Tree Genetics & Genomes (2011) 7:857–868
DOI 10.1007/s11295-011-0380-8

ORIGINAL PAPER

Genetic diversity of the genus *Malus* and implications for linkage mapping with SNPs

Diego Micheletti · Michela Troggio · Andrey Zharkikh · Fabrizio Costa · Mickael Malnoy · Riccardo Velasco · Silvio Salvi

Received: 20 September 2010 / Revised: 27 January 2011 / Accepted: 8 February 2011 / Published online: 19 March 2011

Abstract Knowledge about the sequence-based genetic diversity of a crop species is important in order to develop highly informative genotyping assays, which will eventually positively impact breeding practice. Diversity data were obtained from two pools of 185 and 75 accessions each, representing most of the species belonging to the genus *Malus*, by re-sequencing 27 gene-specific amplicons and by screening 237 *Malus* × *domestica* SNPs using the multiplex genotyping technology SNPlex™. Nucleotide diversity and insertion/deletion rates in *M.* × *domestica* were estimated as π=0.0037 and 1/333 bp, respectively. The SNP frequency was estimated as 0.0194 (1 SNP/52 bp) while within a single apple cultivar an average of one SNP in every 455 bp was found. We also investigated transferability (T_{SNP}) of the heterozygous state of SNPs across the species *M.* × *domestica* and the genus *Malus*. Raw re-sequencing showed that 12–15% of *M.* × *domestica* SNPs are transferable to a second *M.* × *domestica* cultivar, however T_{SNP} rose to ~41% with SNPs selected for high minor allele frequency. T_{SNP} of chosen SNPs averaged ~27% in the two *M.* × *domestica*-related species, *Malus sieversii* and *Malus sylvestris*, but was much lower in more distantly related species. On the basis of T_{SNP}, simulations, and empirical results, we calculated that a close-design, multiplexed genotyping array with at least 2,000 SNPs is required for building a highly saturated linkage maps within any *M.* × *domestica* cross. The same array would gradually lose informativeness in increasingly phylogenetically distant *Malus* species.

Keywords Apple · High-throughput genotyping · SNP

Communicated by E. Dirlewanger

Electronic supplementary material The online version of this article (doi:10.1007/s11295-011-0380-8) contains supplementary material, which is available to authorized users.

D. Micheletti · M. Troggio (✉) · F. Costa · M. Malnoy · R. Velasco · S. Salvi
Istituto Agrario San Michele all'Adige,
Research and Innovation Centre, Foundation Edmund Mach,
Trento, Italy
e-mail: michela.troggio@iasma.it

A. Zharkikh
Myriad Genetics,
Salt Lake City, UT, USA

S. Salvi
Department of Agroenvironmental Sciences and Technologies,
University of Bologna,
Bologna, Italy

Introduction

In many crop species, single nucleotide polymorphism (SNP)-based genotyping has been adopted as the approach of choice for evaluating genetic diversity and for trait genetic mapping (Rafalski 2002; Gupta et al. 2008). The main factors responsible for the success of SNP technologies are the ubiquity of this type of genetic variation, easy standardisation of results thanks to the inherently biallelic nature of SNPs, and the possibility of increasing throughput by multiplexing (Syvänen 2005; Fan et al. 2006; Laframboise 2009). However, there are constraints on the implementation of SNP genotyping assays. The first limitation is the presence of chromosome segment duplications and/or polyploidization, which has been shown to give rise to spurious genotype calls (Hyten et al. 2008; Bérard et al. 2009). Secondly, implementation of SNP assays may be restricted by the need to have a conserved genomic region (the length varying according to the

technology employed) around each target SNP (Drost et al. 2009; Jones et al. 2009). Thirdly, SNPs are the product of the very low mutation rate of the species and may not segregate in many sub-populations, resulting in a negative impact on assay informativeness. The significance of these constraints is magnified by the need for the main SNP-based genotyping technologies to be 'closed systems', that is, the SNPs to be used must be identified a priori of genotyping, as the cost of producing the array precludes re-assembly with a different SNP set in each experiment. The need to define a universal core set of SNPs for genetic studies in a given species has been recognised and extensively investigated in human genetics (for review, see Gu et al. 2008).

An additional key parameter in ensuring the efficiency of trait genetic mapping in a heterozygous fruit crop species is SNP transferability (T_{SNP}), that is, the probability that a nucleotide position known to be heterozygous in one accession (i.e. a SNP) is heterozygous in a second accession (Vezzulli et al. 2008). Availability of heterozygous SNPs is important because most experimental cross populations are generated as F_1 crosses between two heterozygous parents and genetically informative meioses are only produced between heterozygous markers. T_{SNP} across species, especially where inter-specific crossing is common practice for breeding and genetic studies, should also be critically considered.

The cultivated apple (*Malus* × *domestica* Borkh.) is a highly heterozygous, self-incompatible, diploid fruit crop species belonging to the Rosaceae family and is cross fertile with all other genus members (approximately 30–50 species; Janick et al. 1996; Robinson et al. 2001). Because of the wide cross fertility, inter-specific crossing is a relatively common apple breeding practice (Brown 1992). The completion of the genome sequence of the cultivar 'Golden Delicious' (Velasco et al. 2010) revealed that the 17-chromosome 750-Mb genome underwent a relatively recent (50–60 Ma ago) whole-genome duplication. Analysis of molecular genetic distances within the genus *Malus* confirmed the close genetic relationship between the cultivated apple and the wild *Malus sieversii* from central Asia (Juniper and Mabberley 2006; Velasco et al. 2010).

Following analysis of a representative collection of the genus *Malus* and the species *M.* × *domestica*, we provide here estimates of sequence-based genetic diversity, results from the first test of a SNP-based genotyping array (SNPlex™; Tobler et al. 2005), and predictions concerning the number of SNPs required for the construction of linkage maps in apple.

Materials and methods

Plant materials

Two germplasm sets were analysed (Table 1 in the Electronic supplementary materials). Germplasm set-1 (a total of 185 accessions) consisted of 123 cultivars of *M.* × *domestica* (including 119 elite, founders and old cultivars and four rootstock cultivars), 12 accessions of species of the series Pumilae (*Malus sylvestris, M. sieversii, Malus pumila, Malus prunifolia, Malus* × *asiatica, Malus micromalus, Malus orientalis* and *Malus spectabilis*) (Janick et al. 1996), 25 accessions of 17 *Malus* species more distantly related to *M.* × *domestica*, and 25 accessions representing hybrids or uncertain entries of *Malus*. Germplasm set 2 (74 accessions) has already been described in Velasco et al. (2010). Briefly, it consisted of 72 accessions of *Malus*, including 12 *M.* × *domestica* cultivars, ten accessions of *M. sieversii* and 21 of *M. sylvestris* (one shown to be a putative hybrid with *M.* × *domestica*; Velasco et al. 2010), accessions covering 29 of the 34 *Malus* species (according to Janick et al. 1996) and two *Pyrus* species.

