP detection utilizing a three-step process. For the first step, pedigree information was used to select 17 founders of the cultivars (Table 1). For the second step, polymorphism analyses (see below) and genetic distances among these 17 founders were estimated using data from 72 previously reported markers that included 69 SSRs, the *S* locus *RNase*, and two Indel markers (PR33 and PR101) [for information about the markers used, see Olmstead et al. 2008; Electronic supplementary material (ESM) Table 1]. Allelism for all of the marker loci was confirmed by tracing inheritance through known pedigrees. Finally, the last criterion used to select the RosCOS SNP detection panel was based on the contribution of the cultivar to pedigrees in the breeding program and in mapping populations used in this study.

RosCOS SNP detection and haplotype identification for the six cultivars in the SNP detection panel

The intron-flanking primers for 627 RosCOS (Cabrera et al. 2009) were used to amplify the corresponding regions from the six cultivars in the SNP detection panel. Of these, 601 RosCOS were selected that are unambiguously placed on the *Prunus* bin reference map (Cabrera et al. 2009). The remaining 26 RosCOS were thought to map to parts of the *Prunus* genome not covered by the bin map and were placed in so-called orphan bins. PCR fragments were sequenced by Agencourt Bioscience Corporation (Agencourt, Beverly, MA) using the M13 forward primer located 5′ to every forward RosCOS primer. Sequences were examined for the presence of SNPs and Indels using Sequencher software v4.2 (Gene Codes Corporation, Ann Arbor, MI). The presence of a double peak in an otherwise high-quality chromatogram was indicative of the presence of a SNP. Sudden decay of a high-quality chromatogram sequence was indicative of the presence of an Indel.

SNPs and Indels used to generate haplotypes for every polymorphic RosCOS

We counted the number of haplotypes for each RosCOS in the set of six cultivars as the minimum number of haplotypes that explained the different SNP combinations. Because most haplotypes were not confirmed by inheritance, it was not always possible to assign the true haplotype sequence to a particular combination of SNPs. For instance, the presence of two polymorphic SNPs between two individuals that are homozygous can be easily interpreted as two haplotypes in our sweet cherry collection (individual 1—SNP1: A/A, SNP2: G/G; individual 2—SNP1: C/C, SNP2: A/A). A third individual that is heterozygous for both SNPs (SNP1: A/C, SNP2: G/A) was inferred to carry one allele from each of the first two individuals, therefore maintaining two haplotypes in the collection. However, if a recombination between the two SNPs had occurred, the number of haplotypes would increase to 4. Because we could not determine whether a recombination had occurred without segregation data, we assumed the minimum number of

haplotypes. In other cases, the presence of a unique SNP in one individual would indicate a novel haplotype not found in any of the other accessions and was listed as an additional haplotype. Moreover, sequencing over an Indel of a heterozygous RosCOS would not allow the identification of polymorphisms after the Indel. Thus, the estimate of the number of haplotypes per RosCOS should be viewed as the minimum number that we could discern in the six sequenced accessions. Haplotypes were arbitrarily designated with letters (i.e., A, B, C, etc.) in order to indicate the genotype of each of the sequenced sweet cherry founders. When possible, the number of haplotypes of selected RosCOS was confirmed in the set of 36 sweet cherry cultivars. We used the term "RosCOS allele" to refer to a unique RosCOS haplotype (ESM Table 2).

Polymorphism analysis for the set of 36 sweet cherry cultivars for a subset of RosCOS SNPs

An initial subset of 96 RosCOS SNPs representing 77 distinct RosCOS was selected. These RosCOS SNPs were chosen based on three criteria. First, collectively, they would cover the eight *Prunus* linkage groups based on their previously determined *Prunus* bin map location (Cabrera et al. 2009). This set included five 'orphan' RosCOS whose T×E bin map positions could not be determined (Cabrera et al. 2009). Second, the RosCOS SNPs had to be heterozygous in parents of available mapping populations (e.g., 'NY54' × 'EF' and 'Regina' × 'Lapins') to facilitate placement on cherry genetic linkage maps. Finally, RosCOS SNPs that had a quality score provided by Illumina for the Illumina GoldenGate® Assay (Illumina, Inc.) >0.7 were selected. These quality scores were calculated from the analysis of the SNP-harboring sequences using the Illumina Assay Design tool (ADT) (ESM Table 3). For each SNP, the ADT score can range from 0 to 1, and SNPs with scores higher than 0.6 have an increased success rate. This set of 96 SNPs was genotyped for 370 progeny from the four mapping populations and the total set of 36 sweet cherry selections (ESM Table 3) utilizing the GoldenGate®Assay. The SNP intensity data from the BeadArray reader was analyzed using GenomeStudio™ Genotyping Module V2010.2 (Illumina, Inc.) where the genotype calling (Gen Call) threshold was set to 0.25. Pedigree information was used to detect genotyping errors. A final report was generated with the actual SNP base pairs (A/T/G/C) according to Illumina's Top/Bottom approach.

