Theor Appl Genet (2003) 106:1497–1508
DOI 10.1007/s00122-003-1209-0

R. Liebhard · B. Koller · L. Gianfranceschi · C. Gessler

Creating a saturated reference map for the apple (Malus × domestica Borkh.) genome

Received: 26 July 2002 / Accepted: 1 November 2002 / Published online: 2 April 2003

Abstract The availability of a high quality linkage map is essential for the detection and the analysis of quantitative traits. Such a map should cover a significant part of the genome, should be densely populated with markers, and in order to gain the maximum advantage should be transferable to populations or cultivars other than the ones on which it has been constructed. An apple genetic linkage map has been constructed on the basis of a segregating population of the cross between the cultivars Fiesta and Discovery. A total of 840 molecular markers, 475 AFLPs, 235 RAPDs, 129 SSRs and 1 SCAR, were used for the two parental maps constructed with JoinMap and spanning 1,140 cM and 1,450 cM, respectively. Large numbers of codominant markers, like SSRs, enable a rapid transfer of the map to other populations or cultivars, allowing the investigation of any chosen trait in another genetic background. This map is currently the most advanced linkage map in apple with regard to genome coverage and marker density. It represents an ideal starting point for future mapping projects in *Malus* since the stable and transferable SSR frame of the map can be saturated quickly with dominant AFLP markers.

Keywords AFLP · Genetic mapping · Microsatellites · Molecular markers · RAPD · SSR

Communicated by H. Nybom

R. Liebhard · C. Gessler (✉)
Swiss Federal Institute of Technology,
Institute of Plant Science/Phytopathology, Universitätstrasse 2,
8092 Zurich, Switzerland,
e-mail: cesare.gessler@ipw.agrl.ethz.ch
Fax: +41-(1)-632-11-08

B. Koller
ECOGENICS GmbH, Winterthurerstrasse 190,
8057 Zürich, Switzerland

L. Gianfranceschi
Dipartimento di Genetica e di Biologia dei Microorganismi
(DGBM), Università degli Studi di Milano, via Celoria 26,
20133 Milano, Italy

Introduction

Saturated and high-density genetic linkage maps are very useful in fundamental and applied genetic research. Such maps are becoming increasingly available for woody perennials like *Citrus* (Sankar and Moore 2001), *Prunus* (Joobeur et al. 1998), Cacao (Risterucci et al. 2000) and *Malus* (Hemmat et al. 1994; Conner et al. 1997; Seglias and Gessler 1997; Maliepaard et al. 1998; Liebhard et al. 2002). Genetic studies and breeding for high quality apple cultivars has always been complicated by the slow growth, the long juvenile phase and the strong self incompatibility present in this species. These considerations have fostered the investigation of the apple genome by means of molecular markers and the construction of genetic linkage maps.

Linkage maps allow studies of the genome structure, the localisation of genes of interest, and permit the identification of quantitative trait loci (QTLs), by providing the framework to understand the biological basis of complex traits (Tanksley et al. 1989). Finally, as a combination of the described properties, they enable marker-assisted breeding and selection (MAS). MAS is especially promising in perennial tree crops, like apple, where many important traits are expressed only after years of costly field maintenance. Using MAS, the presence of favourable alleles can be determined precociously, and therefore the population size can be drastically reduced at the early stages of selection, theoretically by factor 2^n, where n is the number of loci under selection, with alleles showing a 1:1 segregation.

The maximum advantage of a genetic linkage map is gained when it can be easily transferred to cultivars or populations other than those for which it was originally constructed (Joobeur et al. 1998; Liebhard et al. 2002). A number of apple linkage maps have been published so far (Hemmat et al. 1994; Conner et al. 1997; Seglias and Gessler 1997; Maliepaard et al. 1998; Liebhard et al. 2002), composed mainly of RFLPs, isozymes, RAPDs and microsatellite markers (simple sequence repeats, SSRs). Dominant markers, such as RAPDs, can be used

for map alignments if they are heterozygous in both parents, as shown by Hemmat et al. (1994) and Conner et al. (1997), but their transferability to other maps is limited.

Codominant markers like RFLPs and isozymes are very useful with regard to transferability to other apple cultivars but are rather labour intensive and/or require large amounts of high quality DNA. Additionally, although being theoretically codominant, in practice they segregate to a large extent as dominant markers. In apple, Maliepaard et al. (1998) found that only about 30% of RFLPs and 5% of isozyme markers segregate as highly informative, codominant markers with at least three different alleles. By contrast, 90% and 75% of the SSR markers tested by Maliepaard et al. (1998) and by Liebhard et al. (2002), respectively, were codominant with at least three different alleles.

A large number of SSR markers available in apple have been reported to be useful in map alignment and transferable between cultivars (Liebhard et al. 2002). Due to the large number of SSR alleles present in apple, there is a high chance of finding polymorphisms in and between most cultivars as shown by Gianfranceschi et al. (1998) and Liebhard et al. (2002). The use of PCR-based markers, such as codominant SSRs (Liebhard et al. 2002) and dominant AFLPs (Vos et al. 1995), requiring only small amounts of DNA and being easily automated, offers a simple and fast approach toward new and transferable maps. AFLP markers have been successfully applied for map saturation purposes in apple (Xu and Korban 2000) and other crops (Vuylsteke et al. 1999; Risterucci et al. 2000), producing large numbers of polymorphic markers with only few PCR reactions. However, even though AFLP is a powerful technique for the fast production of many markers in a specific linkage map, the transferability of the map to other crosses relies completely on SSR markers.

Here we present a linkage map of the apple cultivars Fiesta and Discovery, consisting of a robust core of SSR markers, saturated with a large number of RAPDs and AFLPs. This map should be taken as reference for future map constructions in apple, since the large number of SSR markers can be transferred to any apple cultivar and would serve as a frame which can be quickly saturated with AFLPs. A detailed discussion of the mapping procedure, of the problems encountered and of software options used to handle them is provided. The linkage maps are compared with those already published and the results obtained with two mapping programs, JoinMap 2.0 (Stam and van Ooijen 1995) and MAPMAKER/EXP 3.0 (Lander et al. 1987), are included.

Materials and methods

Plant material and DNA extraction

The segregating population of the cross between 'Fiesta' and 'Discovery', described in Liebhard et al. (2003), was used for the construction of the genetic linkage map. A double pseudo-testcross strategy appropriate for highly heterozygous, outbreeding species was applied (Grattapaglia and Sederoff 1994; Weeden et al. 1994).

Leaf material of the parental cultivars and the progeny plants of the cross was kindly provided by Markus Kellerhals, Swiss Federal Research Station, for fruit growing, viticulture and horticulture (FAW), Wädenswil, Switzerland.

DNA of about 300 progeny individuals was extracted according to Koller et al. (2000). Samples were divided into four subsets consisting of between 46 and 90 individuals. These four subsets were maintained throughout the analysis until the final map construction.