Table 1 Survey of *Malus domestica* SNP diversity based on amplicon re-sequencing

Dataset	Amplicons (no.)	Cultivars (no.)	Length per cv.[a] (bp)	SNPs (no.)	S[b] (SNPs/bp)	$S_{>0.1}$[c] (SNPs/bp)	π[d]	π_w[e]	SNPs in founders[f] (no.)	Percent
1	8	123	3,386	74	0.022 (1/45)	0.009 (1/106)	0.0036	0.0035 (1/286)	35	47.3
2	23	12	11,300	181	0.017 (1/59)	0.011 (1/91)	0.0038	0.0018 (1/556)	136	75.1
Overall	27	–	–	255	0.019 (1/52)	0.010 (1/99)	0.0037	0.0022 (1/455)	171	67.1

[a] Length of investigated sequences, as concatenation of different amplicons, per cultivar

[b] Proportion of segregating nucleotides over the total number of nucleotides

[c] Proportion of segregating nucleotides with MAF>0.1 over the total number of nucleotides

[d] Nucleotide diversity

[e] Mean nucleotide diversity of single cultivars, equivalent to the SNP frequency within one cultivar

[f] Founders=the five cultivars (Cox's Orange Pippin, Golden Delicious, Jonathan, McIntosh and Red Delicious) recognised as founders of modern apple breeding (Noiton and Alspach 1996)

DNA preparation and whole-genome amplification

Between 50 and 100 mg of young leaves from each accession were freeze-dried and used for genomic DNA extraction with the DNeasy 96 Plant Mini Kit (Qiagen, Valencia, CA, USA). Ten nanograms of genomic DNA from each sample were then amplified by whole-genome amplification (WGA; Hosono et al. 2003) using the GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Little Chalfont, Buckinghamshire, UK). The success of the WGA reaction and the absence of product in the negative control samples were assessed by agarose gel electrophoresis.

Sequencing analysis and production of datasets 1 and 2

Dataset 1 included sequences and SNPs extracted from eight gene fragments re-sequenced over 119 cultivars of *M.* × *domestica* belonging to Germplasm set 1. Dataset 2 included sequences and SNPs extracted from 23 gene fragments re-sequenced over Germplasm set 2. The two datasets had four amplicons in common out of a total of 27 sequenced amplicons. A list of the amplicons and details are given in Table 2 in the Electronic supplementary materials. Sequences were obtained by re-sequencing gene fragments (one PCR amplicon per gene). Amplicons had an average length of approximately 500 bp and were selected to be approximately evenly distributed on the genetic and genome maps of 'Golden Delicious'. Gene function was not a criterion for selecting the SNP containing amplicons. The PCR primers were designed based on the 'Golden Delicious' genomic sequence using the Primer3 software (Rozen and Skaletsky 2000) in accordance with the following criteria: (1) forward and reverse primers 200–300 bp upstream and downstream of the target SNP; (2) primer size between 18 and 25 bases; (3) primer melting temperature (T_m) between 59 and 61°C. PCR and sequencing was performed as in Vezzulli et al. (2008). The DNA sequences were clustered and analysed with the Pregap4/Gap4 software package (Staden Package; Staden et al. 2000). All sequences were aligned with T-coffee (Notredame et al. 2000) for SNP and insertion/deletion (indel) discovery. The proportion of segregating nucleotides over the total number of investigated nucleotides (S) was computed as defined in Hartl and Clark (2007). Estimations of nucleotide diversity (π) and minor allele frequencies (MAF) were carried out using Arlequin 3.11 (Excoffier et al. 2005). Only one clone per cultivar was considered in the computation of the diversity indexes.

Design of the SNP array (SNPlex™) and dataset 3

Dataset 3 included the results of SNP genotyping of Germplasm set 1 after design and application of five SNPlex™ (Applied Biosystems Inc., Foster City, CA) arrays. At an early stage of sequencing and assembly (coverage=approximately 4×) of the 'Golden Delicious' apple cultivar genome (Velasco et al. 2010), the SNP search was begun by matching the apple ESTs available in Genbank (256,222 ESTs derived from 25 apple cultivars, at SNP search date 31 December 2008) to genomic sequencing reads in order to focus on the gene-containing regions. The chosen reads were then assembled into gene-containing contigs and the contig positions with ambiguous bases were recorded as potential SNP sites. The repetitiveness of each SNP site was estimated by matching the contigs back to the original reads. Only positions with a number of hits <4 were selected. As there is an overlap in the distributions of the coverage of unique and of duplicated regions, a small number of positions located in duplicated segments were also selected. These cases were identified and removed during the mapping stage. An additional filter required that no other SNPs be present in a 40-base window around the selected position and no more than five other SNPs be present in a 200-base window.

A total of approximately 1,400 SNPs were identified in the 'Golden Delicious' genomic contigs corresponding to coding or neighbour-to-coding sequences, as verified by comparison with the *Arabidopsis* genome sequence. The

Table 2 Types of detected sequence polymorphisms from amplicon re-sequencing

	Transition		Transversions				Total
	A↔G	C↔T	A↔C	A↔T	C↔G	G↔T	
Coding							
Dataset 1	8	10	5	4	2	3	32
Dataset 2	12	13	3	3	3	5	39
Overall (%)	27.9	32.3	11.7	10.1	7.0	11.1	100.0
Non-coding							
Dataset 1	5	14	9	6	2	10	46
Dataset 2	35	33	23	19	20	15	145
Overall (%)	17.5	26.6	17.7	13.1	9.1	16.0	100.0

first 260 SNPs of this list and an additional eight SNPs from within the genes of other cultivars under investigation (not shown) were entered into the SNPlex automated multi-step pipeline (https://ms.appliedbiosystems.com/snplex/snplexStart.jsp); briefly, this involved (1) entering a genome sequence of approximately 200 nucleotides surrounding the target SNP (100 upstream and 100 downstream), and (2) checking for the presence of non-target SNPs near the target and/or sequence motifs incompatible with the assay. All other details were as reported in Pindo et al. (2008). The SNPlex™ genotyping assays (Tobler et al. 2005) were carried out on 1 μl (from 45 to 225 ng) of fragmented GenomiPhi amplified gDNA (WGA-DNA) on Germplasm set 1, according to the manufacturer's protocol. The samples were run on a 3730xl DNA Analyzer (Applied Biosystems Inc.) and the data were analysed using the Gene Mapper v.4.0 software (Applied Biosystems Inc.). Genotype analysis was performed according to the SNPlex_Rules_3730 method, in accordance with the manufacturer's default settings, including quality thresholds for genotype calling (http://www3.appliedbiosystems.com/sup/URLRedirect/index.htm?xDoD=4360856).

SNP transferability

T_{SNP} is defined here as the frequency at which a given nucleotide known to be heterozygous in a first accession is heterozygous in a second accession. T_{SNP} can therefore be computed for a given SNP or for a list of SNPs by considering the proportion of germplasm accessions which are heterozygous at the SNPs tested. Cultivar clones were considered when the *M.* × *domestica* T_{SNP} was computed. T_{SNP} is closely related to the standard population genetics parameter H_0 (observed heterozygosity; Hartl and Clark 2007) at the nucleotide level. However, with T_{SNP}, the initial condition of the nucleotide is assumed to be heterozygous. According to the Hardy–Weinberg equilibrium, the probability of detecting a heterozygous individual at a known segregating SNP is simply 2 pq, where p and q are the population frequencies of the two nucleotide alleles; therefore T_{SNP} correlates with the MAF, while the maximum T_{SNP} is 0.5 when p=q=0.5. Datasets 1, 2 and 3 were used for T_{SNP} estimation.

Cultivar resampling for capturing SNP

In order to test whether and to what extent SNP identification can be maximised by selecting the appropriate set of *M.* × *domestica* cultivars, we carried out a re-sampling experiment on dataset 3. The approach used here is similar to the one used by Yan et al. (2009), but with a much higher rate of re-sampling. The SNP progression curve was obtained by computing and plotting averages (with standard deviations) and maximum and minimum numbers of SNPs in each re-sampled subgroup, 10,000 sub-groups having been randomly chosen for each of the different group sizes (ranging from five to 105 cultivars).