Heterozygosity and PIC estimations of RosCOS haplotype markers

Heterozygosity and allele frequency for the 36 sweet cherry selections were calculated for each RosCOS using Popgene 1.31 software (Yeh and Boyle 1997), where the heterozygosity $\widehat{H}$ for a locus l is defined as:

$$\widehat{H}_l = 1 - \sum_{u=1}^{k} P_{luu}$$

where P_{luu} is the frequency of genotypes Aluu at locus l and where the allele u varies from 1 to the total number of alleles k. Besides observed heterozygosity, expected heterozygosity (unbiased estimate of heterozygosity under a small sample size) was calculated in Popgene 1.31 according to the method of Nei (1978). The allele frequency was used to calculate polymorphic information content (PIC) as defined by Botstein et al. (1980), where Plu and Ply are the frequencies of alleles Au and Av, respectively.

$$\mathrm{PIC}_l = 1 - \sum_{u=1}^{k} (P_{lu})^2 - \sum_{u=1}^{k-1} \sum_{v=u+1}^{k} 2\times(P_{lu})^2 \times (P_{lv})^2$$

Linkage map construction

A consensus linkage map was constructed from the segregation data of the RosCOS SNPs, 69 previously mapped SSRs, 2 Indels, and the *S-RNase* locus using the four mapping populations. SNPs from the same gene (less than a hundred base pairs apart) were considered as a single locus for haplotype development due to their close physical location. Linkage analysis was performed using JoinMap® 4.0 for a cross-pollinated population (Van Ooijen 2006). A consensus map was constructed where JoinMap® first estimates the recombination frequency between a given pair of markers from different populations and then applies the appropriate weighting to generate a consensus recombination value. A LOD threshold of 3.0 was used for the majority of the markers. However, the threshold was relaxed to 1.0 for large centimorgan intervals when the

Table 3 Amplification success of RosCOS in a SNP detection panel of six diverse sweet cherry selections (Cristobalina, Emperor Francis, Krupnoplodnaya, Lambert, Regina, and New York 54)

No. of RosCOS attempted	No. of RosCOS that amplified (%)	No.of amplified RosCOS with poor sequence (%)	No.of polymorphic RosCOS (%)	No.of monomorphic RosCOS (%)
627	595 (95%)	29 (5%)	282[a] (45%)	284 (45%)

[a] A total of 268 of these RosCOS has known *Prunus* T×E bin map locations

two markers were known to map to the same linkage group. Kosambi's mapping function (Kosambi 1944) was used to calculate the map distance from the recombination frequency, considering the possibility of interference. For those SNP, SSR, and Indel markers with known locations on the peach physical map (www.rosaceae.org), the order of the loci and the grouping of markers into linkage groups were inferred based on the physical location. In those cases where the consensus map locations did not match the previously determined T×E bin map positions, the marker locations were checked based on segregation in the larger populations, either 'NY' × 'EF' or 'Lapins' × 'Regina'. MapChart (Voorrips 2002) was used to draw the linkage map.

PCO of the 17 founder accessions

PCO was performed using GENALEX software, v.6.4 (Peakall and Smouse 2006). For this analysis, we first determined whether the selected RosCOS were in Hardy–Weinberg equilibrium (HWE) in the entire set of 36 selections. Despite violating nearly all HWE rules (non-random mating, lack of gene flow, small population size), the majority of the markers appeared to fit HWE expectations. This allowed us to remove markers not in HWE that would contribute to spurious correlations between the selections. PCO was conducted using only the RosCOS that fit HWE, and the clusters obtained were confirmed by removing outlier markers with significant high or low fixation index (F_{st}) compared with neutral expectations using the F_{st} outlier detection method from LOSITAN (Antao et al. 2008).

Results

RosCOS SNP reveal levels of diversity in six sweet cherry cultivars

Of the 627 RosCOS primer pairs, 595 (95%) successfully amplified a product from the six sweet cherry cultivars comprising the SNP detection panel (Table 3). In total, 282 out of 566 RosCOS were polymorphic in one or more cultivar based on sequence analysis. 'Cristobalina' contributed the most unique alleles, followed by 'NY54', while 'Regina' contributed the least number of unique alleles, followed by 'EF'. The percentage of successfully sequenced RosCOS that were heterozygous in one or more of the six cultivars ranged from 17% for 'NY54' to 25% for 'EF' (Table 4).