Automation

DNA samples were partially quantified by Pico Green fluorescence measurement according to the manufacturer instructions on a TECAN SpectrafluorPLUS fluorescence reader, allowing a very fast and accurate quantity determination.

All liquid-handling processing, including DNA sample preparation for fluorescence quantification, template pre-arrangement in a 96-well-layout, dilution to PCR concentration, replication of the subsets, as well as non-radioactive PCR mix preparation and distribution, were performed using a TECAN Genesis 150 RSP working platform and a customer-written Gemini 2.0 protocol for the specific liquid handling tasks.

Molecular-marker analysis

RAPD and SSR marker reactions and analysis were carried out as described in Liebhard et al. (2003) on 282 progeny plants.

AFLP reactions were performed on the same 282 individuals as described in Vos et al. (1995) except that the restriction and ligation steps were combined as described in Van der Lee et al. (1997). The restriction enzyme combination *Eco*RI/*Mse*I was used and the pre-amplification was performed with the primer combination E01-M01 having the same selective nucleotide A.

The selective amplification was performed with all 96 possible primer combinations between E31 to E38 and M31 to M42, with two additional selective nucleotides on the parental cultivars and a set of ten progeny plants. Fifty five primer combinations yielding at least eight clear polymorphic bands were then tested on a subset of 50 individuals and the resulting marker bands were mapped. This led to the identification of 46 primer combinations producing markers which saturate or elongate the existing linkage map of the two cultivars (Liebhard et al. 2002). These primer combinations were then tested on all four subsets of the population producing over 400 polymorphic marker bands.

AFLP fragments were scored as dominant, i.e. presence versus absence of bands, and their size was estimated by comparison with the ^{33}P-labelled Gibco 30–330-bp AFLP DNA ladder. All marker bands were scored twice visually and the data files were double checked for mis-scorings or typing errors with the computer program ReadMarkers, developed by one of the authors (B.K.). This program reads the data file and gives an acoustic output of the marker genotype (aa, ab, ac, ad, bd) allowing the scoring person to fully concentrate on the gel image and the markers.

Map construction

Segregation analysis was performed on a total of 267 individuals of the 'Fiesta' × 'Discovery' cross. After the removal of all 15 outcrossed and self-pollinated individuals, the four subsets, generated for DNA extractions and marker reactions, consisted of 44, 68, 70 and 85 individuals. All subsets were analysed independently. In the second stage of the analysis, when data were revised, the locus files were merged and the final map was constructed.

All linkage analysis and map calculations were performed with JoinMap version 2.0 (Stam and van Ooijen 1995) in connection

with JMDesk 3.6 (http://www.ecogenics.ch/software_e.html). Maternal and paternal data were kept separate throughout the analysis. A LOD score of 4.0 was applied for each subset to determine markers belonging to the same linkage groups (LGs). The thresholds for the following steps were set so that all data were included (LOD = 0.001, REC = 0.499) and the Kosambi mapping function was applied. Identification of problematic markers and the verification of unlikely or the correction of false genotypes was performed as described in Liebhard and Gessler (2000) and Liebhard et al. (2002).

Markers were excluded from the analysis when their distorted segregation conflicted with the segregation pattern of neighbouring markers, when they showed strong linkage to two different linkage groups or when their recombination frequencies conflicted with many other markers in the same linkage group.

When the four subset maps were of satisfactory quality and displayed comparable marker orders (i.e. the order of the SSR markers as anchor points had to be the same whereas small inversions in the order of dominant RAPD and AFLP markers, extending over only a few cM, were accepted), the data files of the subsets were merged and split into locus files for each chromosome according to the information obtained from the subset maps. For the final map construction, information of marker pairs with a LOD score below 0.5 and a recombination frequency greater than 0.48 were excluded. The same data set was also analysed with MAPMAKER/EXP 3.0, where the same mapping function (Kosambi) was applied. JoinMap data files for each linkage group were reformated to meet the requirements of a Mapmaker F2 backcross analysis.

The commands 'GROUP', 'COMPARE', 'MAP', 'TRY' and 'RIPPLE' were used to confirm the linkage of the markers, to determine their order within the linkage group, to calculate the distances between the markers, to add more markers to an existing map and to test the map order by permutating local marker sequences.

Map integration

Homologous 'Fiesta' and 'Discovery' linkage-group locus files were merged for the construction of an integrated 'Fiesta' × 'Discovery' linkage map. SSR markers were coded as codominant markers with up to four distinguishable alleles. Markers with less than 100 informative individuals as well as some markers from very dense map regions were excluded to facilitate the map construction. The same thresholds, described for the construction of the final single parent maps, were used.

Nomenclature of markers and linkage groups

RAPD and SSR markers are named as in Liebhard et al. (2003), except that allele size information is provided for all microsatellite markers by their name. Size information of each SSR allele allows the identification of the number of alleles involved and the recognition of the segregation type (i.e. aaxab, abxaa, abxab, abxac, abxcd) in the population. AFLP marker names consist of the name of the primers, as introduced by KeyGene NV, followed by the size of the fragment in basepairs. Linkage group numbering corresponds to the one used in Liebhard et al. (2003) introduced by Maliepaard et al. (1998).

Visual presentation

Drawings of the parental maps were generated with MapChart (Voorrips 2001). The visual presentation of the map also includes linkage phase information of the markers/alleles. Homologous chromosomes within one parent are termed '+' and '–', and marker names/alleles are provided with the symbol (i.e. + or –) of the chromosome on which they are located. With codominant markers, phase information is provided in the order of allele sizes in the marker name, i.e. the first phase information belonging to the first allele (Ex.: –/+CH03a09-125/131 indicates that the allele 125 bp maps on chromosome '–' and the allele 131 bp maps on chromosome '+').

Results

AFLP markers

Of the 96 *Eco*RI/*Mse*I primer combinations tested on ten progeny plants and the parental cultivars, 75 yielded between 1 and 25 unambiguously scorable, polymorphic bands. Forty six primer combinations, selected according to the readability of the bands and the number of polymorphisms, were tested on a set of 52 progeny plants, and 514 marker bands were scored and mapped. Based on the map positions of the markers, a further selection was performed, and 21 primer combinations were tested on the entire population, resulting in 300 scorable marker bands. For the final map construction a total of 475 AFLP markers (300 on all progeny plants, 175 on 52 plants only) from 46 primer combinations were available.

Map construction

A total of 840 AFLP, RAPD, SSR and SCAR markers (439 in 'Fiesta', 499 in 'Discovery', 98 SSRs appearing on both maps) were used for the construction of the genetic linkage map. The map (Fig. 1) covers a length of 1,143.8 and 1,454.6 cM in 'Fiesta' and 'Discovery', respectively, with an average chromosome length of 67.3 and 85.6 cM for the two parental cultivars.

