

Tree Genetics & Genomes (2014) 10:627–639
DOI 10.1007/s11295-014-0709-1

ORIGINAL PAPER

A consensus ‘Honeycrisp’ apple (*Malus × domestica*) genetic linkage map from three full-sib progeny populations

Matthew D. Clark · Cari A. Schmitz ·
Umesh R. Rosyara · James J. Luby · James M. Bradeen

Received: 3 May 2013 / Revised: 11 February 2014 / Accepted: 18 February 2014 / Published online: 11 March 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract The apple cultivar Honeycrisp is emerging in North American markets due to its outstanding eating quality. A set of three ‘Honeycrisp’ progeny populations from the University of Minnesota apple breeding program were utilized to construct parental and consensus ‘Honeycrisp’ linkage maps to enable marker-assisted breeding. Two populations were segregated for fruit texture traits and a third was of interest in examining disease resistance. All available individuals were genotyped with the International RosBREED SNP Consortium (IRSC) apple 8K SNP array v1, for a total of 318 progeny individuals. Three unique ‘Honeycrisp’ parental maps (‘Honeycrisp’ × ‘Monark,’ ‘Honeycrisp’ × ‘Gala,’ and ‘Honeycrisp’ × MN1764) were developed, consisting of 1,018, 1,042, and 1,041 single-nucleotide polymorphism (SNP) markers, respectively. Among all three ‘Honeycrisp’ parental maps, 951 SNP markers were in common. Combining these maps with the MergeMap tool, a consensus ‘Honeycrisp’ linkage map with 1,091 SNP markers was developed with an average distance of 1.36 cM between

consecutive markers. The ‘Honeycrisp’ consensus map was largely in agreement with the physical position of markers in the ‘Golden Delicious’ reference genome sequence (v1.0, as of February 2013). The consensus linkage map is informative for an elite cultivar that is being utilized in breeding programs worldwide for its superb fruit quality traits.

Keywords Genetic map · *Malus × domestica* · Linkage map · JoinMap

Introduction

‘Honeycrisp’ is an emerging apple cultivar with increased importance in North America due to its outstanding flavor and textural traits (Hoover et al. 2000; Luby and Bedford 1992; Tong et al. 1999). Although it is prone to some storage disorders, ‘Honeycrisp’ can maintain crispness for 6–9 months in storage (Luby and Bedford 1992; Tong et al. 1999). ‘Honeycrisp’ has been shown to exhibit field resistance to foliar apple scab infection when grown under organic disease management practices (Berkett et al. 2009), a characteristic important for growers who may be able to reduce fungicide inputs in their orchards. For the apple breeder, using ‘Honeycrisp’ as a parent offers the genetic background for superb fruit quality and disease resistance traits that should be leveraged in breeding (McKay et al. 2011). Identifying the marker-locus-trait associations in ‘Honeycrisp’ progeny will give plant breeders additional tools for marker-assisted breeding (MAB), in developing new cultivars. The development of a ‘Honeycrisp’ linkage map will add to the toolbox available to apple breeders and geneticists.

Fruit quality traits are among the most important characteristics evaluated and the most crucial component of a breeding project as the fruit are the saleable product driven by consumer demand. These quality traits include texture (King

Communicated by E. Dirlewanger

Matthew D. Clark and Cari A. Schmitz contributed equally in the development of this manuscript.

Electronic supplementary material The online version of this article (doi:10.1007/s11295-014-0709-1) contains supplementary material, which is available to authorized users.

M. D. Clark · C. A. Schmitz · J. J. Luby (✉)
Department of Horticultural Science, University of Minnesota, St. Paul, MN, USA
e-mail: lubyx001@umn.edu

U. R. Rosyara
Department of Horticulture, Michigan State University, East Lansing, MI, USA

J. M. Bradeen
Department of Plant Pathology, University of Minnesota, St. Paul, MN, USA

et al. 2000) and its components firmness (Pre-Aymard et al. 2005), juiciness, and crispness. The development of scab-resistant cultivars faces genetic challenges (linkage drag) and marketing challenges. Any new cultivar must be an outstanding alternative or replacement to an existing, consumer-recognized cultivar. Consumer familiarity with a cultivar and previous purchase of a particular apple cultivar rank as top determinants in selecting fruit to purchase (Kelley et al. 2010).

Cultivar development is hindered by long juvenility and self-incompatibility which constrain crossing decisions. The development of a single cultivar can take as long as 20–25 years. Due to the large size of mature trees, orchard space is limiting, and the maintenance of individual trees from juvenility to fruit-bearing age is expensive and requires a large amount of space. The development of genetic markers to screen important traits at the seedling stage and for parental selection will result in the enrichment of the target trait among seedlings that are grown to maturity for phenotypic evaluation. Accurate phenotyping of the traits of interest predates detection of robust marker-trait associations to enable MAB (Luby and Shaw 2001). The traits must be well defined and also objectively measurable.

Genetic studies in apple can be challenging due to its highly heterozygous genome, high levels of inbreeding depression, and self-incompatibility (Lawson et al. 1995). Linkage maps for self-incompatible species, including apple, are created using the two-way pseudo-testcross method within a single progeny (Grattapaglia and Sederoff 1994). In this approach, a map for the first parent is made using markers heterozygous in the first parent and homozygous in the second, and conversely, the second parental map consists of markers homozygous in the first parent and heterozygous in the second. These maps can then be integrated using markers heterozygous in both parents, creating a population map. Genetic mapping using biparental mapping populations is common in apple genetics, especially in developing molecular markers for monogenic traits such as disease resistance (Schenato et al. 2008; Tartarini and Sansavini 2003). A number of linkage maps have been developed and were used to detect quantitative trait loci (QTLs) and map genes for a range of important traits including disease resistance (*Vf* for apple scab) (Gianfranceschi et al. 1996), acidity (*Ma*) (Maliepaard et al. 1998), and growth habit and developmental traits (Lawson et al. 1995). Mapping populations typically use parents divergent for an important trait. This approach is illustrated by recently published microsatellite and single-nucleotide polymorphism (SNP) maps in *Malus* species (Antanaviciute et al. 2012; Fernández-Fernández et al. 2012; Wang et al. 2012). The advantage of a consensus ‘Honeycrisp’ linkage map, constructed across different populations, is that marker alleles in this cultivar would yield a novel map that would be informative for MAB in breeding programs using this cultivar.

Recently, the apple genome was sequenced (Velasco et al. 2010), and additional supporting tools have been developed, including a physical map, BLAST search engine, and genome browser available on Genome Database for Rosaceae (GDR, <http://www.rosaceae.org>) (Jung et al. 2008). High-throughput SNP genotyping allows for efficient genotyping of large numbers of individuals or populations with a relatively low cost per marker. An 8K SNP array v1 was developed by the International RosBREED SNP Consortium (IRSC). Based on the Illumina Infinium platform, the BeadChip is a small, portable, highly repeatable assay that allows for rapid scoring of individuals, providing even coverage throughout the apple genome, including SNPs within putative expressed genes (Chagné et al. 2012). The array was designed using a clustering strategy with a cluster of 4–10 closely positioned SNPs spaced at 1-cM intervals between clusters (Chagné et al. 2012). Clustered markers should provide local information for diverse apple populations representing unique haplotypes, and recombination is rarely expected within a cluster. The result is a SNP array that is not population dependent and is applicable across cultivars and progeny populations (Micheletti et al. 2011).

Linkage maps with dense marker coverage and with markers evenly spaced across the genome are ideal for QTL analysis. Increased density and coverage of markers helps increase power and precision of QTL analysis, thereby helping in gene discovery. Khan et al. (2012) created a highly saturated map of apple by merging five biparental maps by use of single sequence repeat (SSR) and SNP markers shared among the linkage maps. The construction and analysis of genetic linkage maps provide support for the placement of molecular markers into the correct order and position. The correct order and position is very important to precisely locate QTLs. A constraint in map construction is marker checking to validate and correct automated SNP genotyping calls, especially in cases of expected paralogous regions from local or whole-genome duplication events, which are common in plant genomes, including those of *Malus* species (Velasco et al. 2010). A comparative analysis of maps from different populations and development of a consensus map help to determine whether large genome rearrangements are present and to establish consensus order and positions of mapped markers.