When taking into consideration more than one SNP in the same sequence, the analyses of the sequences of the 282 polymorphic RosCOS identified a total of 660 RosCOS haplotypes (ESM Table 2). When considering RosCOS with known bin position (268) and including the monomorphic sequences, the mean RosCOS haplotype number ranged from 1.42 for G2 to 1.81 for G6 (Table 5 and ESM Table 2). When only considering the polymorphic RosCOS, the mean haplotype number per linkage group increased and ranged from 2.3 to 2.5. Across the eight *Prunus* linkage groups, the mean interval length between the polymorphic RosCOS ranged from 1.2 cM for G1 to 2.9 cM for G7 (Table 5), with an overall mean interval of 2.1 cM.

Diversity analysis of the sweet cherry selections

A subset of 96 SNPs representing 77 RosCOS markers spanning the eight *Prunus* linkage groups of the T×E bin map were selected for genotyping 36 sweet cherry selections (Table 1 and ESM Table 3). Four SNPs were excluded due to high inheritance errors (RosCOS0921-061 and RosCOS1197-622) or lack of polymorphism (RosCOS1549-350 and RosCOS2961-047). As a result, 92 SNPs representing 76 RosCOS were analyzed. The mean PIC value for the 76 RosCOS evaluated for all 36 sweet cherry selections was 0.317, ranging from 0.077 to 0.593. The mean observed (H_o) and mean expected (H_e) heterozygosity for the RosCOS ranged from 0.029 to 0.861 and from 0.028 to 0.663, respectively (ESM

Table 4 Comparison of percent heterozygous loci (*H*) and number of UA contributed by each of the six sweet cherry selections in the SNP detection panel for the total set of 566 RosCOS (polymorphic and monomorphic) and the subset of 282 bin-mapped polymorphic RosCOS

Cultivar	% heterozygous loci based on successfully sequenced RosCOS (n=566)	% heterozygous loci based on polymorphic RosCOS (n=282)	No. of unique alleles based on successfully sequenced RosCOS (n=566)
Cristobalina	20.3	40.8	34
Emperor Francis	25.0	50.7	13
Krupnoplodnaya	20.0	40.1	16
Lambert	20.7	41.1	19
New York 54	17.0	34.0	29
Regina	22.8	45.7	10

UA unique alleles

Table 5 Marker density and mean allele number per linkage group for the RosCOS SNP markers in sweet cherry

LG	Length (cM)	No. of T×E bin-mapped RosCOS markers			Mean no. of alleles per marker	
		Polymorphic[a]	Mean interval length (cM/marker)	Total[b]	Polymorphic[a]	Total[b]
1	87.0	71	1.2	119	2.3	1.77
2	50.5	19	2.7	57	2.3	1.42
3	48.4	29	1.7	61	2.4	1.66
4	62.5	27	2.3	46	2.3	1.76
5	49.1	29	1.7	58	2.5	1.72
6	83.7	43	1.9	79	2.5	1.81
7	70.6	24	2.9	63	2.3	1.48
8	55.9	26	2.2	57	2.3	1.58
Total	507.7	268		540		

Map location is based on bin map position in the interspecific *Prunus* reference population derived from a cross between almond cultivar 'Texas' and peach cultivar 'Earlygold' (Cabrera et al. 2009). Calculations are based on the six cherry accessions selected for sequencing (see Table 4)

[a] Numbers are based on 282 polymorphic − 14 unmapped RosCOS, totaling up to 268 genome-wide RosCOS

[b] Numbers are based on 282 polymorphic + 284 monomorphic − 26 unmapped RosCOS, totaling up to 540 genome-wide RosCOS

Table 4). Chi-square analysis of each RosCOS with H_o and H_e led to the removal of 13 out of 76 RosCOS that deviated from HWE in the 36 sweet cherry selections. The 13 outlier RosCOS that were removed from the analysis were distributed across the eight linkage groups and most often showed a higher than expected mean heterozygosity (on average, H_e=0.44, H_o=0.58; Table 6). The allele frequencies of the remaining 63 RosCOS fit HWE expectations.

To determine whether the 17 founders, most of which were of unknown origin, represented different subpopulations or were part of the same genetic pool, PCO was performed using the RosCOS that fit HWE expectations. PCO explained a total of 74% of the variance with the first, second, and third principal coordinates (PC) explaining 30%, 25.9%, and 18.1%, respectively. Because of the relative importance of the first and second PC, the placement of the 17 founders relative to these first two PCs is presented in Fig. 1.