Simulation of SNP number requirements for linkage map construction

We applied a simulation approach in order to extrapolate the number of SNPs required to build genetic linkage maps. The simulation was done with a Microsoft Excel™ VBA-based script purposely written and available upon request. Each simulated map consisted of one chromosome of 100 cM. Populations were assumed to be F_1 intercross type (obtained by crossing two heterozygous parent plants) with 100 individuals (one hundred individuals may be considered the lower dimension bound for gene/QTL mapping; larger populations would have provided greater statistical power for linkage analysis, thus decreasing the number of SNPs required). Map positions of SNPs and crossovers were simulated as randomly distributed across each simulated chromosome. Only backcross type SNP markers (e.g. Aa×aa) were considered; occurrences of poorly informative double-heterozygous SNPs (e.g. Aa×Aa) were included in the simulation and, when recorded, were excluded from the linkage map construction. Four SNP densities (from 25 to 200 per chromosome) and five T_{SNP} levels (from 0.1 to 0.5) were considered in the simulation. For each simulated chromosome, map coverage (expressed as a percentage of the reference chromosome) was computed on the basis of chromosome segments with at least two linked SNPs and the total cM distance between the top and the bottom polymorphic SNPs, excluding any intermarker distance of >30 cM. Five hundred maps were simulated per combination of number of SNPs and T_{SNP}.

Results

DNA sequence diversity within *M.* × *domestica*

Dataset 1 included the aligned nucleotide sequence of eight amplicons (representing 3,386 bp per cultivar), each derived from a different gene (Table 2 in the Electronic supplementary materials), from each of 119 apple cultivars, for a total of approximately 390,000 bp. Within this dataset, the proportion of segregating nucleotide sites over the total number of nucleotides investigated (S) was 0.0224, equivalent to a SNP frequency of 1 SNP/45 bp (Table 1). Nucleotide diversity (π), that is the likelihood of nucleotide substitution when two allelic chromosomes are randomly drawn from the population and compared, was 0.0036. When averaged with dataset 2, S and π were 0.0194 (1 SNP/

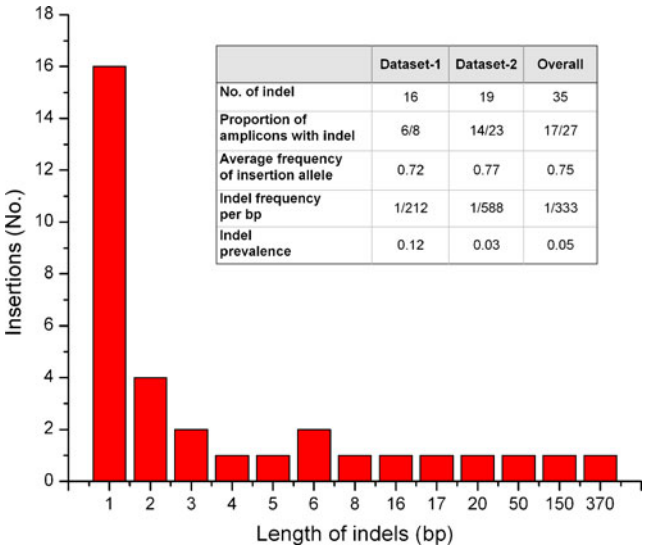
52 bp) and 0.0037, respectively, while within-cultivar π (π_W), that is the SNP frequency observed in a single *M.* × *domestica* cultivar, averaged 0.0022 (1 SNP/455 bp). When only SNPs with MAF>0.1 were considered, overall SNP frequency reduced to 1 SNP/99 bp. Gene-wise, MDE00525.3 was the most diverse amplicon (π=0.0062) and MDE01960.1 was the least diverse (π=0.0020).

The genomic sequence investigated consisted of 1,546 (46%) and 1,840 bp (54%) coding and non-coding sequences respectively. An average of one SNP every 40 bp was observed in non-coding regions whereas one SNP every 48 bp was observed in coding regions. . The most frequent type of polymorphism was the C–T transition, both for coding and non-coding SNPs (32.3% and 26.6% of SNPs, respectively), and the least frequent was the C–G transversion, again for both coding and non-coding SNPs (7.0% and 9.1%, respectively) (Table 2).

We also identified a number of indels. Single-base indels were by far the most frequent (16 out of a total of 35 indels; Fig. 1). Six amplicons in dataset 1 and 14 in dataset 2 contained at least one indel. For each gene fragment, we identified one indel every four accessions, the overall indel frequency being 1/333 bp. Ca. 5% of the surveyed genomic sequence was involved in insertion or deletion polymorphisms in at least one accession.

Design and application of SNP arrays in apple

A list of 260 SNPs was obtained from a preliminary assembly of the 'Golden Delicious' genome sequence. Of these SNPs, 237 (88%) passed the SNPlex assay design pipeline and were used to produce five SNPlex arrays (Table 3). The five arrays were used to genotype the 185 *Malus* samples of Germplasm set 1, including 123 *M.* × *domestica* cultivars. Overall, 187 out of 237 SNPs (78.9%) satisfied the SNPlex™ quality evaluation and were therefore considered successful, while 50 (21.1%) failed. When only *M.* × *domestica* accessions were considered, 199 (84%) SNPs were successful and 38 (16%) failed. Of the failed SNPs, four (~2%) were false SNPs erroneously predicted at the SNP identification phase (two SNPs called as homozygous in all samples, including the three 'Golden Delicious' *M.* × *domestica*; two additional SNPs were found to be heterozygous in all accessions and therefore probably originated from paralogous sequences). Preferential annealing of the probes triggered by additional unknown SNPs near the target site was one of the probable causes of the failure of the remaining 34 SNPs (~14%) (Pindo et al. 2008).



	Dataset-1	Dataset-2	Overall
No. of indel	16	19	35
Proportion of amplicons with indel	6/8	14/23	17/27
Average frequency of insertion allele	0.72	0.77	0.75
Indel frequency per bp	1/212	1/588	1/333
Indel prevalence	0.12	0.03	0.05

Fig. 1 Frequency distribution and statistics for insertion/deletion (indel) polymorphisms detected by amplicon re-sequencing within *M.* × *domestica*

SNP allele frequencies

We computed MAF frequencies as a further parameter of SNP informativeness. A low MAF (MAF<0.2) was observed in a high proportion of the SNPs identified by re-sequencing (74% and 58% of datasets 1 and 2, respectively) (Fig. 2). On the other hand, *M.* × *domestica* SNPs in the SNP arrays (dataset 3) generally had a much higher MAF (only 25% with MAF<0.2). High proportions (36.3%, 44.6% and 23.1% for datasets 1, 2 and 3, respectively) of SNP genotype frequencies did not fit the Hardy–Weinberg equilibrium.

Transferability of *M.* × *domestica* SNPs

Within dataset 3, T_{SNP} of *M.* × *domestica* SNPs was extremely high (>94%) in independent 'Golden Delicious' plants or clones ('PinkGold™) and in the 'Golden Delicious'-related cultivar 'Mutsu' (Fig. 3 and Table 3 in the Electronic supplementary materials). T_{SNP} averaged 40.9% across the apple cultivars, with a range between 67.8% and 25.7% (cultivars 'Initial' and 'Wagener', respectively). Accessions of species phylogenetically closer to *M.* × *domestica* and belonging to Series Pumilae, such as *M. sieversii*, *M. sylvestris*, *M.* × *asiatica*, *M. micromalus*, *M. orientalis* and *M. prunifolia* (classified according to Janick et al. 1996), displayed a lower T_{SNP} (mean T_{SNP} 25.1%, range 35.7–17.6%; P<0.01). Specifically, T_{SNP} averaged 26.7% and 26.6% for *M. sieversii* and *M. sylvestris*, respectively. Much lower T_{SNP} were computed for the other species of the genus *Malus*, including more distantly related species such as *Malus coronaria*, *Malus florentina*, *Malus yunnanensis* and others (mean T_{SNP}=7.2%; P<0.01. Fig. 3 and Table 3 in the Electronic supplementary materials). T_{SNP} ranged