The objective of this study was to develop a high-density SNP consensus linkage map for ‘Honeycrisp’ utilizing several ‘Honeycrisp’ full-sib progeny populations that segregate for fruit quality and apple scab resistance. This map will provide the framework for future genetic studies in ‘Honeycrisp’-specific progeny to identify marker-locus-trait associations for important fruit quality and disease resistance traits, thus enabling MAB. It will also provide additional support in map construction (marker order and position) for pedigree-based analysis and in resolving potential issues in the apple physical

map v1 (GDR database: <http://www.rosaceae.org>) (Jung et al. 2008).

Materials and methods

Plant materials

A portion of the genotypic data in this study was produced as part of the RosBREED crop reference set (rosbreed.org). The corresponding apple genotypes, hereafter referred to as “RosBREED samples,” included the parents (‘Honeycrisp,’ ‘Gala,’ ‘Monark,’ MN1764, and 21 individuals of the ‘Honeycrisp’ × ‘Monark’ population (described below). The majority of individuals described in this paper were genotyped independently of the RosBREED crop reference set, and these individuals are hereafter referred to as “UMN samples.”

The UMN samples comprise three full-sib families sharing ‘Honeycrisp’ as a common parent and were utilized in the development of the ‘Honeycrisp’ consensus map. Two ad hoc populations [‘Honeycrisp’ × MN 1764 ($n=130$) and ‘Honeycrisp’ × ‘Monark’ ($n=88$)] were selected from breeding populations growing at the University of Minnesota Horticultural Research Center (Excelsior and Chanhassen, MN) that were developed from crosses made in 1992–1998. These ad hoc populations have been described previously by McKay et al. (2011). A third population was created in 2010 from a cross of ‘Honeycrisp’ × ‘Twin Bee Gala’ ($n=128$, this population is referred to as ‘Honeycrisp’ × ‘Gala’ throughout) and grown in greenhouses at the University of Minnesota-Twin Cities (St. Paul, MN).

DNA extraction protocol

For RosBREED samples, stems with newly expanding leaf tissue were collected in the field in 2010 and 2011 and placed in labeled plastic bags on ice. Into a cluster tube (Corning, Tewksbury, MA), 30 to 50 mg of leaf tissue was later harvested. These RosBREED tissue samples were frozen in liquid nitrogen and held at -80°C until DNA extraction. For the UMN samples, newly expanding or youngest leaves were collected from individual trees in 2012, frozen at -80°C , lyophilized, and held at -80°C until DNA extraction. Approximately 10 to 15 mg of lyophilized leaf tissue from each sample was placed into a cluster tube.

The day of DNA extraction, leaf tissue was homogenized by grinding lyophilized (UMN) or frozen (RosBREED) samples. A 4-mm stainless steel bead (McGuire Bearing Company, Salem, OR) was added to each cluster tube, caps were applied, and the 96-tube rack was submerged in liquid nitrogen. The rack was then placed into a Retsch MM301 Mixer Mill (Retch, Haan, Germany) and shaken for 30 s. Sample racks were re-submerged in liquid nitrogen and

shaken two additional times, disrupting the leaf tissue into a fine powder. The homogenized RosBREED and UMN tissue was stored at -80°C until 10 min prior to extraction.

Extraction was conducted using the E-Z 96[®] Plant DNA Kit (Omega Bio-Tek, Norcross, GA) with modifications (Gilmore et al. 2011). Modifications included using SP1 solution equilibrated to 65°C in a water bath. The supernatant (580 μL) for each sample was transferred in one step to a new cluster tube containing 10 μL of RNase solution (2.5 μL RNase and 7.4 μL Tris-EDTA (TE) pH 8). After the drying step, DNA was eluted in 100 μL elution buffer, and samples were stored at 4°C and quantitated within seven days or stored at -20°C . DNA samples were quantified using the Quant-iT[™] PicoGreen[®] dsDNA Assay Kit (Invitrogen, Eugene, OR) and a VICTOR multiplate reader (PerkinElmer Inc., San Jose, CA). Samples with DNA concentrations of $>100\text{ ng}/\mu\text{L}$ were diluted with the addition of an equal volume of TE to achieve concentrations between 50 and $100\text{ ng}/\mu\text{L}$. Fifteen microliters of each DNA sample was aliquoted into PCR plates comprising 96–0.2-mL wells, sealed with adhesive aluminum foil seals, and shipped to the genotyping facility on dry ice.

Marker data generation and analysis

The UMN DNA samples were submitted to the SNP Genotyping Facility at Michigan State University (East Lansing, MI). The RosBREED samples were analyzed at the University of Western Cape, South Africa. Using previously published protocols (Illumina 2006), samples were hybridized onto the IRSC apple 8K SNP array v1 (Chagné et al. 2012) following a whole-genome amplification reaction. BeadChips were imaged by the iScan system and converted into intensity data. The intensity data from the two data sets were combined for analysis and interpretation in the genotyping module of GenomeStudio for genotype clustering (Illumina Inc. 2010a).

The iScan data from both genotyping facilities were loaded into a single project file for data analysis. SNP genotype scoring employed the genotyping module of GenomeStudio (Illumina Inc. 2010b) software version v2010.3.0.30128. The software normalizes the intensity values across BeadChips to allow for uniformity in allele calling. To ensure high-quality reads, stringent initial parameters were set as follows: GenTrain >0.60 and AB Freq from 0.45 to 0.55. The SNPs were clustered by marker locus using the clustering algorithm Gentrain2 (Illumina Inc. 2010c), and all SNPs were visually examined for an expected maximum of three clusters (AA, AB, and BB) and then classified as failed, monomorphic, or polymorphic.

Automated allele calling with visual checking to confirm clustering of individuals into appropriate classes was utilized. Manual clustering was performed for some markers, when automated clustering was not satisfactory. Markers with more

than three distinctly spaced clusters, presumably the result of annealing to more than one genomic region (i.e., paralogs), were excluded. The ‘Honeycrisp’ × MN1764 population, with the largest number of progeny, was utilized to select nearly 2,000 high-quality markers for the development of a saturated linkage map as suggested by Micheletti et al. (2011). A preliminary map was developed to evaluate genome coverage and relative positions in comparison to the physical map (Clark et al. 2013). For the preliminary map, the default settings of the maximum likelihood method in JoinMap 4.1 (Kyazma B.V., Wageningen, Netherlands) (Van Ooijen 2006) were used to map 1,952 SNP markers. Marker grouping during map construction utilized a published SNP map (Antanaviciute et al. 2012). These ~2,000 markers were then scored for the ‘Honeycrisp’ × ‘Gala’ and ‘Honeycrisp’ × ‘Monark’ populations.

Marker loci at which missing parental genotypes could not be positively determined based on progeny segregation in two or more families were removed. Markers with >10 % missing data were eliminated. Progeny that did not conform to the parental genotypes was removed, as they were expected to be outcrosses, non-progeny, or contaminated samples. The identity by descent (IBD) analysis program within FlexQTL (Bink et al. 2008) was used to identify miscalled alleles and impute parental genotypes using the ‘Golden Delicious’ physical map positions. This tool allowed for the aggressive detection of errors (missing markers, null alleles, other anomalies), but required additional manual correction or imputation of parental genotypic scores based on the progeny SNP calls.

Linkage mapping

The codominant SNP markers from each outbreeding, full-sib population were coded for linkage map construction according to JoinMap 4.1 conventions as heterozygous in either first or second parent (<nn × np>, <lm × ll>) or both parents (<hk × hk>) (Van Ooijen 2011). The three populations were mapped separately using the high-quality polymorphic SNPs (‘Honeycrisp’ × ‘Monark,’ 1,428; ‘Honeycrisp’ × ‘Gala,’ 1,421; and ‘Honeycrisp’ × MN1764, 1,885). The initial grouping procedure in JoinMap was completed using the published M432 progeny linkage map (Antanaviciute et al. 2012), resulting in a large proportion of the called SNPs remaining ungrouped. The strongest crosslink values (SCLs) were applied repeatedly using restrictively lower values in an iterative process to assign ungrouped loci to the correct linkage group (Van Ooijen 2006). Markers with suspected linkage (recombination frequency estimate >0.6) were removed before mapping. Then, map order was calculated using the maximum

likelihood option which calculates both parental maps and an integrated map. For this study, only the single parent ‘Honeycrisp’ map from each population was used for the construction of the consensus map. Maps were then constructed de novo using the independence logarithm of odds (LOD) calculation for grouping based on pairwise recombination frequencies between loci. The significant LOD threshold for grouping was LOD=3 for ‘Honeycrisp’ × ‘Gala’ and ‘Honeycrisp’ × ‘Monark’ populations and LOD=4 for the ‘Honeycrisp’ × MN1764 population.