Three of the founders clustered separately from the other accessions which were derived from North and Central Europe. These were the wild mazzard 'NY54' from Germany and two landrace selections, 'Ambrunes' and 'Cristobalina', from Spain (Fig. 1). Identification of outlier loci with significantly high or low F_{st} values compared with neutral expectations between 'NY54', 'Ambrunes', and 'Cristobalina' compared with the other 14 accessions led to the identification of 35 candidate RosCOS that might be under selection (ESM Table 5). After removing these markers, PCO showed a less clear separation (ESM

Table 6 RosCOS haplotypes that deviated from Hardy–Weinberg equilibrium evaluated over the 36 sweet cherry selections

RosCOS number and SNP position(s)	T×E bin map position	Cherry genetic linkage group: cM location	Deviation from HWE
RC1549-381	1:50	1: 127.1	Higher
RC1169-212	1:73	1: 149.9	Higher
RC0541-317	2:08	Not mapped	Lower
RC1310 (069, 102)	2:38	2: 30.7	Higher
RC3761 (522, 558)	3:04	3: 0.0	Higher
RC1311-281	3:12	3: 7.1	Higher
RC1207-093	3:49	3: 27.2	Higher
RC1132-199	4:46	4: 31.6	Higher
RC1616-297	5:05	5: 6.2	Higher
RC2586-117	6:65	6: 85.9	Higher
RC1250-274	7:56	7: 62.1	Lower
RC1338 (411,530)	8:41	8: 57.9	Higher
RC3732-078	8:41	8: 62.6	Higher

The RosCOS number is followed by the position of the SNP. The number in parenthesis reflects the position of two SNPs in the same RosCOS

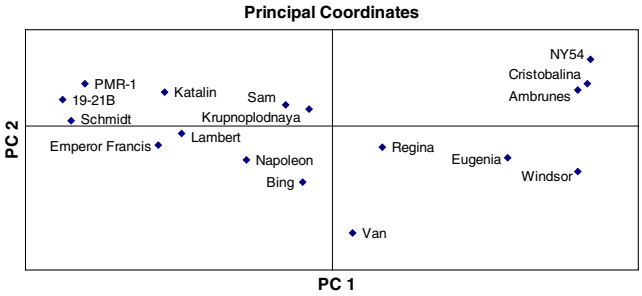

Fig. 1 Locations of 17 sweet cherry founder accessions on the first two principal coordinates determined from the 63 RosCOS totaling 142 haplotypes (50 RosCOS with two haplotypes, 10 RosCOS with three haplotypes, 3 RosCOS with four haplotypes)

Fig. 1); however, the F_{st} value between the two clusters was still significant (F_{st}=0.09, p=0.01), confirming the genetic differentiation between the clusters.

Although the *S* locus genotype was not used in the PCO, the separation of the founders was consistent with the presence or absence of certain *S* alleles. For example, the S_6 allele was only present in those selections clustered in the first quadrant ['Ambrunes' (S_3S_6), 'Cristobalina' (S_2S_6), 'NY54' (S_2S_6)]; the S_9 allele was only present in the second quadrant in three of the five selections ['Krupnoplodnaya' (S_5S_9), 'PMR-1' (S_4S_9), '19-21B' (S_2S_9)]; the S_4 allele was only present in selections in the third quadrant ['Bing' (S_3S_4), 'EF' (S_3S_4), 'Lambert' (S_3S_4), 'Napoleon' (S_3S_4)]; and the S_1 allele was only present in the fourth quadrant in three of the four selections ['Regina' (S_1S_3), 'Van' (S_1S_3), 'Windsor' (S_1S_3)]. The S_3 allele was not found to be specific to a single quadrant.

Linkage map construction

In total, 81 SNPs representing 68 RosCOS were placed on the consensus sweet cherry linkage map that included previously reported SSR, Indel, and *S-RNase* markers (Fig. 2). This new consensus linkage map had a length of 779.4 cM; the average distance between markers was 5.4 cM. The largest gap between markers was present on G1 between CPPCT016 and EMPA001 (46.5 cM, LOD score=1.1). On the interspecific cherry map (*P. avium* × *P. nipponica* map), these two markers mapped 29.7 cM apart (Clarke et al. 2009). The two other largest gaps were on G4 between PMS3 and BPPCT040 (29.5 cM) and on G8 between RosCOS3030 and RosCOS1154 (31.0 cM).

