Tree Genetics & Genomes (2015) 11: 115
DOI 10.1007/s11295-015-0941-3

ORIGINAL ARTICLE

A RAD-based linkage map of kiwifruit (*Actinidia chinensis* Pl.) as a tool to improve the genome assembly and to scan the genomic region of the gender determinant for the marker-assisted breeding

Davide Scaglione[1] · Alice Fornasiero[1] · Catalina Pinto[2] · Federica Cattonaro[1] · Alessandro Spadotto[1] · Rodrigo Infante[2] · Claudio Meneses[3] · Rachele Messina[4] · Orietta Lain[4] · Guido Cipriani[4] · Raffaele Testolin[4]

Received: 15 July 2015 / Revised: 24 September 2015 / Accepted: 30 September 2015 / Published online: 9 October 2015

Abstract Kiwifruit breeding still largely relies on phenotypic observation of cross progeny grown in the field to fruiting maturity, without any selection prior to the juvenility being overcome. Developing markers for the selection of traits of interest would greatly help breeders to rapidly screen breeding populations. With the aim of mapping several traits of interest in kiwifruit, a F1 population of diploid ($2n=58$) *Actinidia chinensis* was produced by combining parents with contrasting phenotypic traits. Ninety-four individuals were preliminarily analyzed to obtain a saturated genetic map based on 167 SSRs from the literature and 12,586 segregating restriction-site-associated DNA (RAD) loci obtained through an approach known as genotyping-by-sequencing (GBS) based on haplotype calling of SNP markers identified by a modified double digest restriction-associated DNA sequencing (ddRADseq) protocol as proposed by Peterson et al. (2012). To improve the accuracy of genotype calling, restriction-site-associated reads were aligned to the scaffolds of the recently published kiwifruit genome (Huang et al. 2013). This strategy provided genetic anchoring to 557 Mbp (90 %) of the assembly, helping also to anchor some 120 unmapped Mbp and to identify some mis-joined scaffolds. The analysis of the region controlling the dioecy in kiwifruit, spanning 16 scaffolds in the pseudomolecule 25 of the genome assembly (approximately 4.9 Mbp), with RAD markers that co-segregated with the gender determinant, allowed to sort out markers suitable for marker-assisted selection for the gender in the mapping population with successful extension to further controlled crosses having parents at different ploidy level and belonging to the *A. chinensis*/*Actinidia deliciosa* complex.

Keywords Genotyping-by-sequencing · Next-generation sequencing · Single-nucleotide polymorphism · Genetic map · Marker-assisted selection

Communicated by M. Wirthensohn

Electronic supplementary material The online version of this article (doi:10.1007/s11295-015-0941-3) contains supplementary material, which is available to authorized users.

✉ Davide Scaglione
dscaglione@igatechnology.com

[1] IGA Technology Services, Parco Scientifico e Tecnologico, Via J. Linussio 51 Z.I.U., 33100 Udine, Italy

[2] Departamento de Producción Agrícola, University of Chile, Santiago, Chile

[3] Centro de Biotecnología Vegetal, Universidad Andrés Bello, Santiago, Chile

[4] Dipartimento di Scienze agrarie e ambientali, University of Udine, Via delle Scienze 208, 33100 Udine, Italy

Introduction

Kiwifruit breeding is still based on phenotypic observation of traits in cross populations, but it should take great advantage from a selection based on molecular markers associated with traits of interest. Such an approach, popularly known as marker-assisted selection (or MAS), could speed up the progeny screening. Kiwifruit (*Actinidia chinensis*, diploid, $2n=58$ and related diploid and polyploidy *Actinidia* species) is dioecious, and males, which represent 50 % of the progeny in any species and at any ploidy level, could be easily discarded by the analysis of a gender-associated molecular marker without waiting until the juvenility of seedlings is over (Testolin et al. 1995; Fraser et al. 2009; McNeilage et al. 2012). In recent