Each of the three ‘Honeycrisp’ maps based on the M432 grouping and the corresponding progeny genotypic data sets was assembled for analysis in FlexQTL (Bink et al. 2008) to detect observed double recombinants versus expected double recombinants provided the newly constructed linkage map. The FlexQTL program calculated observed double recombinants (oDRs) minus the expected double recombinants (eDRs) for the two parents at each marker position. This helped to identify markers that had high genotyping error rates or that are misplaced by the mapping algorithm. Markers with oDRs–eDRs≥0.03 were removed from the subsequent round of JoinMap mapping, eliminating 100 (‘Honeycrisp’ × MN1764), 80 (‘Honeycrisp’ × ‘Gala’), and 105 (‘Honeycrisp’ × ‘Monark’) spurious markers. After two rounds of mapping and removal of suspect markers identified with FlexQTL, maps were inspected for large gaps (> 15 cM) and markers creating unusual large gaps were identified. Markers creating unusually large gaps at linkage group ends were referred to as “lone wolf” markers as the gaps suggest poor linkage to the marker group. If a large gap existed at the end of a linkage group (LG) in a single population map and the causative marker was not found in the corresponding LG in either of the other two maps, it was removed. After marker removal from any map, the map was recalculated in JoinMap 4.1. The resulting three ‘Honeycrisp’ maps were combined into a consensus map with the MergeMap (2012 version) software tool (Wu et al. 2011). Maps were weighted based on population size (‘Honeycrisp’ × ‘Monark,’ 0.255; ‘Honeycrisp’ × ‘Gala,’ 0.393; ‘Honeycrisp’ × MN1764, 0.352).

The consensus ‘Honeycrisp’ linkage map was compared to the available physical map of apple. SNP map positions for each of the 17 linkage groups were plotted against marker positions in the respective pseudo-chromosomes of the ‘Golden Delicious’ genome sequence with R v2.15.1 (R Core Team 2012). Base pair positions were those of the mapped IRSC apple markers, and these data are available at the Genome Database for Rosaceae (<http://www.rosaceae.org>; accessed 28 Feb 2013). Each marker included in the consensus ‘Honeycrisp’ map was checked for significant

segregation distortion (χ^2 , $p < 0.005$) in each of the three families using JoinMap.

Results

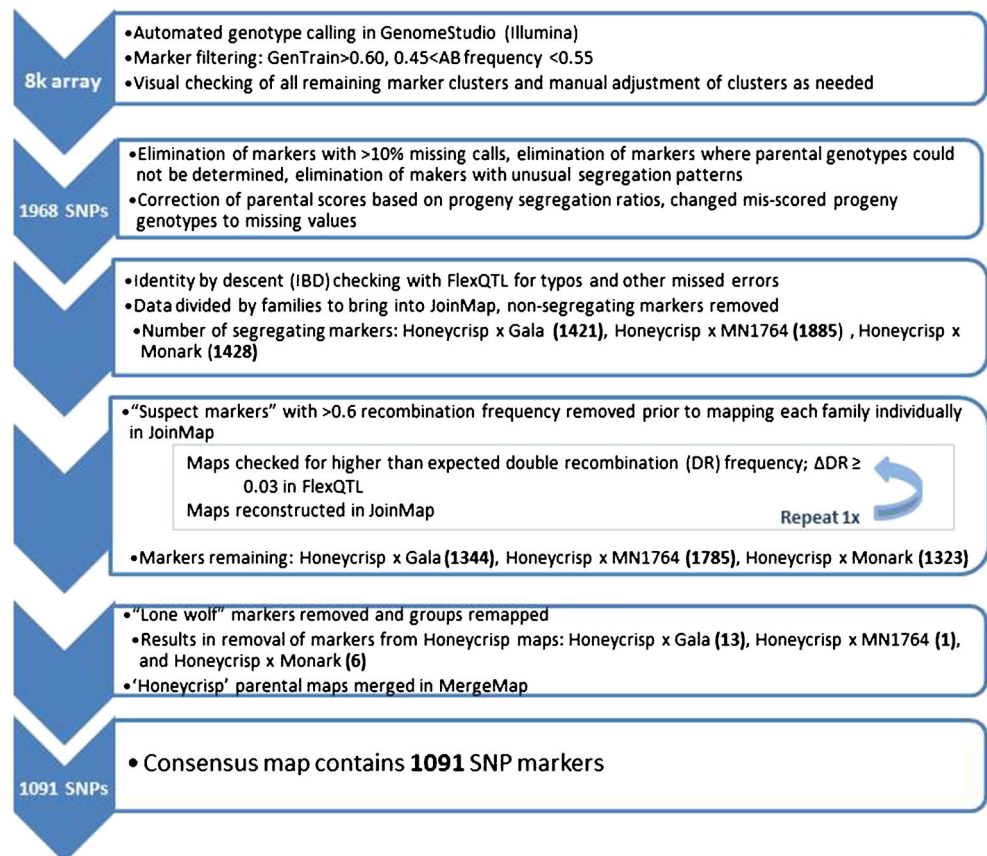
A diagram describing the work flow and remaining high-quality SNP markers at each phase are shown in Fig. 1. For each of the populations, individuals were removed whose genotype did not conform to the parental genotypes, suggesting it was an outcross, non-progeny, or contaminated sample ('Monark,' 7; 'Gala,' 3; and MN1764, 18). Three 'Honeycrisp' linkage maps were constructed which varied due to different high-quality markers in each population. Figure 2 details heterozygosity for each parent ('Honeycrisp' is heterozygous at each marker position) along the consensus map. MN1764 had the lowest proportion of heterozygous markers (34.5 %) in the corresponding 'Honeycrisp' parental map, and MN1764 additionally had the lowest proportion of heterozygous markers in the consensus map (32.9 %, Table 1). The highest proportion of heterozygous markers was in the

'Honeycrisp' × 'Monark' population with 48.4 % in the parental map and 45.2 % in the consensus map. The 'Honeycrisp' × 'Gala' population had 33.0 % heterozygous markers in the parental map and 31.0 % in the consensus map.

Parental linkage maps

Three 'Honeycrisp' (single parent) linkage maps were constructed from segregating populations with SNP markers using grouping methods in JoinMap that utilized the previously published M432 linkage map (Antanaviciute et al. 2012) and a de novo grouping algorithm. Both methods produced identical assignments of SNPs to linkage groups, and the M432 grouped maps were selected for the remaining mapping procedures. The final 'Honeycrisp' linkage map for each population is visualized with homology among maps in Fig. 3 (markers and positions provided in Table S1). The maps each contain 17 linkage groups representing the 17 known chromosomes that comprise the *Malus* × *domestica* genome. The shortest map was 1,097.55 cM and was constructed from the 'Honeycrisp' × 'Gala' population from 1,042 markers with an average spacing of 1.05 cM between markers. The next

Fig. 1 Work flow describing the mapping process including the number of SNP markers retained at each stage