Six pairs of previously unlinked chromosome fragments could be connected. Four of them were linked in at least one of the four subsets (with a LOD score of >4.0) and the fragments of linkage groups D3 and D6 were linked via the codominant SSR markers in the other parent. Six segments larger than 15 cM were added to the ends of linkage groups. The overall number of gaps larger than 20 cM was not reduced due to the connection of previously unlinked chromosome fragments, although 8 of the 11 gaps on chromosomes or fragments in the previous map were filled with new markers.

Two RAPD markers from the previous map had to be discarded since they did not meet the stringent requirements with respect to reproducibility and unambiguous scoring. Ten markers remained unlinked and 11 markers, showing distorted segregation, were also excluded, either because they map on two linkage groups or because the molecular data conflicted with those of the neighbouring markers of the linkage group.

The comparison of the four subset maps (data not shown) and the final map showed that the marker distances and the linkage group lengths varied considerably, whereas the marker order generally remained the same. Taking the final map as a reference the linkage

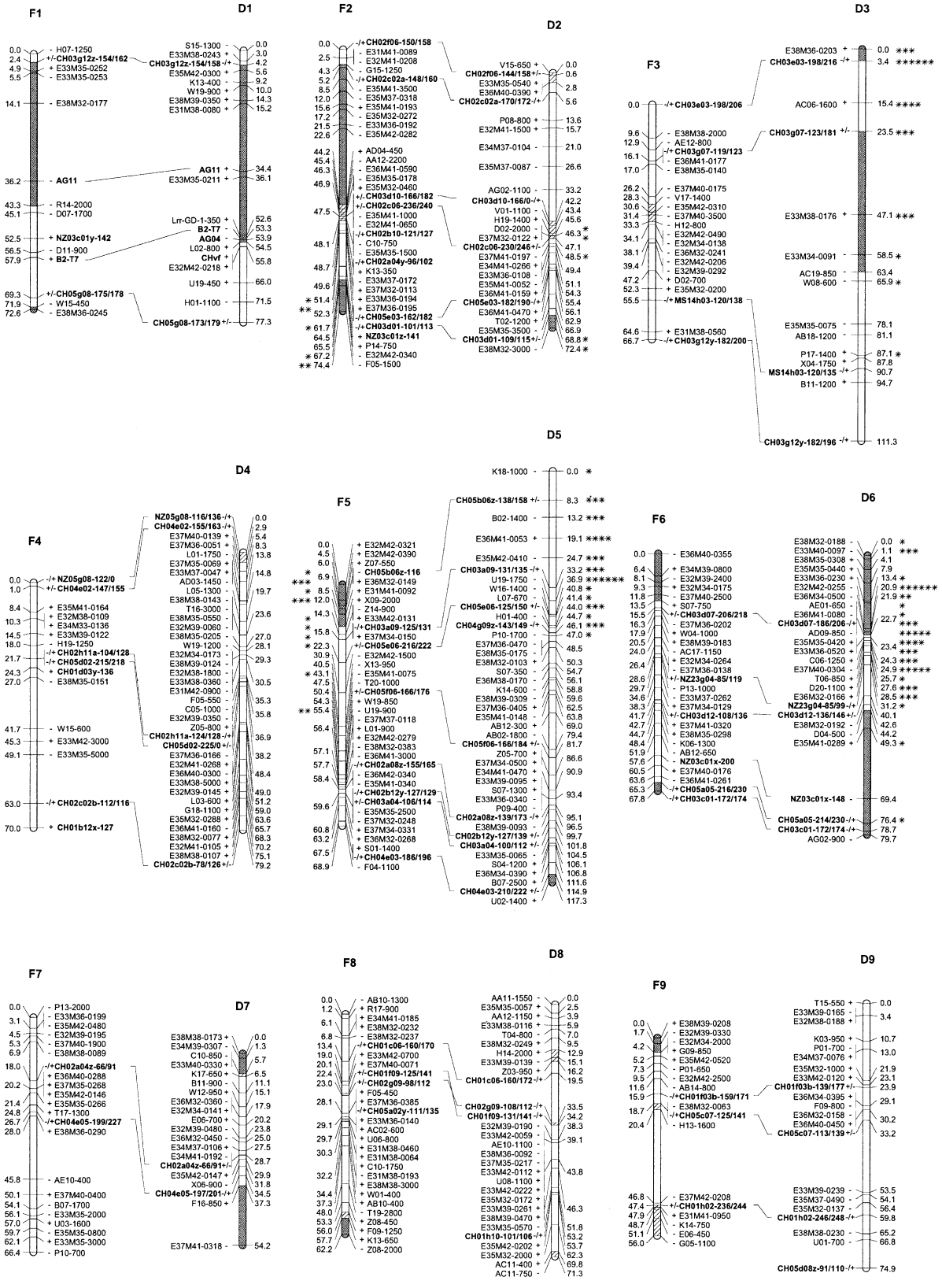
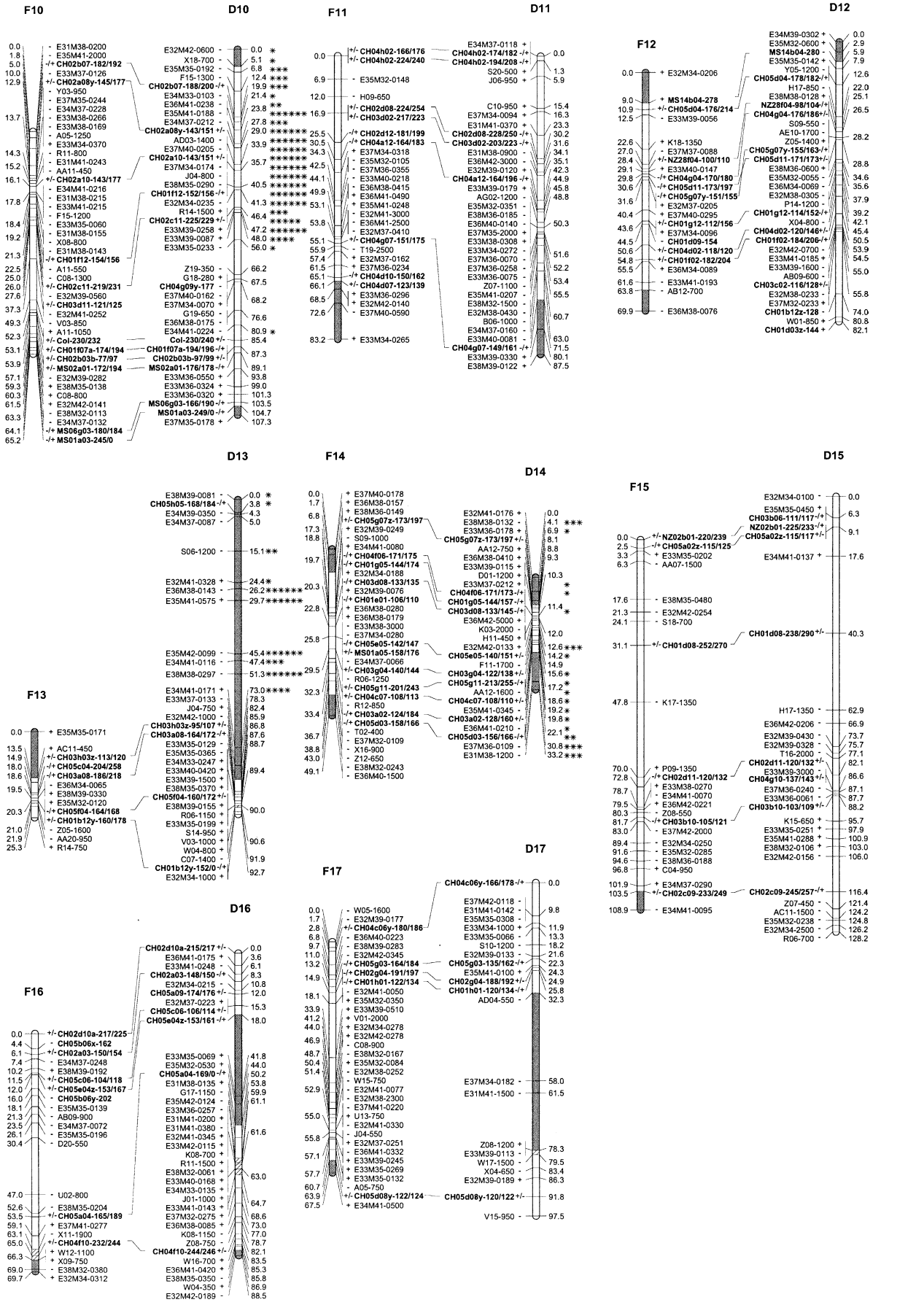