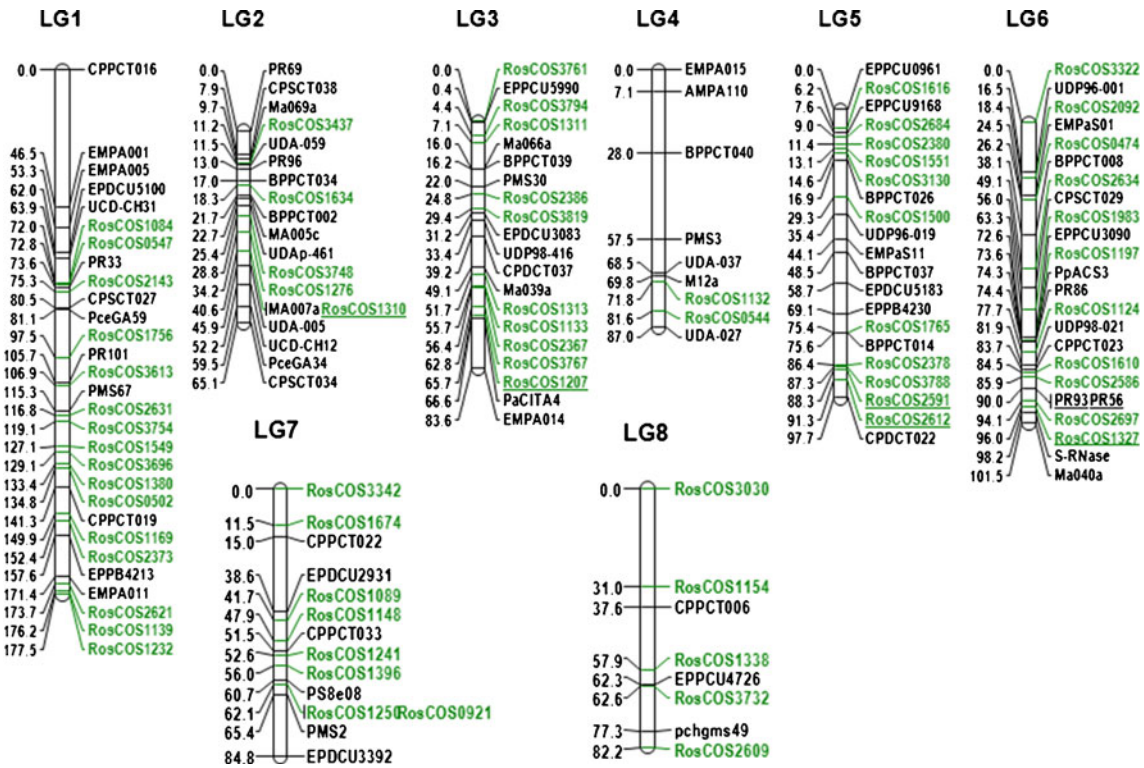

Fig. 2 Consensus sweet cherry linkage map constructed from four populations (Table 2) using RosCOS SNP and SSR markers. *Underline* denotes those markers where the consensus map locations deviated significantly from the map locations determined based on segregation in an individual population. The locations presented for RosCOS1310, RosCOS2591, RosCOS2612, RosCOS1327, PR93, and PR56 are based on segregation in 'Emperor Francis' × 'New York 54' only, and the location presented for RosCOS1207 is based on segregation in 'Lapins' × 'Regina' only. The following four markers were not placed on the linkage map due to severe segregation distortion or inconclusive map position: RosCOS1201, RosCOS0478, RosCOS0546, and RosCOS0483

Positions of the RosCOS markers placed on the cherry linkage map were compared with their previously determined *Prunus* T×E bin map locations. Only 5 out of the 65 RosCOS with known T×E bin locations mapped to different positions on the integrated consensus cherry linkage groups (RosCOS1310 on G2; RosCOS1207 on G3; RosCOS2612 and RosCOS2591 on G5; RosCOS1327 on G6). However, when the positions of these markers were tested in only one population, either 'NY54' × 'EF' or 'Regina' × 'Lapins', they mapped to the expected T×E bin positions. In addition, five of the RosCOS included in the set of 76 had previously been identified as 'orphan' markers (RosCOS1133, 1148, 1765, 2367, and 3322) because they could not be assigned to a T×E bin (Cabrera et al. 2009). By traditional linkage mapping, these five RosCOS were placed on the sweet cherry map. RosCOS1133 and RosCOS2367 mapped together on G3, flanked by two RosCOS markers that were previously mapped to the T×E bin 3:37. RosCOS1765 mapped to G5 flanked by markers that mapped to bins 5:46 and 5:49. RosCOS3322 mapped to the top of G6 where the closest RosCOS marker (RosCOS2092) was located in bin 6:25. Finally, RosCOS1148 mapped to G7 flanked by markers that mapped to bins 7:31 and 7:41.