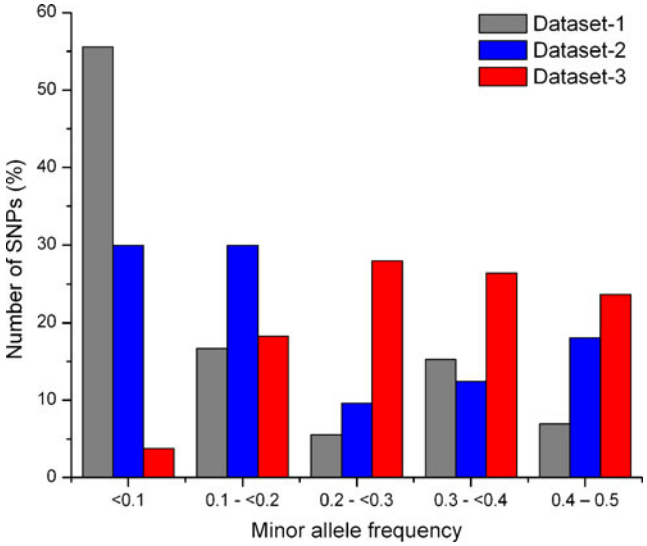


Fig. 2 Distributions of SNP allele frequencies computed on the basis of amplicon re-sequencing (datasets 1 and 2) or SNPlex™ genotyping (dataset 3)

between 48.7% and 12.0% for a number of hybrids and unclassified *Malus* accessions.

With respect to *M.* × *domestica* SNPs discovered from crude re-sequencing of M. x domestica and other species accessions (datasets 1 and 2), T_{SNP} in *M.* × *domestica* averaged 14.8 and 12.1%, respectively (Fig. 3a, b and Table 3a, b in the Electronic supplementary materials). However, by only including those SNPs with MAF>0.1 within *M.* × *domestica*, T_{SNP} increased to 30.2% and 15.1% for datasets 1 and 2, respectively. As observed for dataset 3, T_{SNP} resulted significantly lower for *Malus* accessions belonging to the series Pumilae (T_{SNP}=6.6% or 9.0% with MAF>0.1; P<0.01). Specifically, the mean T_{SNP} for *M. sieversii* accessions was 6.4 (8.1% with MAF>0.1), while for *M. sylvestris* accessions was 6.1% (8.5% with MAF> 0.1; P=n.s.). T_{SNP} was extremely low for *Pyrus communis* and *P. pyrifolia* (3.0% and 1.2%, respectively).

Capturing SNP variability

By means of simulated re-sampling, we evaluated the proportion of SNPs which could be captured by sub-groups of *M.* × *domestica* cultivars of increasing dimensions (Fig. 4) using data from dataset 1. On average, sub-groups of 55 cultivars captured 90% of SNPs (corresponding to 66 out of 74 SNPs). The same degree of variability was captured by a single group of just 15 cultivars ('Cox Orange Pippin', 'Sonya', 'Ingrid Marie', 'Pearmain Dorè', 'Hapke', 'Rosana', 'Crimson Snow', 'Sunkrisp', 'Ariwa', 'Ananas Reinette', 'Jonafree', 'Antonovka', 'Delgrared', 'Renora' and 'Sila'). Thirty-five cultivars captured an average of 80% of SNPs, although the latter were also captured by a specific group of just ten cultivars ('Freedom', 'Delgared', 'Jonathan', 'Durello di Forlì', 'Crimson Snow', 'Tunda', 'San Lugano', 'Renora', 'Ariwa' and 'Red Field'). 'Golden Delicious', 'Red Delicious', 'Jonathan', 'McIntosh' and 'Cox's Orange Pippin', the five cultivars recognised as the founders of modern apple breeding (Noiton and Alspach 1996), would have enabled us to capture 35 SNPs, corresponding to 47% of all SNPs. This number of SNPs corresponds to the expected number captured by a random five-cultivar group and is significantly lower than the number of SNPs (56=76% of total number of SNPs) captured by the best 5-cultivar group, which included 'Reanda', 'Laxton's Superb', 'Ingrid Marie', 'Priscilla' and 'Mela della Serla').

Number of SNPs for linkage mapping in *M.* × *domestica*

Simulation was used to evaluate the influence of (1) the number of SNPs and (2) the different levels of T_{SNP}, on the efficiency of linkage map construction expressed as obtained map coverage. We hypothesized to use a close-design (that is, SNPs must be chosen in advance of industrial production) genotyping array with increasing numbers of *M.* × *domestica* SNPs. Fig. 5 summarises the results of the simulation. T_{SNP}=0.4 with 100 SNPs per chromosome of (100 cM) provided a map coverage of 92% (modal value) and a coverage of >84% at P= 0.9. Two hundred SNPs per chromosome increased the map coverage to 96% (modal value) and a coverage of >92% at P=0.9.

Therefore, given the apple genome (17 chromosomes, with an average genetic length of 71 cM each. Velasco et al. 2010), approximately 120 SNPs per apple chromosome (corresponding to approximately 50 polymorphic SNPs per chromosome) or a total of 2,000–2,100 SNPs, where SNPs have an average T_{SNP} of 0.4 (such as those derived from 'Golden Delicious'), will be required to obtain a linkage map of each parent covering approximately 90% (at P > around 0.9) in any experimental cross from a cultivated apple background. When the parent to be mapped is not a *M. x domestica* cultivar and T_{SNP} averages are lower—such as 0.251 (other Series Pumilae species) or 0.072 (other *Malus* species)—150 SNPs per chromosome (approximately 2,600 SNPs total) or 430 SNPs per chromosome (approximately 7,300 SNPs total) using cultivar-derived SNPs would be required, respectively. Alternatively, using SNPs determined from re-sequencing of a mapping parent itself would translate into a T_{SNP} close to 100%, (although only approximately 60% of the SNPs will be fully informative backcross markers due to the heterozygosity in the second parent). Thirty-five SNPs per chromosome (approximately 590 SNPs total) would be required to build a map in this case. The predictions based on our simulation were empirically confirmed by inspecting the cM coverage of