Fig. 1 Genetic linkage map of the apple progeny 'Fiesta' (*F*) × 'Discovery' (*D*). Linkage groups are numbered from F1 to F17 and from D1 to D17. SSR markers are printed in *bold*. Newly mapped chromosome segements are *grey*, regions with potentially erroneous marker orders are *hatched*, linkage phase infromation is provided as + or –, or both, indicating on which of the homologous chromosomes the marker/allele is located. Order of the phase information corresponds to the order of allele size with codominant SSR markers. Segregation distortion is indicated by means of significance level *p* of the chi-square test: * = 0.05, ** = 0.01, *** = 0.005, **** = 0.001, ***** = 0.0005, ****** = 0.0001


F10
D10
F11
D11
F12
D12
F13
D13
F14
D14
F15
D15
F16
D16
F17
D17

groups lengths between the subsets ranged from 47% (LG 13) to 133% (LG 16) in 'Fiesta' and from 35% (LG 14) to 125% (LG 2) in 'Discovery'.

Various small inversions in marker order appeared throughout the map when compared with the previously published ones. Although these inversions are considered to be unproblematic artefacts of the mapping procedure that only extended over a few cM, these regions are indicated in Fig. 1 in order to be aware of marker orders which are possibly not fully correct.

The number and type of markers, map lengths, marker density and gaps per linkage group and for the entire parental maps are presented in Table 1A and C, as well as the relative increase of the specified measures compared to the previous map (Liebhard et al. 2003) in Table 1B.

Phase information on markers and alleles, as presented in Fig. 1, allows the identification of markers and QTLs in coupling, i.e. on the same parental chromosomes and of markers in repulsion, i.e. on different homologous parental chromosomes. Allele size information and recognition of the number of involved alleles is useful since the information content of the marker correlates with the accuracy with which the markers can be mapped (Maliepaard et al. 1997).

Several linkage groups turned out to carry markers with heavily distorted segregation. These markers were generally not discarded, since they represent the occurrence of natural selection in the otherwise unselected population. The markers are labelled in Fig. 1 with an asterisk (*), representing the significance level of the χ^2 goodness-of-fit test. Thanks to the linkage phase infor-

Table 1A–B **A** Number and type of markers on the linkage map of 'Fiesta' and 'Discovery' (Fig. 1), map length, marker density, gaps, average linkage group length and average number of markers per linkage group. **B** Increase/reduction of the mentioned measures relative to the previous map (Liebhard et al. 2002)

A:					B:			
Marker type	Updated map				Increase			
	Total	Fiesta	Discovery	in common	Total	Fiesta	Discovery	in common
RAPD	235	104	131	–	21	10(+10.6%)	11(+9.2%)	
SSR	129	115	112	98	14	9(+8.5%)	12(+12%)	7
AFLP	475	220	255	–	475	220	255	
SCAR	1	–	1	–	1	1		
Total	840	439	499	–	511	239(+119.5%)	279(+126.8%)	
Lengths								
Length in cM		1143.8	1454.6			229.7(+25.1%)	439.6(+43.3%)	
Marker density in markers/cM		0.38	0.34			0.16	0.12	
Average distance between markers		2.61	2.92			–1.96 (–25%)	–1.69 (–42%)	
Largest gap		26.4	25.7					
Ø cM/linkage group		67.3	85.6			13.5 (+25%)	25.9 (+43%)	
Ø Markers/linkage group		26	29			14 (+116%)	16 (+123%)	

Table 1C Detailed presentation of the number and type of marker, length, marker density and gaps for each linkage group of the updated map

C:																	
Marker type	Linkage groups																
	F1	D1	F2	D2	F3	D3	F4	D4	F5	D5	F6	D6	F7	D7	F8	D8	F9
RAPD	5	6	7	7	4	7	2	10	10	17	6	7	6	7	13	9	7
SSR	5	6	9	6	4	4	7	5	8	9	6	6	2	2	4	4	3
AFLP	4	6	19	14	13	4	7	24	18	16	15	14	14	11	11	16	8
SCAR		1															
Total	14	19	35	27	21	15	16	39	36	42	27	27	22	20	28	29	18
Lengths																	
Length in cM	72.6	77.3	74.4	72.4	66.7	111.3	70.0	79.2	68.9	117.3	67.8	79.7	66.4	54.2	62.2	71.3	56.0
Marker density in markers/cM	0.2	0.2	0.5	0.4	0.3	0.1	0.2	0.5	0.5	0.4	0.4	0.3	0.3	0.4	0.5	0.4	0.3
Average distance between markers	5.2	4.1	2.1	2.7	3.2	7.4	4.4	2.0	1.9	2.8	2.5	3.0	3.0	2.7	2.2	2.5	3.1
Largest gap	22.1	20.9	21.6	9.0	9.6	23.6	14.7	10.2	9.6	10.4	6.4	20.1	17.8	16.9	10.7	14.0	26.4

mation, it is possible to clearly identify the chromosome regions under selection, carrying the under-represented alleles: F2-, D2+, D3+, F5-, D5+, D6+, D10+, D13+ and D14+.