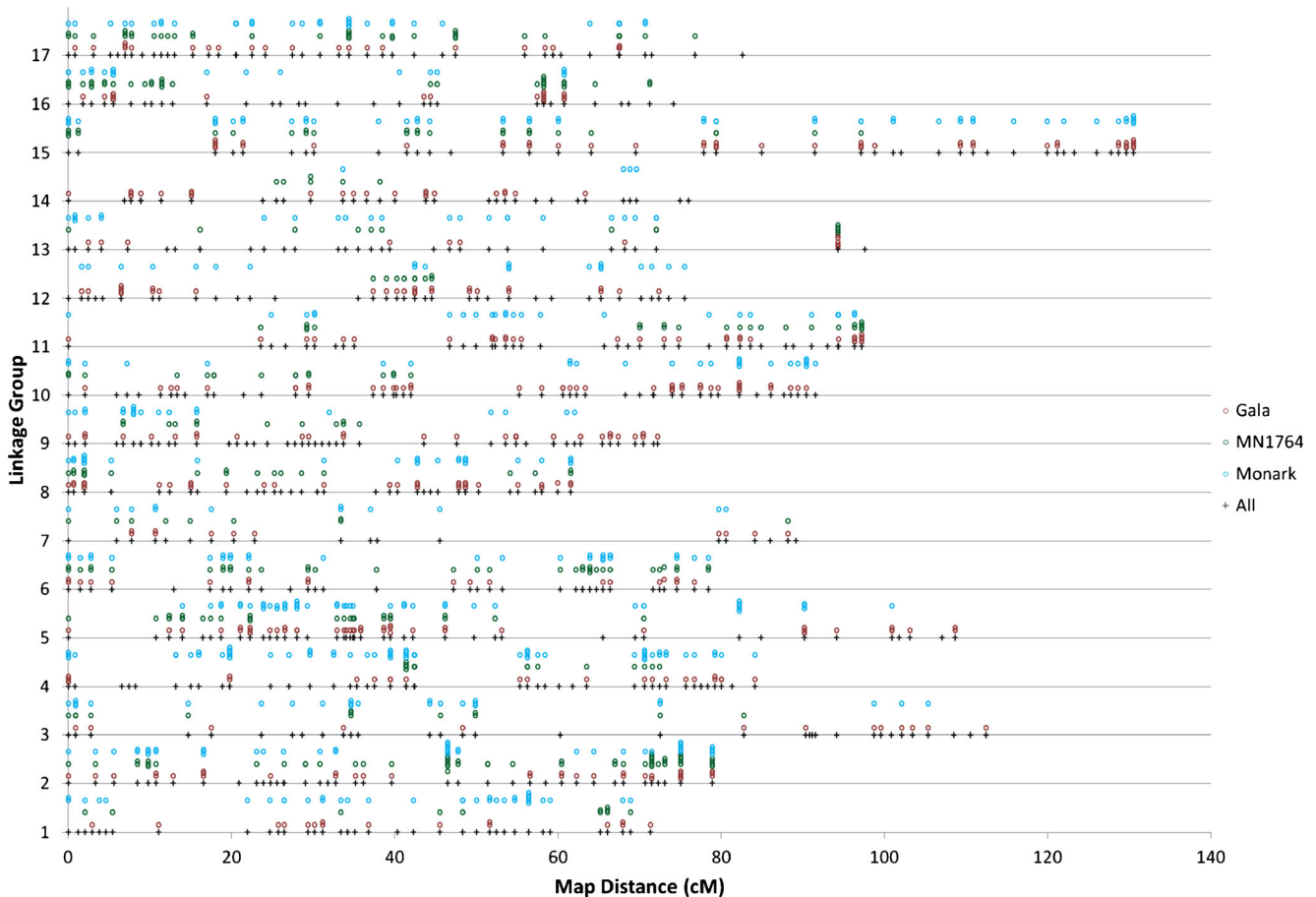


Fig. 2 Homozygosity plot indicating polymorphism in the parents from the three mapping populations (‘Honeycrisp’ × ‘Gala,’ ‘Honeycrisp’ × MN1764, and ‘Honeycrisp’ × ‘Monark’) plotted on the consensus map

(X-axis). ‘Honeycrisp’ is heterozygous at all loci. Multiple *open circles* at a locus indicate more than one SNP marker mapped to that locus for the given parent

longest map was 1340.20 cM and was constructed from the ‘Honeycrisp’ × ‘Monark’ population with 1,018 SNP markers and an average marker spacing of 1.32 cM between markers (Table 2). The ‘Honeycrisp’ × MN1764 map was 1,350.29 cM in length and was constructed from 1,041 SNP markers, with an average marker spacing of 1.30 cM. The marker coverage for the linkage groups ranged from 23 markers [LG7 (‘Honeycrisp’ × MN1764)] to 88 markers [LG4

(‘Honeycrisp’ × MN1764 and ‘Honeycrisp’ × ‘Monark’)]. The maximum gap size for any linkage group ranged from 5.13 cM (LG9 ‘Honeycrisp’ × ‘Gala’) to 129.64 cM (LG17 ‘Honeycrisp’ × ‘Monark’). The “lone wolf” marker on LG17 (refer to Fig. 3) was retained as it met the parameters described above and was resolved in the consensus map.

Consensus linkage map

The three ‘Honeycrisp’ linkage maps were merged to create one consensus linkage map comprising markers segregating in one or more of the ‘Honeycrisp’ mapping populations (Fig. 4). The consensus map was constructed using 1,091 SNP markers (13.9 % of the IRSC 8K SNP array v1, Table 2, Table S2). Figure 5 details the 951 markers in common across all three populations and the 140 SNP markers segregating in only one or two populations. The consensus map is 1,481.72 cM with an average distance of 1.36 cM between markers (Table 2). The sizes of the linkage groups range from 61.58 cM (LG8) to 130.48 cM (LG15). The largest gap in the consensus map was 34.21 cM on LG7.

Table 1 Number and percentage of heterozygous markers of the non-‘Honeycrisp’ parent for its corresponding parental map and in the consensus map for three mapping populations (‘Honeycrisp’ × ‘Gala,’ ‘Honeycrisp’ × MN1764, and ‘Honeycrisp’ × ‘Monark’). ‘Honeycrisp’ is heterozygous at all mapped loci

Parent	Heterozygous markers in parent	Markers in parental map	Proportion of parental map (%)	Proportion of consensus map (%) (1,091 markers)
Gala	448	1,042	43.0	41.1
MN1764	359	1,041	34.5	32.9
Monark	493	1,018	48.4	45.2

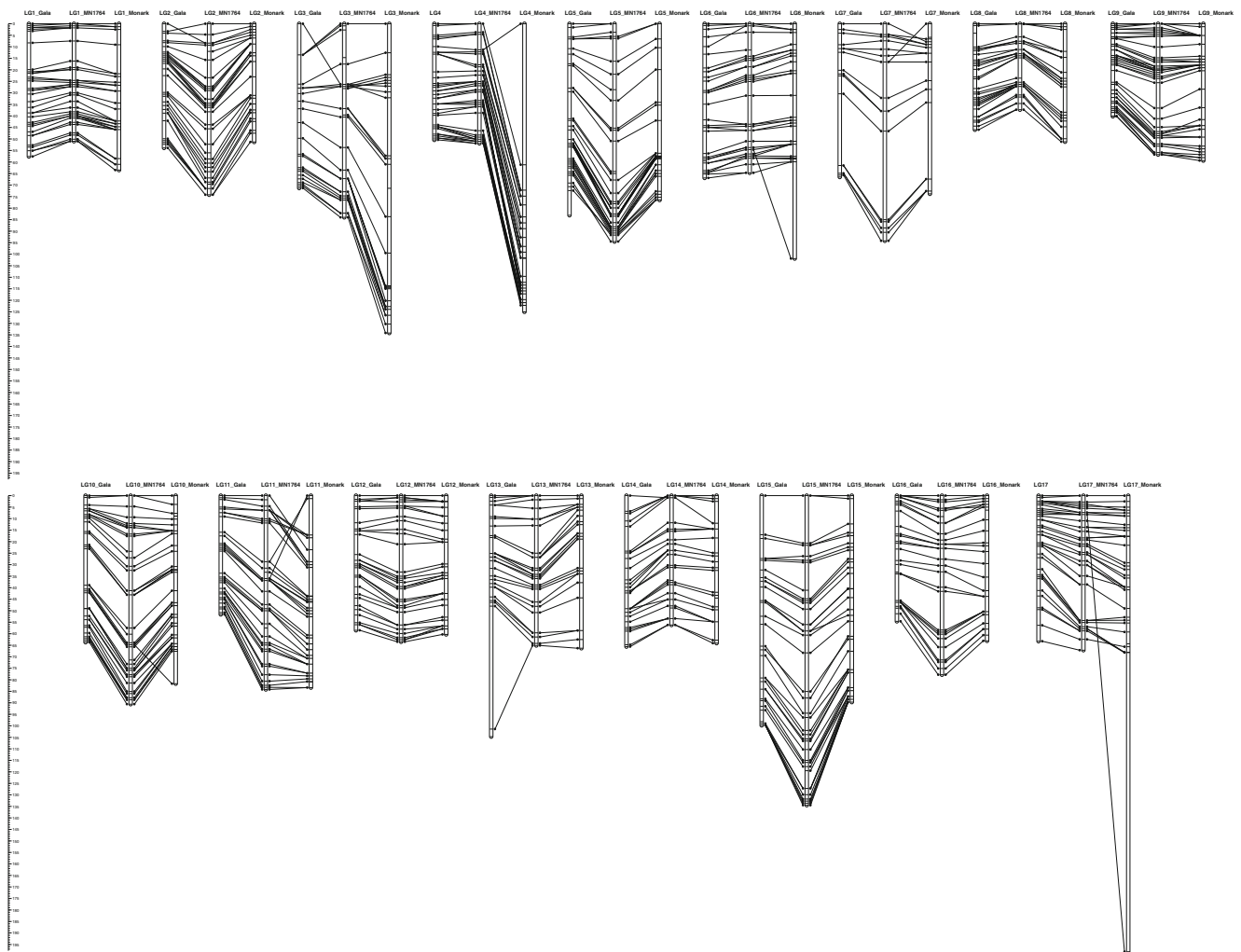


Fig. 3 Three ‘Honeycrisp’ parental maps (‘Honeycrisp’ × ‘Gala,’ ‘Honeycrisp’ × MN1764, and ‘Honeycrisp’ × ‘Monark’) utilized in consensus map construction. Lines between linkage groups show homology between maps within that linkage group