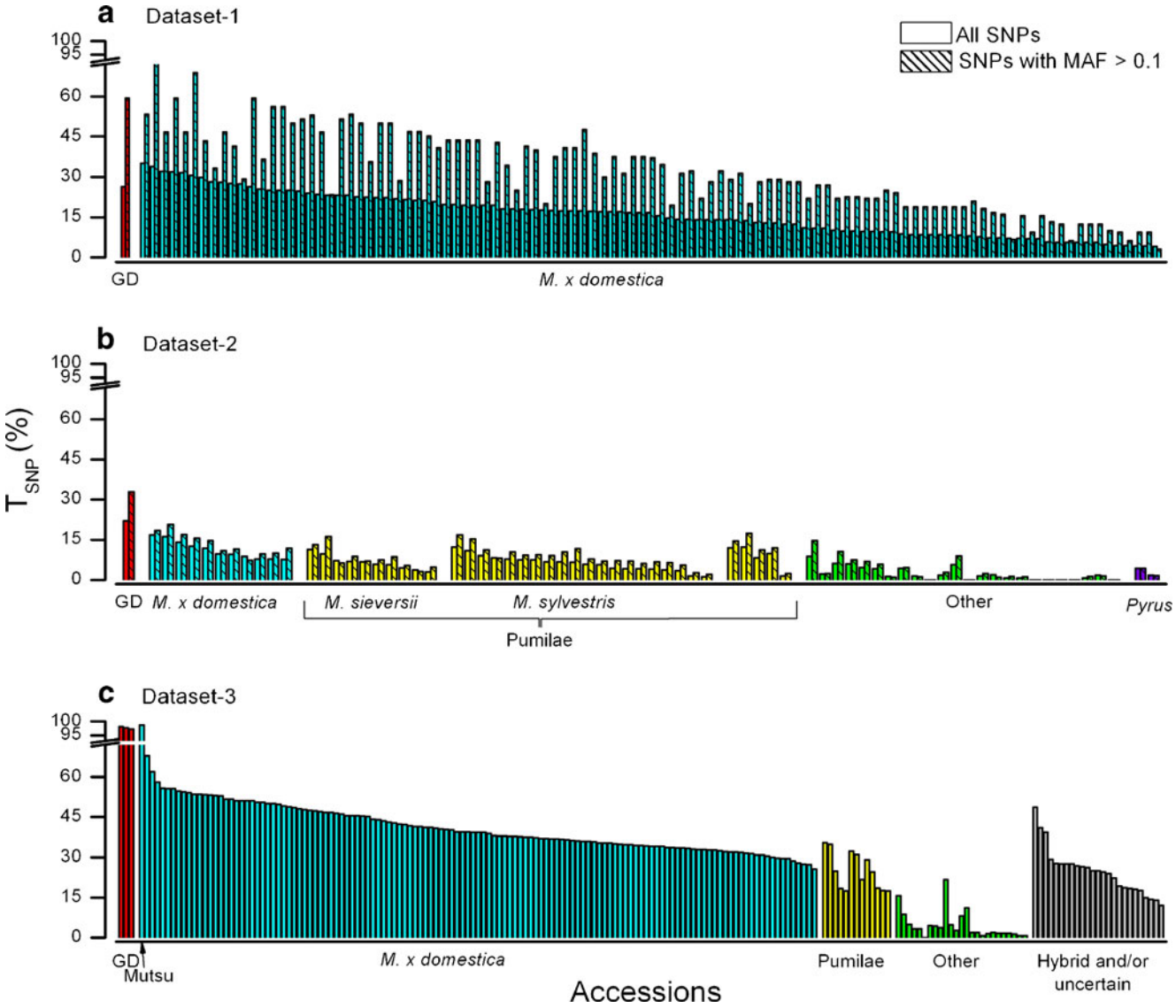


Fig. 3 Percentages of *Malus* × *domestica* SNP transferability (T_{SNP}) in different *Malus* accessions based on amplicon re-sequencing (datasets 1 and 2) or genotyping results from SNPlex™ (dataset 3). 'GD' (*red*): 'Golden Delicious' clone B, used for the apple genome sequencing (Velasco et al. 2010). 'Mutsu': triploid *M.* × *domestica* cultivar originating from unreduced egg cell of 'Golden Delicious'. 'Pumilae' (*yellow*): accessions belonging to genus *Malus*, Series Pumilae (*Malus* × *asiatica*, *Malus micromalus*, *Malus orientalis*, *Malus prunifolia*, *Malus sieversii* and *Malus sylvestris*) with the exception of *M.* × *domestica*. 'Other' (*green*): accessions of species belonging to other series of the genus *Malus*. 'Pyrus': *Pyrus communis* and *Pyrus pyrifolia*. 'Hybrid and/or uncertain': *Malus* accessions belonging to hybrid species or with ambiguous classification. Detailed T_{SNP} for all accessions is provided in Table 3 in the Electronic supplementary materials

three apple linkage maps being build internally. For instance, testing 693 SNPs in a 93 plants cross involving two apple cultivars with T_{SNP}=0.39 and 0.51, provided single parental maps of 818 and 690 cM, corresponding to 60% and 51% coverage of the apple reference maps (Table 4).

Discussion

Information regarding the degree of genetic diversity at the DNA sequence level is a cornerstone of molecular genetics and the breeding and management of genetic resources of a crop species (Glaszmann et al. 2010). SNP frequency has been extensively investigated on the basis of large re-sequencing efforts in annual crops, where it appears to vary considerably between species (e.g. from 1 SNP/61 bp to 1 SNP/5,700 bp for maize and rice japonica cultivars, respectively; Jones et al. 2009; Yamamoto et al. 2010). As expected, high SNP frequencies have been observed in forest tree species (e.g. 1 SNP/63 bp and 1 SNP/130 bp in *Pinus taeda* and *Populus trichocarpa*, respectively; Brown et al. 2004; Gilchrist et al. 2006). However, information is

Table 3 Statistics of five apple SNP arrays (based on SNPlex™—applied Biosystems) after designing and application on 185 *Malus* accessions

SNPset	SNPs on array (no.)	Passed SNPs[a] (no.)	Failed SNPs (no.)	Passed SNPs *M.* × *domestica*[b] (No.)	False SNPs[c] (no.)
W0710109438_0001	48	33	14	33	2
W0711109485_0001	47	37	9	37	1
W0711109485_0002	47	36	10	38	1
W0711109485_0003	48	42	6	45	0
W0711109485_0004	47	39	8	46	0
Total	237 (100%)	187 (79.3%)	47 (19.8%)	199 (84%)	4 (1.8%)

[a] SNPs satisfacting the SNPlex™ quality value when all 185 accessions were contemporarily analysed

[b] SNPs satisfacting the SNPlex™ quality value when only 119 accessions belonging to *M.* × *domestica* were analysed

[c] Assays showing lack of polymorphism at the target nucleotide position for all 185 accessions (including Golden Delicious)

only beginning to accumulate for perennial fruit crop species such as *Citrus* (1 SNP/164 bp, based on database ESTs; Jiang et al. 2009), almond (1 SNP/114 bp based on EST-based amplicon re-sequencing; Wu et al. 2008), or cacao (1 SNP/71–223 bp, based on database ESTs; Lima et al. 2009). Grape is an exception as its SNP frequency has been estimated in various studies as 1 SNP/117 bp (based on the Pinot Noir genome sequence; Velasco et al. 2007) or 1 SNP/64–104 bp (based on multi-locus analysis of cultivars; Lijavetzky et al. 2007; Vezzulli et al. 2008).

With respect to *M.* × *domestica*, previous works have estimated 1 SNP/149 (based on database ESTs; Chagné et al. 2008), 1 SNP/225 bp (based on BAC end sequences; Han and Xue 2003) and 1 SNP/227 bp (based on the genome sequence of 'Golden Delicious'; Velasco et al. 2010). Our findings of an average of 1 SNP/455 bp in a single cultivar and 1 SNP/52 bp across germplasm confirm that apple is one of the most genetically polymorphic agricultural species analysed so far and is more than one order of magnitude more polymorphic than human (1 SNP/1,910 bp; Sachidanandam et al. 2001). Our estimate of π=0.0038 for *M.* × *domestica* confirms this relatively high polymorphism and is in the lower range of values recorded for maize (π= 0.0096 or 0.0036, according to Tenaillon et al. 2001 or Jones et al. 2009, respectively), is similar to that observed for grape (π=0.0051; Lijavetzky et al. 2007) and barley (π= 0.003–0.004; Kota et al. 2008), and is higher than for rice (π =0.0010–0.0027; Zhu et al. 2007) and human (π=0.00075; Sachidanandam et al. 2001).

Indels in plant species have been the target for the development of molecular markers (Bhattramakki et al. 2002), including highly parallel approaches such as single feature polymorphisms (SFP) (Borevitz et al. 2003; Drost et al. 2009). The frequency of indels in apple that we observed (1 indel/333 bp) is comparable to that observed in grape (1/ 450 bp; Velasco et al. 2007), maize (1/126–1/186 bp; Ching et al. 2002; Jones et al. 2009) and sunflower (1/277 bp; Kolkman et al. 2007), and lower than that observed in melon (1/3 kb; Deleu et al. 2009). All in all, indel frequency in apple would seem to be suitable for the implementation of marker technologies such as SFP.