Map integration

The integrated 'Fiesta' × 'Discovery' map (data not shown) consists of a total of 643 markers of which 550 are dominant for one parent, four are codominant with only two different alleles (abxab) and 89 markers segregate for three (abxac) or four (abxcd) different alleles and thus are fully informative. A genetic length of 1,371 cM is covered by the map and the average linkage group spans 81 cM with 38 markers. No significant changes in the marker order was observed and the succession of the codominant SSR markers was identical to the ones in the maps of the single parents, which is interpreted as another indication of the robustness of the genetic map.

Integration of parental maps by means of codominant markers has already been performed and discussed in detail by Maliepaard et al. (1998). Although the single parent maps can be used, an integrated map is advantageous for the analysis of quantitative trait loci especially where both parents contribute to the investigated trait (Knott and Haley 1992; Maliepaard and Van Ooijen 1994).

Comparison with MAPMAKER

Marker orders obtained with MAPMAKER generally corresponded with those obtained with JoinMap. Slight differences observed in marker order involved: (1) regions already identified by different orders compared to the previous 'Fiesta' × 'Discovery' map (Liebhard et al. 2003), indicated in Fig. 1., (2) markers tested on only a few (less than 50) individuals and therefore not being unambiguously positioned, and (3) regions densely populated with markers. However, the alternative marker orders obtained with the two mapping programs were only a little less likely. For the JoinMap order MAPMAKER indicates a log likelihood of between 0 and –1.5 compared to its own best order, i.e. equally likely to be approximately 30-times less likely.

Marker distances and linkage group lengths, however, were consistently larger with MAPMAKER/EXP than with JoinMap, using the same mapping function (Kosambi). Linkage group lengths were increased by MAPMAKER/EXP for up to 83% in 'Fiesta' and 96% in 'Discovery' with an average of 36% and 38%, respectively. This resulted in total MAPMAKER/EXP-genome lengths of 1,551 cM for 'Fiesta' and 1,845 cM for 'Discovery'.

Discussion

Automation

Using a Tecan pipetting robot proved to be extremely useful in high throughput applications like genetic mapping projects, where a large work load of routine tasks has to be accomplished. The biggest advantages of this mechanical pipetting help were its reliability, precise pipetting and no oversights in extensive routine tasks, as well as its speed. Multiple DNA quantifications, sample layout and PCR preparations could be executed simultaneously in a fraction of the time necessary to perform those tasks manually.

Extension and saturation of the genetic map

Apple linkage maps published to-date (Hemmat et al. 1994; Conner et al. 1997; Seglias and Gessler 1997; Maliepaard et al. 1998; Liebhard et al. 2003) consist of from 16 to 24 linkage groups. The most elaborated maps

Table 1C (continued)

C:																
Linkage groups																
D9	F10	D10	F11	D11	F12	D12	F13	D13	F14	D14	F15	D15	F16	D16	F17	D17
5	11	8	2	6	2	9	4	7	6	6	6	6	6	8	7	6
4	12	12	9	6	10	12	5	6	12	10	6	8	8	7	5	5
12	22	21	17	23	10	13	4	21	15	12	12	15	9	22	22	11
21	45	41	28	35	22	34	13	34	33	28	24	29	23	37	34	22
74.9	65.2	107.3	83.2	87.5	69.9	82.1	25.3	92.7	49.1	33.2	108.9	128.2	69.7	88.5	67.5	97.5
0.3	0.7	0.4	0.3	0.4	0.3	0.4	0.5	0.4	0.7	0.8	0.2	0.2	0.3	0.4	0.5	0.2
3.6	1.4	2.6	3.0	2.5	3.2	2.4	1.9	2.7	1.5	1.2	4.5	4.4	3.0	2.4	2.0	4.4
20.3	12.0	10.2	10.6	9.5	10.1	18.2	13.5	21.7	10.5	8.7	22.2	22.7	16.6	23.8	15.8	25.7

Table 2 Comparison of published apple genetic linkage maps with regard to the year of publication, the varieties mapped, map lengths, the average marker distance and the mapping software used

Authors and year of publication	Apple varieties	Map lengthin cM	Av. marker distance	Mapping software used
Weeden et al. 1994	White Angel	950	3.8	MAPMAKER
	Rome Beauty	950	6.1	
Conner et al.1997	Wijcik McIntosh	1206 (integrated)	5.1	JoinMap
	NY 75441-58	898	8.2	
	NY 75441-67	692	3.8	
Seglias and Gessler 1997	Iduna	548	6.1	JoinMap
	A679-2	690	4.6	
Maliepaard et al. 1998	Prima	842	4.3	JoinMap
	Fiesta	984	6.0	
Liebhard et al. 2002	Fiesta	914	4.6	JoinMap
	Discovery	1015	4.6	
Map Fig. 1	Fiesta	1,144	2.6	JoinMap
	Discovery	1,455	2.9	
	F×D Integrated	1,371	2.1	

are considered to be the ones organised in 17 linkage groups (Maliepaard et al. 1998; Liebhard et al. 2003), corresponding to the 17 apple chromosomes. Including a total of 329 markers (200 in 'Fiesta', 220 in 'Discovery' of which 91 appear on both maps) and covering 914.2 cM and 1,015.0 cM in 'Fiesta' and 'Discovery', respectively; the map of Liebhard et al. (2003) has so far been the most comprehensive linkage map of apple.

A total of 511 new AFLP, RAPD, SCAR and SSR markers (239 in 'Fiesta', 279 in 'Discovery', increasing the number of common markers from 91 to 98) were added to this map, increasing the marker density and overall map length. An increase of 229.7 cM (+25.1%) for 'Fiesta' and 439.6 cM (+43.3%) for 'Discovery' was achieved with the number of markers more than doubled. The relatively small map-length increase compared to the increase in the number of markers indicates that the present map is very close to full genome coverage.

Map construction

On the basis of the existing linkage map of the two apple cultivars Fiesta and Discovery (Liebhard et al. 2003), an updated map was constructed. To achieve such a genetic linkage map, a strategy allowing the detection of problematic linkage groups and a stepwise detection, correction and/or elimination of markers and individuals was applied (Liebhard and Gessler 2000).

Problematic linkage groups, requiring a third round in the JMMAP module, often corresponding with a high mean χ^2 goodness-of-fit, were identified. Erroneous marker scores within such linkage groups were identified by their LOD and χ^2 contribution or their distorted segregation, as revealed with the JMSLA module. Falsely genotyped individuals were identified using the JMCHK module which detects unlikely double crossovers or with the 'Find Recombinants' option of JMDesk which identifies clustered recombinants generally caused by gel-loading mistakes. Once those problematic linkage groups were detected, the markers were scored once more from the gel pictures and data were double-checked and compared with the original data files. When such results could be identified as mis-scorings or scorings of doubtful genotypes, they were corrected or excluded, respectively.