Comparison of genetic positions to physical map

The genetic positions of markers in the consensus ‘Honeycrisp’ map were plotted against the physical positions of marker loci on the ‘Golden Delicious’ genome (Fig. 6). Generally, there was agreement in the placement of the markers between the ‘Honeycrisp’ map and the genome sequence as evidenced by the linearity in the plots. The majority of the markers revealed direct correspondence between the linkage groups and the ‘Golden Delicious’ pseudo-chromosomes. Across the linkage map, 110 (10.1 %) markers mapped to linkage groups other than the corresponding pseudo-chromosome. Eight markers that were placed in the consensus ‘Honeycrisp’ map were classified as “unanchored” in the physical map. Areas of high recombination, indicated by large horizontal gaps in Fig. 6, were detected along several of the LGs including LG1, LG6, LG7, and LG10. Areas of low recombination are also evident as marker clusters.

Segregation distortion

Of the markers included in the consensus linkage map, 57 showed significant ($p < 0.005$) segregation distortion in the ‘Honeycrisp’ × ‘Gala’ progeny, 58 were significantly distorted in the ‘Honeycrisp’ × MN1764 progeny, and 41 were significantly distorted in ‘Honeycrisp’ × ‘Monark’ progeny. In total, only nine markers showed a significant segregation distortion in two families (black points, Fig. 6) and 138 markers showed a significant segregation distortion in only one family (gray points, Fig. 6). None of the markers of the consensus map showed a significant segregation distortion at the 0.005 level in all three progenies. Because segregation distortion was not used as a quality control measure during marker selection or map construction, 13.5 % of mapped markers showed a significant segregation distortion in at least one of the three families. Significant distortion was primarily clustered to regions on LG2, LG5, LG6, LG13, LG14, and LG17.

Table 2 Details from the genetic linkage maps of three ‘Honeycrisp’ parental maps from three full-sib populations (‘Honeycrisp’ × ‘Gala,’ ‘Honeycrisp’ × MN1764, and ‘Honeycrisp’ × ‘Monark’). Number of markers per linkage group, map size (centimorgans), density, and largest gap are given. Details from the consensus map constructed from the integration of the three ‘Honeycrisp’ parental maps are also shown in italics

	LG1	LG2	LG3	LG4	LG5	LG6	LG7	LG8	LG9	LG10	LG11	LG12	LG13	LG14	LG15	LG16	LG17	Total
Number of markers	Gala 55	77	48	85	82	51	24	56	90	66	59	50	48	42	84	57	68	1,042
	MN1764 54	80	50	88	82	51	23	55	84	67	62	51	50	38	84	55	67	1,041
	Monark 54	77	49	88	69	49	24	51	87	68	60	51	48	38	85	55	65	1,018
	Consensus 56	83	53	90	83	52	26	57	91	72	64	52	51	44	91	58	68	1,091
Size (cM)	Gala 57.60	53.95	71.32	50.36	83.08	66.91	66.47	45.98	40.18	63.71	51.61	58.51	104.70	65.63	99.83	54.64	63.07	1,097.55
	MN1764 50.99	74.21	134.16	125.03	94.46	64.73	73.84	37.32	56.76	90.56	84.22	63.48	65.05	56.19	134.47	77.85	66.98	1,350.29
	Monark 63.53	51.35	84.02	51.86	76.54	101.90	94.16	51.16	59.37	81.72	83.35	60.34	66.34	64.07	89.72	63.11	197.66	1,340.20
	Consensus 71.34	78.86	112.40	84.11	108.66	78.39	89.14	61.58	72.23	91.54	97.18	75.51	97.60	75.95	130.48	74.12	82.63	1,481.72
Average marker distance (cM)	Gala 1.05	0.70	1.49	0.59	1.01	1.31	2.77	0.82	0.45	0.97	0.87	1.17	2.18	1.56	1.19	0.96	0.93	1.05
	MN1764 0.94	0.93	2.68	1.42	1.15	1.27	3.21	0.68	0.68	1.35	1.36	1.24	1.30	1.48	1.60	1.42	1.00	1.30
	Monark 1.18	0.67	1.71	0.59	1.11	2.08	3.92	1.00	0.68	1.20	1.39	1.18	1.38	1.69	1.06	1.15	3.04	1.32
	Consensus 1.27	0.95	2.12	0.93	1.31	1.51	3.43	1.08	0.79	1.27	1.52	1.45	1.91	1.73	1.43	1.28	1.22	1.36
Maximum gap size (cM)	Gala 11.66	7.78	13.54	7.77	11.66	9.68	31.18	5.94	5.13	15.87	9.68	8.72	53.36	10.66	16.98	11.66	6.85	
	MN1764 8.76	7.71	13.21	7.70	13.21	12.43	38.63	8.76	10.94	14.38	16.82	12.06	12.06	10.94	20.72	16.82	15.59	
	Monark 12.57	8.04	25.14	61.24	14.20	41.96	33.15	11.00	8.02	14.12	15.85	9.49	14.20	9.49	11.00	6.59	129.64	
	Consensus 10.87	6.88	12.24	12.92	12.39	9.48	34.21	6.33	7.91	13.34	23.54	10.09	22.26	8.68	16.77	12.19	5.82	