Discussion

Moderate levels of SNP polymorphism in the sweet cherry breeding germplasm

The narrow genetic base of the sweet cherry germplasm used in North America has been attributed to the intense utilization of only five founding clones (Choi and Kappel 2004) and subsequent use of one genetic source derived from these founders, 'Stella', a self-compatible mutant that is utilized to avoid natural self-incompatibility (Hedrick et al. 1915). A similar narrow genetic base of breeding germplasm was found in the analysis of a broad set of sweet cherry germplasm (Mariette et al. 2010). The moderate level of polymorphism identified in our study (between 17% and 25% of random markers are heterozygous in one parent) is similar to previously reported SSR polymorphism levels in a sweet cherry population (26%) (Olmstead et al. 2008) and slightly lower in a larger set of 76 sweet cherry cultivars (38%; Wunsch and Hormaza 2002). Yet, many alleles contributing to heterozygosity are shared among the 36 selections evaluated and therefore not contributing to the diversity in its germplasm. The exceptions, 'NY54' (a wild accession) and 'Cristobalina' (a Spanish landrace), have the highest numbers of unique alleles. It is conceivable that the unique 'NY54' alleles correspond predominantly to those found in other wild cherry relatives. A genetic diversity analysis of wild, landrace, and modern sweet cherry accessions showed a loss of genetic diversity between the wild and landrace selections compared with modern cultivars, suggestive of a bottleneck due to breeding (Mariette et al. 2010). This finding is consistent with our study in that the wild and landrace Spanish cultivars share many alleles not found in the modern cultivars.

Significant amount of genetic diversity is still conserved despite the lack of population structure

Overall, the mean observed heterozygosity for the 76 RosCOS haplotypes across the genome of 17 founders was higher than expected (H_e=0.39, H_o=0.45), which is consistent with previously reported excess of heterozygosity in landraces (H_e=0.64, H_o=0.66) and modern varieties (H_e= 0.56, H_o=0.59; Mariette et al. 2010). H_o and H_e were used to estimate whether these loci were in Hardy–Weinberg equilibrium. We found that the vast majority, 83% (63) of the RosCOS, were in HWE for the 36 selections. This suggests that most loci and their alleles behave as neutral markers and that a significant amount of genetic diversity is still conserved in cherry despite the limited germplasm used in breeding, a finding supported by others (Mariette et al. 2010). Most of the markers deviating from HWE were located randomly throughout the genome and showed higher than expected heterozygosity. Although we could not find clear evidence of genomic regions that were under balancing selection, there was one region on G8 where two RosCOS deviated from HWE, suggesting that these two RosCOS are under balancing selection.

Using the remaining neutral RosCOS, PCO showed that 'NY54', 'Cristobalina', and 'Ambrunes' separate from the other selections in the breeding set. The F_{st} estimate between these three accessions and the remaining 14 accessions is significant, suggesting that the Spanish accessions offer novel alleles not found in other sweet cherry germplasm. The trend of many unique alleles and population differentiation of the Spanish germplasm is consistent with a previous study where central European and Spanish sweet cherry landraces were shown to contribute the highest proportion of rare alleles compared with northern and Central European germplasm (Wunsch and Hormaza 2002). Interestingly, the genotype distribution of the 17 founders across the four quadrants of the PCO showed a pattern associated with the alleles for the *S* locus, suggesting that this locus might drive genome-wide differentiation together with diversification in cherry, possibly before domestication. This differentiation might still be conserved due to limited breeding efforts and the intense clonal propagation that protect diversity at neutral loci (Balloux et al. 2003).

While sweet cherry germplasm is moderately diverse, we could not detect a clear population structure, implying that the cultivated accessions were behaving as if they were part of a single population with moderate levels of outcrossing among one another. STRUCTURE (Pritchard et al. 2000) and dendrogram cluster analyses to detect population structure within the set of 17 founders, as had been done with other sweet cherry accessions (Guarino et al. 2009; Mariette et al. 2010), did not lead to the identification of the optimum *K* value or significant bootstrap values, suggesting a lack of population structure in the breeding germplasm evaluated in this study. This finding is consistent with results from Wunsch and Hormaza (2002) where multiple alternative dendrograms and low bootstrap values were also obtained when analyzing European and American sweet cherry germplasm. In addition, Mariette et al. (2010) detected lack of population structure for wild cherry material from France and just two populations that were significantly different when considering modern landraces and wild cherries together.

RosCOS SNP markers significantly improved the sweet cherry genetic map

Construction of a consensus map is important for pedigree-based QTL mapping approaches that simultaneously use data from multiple populations (Bink et al. 2002, 2008). The 779.4-cM length of the sweet cherry consensus map constructed with the additional 68 RosCOS was largely consistent with the two most complete sweet cherry linkage maps reported. These include the 711-cM 'EF' parental linkage map (Olmstead et al. 2008) and the 680-cM interspecific 'Napoleon' (*P. avium*) × *P. nipponica* consensus map (Clarke et al. 2009). The increase in our map's genetic length resulted from the addition of markers at the end of a linkage group and increased map distances between markers. In addition, we describe the linkage map locations for five RosCOS for which a *Prunus* bin map location could not be identified. There were a few RosCOS that could not be placed with high confidence on the integrated map. Those RosCOS (five in total) were placed on the linkage map using one of the two larger mapping populations. It is likely that multiple factors contributed to the consensus map positions not matching the map positions calculated based on a single population alone. These include different sets of markers segregating in different populations and differing population sizes.