The division into subsets of less than 100 individuals proved to be very useful since the data files were kept small in size and easy to investigate. Additionally, the four subsets could be regarded as replications of the same experiment where the identical outcome greatly increases the certainty of the results.

JMDesk provides an option to automatically merge pairwise data files (*.pwd), which are the output of JMREC. The possibility to test data heterogeneity, analysing the merged subset files with the JMHET module of JoinMap, proved to be another advantage offered by the separate analysis of the subsets.

JMHET indicates differences in recombination frequencies of the same marker pair in the different subsets. With this module, two markers were identified that were positioned very close to each other in three subsets and quite far apart in the fourth, without causing a third round map, nor a large mean χ^2, nor any unlikely double crossovers, but only the different marker order in one subset. The problematic linkage group was identified only after comparing the subset maps and the recombination data of the involved markers using JMHET. A close investigation of the actual recombinants, showed them clustered and revealed an undetected loading mistake: eight samples, coinciding with the recombinants were loaded twice. Working with only one data set would never have allowed the detection of this kind of error.

The many options provided by JoinMap and JMDesk to investigate and refine input and output files during the mapping procedure are very powerful and improve the quality and reliability of the final genetic linkage map.

The final map presented in this paper is stable over a wide range of settings. But this was not the case for the first map calculated from a data set containing incorrect genotypes. During the analysis we observed cases where

faulty classification of a single individual caused an entire chromosome segment to change its orientation, illustrating the importance of high quality raw data files.

Marker distances and map length

Marker distances and linkage group lengths between the different subsets varied considerably whereas the order of markers, which is important, was generally the same. Although an increase in length of 230 cM and 440 cM in the map of 'Fiesta' and 'Discovery', respectively, has been achieved, the exact marker intervals, linkage groups and map lengths are considered to be of minor importance. It is well known that even small error rates in genotyping can lead to severe map inflation (Lincoln and Lander 1992). In the present study, changing one single data point in the matrix of 40 markers × 250 individuals (i.e. one linkage group; 40 × 250 = 10,000 data points) had considerable effects on the recombination frequencies and therefore on map distances, especially when operating close to the set thresholds or with high-density maps.

Since each marker can influence all distances in the linkage group, the length of the linkage group is a very sensitive measure. This is illustrated in Fig. 1 (compared with Fig. 1 in Liebhard et al. 2003) where linkage groups, to which terminal segments have been added, are still shorter than they were on the original map (D5), while other linkage groups increased in length without the addition of markers at their ends (F10, D15).

A comparison of map lengths and marker density with previously published apple maps is presented in Table 2. Good correspondence is observed between the number of markers tested (expressed as marker density) and map length, in fact a higher degree of saturation is often associated with increased cM-coverage. A comparison with maps from other *Rosaceae* is difficult since we did not include RFLP markers in the current apple map, and the few published maps from *Rosaceae* species including peach (Foolad et al. 1995; Dirlewanger et al. 1998; Joobeur et al. 1998), almond (Foolad et al. 1995; Viruel et al. 1995; Joobeur et al. 1998) and cherry (Wang et al. 1998) do not contain SSR markers which could be transferred to *Malus*.

Thresholds

A large number of thresholds have to be set by the user during the different steps of the map construction. Initially, we used a LOD score of 4.0 to group markers belonging to the same linkage group. That value is considered to set rather stringent conditions, especially for small data sets such as the 44-individuals subset. In fact, in such a data set a LOD score of 4.0 represents a recombination frequency of 0.19 whereas in a subset of 90 individuals a LOD score of 4.0 represents a recombination frequency of approximately 0.27. While other authors (Hemmat et al. 1994; Seglias and Gessler 1997) used a LOD score of 3.0, we preferred to set more stringent conditions and to use a LOD score of 4.0 as previously published by Conner et al. (1997) and Maliepaard et al. (1998).

On the other hand, during the subsequent mapping procedure the thresholds were very loose (LOD = 0.001; REC = 0.499) in order to use all the information available. This allows markers from one end of the linkage group to influence distances between markers at the other end. With dense maps this will inevitably lead to conflicts since markers can hardly be ordered without contradictions. However, if we would set more restrictive thresholds, information outside the thresholds (i.e. rec. freq. > threshold, LOD score < threshold) would be neglected. As a consequence, large but true recombination frequencies would be discarded and markers would be placed too close to each other. This too would lead to conflicts.

Saturation mapping and genome coverage estimation

The use of the AFLP technique to saturate apple linkage maps has been reported before (Xu and Korban 2000) and has been successfully applied here to update and integrate the previously published linkage maps of 'Fiesta' and 'Discovery'. In a relatively short time a large number of primer combinations were screened, the promising combinations identified, tested and mapped on the entire progeny. The present map will considerably facilitate future mapping in apple. The core of the map consists of a selection of codominant SSR markers; their position is known, and the gaps between markers can easily and quickly be filled with dominant AFLP markers, considerably reducing the time required for the construction of a new linkage map.

Expressions like 'saturated' and 'high-density' when referring to genetic linkage maps are used somewhat inconsistently. Maps described as 'saturated', range from an average marker distance of 8 cM in rubber tree (Lespinasse et al. 2000) to 2 cM in *Prunus* (Joobeur et al. 1998). The same is true for the term 'high density' used to describe map densities ranging from an average marker distance of 2.6 cM in *Brassica* (Sebastian et al. 2000), over 2.1 cM in Cacao (Risterucci et al. 2000) to less than 1 cM in maize (Vuylsteke et al. 1999) and tomato (Haanstra et al. 1999). However knowing, that such average measures are of limited use, since linkage maps tend to be denser in centromeric regions than in regions close to the chromosome telomeres, the term 'saturated' should be regarded as 'completely covered with markers', emphasising a good coverage of the linkage group ends rather than gaps.

Several methods to estimate genome coverage have been proposed (Hulbert et al. 1988; Chakravarti et al. 1991). Since one of the assumptions for the estimation is the uniform distribution of the marker loci and incomplete genome coverage, the estimate will always be larger than the covered map distance, no matter how close to complete coverage the actual map is.