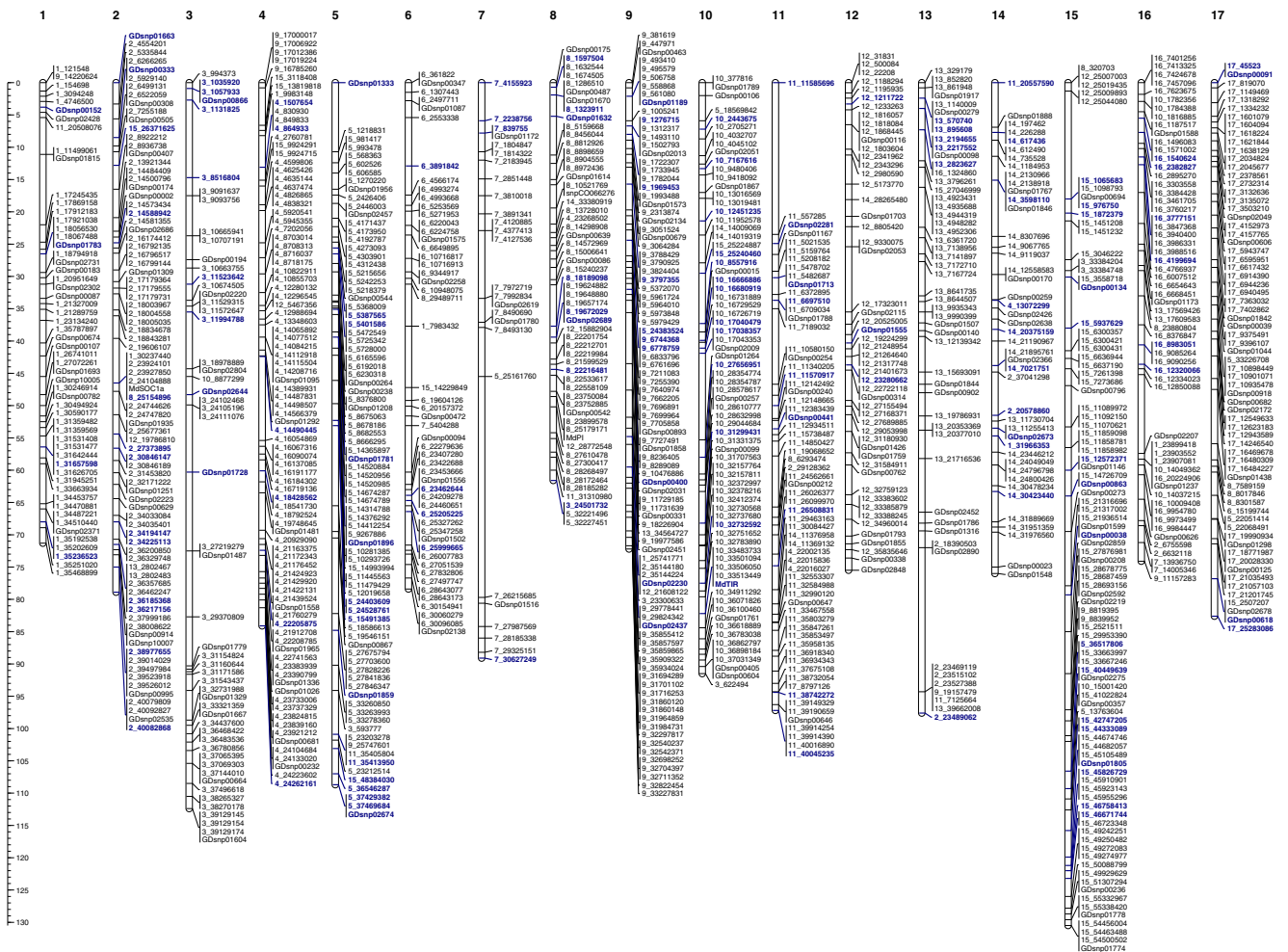


Fig. 4 Consensus ‘Honeycrisp’ linkage map constructed from three ‘Honeycrisp’ parental maps (‘Honeycrisp’ × ‘Gala’, ‘Honeycrisp’ × MN1764, and ‘Honeycrisp’ × ‘Monark’). Markers shown in blue were not common to all three parental maps

Discussion

We have developed a consensus ‘Honeycrisp’ linkage map spanning 17 linkage groups representing the 17 chromosomes

in the apple genome using the high-throughput IRSC 8K SNP array v1 (Chagné et al. 2012) and three mapping populations. The strategy utilized stringent data checking steps to ensure quality marker data including selection of high-quality SNP reads, removal of markers demonstrating a high frequency of double recombination, and examination of “lone wolf” markers. FlexQTL identified problematic markers that exceeded the threshold for observed double recombinants based on the expected frequency in each of the families. We were not able to position these markers elsewhere in the map using JoinMap. The double recombination pattern was visualized in Map Chart v2.2 (Voorrips 2002) and also provided a quick graphical interpretation after each round of mapping. This method was convenient and intuitive without the added complexity of graphical genotyping for ordering markers and identifying spurious markers. This methodology utilizes files that can be used in QTL analysis with FlexQTL, thus reducing the burden of creating new files or data for other interfaces.

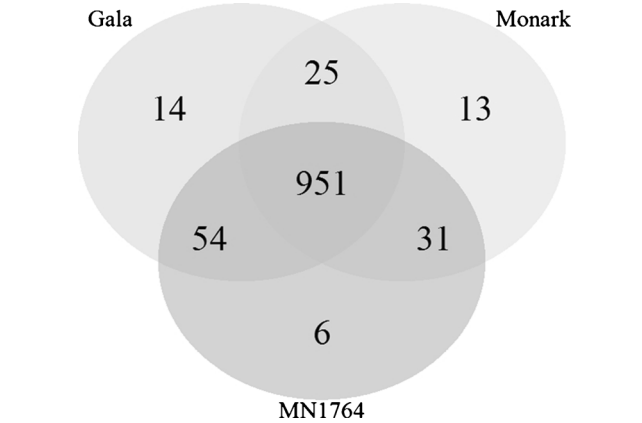


Fig. 5 Venn diagram showing the number of markers shared in the ‘Honeycrisp’ consensus map (1,091 total SNP markers) and those unique to each population

The mapping approach outlined here drastically reduced the number of SNP markers to only 13.8 % of those on the 8K

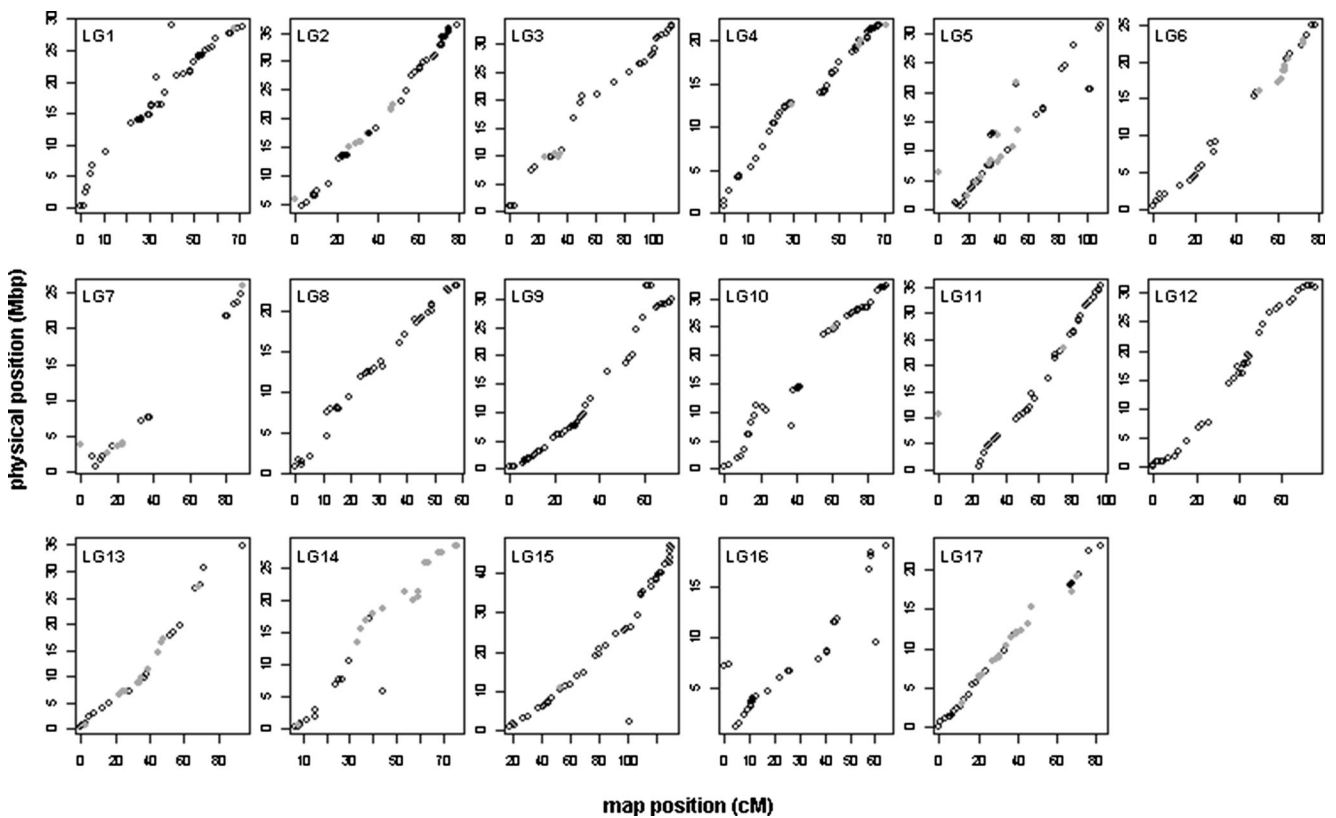


Fig. 6 Comparison of ‘Honeycrisp’ consensus map to physical position on ‘Golden Delicious’ genome sequence for each of the 17 linkage groups. Each plot directly compares the linkage group (LG1–LG17) to the pseudo-chromosome (1–17) available in the Genome Database for Rosaceae (www.rosaceae.org). Markers showing segregation distortion (p value <0.005) are indicated as follows: *open circles*, no significant

distortion in any of the three families; *gray filled circles*, significant distortion in one family; and *black filled circles*, significant distortion in two families. No markers in the consensus ‘Honeycrisp’ linkage maps showed a significant segregation distortion in all three of the mapping populations

IRSC SNP array. The first reduction to ~2,000 SNP markers was based on stringent parameters to identify high-quality reads with visually distinguishable clusters in the GenomeStudio software. These markers were then scored for all three populations in accordance with other reports using similar numbers of markers for linkage mapping in apple (Antanaviciute et al. 2012; Micheletti et al. 2011). The FlexQTL inheritance checking algorithm efficiently identified problematic markers or inheritance errors. Data free of genotyping errors are very important for the construction of genetic maps to ensure proper marker ordering.