In apple, as in many other fruit crop species, linkage mapping of genes and QTLs is commonly carried out on the basis of F_1 experimental crosses, where the only informative markers are those that are heterozygous in the parent plants (Lynch and Walsh 1997). One of the aims of our study was to estimate the rate of transferability (T_{SNP}) of the heterozygous state of *M.* × *domestica* SNPs. Estimates of T_{SNP} were obtained based on comparison of all SNPs identified in simple gene-based re-sequencing experiments (datasets 1: from 119 M. x domestica cultivars; dataset 2: from 72 diverse accessions of *Malus* and two of *Pyrus*) or from genotyping based on a *M.* × *domestica* SNP whole-genome-based array (dataset 3: primarily from

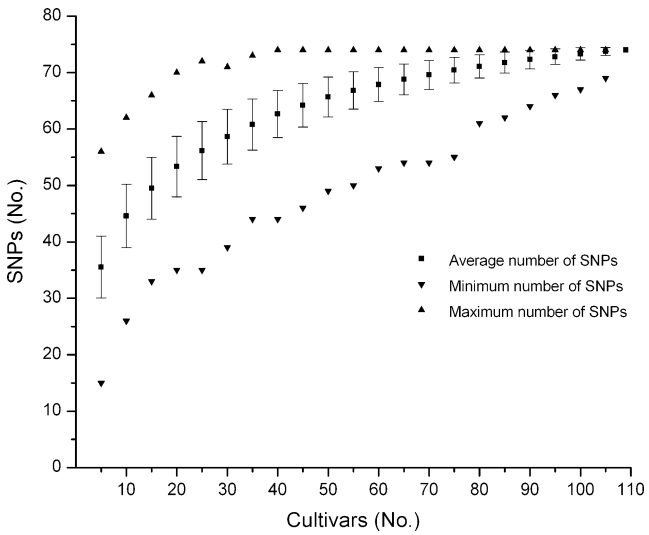


Fig. 4 Number of SNPs captured by apple cultivars (*Malus* × *domestica*) in groups of increasing dimensions, sampled from a collection of 123 cultivars

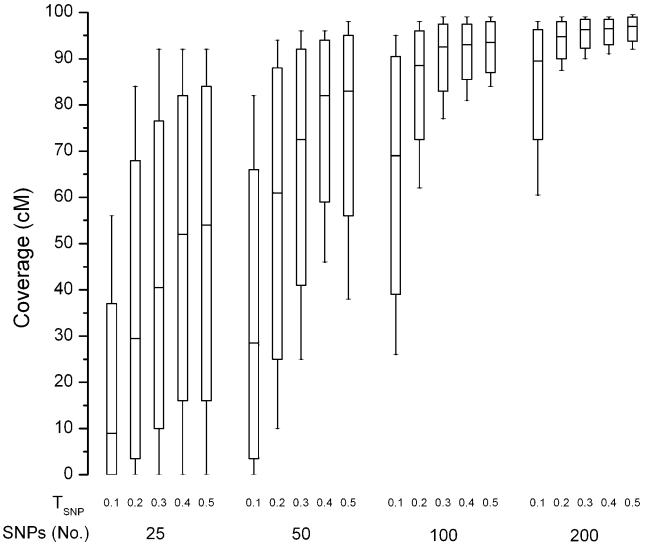


Fig. 5 Coverage of simulated chromosome maps based on increasing numbers of SNPs. Computations are based on simulated single parent linkage maps of one chromosome (100 cM), 100 F_1 progenies and markers in backcross design (Aa×aa) only. T_{SNP}: rate of SNP transferability considered. *Boxes* indicate 25–75 percentiles, *horizontal lines in the boxes* indicate median values and *vertical bars* indicate minimum and maximum values

'Golden Delicious' SNPs) and were much higher in the latter dataset. Besides sampling of SNPs and cultivars, two factors were probably responsible for this difference. The first factor was the bias arising by the different origins of the SNPs. While no SNP selection step was applied for datasets 1 and 2, the T_{SNP} based on the SNP array was derived from SNPs pre-selected at the discovery stage. These SNPs were 'Golden Delicious' SNPs further confirmed by SNPs within EST reads. Because the EST sequences were derived from 25 different cultivars, this step implicitly selected SNPs with a higher MAF while SNPs unique to 'Golden Delicious' (and therefore with a potentially lower MAF) were generally counter selected. A second factor that may have been responsible for the higher T_{SNP} in dataset 3 was the different cultivar assortments. T_{SNP} data from dataset 2 was based on largely unrelated cultivars (including the five cultivars considered the founders of modern apple breeding) with just two exceptions: 'Fuji', an offspring of 'Red Delicious', and 'Clivia', an offspring of 'Cox's Orange Pippin' (Noiton and Alspach 1996). It is therefore very likely that the T_{SNP} obtained in this experiment underestimates the actual T_{SNP} within *M.* × *domestica* due to the unrelatedness of the accessions. This explanation is supported by the observation that dataset 1, which included the same 12 cultivars within a larger set (119 cultivars), provided a higher T_{SNP}. In addition, filtering SNPs with a higher MAF (MAF>0.1) remarkably increased T_{SNP} in datasets 1 and 2, confirming the positive effect on T_{SNP} of a selection step for higher MAF.

As expected, in dataset 3 T_{SNP} was close to 100% when 'Golden Delicious' or 'Golden Delicious' clones ('Pink Gold™') were analysed, indicating high reliability of the genotyping technique. Transferability was similarly high

Table 4 Coverage of linkage maps obtained based on SNP genotyping arrays only

Cross	Parent	Plants (no.)	SNPs tested[a] (no.)	Pol. SNP[b] (no.)	T_{SNP} (%)	SNPs utilized (no.)[c]	SNPs mapped (no.)[d]	Map (cM)[e]	Map (%)[f]
A×B		192	355						
	A			179	50.5	108	104	665	49
	B			139	39.1	70	69	392	29
C×D		93	693						
	C			243	38.7	129	119	818	60
	D			317	50.7	190	141	690	51
B×*Malus* [g]		121	269						
	B			83	30.9	80	73	470	35
	Malus			5	4.7	–	–	–	–

[a] Total of SNPs providing unambiguous scores

[b] Number of polymorphic (= heterozygous) SNPs

[c] Number of polymorphic SNPs segregating as backcross type (e.g., Aa×aa)

[d] Number of SNPs eventually included in linkage maps based on SNP markers only

[e] Computed as total length in cM of all linkage groups after projection of the most external markers per linkage group on the apple integrated linkage map (http://genomics.research.iasma.it/cgi-bin/cmap/viewer)

[f] Computed as % of the apple integrated linkage map (http://genomics.research.iasma.it/cgi-bin/cmap/viewer) as indicated in footnote (e)

[g] Experimental population obtained from a cross between a *M.* × *domestica* cultivar and a wild *Malus* species

for 'Mutsu', a cultivar known to be a triploid accession, the genome of which includes a complete 'Golden Delicious' genome resulting from an unreduced egg cell. Ranking of cultivars based on T_{SNP} shows that the first 15 cultivars were first- or second-generation seedlings of 'Golden Delicious' (with 'Golden Delicious' often present several times in the pedigree) or triploids, or both (not shown). For instance, 'Initial', the cultivar with the highest T_{SNP}, is a triploid cultivar originating from the cross 'Gala' (a seedling of 'Golden Delicious')×'Red Free' (a complex pedigree apple which includes 'Golden Delicious' as an ancestor) (Laurens et al. 2000). The high T_{SNP} of these cultivars seems due, therefore, to their common 'Golden Delicious' genetic background. In addition, the relatively high T_{SNP} of triploid cultivars observed (Table 2 in the Electronic supplementary materials) may well have been due to the increased probability of these cultivars having one of their three nucleotides different, which were called as heterozygous by the SNPlex™ technology. As confirmation, the 15 cultivars with the lowest T_{SNP} did not share 'Golden Delicious' in their pedigree and were all diploid. Finally, the T_{SNP} between 'Golden Delicious' and 'Royal Gala' observed by Chagné et al. (2008) were found to be remarkably similar (53.7% and 51.8–54.0%, respectively).