An estimation of the genome length according to Hulbert/Chakravarti (method 3), based on the data of the first two subsets without AFLPs used for the construction of the map in Liebhard et al. (2003), predicted a map length of 1,500 cM and 1,600 cM for 'Fiesta' and 'Discovery', respectively. Entering the data of the current map, the estimated lengths are 1,850 cM and 2,100 cM for the two cultivars. Genome length estimation according to Remington et al. (1999), which corrects the overestimation of map length at the chromosome ends in the Hulbert/Chakravarti formula, assumes complete coverage already in the first map (Liebhard et al. 2003). Estimates were 930 cM for 'Fiesta' and 1,100 cM for 'Discovery', while 915 cM and 1,015 cM were actually achieved with JoinMap for the previous map. An estimation according to Remington with the present data predicts nearly the same (940 cM for 'Fiesta' and 1,100 for 'Discovery'), even though a considerable increase in map length was achieved with the newly generated markers (Fig. 1, Table 2). The demonstrated variation in estimates clearly shows that not only map length has to be taken cautiously but also genome length estimates are not beyond doubt.

Segregation distortion and differences in male and female recombination frequencies

Markers with distorted segregation clustered on particular linkage groups have repeatedly been reported in a variety of crops (Landry et al. 1991; Jarrell et al. 1992; Prince et al. 1993; Cai et al. 1994; Vuylsteke et al. 1999). This clustering is not surprising if it is assumed that sub-lethal genes are present on those chromosomal regions.

Distorted segregation of an entire set of markers can therefore be regarded as the expression of the phenotypic trait 'low viability'. Genes coding for these traits are located where the distorted markers map. The effect of such disadvantageous alleles can be estimated by the extent of the distortion. In our case, genetic factors on LG D10 and D13 seem to have a larger effect than the others observed. It is also clear that distortions sometimes affect only one parent since some traits are inherited from one parent (D3, D6, D10, D14) and some traits from both (F2-D2, F5-D5).

Whether the under-representation of certain alleles is the expression of genes conferring "low viability", on which level this low viability is expressed (zygote, embryo, seedling), or whether egg and pollen formation or function are concerned cannot be answered here. This would require an in-depth-analysis of embryogenesis and cytological studies during meiosis and pollen development. It is noteworthy that the segregation distortions do not coincide with the reported self-incompatibility locus S on LG 17 (Maliepaard et al. 1998).

Jarrell et al. (1992) stated that linkage group segments with highly distorted markers may be mapped inaccurately. However, this cannot be confirmed with the results obtained here. Linkage groups with distorted markers often have a non-distorted, accurately mapped counterpart (F3-D3, F6-D6, F10-D10, F14-D14) that not only displays an identical order of codominant SSR markers but also corresponding distances between them. Only the upper part of linkage group D13, which represents the largest terminal extension of a chromosomal segment from the previous map, has to be treated with caution since no counterpart exists to verify its accuracy.

Some researchers discard markers deviating significantly from the expected Mendelian ratio (Ellis et al. 1992; Castiglioni et al. 1998; Marques et al. 1998). However, this seems to be unnecessary and in fact a waste of valuable data. The exclusion of distorted marker data would have led to the loss of half of linkage group D10, which seems to be mapped correctly, according to the comparison with the corresponding marker orders and distances of the undistorted 'Fiesta' homologue.

The identical order of SSR markers on both parental maps implies that their positions are correct and transferable between cultivars (Liebhard et al. 2003). The use of the molecular markers allows exploration of recombination frequencies in corresponding genetic regions of different cultivars and between male and female gametes. No discrepancies in recombination frequencies between the two parental cultivars were observed. Most of the differences in distance between corresponding markers in 'Fiesta' and 'Discovery', visible on the map (Fig. 1), are caused by the algorithms used in the map construction (mapping artefacts) and they do not represent actual differences in recombination frequencies. For example CH04e03 and CH02h11a on LG 4 have map distances of 20.7 cM in 'Fiesta' and 34 cM in 'Discovery' but their effective recombination frequencies are 0.21 and 0.26. The map distances are influenced by neighbouring markers that in one case compress and in the other case stretch the corresponding map interval.

Comparison of JoinMap and MAPMAKER

MAPMAKER has been reported to produce systematically larger linkage groups with identical data by other authors (Cai et al. 1994; Castiglioni et al. 1998; Vuylsteke et al. 1999). One possible explanation is the use of different mapping algorithms, different ways of applying the same mapping function, and usage of different amounts of available information for map calculations. In addition, slightly different marker orders and distances should be expected when using different mapping software and even with different settings of the same program, and this was the case in the present study.

In large data sets, like those used in mapping projects, small changes of the thresholds affect the statistical calculations and the output results.

In our experience the settings that were optimal for one program were suboptimal for the other. However, only minor differences occurred in marker orders, and the differences in length were systematic. We considered the data set and the output maps to be robust and reliable and concluded that the two programs deliver comparable

results. The major and most relevant differences are the many possibilities provided by JoinMap, to check input files and analyse intermediate data, especially when used in conjunction with JMDesk. These options are a great advantage with large data files where small errors can hardly be found without such functions.

The question arises whether an indication of the achieved map length is truly meaningful. As we have seen, many factors (segregation type, error rate in data files, mapping function, mapping algorithm, thresholds) are influencing the marker distances and therefore the map length, making it an extremely sensitive estimate. Much more important is the degree of saturation of a map and the fraction of the genome covered with markers.

Acknowledgements We thank Dr. Markus Kellerhals from FAW, Wädenswil, for providing leaf material of the progeny plants and the parental cultivars. We also acknowledge Andrea Jegerlehner, Kirstin Knapp, Daniel Füger, Giovanni Broggini, Angela Tenzer, Georg Feichtinger, Eve Dilworth and Barbara Bigler for their help in the lab, and Tecan Switzerland AG, Männedorf, for providing the liquid handling and fluorescence measurement devices and the support with their use. Eve Dilworth and Dr. Graham King are acknowledged for critically reviewing and linguistically improving the manuscript. This research was funded by the Swiss National Foundation for scientific research, SPP-Biotechnology, grant No. 5002-045052 and No. 5002-057830.

References

Cai Q, Guy CL, Moore GA (1994) Extension of the linkage map in *Citrus* using random amplified polymorphic DNA (RAPD) markers and RFLP mapping of cold-acclimation-responsive loci. Theor Appl Genet 89:606–614

Castiglioni P, Pozzi C, Heun M, Terzi V, Müller KJ, Rohde W, Salamini F (1998) An AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley. Genetics 149:2039–2056

Chakravarti A, Kasher LK, Reefer JF (1991) A maximum likelihood method for estimating genome length using genetic linkage data. Genetics 128:175–182

Conner JP, Brown SK, Weeden NF (1997) Randomly amplified polymorphic DNA-based genetic linkage maps of three apple cultivars. J Am Soc Hort Sci 122:350–359

Dirlewanger E, Pronier V, Parvery C, Rothan C, Guye A, Monet R (1998) Genetic linkage map of peach [*Prunus persica* (L.) Batsch] using morphological and molecular markers. Theor Appl Genet 97:888–895

Ellis THN, Turner L, Hellens RP, Lee D, Harker CL, Enard C, Domoney C, Davies DR (1992) Linkage maps in pea. Genetics 130:649–663