The detection of functional ‘Honeycrisp’ haplotypes will provide utility in genetic studies of progeny populations with the aim of identifying genetic contributions specific to this parent. The consensus map has an average interval of 1.36 cM between markers, a much higher marker density than has been achieved with SSR linkage maps, and provides sufficient marker coverage for moderate-sized QTL mapping populations. The often touted advantage of a high-throughput SNP array is the reduced price per marker. But marker quality and usefulness are not uniform across all loci or populations. Homozygosity at a marker locus, genotyping quality, and

genotyping errors all contribute to increasing the cost per informative marker. Chagné et al. (2012) showed that of the 8K array, only 70.6 % of the markers were polymorphic in >1,600 individuals, accessions, and segregating populations that were evaluated. The development of a reduced array that retains those highly informative polymorphic markers across the genome could increase the efficiency for MAB.

Linkage mapping in JoinMap 4.1 utilizing the published M432 map (Antanaviciute et al. 2012) for the grouping step was computationally efficient and produced equivalent grouping assignments as the de novo algorithm. Due to high levels of sequence similarity resulting from the genome-wide duplication in *Malus*, SNP markers may map to more than one genomic location. This phenomenon was not evident with the high-quality SNPs used in this study. The method employed here was designed to select markers for consensus map construction that consistently placed SNPs to common groups. The multipoint maximum likelihood method for mapping was faster than the regression mapping and was thus utilized in this study of outcrossing populations (Van Ooijen 2011). The construction of two parental maps and an integrated map for each population was useful in determining the fate of “lone

wolf” markers although only the ‘Honeycrisp’ parental map was retained for consensus map construction.

The three ‘Honeycrisp’ parental maps each comprised shared and unique markers due to observed differences in heterozygosity in the parents and the quality of SNP calls. For example, in the consensus map, the distal end of LG15 was greatly extended by the inclusion of the ‘Gala’ and ‘Monark’ populations since the MN1764 population was uninformative in that region (Fig. 2). Low-quality markers may have been discarded differently within GenomeStudio among the populations due to the quality of the reads. Low levels of heterozygosity were observed even in the consensus map in some areas such as LG7, similar to the M432 map (Antanaviciute et al. 2012). To increase the coverage in these regions, one could return to GenomeStudio and use less stringent quality parameters for SNP calls. Additionally, markers developed specifically from pseudo-chromosome 7 could be scored and added to the maps. Genomic regions with high levels of homozygosity shared among cultivars could be an artifact of domestication, genetic drift, other intentional or unintentional selection, or a bottleneck. An exploration of these genome areas among other cultivars and *Malus* species linkage maps could provide insight into the genes that reside in these areas.

The clustering strategy that was utilized in the development of the IRSC 8K SNP array resulted in many SNP markers mapping to the same locus. Low recombination in these areas makes it difficult to assign the correct map order. Observed differences in local homology between the parental maps may be the result of within-cluster ordering. Using the physical map to order the markers would be one strategy to resolve this issue; however, the ordering of the physical map may also be incorrect. Additionally, the physical order of markers may be different between the three populations due to disruption in microsynteny and structural variations (Khan et al. 2012). Because the recombination frequency is so small within a cluster or tightly mapped clusters/markers, the precise order may not serve as a barrier to QTL detection. This is especially true in a pedigree-based approach, in which markers within a cluster may have different utility for individuals of different subpopulations. That is, any given individual SNP marker within a cluster at a single marker locus may segregate for some individuals or subpopulations and not others, but the map position is not lost for the entire pedigreed population. Additionally, local marker order may not be important in establishing functional haplotypes in a cluster in which low frequencies of recombination events occur.

Antanaviciute et al. (2012) compared map positions of an integrated apple rootstock linkage map to the ‘Golden Delicious’ genome sequence, reporting that 13.7 % of genetically mapped markers did not associate with the predicted pseudo-chromosome. Our results are consistent with this finding and are supported with the de novo grouping and the use

of the M432 map for grouping of markers to linkage groups. Over 10 % of markers in the consensus ‘Honeycrisp’ map were placed to linkage groups other than the corresponding pseudo-chromosome. These markers should be evaluated for known homology in the *Malus × domestica* genome (specifically known genome duplications and possible misalignments of contigs in the development of the ‘Golden Delicious’ genome sequence). Of the 110 markers of the consensus map (1091 total SNPs) that mapped to alternate pseudo-chromosomes, only 10 (9.1 %) were associated with potential homeologous chromosomes from the genome-wide duplication event (Velasco et al. 2010). For instance, a cluster of markers initially associated with pseudo-chromosome 9 of ‘Golden Delicious’ maps to the top of LG4 in both the M432 and ‘Honeycrisp’ maps. However, had our data not supported these placements, it is likely the markers would have been identified as “suspect linkages” during mapping and thus been discarded.

A significant segregation distortion was observed for 13.5 % of the markers in the final ‘Honeycrisp’ consensus map when no quality control measures regarding segregation distortion were used during marker checking or linkage map construction. The choice not to use segregation distortion as a quality control measure was made because marker segregation distortion could represent real, biologically relevant segregation distortion and the inclusion of these markers could be useful in QTL detection. Largely supporting this hypothesis is the observation that markers exhibiting segregation distortion mapped in cohesive clusters along only several linkage groups. Biological reasons for segregation distortion are those that impose selection upon the population such as selective fertilization (apple’s gametophytic self-incompatibility), abortion of gametes (Liebhard et al. 2003), and other unavoidable natural selective pressures such as field environment (e.g., winter hardiness) that are inadvertently imposed upon the breeding populations (i.e., the ad hoc mapping populations utilized in this study). Markers with observed segregation distortion need not be within the survival gene, and they may be linked with the gene conferring survivorship. Segregation distortion observed in this study was not found in the same linkage groups as that reported by Antanaviciute et al. (2012) with the exception of that on LG17 which contains the *S*-locus (Maliapaard et al. 1998).

The GenomeStudio software and manual calling of SNPs into biallelic clusters (AA, AB, BB) are constrained by the quality of reads. Inherent in difficulty with read quality are errors resulting from DNA quality, contamination, DNA hybridization and extension, and fluorescence signal. Recent whole-genome duplication, segmental duplication, and a high degree of homology between some markers result in SNP markers exhibiting segregation behavior similar to that of polyploids in the cluster plots (Voorrips et al. 2011, personal observation). DNA from different genomic regions may hybridize to the same marker, typically resulting in more than

three clusters. However, not all of these occurrences may be detected manually or within the automated calling. The spread of a cluster in automated/manual calling of multiple populations (pedigrees, diverse sets) may provide statistical support of a single cluster, but may mask the presence of more than three clusters within a single population that would have been identified as a potential homolog and removed.

A high degree of colinearity was observed between the consensus map and the physical positions along the ‘Golden Delicious’ pseudo-chromosomes. Large linkage gaps were observed in regions of low marker coverage, presumably centromeric and telomeric regions. Colinearity supports the physical ordering of markers and strengthens the development of meaningful haplotypes that represent true chromosome position. Markers that do not align may result in haplotypes that are a mosaic of different chromosome segments.

The consensus ‘Honeycrisp’ linkage map developed from three progeny populations consists of 1,091 SNP markers distributed across the apple genome. These markers were developed from exonic regions from the ‘Golden Delicious’ genome sequence which adds to their utility in predicting function in marker-locus-trait associations (Chagné et al. 2012). More importantly, these markers are informative in an elite cultivar that is being utilized in breeding programs worldwide for its superb fruit quality traits. QTL analysis in ‘Honeycrisp’ will not only focus on identifying the haplotypes associated with crispness, firmness, and juiciness but will also focus on identifying deleterious associations with post-harvest disorders such as soft scald, internal browning, and bitter pit, to which ‘Honeycrisp’ is prone.

Acknowledgments This work was partially funded by USDA's National Institute of Food and Agriculture—Specialty Crop Research Initiative project, “RosBREED: Enabling marker-assisted breeding in Rosaceae” (2009-51181-05808). The authors gratefully acknowledge Elisabeth Alperin, April Nyberg, and Nahla Bassil (National Clonal Germplasm Repository, Corvallis, OR); Amy Iezzoni (Department of Horticulture, Michigan State University), the Michigan State University Infinium SNP Genotyping Lab, and the Minnesota Supercomputing Institute for providing advice, resources, and services integral to this project.