The addition of the subset of 68 RosCOS markers to a consensus sweet cherry linkage map significantly increased the overall marker density. Most notably, the marker density on sweet cherry G8 was substantially improved and now comparable to that of the other seven linkage groups. This is in contrast to peach in which none of the 181 molecular markers was assigned to G8 (Dirlewanger et al. 2006). Similarly, in a *P. persica* × *P. davidiana* BC_2 population, only one G8 marker was polymorphic (Quilot et al. 2004). Therefore, the subset of RosCOS selected for the construction of the genetic linkage map cover the genome of sweet cherry better than any other type of marker used in prior studies.

Finally, this study represents the first example of the use of a high-throughput SNP genotyping assay in cherry, therefore facilitating rapid linkage map construction. Specifically, a GoldenGate Custom Genotyping Array was designed based on our knowledge of the RosCOS cherry SNPs and 60 bp of DNA sequences flanking the SNP positions. The increasing quantity of DNA sequence forthcoming for a wide array of cherry selections will allow for the identification of a vast number of SNPs. This will significantly increase the ability to design high-density genotyping arrays using the GoldenGate® Custom Genotyping Array.

Conclusion

We identified a set of 282 RosCOS genome-wide markers that are polymorphic in sweet cherry breeding germplasm. These markers represent a valuable resource to anchor the sweet cherry genome to that of apple, peach, and *Fragaria* and enable high-throughput genome scans in sweet cherry for QTL discovery.

Acknowledgments This research was funded by USDA-NRI NIFA 2008-02259 and 2005-00743 and the Washington Tree Fruit Research Commission. We thank Dr. Andrew P. Michel (Ohio State University) for advice on the population genetic analyses and Dr. James Olmstead (University of Florida) for providing leaf tissue of the WSU selections.

References

Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. Theor Appl Genet 119:507–517

Antao T, Lopes A, Lopes R, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{st}-outlier method. BMC Bioinform 9:323

Balloux F, Lehmann L, de Meeus T (2003) The population genetics of clonal and partially clonal diploids. Genetics 164:1635–1644

Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS (2007) SNP discovery via 454 transcriptome sequencing. Plant J 51:910–918

Bink M, Uimari P, Sillanpää M, Janss L, Jansen R (2002) Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theor Appl Genet 104:751–762

Bink M, Boer M, ter Braak C, Jansen J, Voorrips R, van de Weg W (2008) Bayesian analysis of complex traits in pedigreed plant populations. Euphytica 161:85–96

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Cabrera A, Kozik A, Howad W, Arus P, Iezzoni A, Van der Knaap E (2009) Development and bin mapping of a Rosaceae Conserved Ortholog Set (COS) of markers. BMC Genomics 10:562

Choi C, Kappel F (2004) Inbreeding, coancestry, and founding clones of sweet cherries from North America. J Am Soc Hortic Sci 129:535–543

Clarke J, Sargent D, Bošković R, Belaj A, Tobutt K (2009) A cherry map from the inter-specific cross *Prunus avium* 'Napoleon' × *P. nipponica* based on microsatellite, gene-specific and isoenzyme markers. Tree Genet Genomes 5:41–51

Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howad W, Arus P (2004) Comparative mapping and marker-assisted selection in Rosaceae fruit crops. Proc Natl Acad Sci 101:9891–9896

Dirlewanger E, Cosson P, Boudehri K, Renaud C, Capdeville G, Tauzin Y, Laigret F, Moing A (2006) Development of a second-generation genetic linkage map for peach [*Prunus persica* (L.) Batsch] and characterization of morphological traits affecting flower and fruit. Tree Genet Genomes 3:1–13

Guarino C, Santoro S, De Simone L, Cipriani G (2009) *Prunus avium*: nuclear DNA study in wild populations and sweet cherry cultivars. NRC Res Press 52:320–337

Hedrick UP, Howe GH, Taylor OM, Tubergen CB, Wellington R (1915) The cherries of New York report of the New York agricultural experiment station for the year 1914. J.B. Lyon Co., Albany

Hyten DL, Song Q, Choi I, Yoon M, Specht JE, Matukumalli LK, Nelson RL, Shoemaker RC, Young ND, Cregan PB (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. Theor Appl Genet 116:945–952