Foolad MR, Arulsekar S, Becerra V, Bliss FA (1995) A genetic map of *Prunus* based on an interspecific cross between peach and almond. Theor Appl Genet 91:262–269

Gianfranceschi L, Seglias N, Tarchini R, Komjanc M, Gessler C (1998) Simple sequence repeats for the genetic analysis of apple. Theor Appl Genet 96:1069–1076

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

Haanstra JPW, Wye C, Verbankel H, Meijer-Dekens F, van den Berg P, Odinot P, van Heusden AW, Tanksley S, Lindhout P, Peleman J (1999) An integrated high-density RFLP-AFLP map of tomato based on two *Lycopersicon esculentum* ×*L. pennellii* F2 populations. Theor Appl Genet 99:254–271

Hemmat M, Weeden NF, Manganaris AG, Lawson DM (1994) Molecular marker linkage map for apple. J Hered 85:4–11

Hulbert SH, Ilott TW, Legg EJ, Lincoln SE, Lander ES, Michelmore RW (1988) Genetic analysis of the fungus *Bremia lactucae*, using restriction fragment length polymorphism. Genetics 120:947–958

Jarrell DC, Roose ML, Traugh SN, Kupper RS (1992) A genetic map of *Citrus* based on the segregation of isozymes and RFLPs in an intergeneric cross. Theor Appl Genet 84:49–56

Joobeur T, Viruel MA, de Vicente MC, Jauregui B, Ballester J, Dettori MT, Verde I, Truco MJ, Messeguer R, Batlle I, Quarta R, Dirlewanger E, Arus P (1998) Construction of a saturated linkage map of *Prunus* using an almond ×peach F2 progeny. Theor Appl Genet 97:1034–1041

Knott SA, Haley CS (1992) Maximum likelihood mapping of quantitative trait loci using full-sib families. Genetics 132:1211–1222

Koller B, Tenzer I, Gessler C (2000) SSR analysis of apple scab lesions. Integrated control of pome fruit diseases. IOBC/WPRS Bulletin Vol 23: pp 93–98

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

Landry BS, Hubert N Etoh T, Harada JJ, Lincoln SE (1991) A genetic map for *Brassica napus* based on restriction fragment length polymorphisms detected with expressed DNA sequences. Genome 34:543–552

Lespinasse D, Rodier-Goud M, Grivet L, Leconte A, Legnate H, Seguin M (2000) A saturated genetic linkage map of rubber tree (*Hevea* spp.) based on RFLP, AFLP, microsatellite and isozyme markers. Theor Appl Genet 100:127–138

Liebhard R, Gessler C (2000) Possible errors in genome mapping. Integrated control of pome fruit diseases. IOBC/WPRS Bulletin 23:127–135

Liebhard R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van de Weg E, Gessler C (2002) Development and characterisation of 140 new microsatellites in apple (*Malus* ×*domestica* Borkh.) Mol Breed 10:217–241

Lincoln ES, Lander ES (1992) Systematic detection of errors in genetic linkage data. Genomics 14:604–610

Maliepaard C, Van Ooijen JW (1994) QTL mapping in a full-sib family of an outcrossing species. In: Van Ooijen JW, Jansen J (eds) Biometrics in plant breeding: applications of molecular markers, Proc 9th Meeting EUCARPIA Section Biometrics in Plant Breeding, 6–8 July 1994, Wageningen, The Netherlands, pp 140–146

Maliepaard C, Jansen J, Van Ooijen JW (1997) Linkage analysis in a full-sib family of an outbreeding species: overview and consequences for applications. Genet Res 70:237–250

Maliepaard C, Alston FH, Van Arkel G, Brown LM, Chevreau E, Dunemann F, Evans KM, Gardiner S, Guilford P, van Heusden AW, Janse J, Laurens F, Lynn JR, Manganaris AG, Den Nijs APM, Periam N, Rikkerink E, Roche P, Ryder C, Sansavini S, Schmidt H, Tartarini S, Verhaegh JJ, Vrielink-Van Ginkel M, King GJ (1998) Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theor Appl Genet 97:60–73

Marques CM, Araujo JA, Ferreira JG, Whetten R, O'Malley DM, Lui BH, Sederoff R (1998) AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. Theor Appl Genet 96:727–737

Prince JP, Pochard E, Tanksley SD (1993) Construction of a molecular linkage map of pepper and a comparison of synteny with tomato. Genome 36:404–417

Remington DL, Whetten RW, Liu BH, O'Malley DM (1999) Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. Theor Appl Genet 98:1279–1292

Risterucci AM, Grivet L, N'Goran JAK, Pieretti I, Flament MH, Lanaud C (2000) A high-density linkage map of *Theobroma cacao* L. Theor Appl Genet 101:948–955

Sankar AA, Moore GA (2001) Evaluation of inter-simple sequence repeat analysis for mapping in *Citrus* and extension of the genetic linkage map. Theor Appl Genet 102:206–214

Sebastian RL, Howell EC, King GJ, Marshall DF, Kearsey MJ (2000) An integrated AFLP and RFLP *Brassica oleracea* linkage map from two morphologically distinct doubled-haploid mapping populations, Theor Appl Genet 100:75–81

Seglias NP, Gessler C (1997) Genetics of apple powdery mildew resistance from *Malus zumi* (Pl2). Integrated control of pome fruit diseases. IOBC/WPRS Bulletin 20:195–208

Stam P, Van Ooijen JW (1995) JoinMap version 2.0: software for the calculation of genetic linkage maps. Plant Research International, Wageningen, The Netherlands

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Biotechnology 7:257–264

Van der Lee T, De Witte I, Drenth A, Alfonso C, Govers F (1997) AFLP linkage map of the oomycete *Phytophthora infestans*. Fungal Genet Biol 21:278–291

Viruel MA, Messegeur R, de Vicente MC, Garcia-Mas J, Puigdomenech P, Vargas F, Arus P (1995) A linkage map with RFLP and isozyme markers for almond. Theor Appl Genet 91:964–971

Voorrips RE (2001) MapChart version 2.0: Windows software for the graphical presentation of linkage maps and QTLs. Plant Research International, Wageningen, The Netherlands

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23, pp 4407–4414

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lübberstedt T, Xia XC, Stam P, Zabeau M, Kuiper M (1999) Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet 99:921–935

Wang D, Karle R, Brettin TS, Iezzoni AF (1998) Genetic linkage map in sour cherry using RFLP markers. Theor Appl Genet 97:1217–1224

Weeden NF, Hemmat M, Lawson DM, Lodhi M, Bell RL, Manganaris AG, Reisch BI, Brown SK, Ye GN (1994) Development and application of molecular marker linkage maps in woody fruit crops. Euphytica 77:71–75

Xu ML, Korban SS (2000) Saturation mapping of the apple scab resistance gene *Vf* using AFLP markers. Theor Appl Genet 101:844–851