T_{SNP} of other species decreased progressively with distance from *M.* × *domestica* species and in general with the known phylogenetic distances, confirming the close relationship of *M.* × *domestica* with the species belonging to series Pumilae (Janick et al. 1996).

The high T_{SNP} values observed for some of the unclassifieds/hybrids are probably the consequence of (1) one of the hybrid's parents being *M.* × *domestica* (the T_{SNP} of the hybrid between a cultivated apple and a phylogenetically distant *Malus* is expected to be approximately half of the mean T_{SNP} within *M.* × *domestica*), (2) a higher ploidy or (3) a combination of the two causes. Indeed, within this group of accessions, those with the higher T_{SNP} were *Malus* × *zumi* (a triploid hybrid between *Malus baccata* and *Malus sieboldii*), *M.* × *purpurea* (Barbier) Rehd. Lemoinei (a hybrid between *M. sieversii* Niedzwetzkyana and *Malus* × *atrosanguinea*; Janick et al. 1996), and *Malus* × *toringoides* (an often triploid hybrid between *Malus transitoria* and *Malus kansuensis*; Feng et al. 2007).

Our analyses provide an empirical estimate of the efficiency of the process from SNP discovery to SNP genotyping using a multiplexed technology. The application of the SNPlex™ design rules passed 88% of the starting SNPs to the SNPlex™ arrays. In comparison, 86% of initial SNPs passed this step in grape (Pindo et al. 2008). After genotyping, analysis of the full set of *Malus* accessions and of a restricted set of cultivars resulted in 79.3% and 84.0%, respectively, of the SNPlex™ SNPs being successfully called. The higher failure rate observed when a more genetically diverse germplasm collection was analysed was probably due to the higher frequency of additional unknown SNPs or indels around the target SNP, which can disturb the correct biochemistry of the assay, as already observed in species with comparable nucleotide diversity such as grape (Pindo et al. 2008; Vezzulli et al. 2008) or maize (Jones et al. 2009).The effective calling rate of the same SNPlex™ arrays was a little higher (90%) when single biparent apple map populations (Table 3) were analysed, as in some cases it was possible to resolve ambiguities due to additional genotype groups in the SNPlex™ genotype outputs by assuming segregation of unknown SNPs at the probe annealing site (data not shown). Our call rates were therefore similar to those observed in other species (Vezzulli et al. 2008; Lijavetzky et al. 2007; Bérard et al. 2009). A global estimate of the efficiency of SNP genotyping using the SNPlex™ technology is given by the conversion rate (% of initially submitted SNPs which it was possible to successfully call). For *M.* × *domestica*, the SNP conversion rates were ~77% and ~82% for the germplasm and mapping population data respectively, which are close to the rate reported by the manufacturer (≥80%; SNPlex™ Product Bulletin Applied Biosystems). Finally, it appears that a thorough initial step of selection for single copy SNP target sites was sufficient to circumvent the problem of the presence of duplicated regions.

Our simulations indicate that 2,000–2,100 randomly distributed SNPs are required to build a linkage map involving an F_1 cross between apple cultivars and a progeny size as small as one hundred plants (larger populations would require fewer SNPs for simple linkage map construction). This estimate is similar to the prediction for grape (around 2,000 SNPs for a 1,500 cM linkage map. Lijavetzky et al. 2007). Empirical data from our internal molecular genetics programme and based on the SNP-only map coverages of three F_1 crosses (Table 4) support these predictions. Because of the slightly lower T_{SNP}, the same SNP array seems also to be suitable, although less efficient, for map construction in crosses involving *M. sieversii*, *M. sylvestris, M. orientalis* or *M.* × *asiatica*. On the other hand, when T_{SNP} reaches 0.1 or lower (typically observed for *M.* × *domestica* SNPs in phylogenetically distant *Malus* species) >5,000 SNPs would be required for good linkage map coverage. Moreover, predictions regarding the efficiency of genotyping arrays at such low T_{SNP} rates could be highly biased by other factors, such as (1) the non-homogenous distribution of markers on the chromosomes, (2) the presence of large chromosome rearrangements, and (3) the increasingly disrupting effects of DNA polymorphisms near the locus targeted by the SNP assays (see above), and should, in our opinion, be treated conservatively.

Conclusions

It is expected that much larger-scale SNP discovery and genotyping efforts will be needed to make the next obvious step in apple molecular genetics, that is, the deployment of genome-wide association mapping (Rafalski 2010). The main 'operational' issues remaining to be dealt with, that is the extension of linkage disequilibrium (directly influencing the number of SNPs to be used) and description of the population structure of *M.* × *domestica*, are beyond the scope of this work. However, our study provides useful indications. For instance, it appears that most SNPs of *M.* × *domestica* can be recruited by surveying a relatively small sample of cultivars since as few as 15 cultivars captured 90% of the SNPs occurring in 123 cultivars. Our data also show that an array based on *M.* × *domestica* SNPs will still be relatively informative when accessions belonging to *M. sieversii*, *M. sylvestris*, *M. orientalis*, *M.* × *asiatica*, *M. micromalus*, *M. prunifolia* and *M. spectabilis* are analysed (also according to the standard taxonomy of the genus *Malus*. Janick et al. 1996). However, accessions belonging to these species with extremely low T_{SNP} were also found and T_{SNP} close to null was observed for more phylogenetically distant species. Therefore, dedicated SNP discovery efforts appear to be advisable for better application of SNP-based genotyping tools to these species.

Acknowledgements We thank Giuseppina Coppola, Pierluigi Magnago and Massimo Pindo for technical support. We also thank Phil Forsline, Valerio Gallerati, Anders S. Larsen and Andreas Peil for providing germplasm accessions. We thankfully acknowledge the financial support of Provincia Autonoma di Trento, Italy.

References

Bérard A, Le Paslier MC, Dardevet M, Exbrayat-Vinson F, Bonnin I, Cenci A, Haudry A, Brunel D, Ravel C (2009) High-throughput single nucleotide polymorphism genotyping in wheat (*Triticum* spp.). Plant Biotechnol J 7:364–374

Bhattramakki D, Dolan M, Hanafey M, Wineland R, Vaske D, Register JC 3rd, Tingey SV, Rafalski A (2002) Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. Plant Mol Biol 48:539–547

Borevitz JO, Liang D, Plouffe D, Chang HS, Zhu T, Weigel D, Berry CC, Winzeler E, Chory J (2003) Large-scale identification of single-feature polymorphisms in complex genomes. Genome Res 13:513–523

Brown SK (1992) Genetics of apple. Plant Breed Rev 9:333–366

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc Natl Acad Sci USA 101:15255–15260

Chagné D, Gasic K, Crowhurst RN, Han Y, Bassett HC, Bowatte DR, Lawrence TJ, Rikkerink EH, Gardiner SE, Korban SS (2008) Development of a set of SNP markers present in expressed genes of the apple. Genomics 92:353–358

Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. BMC Genet 3:19

Deleu W, Esteras C, Roig C, González-To M, Fernández-Silva I, Gonzalez-Ibeas D, Blanca J, Aranda MA, Arús P, Nuez F, Monforte AJ, Picó MB, Garcia-Mas J (2009) A set of EST-SNPs for map saturation and cultivar identification in melon. BMC Plant Biol 9:90

Drost DR, Novaes E, Boaventura-Novaes C, Benedict CI, Brown RS, Yin T, Tuskan GA, Kirst M (2009) A microarray-based genotyping and genetic mapping approach for highly heterozygous outcrossing species enables localization of a large fraction of the unassembled *Populus trichocarpa* genome sequence. Plant J 58:1054–1067