Iezzoni A, Schmidt A, Albertini A (1991) Cherries (*Prunus* spp.). In: Moore JN (ed) Genetic resources of temperate fruit and nut. ISHS, Wageningen, the Netherlands, pp 111–176

Illa E, Sargent DJ, Girona EL, Bushakra J, Cestaro A, Crowhurst R, Pindo M, Cabrera A, van der Knaap E, Iezzoni A, Gardiner S, Velasco R, Arus P, Chagne 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

Jung S, Staton M, Lee T, Blenda A, Svancara R, Abbott A, Main D (2008) GDR (Genome Database for Rosaceae): integrated web-database for Rosaceae genomics and genetics data. Nucleic Acids Res 36:D1034–D1040

Kosambi D (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Mariette S, Tavaud M, Arunyawat U, Capdeville G, Millan M, Salin F (2010) Population structure and genetic bottleneck in sweet cherry estimated with SSRs and the gametophytic self-incompatibility locus. BMC Genet 11:77

McNally KL, Childs KL, Bohnert R, Davidson RM, Zhao K, Ulat VJ, Zeller G, Clark RM, Hoen DR, Bureau TE, Stokowski R, Ballinger DG, Frazer KA, Cox DR, Padhukasahasram B, Bustamante CD, Weigel D, Mackill DJ, Bruskiewich RM, Rätsch G, Buell CR, Leung H, Leach JE (2009) Genome wide SNP variation reveals relationships among landraces and modern varieties of rice. Proc Natl Acad Sci U S A 106:12273–12278

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Olmstead JW, Sebolt AM, Cabrera A, Sooriyapathirana SS, Hammar S, Iriarte G, Wang D, Chen CY, van der Knaap E, Iezzoni A (2008) Construction of an intra-specific sweet cherry (*Prunus avium* L.) genetic linkage map and synteny analysis with the *Prunus* reference map. Tree Genet Genomes 4:897–910

Peakall ROD, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Blackwell Science, Oxford, pp 288–295

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Quilot B, Wu BH, Kervella J, Genard M, Foulongne M, Moreau K (2004) QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species *P. davidiana*. Theor Appl Genet 109:884–897

Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP, Burns P, Davis TM, Slovin JP, Bassil N, Hellens RP, Evans C, Harkins T, Kodira C, Desany B, Crasta OR, Jensen RV, Allan AC, Michael TP, Setubal JC, Celton J-M, Rees DJG, Williams KP, Holt SH, Rojas JJR, Chatterjee M, Liu B, Silva H, Meisel L, Adato A, Filichkin SA, Troggio M, Viola R, Ashman T-L, Wang H, Dharmawardhana P, Elser J, Raja R, Priest HD, Bryant DW, Fox SE, Givan SA, Wilhelm LJ, Naithani S, Christoffels A, Salama DY, Carter J, Girona EL, Zdepski A, Wang W, Kerstetter RA, Schwab W, Korban SS, Davik J, Monfort A, Denoyes-Rothan B, Arus P, Mittler R, Flinn B, Aharoni A, Bennetzen JL, Salzberg SL, Dickerman AW, Velasco R, Borodovsky M, Veilleux RE, Folta KM (2011) The genome of woodland strawberry (*Fragaria vesca*). Nat Genet 43:109–116

Stockinger EJ, Mulinix CA, Long CM, Brettin TS, Iezzoni A (1996) A linkage map of sweet cherry based on RAPD analysis of a microspore-derived callus culture population. J Hered 87:214–218

Syvanen A-C (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nat Rev Genet 2:930–942

Van Ooijen JW (2006) JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Watkins R (1976) Cherry, plum, peach, apricot and almond. *Prunus* ssp. In: Simmonds NW (ed) Evolution of crop plants. Longman, London

Wu X, Ren C, Joshi T, Vuong T, Xu D, Nguyen H (2010) SNP discovery by high-throughput sequencing in soybean. BMC Genomics 11:469

Wunsch A, Hormaza JI (2002) Molecular characterisation of sweet cherry (*Prunus avium* L.) genotypes using peach [*Prunus persica* (L.) Batsch] SSR sequences. Heredity 89:56–63

Yan J, Yang X, Shah T, Sanchez-Villeda H, Li J, Warburton M, Zhou Y, Crouch JH, Xu Y (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breed 25:441–451

Yang W, Bai X, Kabelka E, Eaton C, Kamoun S, van der Knaap E, Francis D (2004) Discovery of single nucleotide polymorphisms in *Lycopersicon esculentum* by computer aided analysis of expressed sequence tags. Mol Breed 14:21–34

Yeh F, Boyle TB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg J Bot 129:157