Ethical standards Experimental procedures comply with the current state and federal laws.

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

Data archiving statement Relevant data, including linkage maps and SNP markers utilized in map construction, have been submitted to and archived in the Genome Database for Rosaceae (GDR, http://www.rosaceae.org/tgg/data_submission).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Antanaviciute L, Fernández-Fernández F, Jansen J, Banchi E, Evans KM, Viola R, Velasco R, Dunwell JM, Troglio M, Sargent D (2012) Development of a dense SNP-based linkage map of an apple rootstock progeny using the *Malus* Infinium whole genome genotyping array. *BMC Genomics* 13:203. doi:10.1186/1471-2164-13-203
- Berkett L, Garcia M, Moran R, Darby H, Parsons R, Hayden J, Bradshaw T, Kingsley-Richard S, Cromwell M (2009) The OrganicA Project: Organic Disease Management in Orchards with Newer Cultivars. International Organization for Biological and Integrated Control Bulletin <http://www.uvm.edu/~organica/OrganicAProject/IOBCPoster.pdf> Accessed 4 Mar 2014
- Bink MCMA, Boer MP, Ter Braak CJF, Jansen J, Voorrips RE, van de Weg WE (2008) Bayesian analysis of complex traits in pedigreed plant populations. *Euphytica* 161:85–96
- Chagné D, Crowhurst RN, Troglio M et al (2012) Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS ONE* 7:e31745. doi:10.1371/journal.pone.0031745
- Clark MD, Schmitz, Cari A., Rosyara UR, Luby JJ, Bradeen JM (2013) A high resolution SNP linkage map of a “Honeycrisp” apple (*Malus × domestica*) progeny. Plant and Animal Genome XXI Conference, San Diego, CA. 14 January 2013. <https://pag.confex.com/pag/xxi/webprogram/Paper7625.html> Accessed 10 April 2013
- Fernández-Fernández F, Antanaviciute L, van Dyk MM, Tobutt KR, Evans KM, Rees DJG, Dunwell JM, Sargent DJ (2012) A genetic linkage map of an apple rootstock progeny anchored to the *Malus* genome sequence. *Tree Genet Genomes* 8:991–1002
- Gianfranceschi L, Koller B, Seglias N, Kellerhals M, Gessler C (1996) Molecular selection in apple for resistance to scab caused by *Venturia inaequalis*. *Theor Appl Genet* 93:199–204
- Gilmore B, Hummer K, Bassil N (2011) DNA extraction protocols from dormant buds of twelve woody plant genera. *J Am Pomol Soc* 65: 201–207
- Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137:1121–1137
- Hoover E, Bedford D, Foulk D (2000) Apples for Minnesota and their culinary uses. Univ. of Minnesota, St Paul, MN <http://www.extension.umn.edu/distribution/horticulture/dg1111.html> Accessed 10 Apr 2013
- Illumina (2006) Illumina SNP genotype: Infinium II assay workflow. Illumina, Inc., Towne Centre Drive, San Diego, CA. http://www.illumina.com/Documents/products/workflows/workflow_infinium_ii.pdf Accessed 10 Apr 2013
- Illumina Inc. (2010a) GenomeStudio Genotyping Module v1.0, User Guide. Illumina, Inc. Towne Centre Drive, San Diego, CA. http://support.illumina.com/documents/MyIllumina/84861395-800d-4568-a38c-8efa6fda5015/GenomeStudio_GT_Module_User_Guide_RevA.pdf Accessed 10 Apr 2013
- Illumina Inc. (2010b) Infinium genotyping data analysis: a guide for analyzing genotyping data using the Illumina GenomeStudio Genotyping Module. Illumina Inc., Towne Centre Drive, San Diego, CA. http://www.illumina.com/Documents/products/technotes/technote_infinium_genotyping_data_analysis.pdf Accessed 10 Apr 2013
- Illumina Inc. (2010c) Improved cluster generation with GenTrain2, Tech Note. Illumina Inc., Town Centre Drive, San Diego, CA. http://www.illumina.com/Documents/products%5Ctechnotes%5Ctechnote_gentrain2.pdf Accessed 10 Apr 2013
- Jung S, 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
- Kelley K, Hyde J, Travis J, Crassweller R (2010) Accessing consumer preferences of scab-resistant apples: a sensory evaluation. *HortTechnology* 20:885–891

- Khan MA, Han Y, Zhao YF, Troglio M, Korban SS (2012) A Multi-population consensus genetic map reveals inconsistent marker order among maps likely attributed to structural variations in the apple genome. *PLoS ONE* 7:e47864. doi:10.1371/journal.pone.0047864
- King GJ, Maliepaard C, Lynn JR, Alston FH, Durel CE, Evans KM, Griffon B, Laurens F, Manganaris AG, Schrevels E, Tartarini S, Verhaegh J (2000) Quantitative genetic analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (*Malus pumila* Mill.). *Theor Appl Genet* 100:1074–1084
- Lawson DM, Hemmat M, Weeden NF (1995) The use of molecular markers to analyze the inheritance of morphological and developmental traits in apple. *J Am Soc Hortic Sci* 120:532–537
- Liebhart R, Koller B, Gianfranceschi L, Gessler C (2003) Creating a saturated reference map for the apple (*Malus × domestica* Borkh.) genome. *Theor Appl Genet* 106:1497–1508
- Luby JJ, Bedford D (1992) Honeycrisp apple. Minnesota Report 225-1992 (AD-MR-5877-B). Univ. of Minnesota, St. Paul, MN http://www.extension.umn.edu/distribution/horticulture/components/5877_01.html. Accessed 10 Apr 2013
- Luby JJ, Shaw DV (2001) Does marker-assisted selection make dollars and sense in a fruit breeding program? *HortSci* 35:872–879
- Maliepaard C, Alston FH, Van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner SE, Guilford P, Van Heusden AW (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theor Appl Genet* 97:60–73
- McKay SJ, Bradeen JM, Luby JJ (2011) Prediction of genotypic values for apple fruit texture traits in a breeding population derived from ‘Honeycrisp’. *J Am Soc Hortic Sci* 136:408–414
- Micheletti D, Troglio M, Zharkikh A, Costa F, Malnoy M, Velasco R, Salvi S (2011) Genetic diversity of the genus *Malus* and implications for linkage mapping with SNPs. *Tree Genet Genomes* 7:857–868. doi:10.1007/s11295-011-0380-8
- Pre-Aymard C, Fallik E, Weksler A, Lurie S (2005) Sensory analysis and instrumental measurements of ‘Anna’ apples treated with 1-methylcyclopropene. *Postharvest Biol Technol* 36:135–142. doi:10.1016/j.postharvbio.2004.12.007
- R Core Team (2012) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org> Accessed 4 Mar 2014
- Schenato PG, Valdebenito-Sanhueza RM, Duarte V (2008) Determinação da presença da raça 1 de *Venturia inaequalis* no Sul do Brasil. *Trop Plant Pathol* 33:281–287
- Tartarini S, Sansavini S (2003) The use of molecular markers in pome fruit breeding. *Acta Hort* 622:129–140
- Tong C, Krueger D, Vickers Z, Bedford D, Luby J, El-Shiekh A, Shackel K, Ahmadi H (1999) Comparison of softening-related changes during storage of ‘Honeycrisp’ apple, its parents, and ‘Delicious’. *J Am Soc Hortic Sci* 124:407–415
- Van Ooijen JW (2006) JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Wageningen, Kyazma BV
- Van Ooijen JW (2011) Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet Res* 93:343–349. doi:10.1017/S0016672311000279
- Velasco R, Zharkikh A, Affourtit J et al (2010) The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet* 42:833–839
- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* 93:77–78
- Voorrips RE, Gort G, Vosman B (2011) Genotype calling in tetraploid species from bi-allelic marker data using mixture models. *BMC Bioinforma* 12:172. doi:10.1186/1471-2105-12-172
- Wang A, Aldwinckle HS, Forsline P, Main D, Brown S, Xu K (2012) EST contig-based SSR linkage maps for *Malus × domestica* cv Royal Gala and an apple scab resistant accession of *M. sieversii*, the progenitor species of domestic apple. *Mol Breeding* 29:379–397. doi:10.1007/s11032-011-9554-1
- Wu Y, Close TJ, Lonardi S (2011) Accurate construction of consensus genetic maps via integer linear programming. *IEEE/ACM Trans Comput Biol Bioinform* 8:381–394. doi:10.1109/TCBB.2010.35