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform 1:47–50

Fan JB, Chee MS, Gunderson KL (2006) Highly parallel genomic assays. Nat Rev Genet 7:632–644

Feng TT, Zhou ZQ, Tang JM, Cheng MH, Zhou SL (2007) ITS sequence variation supports the hybrid origin of *Malus toringoides* Hughes. Can J Bot 85:659–666

Gilchrist EJ, Haughn GW, Ying CC, Otto SP, Zhuang J, Cheung D, Hamberger B, Aboutorabi F, Kalynyak T, Johnson L, Bohlmann J, Ellis BE, Douglas CJ, Cronk QC (2006) Use of Ecotilling as an efficient SNP discovery tool to survey genetic variation in wild populations of *Populus trichocarpa*. Mol Ecol 15:1367–1378

Glaszmann JC, Kilian B, Upadhyaya HD, Varshney RK (2010) Accessing genetic diversity for crop improvement. Curr Opin Plant Biol 13:167–173

Gu CC, Yu K, Ketkar S, Templeton AR, Rao DC (2008) On transferability of genome-wide tagSNP. Genet Epidemiol 32:89–97

Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. Heredity 101:5–18

Han B, Xue Y (2003) Genome-wide intraspecific DNA-sequence variations in rice. Curr Opin Plant Biol 6:134–138

Hartl DL, Clark AG (2007) Principles of population genetics, Fourth editionth edn. Sinauer Associates, Sunderland

Hosono S, Faruqi A, Dean FB, Du Y, Sun Z, Wu X, Du J, Kingsnore SF, Egholm M, Lasken R (2003) Unbiased whole-genome amplification directly from clinical samples. Genome Res 13:954–964

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

Janick J, Cummins JV, Brown SK, Hemmat M (1996) Apples. In: Janick J, Moore JN (eds) Fruit breeding. Wiley, New York, pp 1–66

Jiang D, Ye Q, Wang F, Cao L (2009) The mining of *Citrus* EST-SNP and its application in cultivar discrimination. Agric Sci China 9:179–190

Jones E, Chu WC, Ayele M, Ho J, Bruggeman E, Yourstone K, Rafalksi A, Smith OS, McMullen MD, Bezawada C, Warren L, Babayev J, Basu S, Smith S (2009) Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (*Zea mais* L.) germplasm. Mol Breed 24:165–176

Juniper BE, Mabberley DJ (2006) The story of the apple. Timber Press, Inc., Portland

Kolkman JM, Berry ST, Leon AJ, Slabaugh MB, Tang S, Gao W, Shintani DK, Burke JM, Knapp SJ (2007) Single nucleotide polymorphisms and linkage disequilibrium in sunflower. Genetics 177:457–468

Kota R, Varshney RK, Prasad M, Zhang H, Stein N, Graner A (2008) EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. Funct Integr Genomics 8:223–233

Laframboise T (2009) Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucl Acids Res 37:4181–4193

Laurens F, Lespinasse Y, Fouillet A (2000) A new scab resistant apple: 'Initial'. Acta Hort ISHS 538:707–710

Lijavetzky D, Cabezas JA, Ibanez A, Rodriguez V, Martinez-Zapater JM (2007) High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. BMC Genomics 8:424

Lima L, Gramacho K, Carels N, Novais R, Gaiotto F, Lopes U, Gesteira A, Zaidan H, Cascardo J, Pires J, Micheli F (2009) Single nucleotide polymorphisms from *Theobroma cacao* expressed sequence tags associated with witches' broom disease in cacao. Gen Mol Res 8:799–808

Lynch M, Walsh B (1997) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland

Noiton DAM, Alspach PA (1996) Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. J Am Soc Hortic Sci 121:773–782

Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. J Mol Biol 302:205–217

Pindo M, Vezzulli S, Coppola G, Cartwright DA, Zharkikh A, Velasco R, Troggio M (2008) SNP high-throughput screening in grapevine using the SNPlex genotyping system. BMC Plant Biol 8:12

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

Rafalski JA (2010) Association genetics in crop improvement. Curr Opin Plant Biol 13:174–180

Robinson JP, Harris SA, Juniper BE (2001) Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. Plant Syst Evol 226:35–58

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

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933

Staden R, Beal KF, Bonfield JK (2000) The Staden package, 1998. Meth Mol Biol 132:115–130

Syvänen AC (2005) Toward genome-wide SNP genotyping. Nat Genet 37:S5–S10

Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci USA 98:9161–9166

Tobler A, Short S, Andersen M, Paner T, Briggs J, Lambert S, Wu P, Wang Y, Spoonde A, Koehler R, Peyret N, Chen C, Broomer A, Ridzon D, Zhou H, Hoo B, Hayashibara K, Leong L, Ma C, Rosenblum B, Day J, Ziegle J, De La Vega F, Rhodes M, Hennessy K, Wenz H (2005) The SNPlex genotyping system: a flexible and scalable platform for SNP genotyping. J Biomol Tech 16:398–406

Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, Salvi S, Pindo M, Baldi P, Castelletti S, Cavaiuolo M, Coppola G, Costa F, Cova V, Dal Ri A, Goremykin V, Komjanc M, Longhi S, Magnago P, Malacarne G, Malnoy M, Micheletti D, Moretto M, Perazzolli M, Si-Ammour A, Vezzulli S, Zini E, Eldredge G, Fitzgerald LM, Gutin N, Lanchbury J, Macalma T, Mitchell JT, Reid J, Wardell B, Kodira C, Chen Z, Desany B, Niazi F, Palmer M, Koepke T, Jiwan D, Schaeffer S, Krishnan V, Wu C, Chu VT, King ST, Vick J, Tao Q, Mraz A, Stormo A, Stormo K, Bogden R, Ederle D, Stella A, Vecchietti A, Kater MM, Masiero S, Lasserre P, Lespinasse Y, Allan AC, Bus V, Chagné D, Crowhurst RN, Gleave AP, Lavezzo E, Fawcett JA, Proost S, Rouzé P, Sterck L, Toppo S, Lazzari B, Hellens RP, Durel CE, Gutin A, Bumgarner RE, Gardiner SE, Skolnick M, Egholm M, Van de Peer Y, Salamini F, Viola R (2010) The genome of the domesticated apple (*Malus x domestica* Borkh.). Nat Genet 42(10):833–839. doi: 10.1038/ng.654

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, FitzGerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma TM, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Demattè L, Mraz A, Battilana J, Stormo K, Costa F, Tao Q, Si-Ammour A, Harkins T, Lackey A, Perbost C, Taillon B, Stella A, Solovyev V, Fawcett JA, Sterck L, Vandepoele K, Grando MS, Toppo S, Moser C, Lanchbury J, Bogden R, Skolnick M, Sgaramella V, Bhatnagar SK, Fontana P, Gutin A, Peer Y, Salamini F, Viola R (2007) High quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE 2:e1326

Vezzulli S, Micheletti D, Riaz S, Pindo M, Viola R, This P, Walker MA, Troggio M, Velasco R (2008) A SNP transferability survey within the genus *Vitis*. BMC Plant Biol 8:128

Wu SB, Wirthensohn M, Hunt P, Gibson JP, Sedgley M (2008) High resolution melting analysis of almond SNPs derived from ESTs. Theor Appl Genet 118:1–14

Yamamoto T, Nagasaki H, Yonemaru J, Ebana K, Nakajima M, Shibaya T, Yano M (2010) Fine definition of the pedigree haplotypes of closely related rice cultivars by means of genome-wide discovery of single-nucleotide polymorphisms. BMC Genomics 11:267

Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J (2009) Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. PLoS ONE 4:e8451

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S (2007) Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. Mol Biol Evol 24:875–888