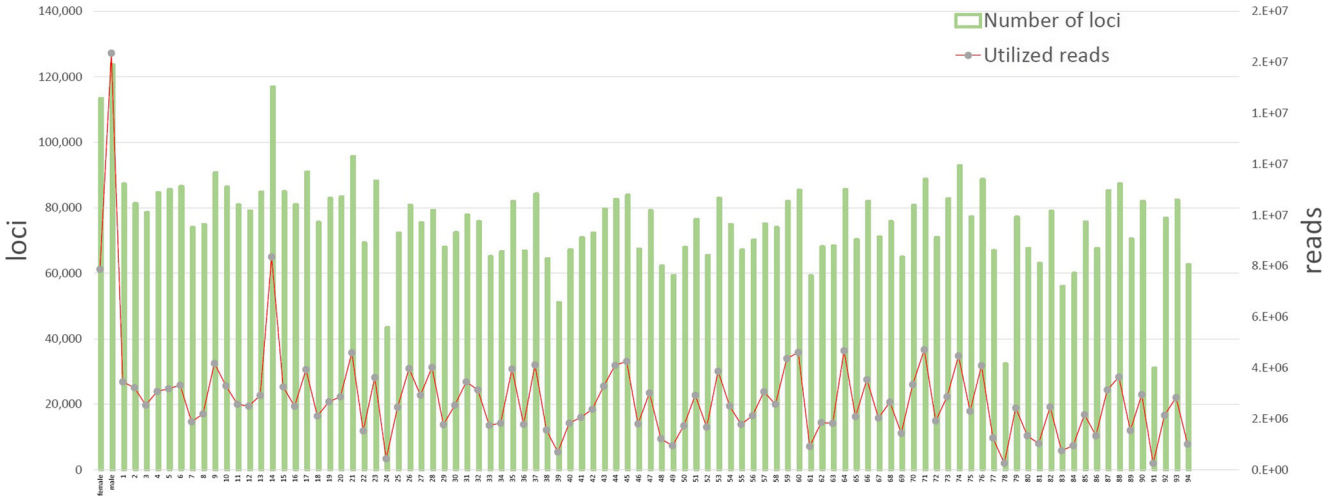

Fig. 1 Distribution of sequenced reads across samples (*red line*, *right axis*) along with number of matched loci against the catalogue of polymorphic RAD sites as detected by parents analysis (*green bar*, *left axis*), Parents are the first two entries of the bar chart (female first)

papers, four molecular markers have been reported being suitable for gender screening (Fraser et al 2009; McNeilage et al. 2012), but either primers were not reported or markers where monomorphic or did not produce PCR products in the population we tested.

Not only the gender but also any trait, no matter it is Mendelian or polygenic, can be early selected before it becomes phenotypically evident by the analysis of associated markers. Moreover, large marker collections well scattered along the genome would pave the way to the genomic estimation of breeding value (GEBV) that represents the frontier in the so-called genomic selection (Meuwissen et al. 2001; Kumar et al. 2012; McNeilage et al. 2012; Testolin 2012, 2013).

SNPs are the markers of choice for massive genotyping, given their ease of detection and low cost of screening, while simple sequence repeats (SSR) markers are ideal markers for the generation of consensus maps, relying on a much higher level of polymorphism in conserved loci, even in genotypes far related from those from which markers were isolated. Hundreds of kiwifruit SSR markers can be retrieved from the literature and EST database collections (Huang et al. 1998; Crowhurst et al. 2008; Fraser et al. 2009), while large-scale

Table 1 Summary of parental SNP detection and statistics of *Actinidia* female and male linkage maps

Parameters	Unit	Female parent	Male parent
RAD fragment size	bp	220–370	220–370
Reads length	bp	85	85
Reads mapped to the reference genome	%	93.9	95.7
SNPs identified	*n*	33,573	28,069
Candidate segregating loci	*n*	6356	6479
- of which RAD haplotypes	*n*	6214	6365
- of which SSR	*n*	142	114
Mapped loci	*n*	6244	6371
- of which RAD haplotypes		6112	6262
- of which SSR		132	109
Linkage groups	*n*	29	29
No. of markers in the map representation	*n*	1390	1429
- of which bin RAD	*n*	1,260	1322
- of which SSR	*n*	130	107
Total map length	cM	3614	3276
Mean distance between genetic bins	cM	2.78	2.40
Mean marker density	cM	0.58	0.51

RAD restriction-site-associated DNA, *SNP* single-nucleotide polymorphism, *SSR* simple sequence repeats

SNP detection requires massive sequencing and requires also a reference sequence to which to map the reads. Alternatively, SNP can be produced in the absence of a reference genome and, more interestingly, simultaneously discovered and mapped, leveraging next-generation sequencing (NGS) technologies coupled with reduced representation of individual genomes with restriction enzymes. These methods permit a fine-tuning over the abundance of fragments obtained by the digestion (Peterson et al. 2012) and thus providing adequate multiplexing for cost-effectiveness. A SNP-based genetic map of 4301 markers has been recently produced through the SLAF-Seq approach with the aim to anchor the scaffolds in a kiwifruit genome sequencing project (Huang et al. 2013).

We adopted such an approach to saturate a SSR-based framework linkage map of diploid kiwifruit. As a side product, we also obtained valuable data points which can be used to further improve the current genome assembly (Huang et al. 2013), both in terms of anchoring-ordering and within-scaffold miss-assembly correction. Moreover, we were able to explore the region of the chromosome 25 where the gender-controlling locus maps and to identify markers suitable for marker-assisted selection for the gender in kiwifruit.

Materials and methods

Plant material and sequencing

A F1 population (C8 x A54.19) was produced using a pair of parents genetically and geographically unrelated of diploid ($2n=58$) *A. chinensis*. The female parent, coded 'C8', is a selection of Fruit & Tea Institute, Hubei province, China, collected in the Fang County in 1980 and carrying large fruits compared with most diploid accessions; the male parent, coded A54.19, is a male seedling introduced from the Beijing Botanical Garden. Ninety-four individuals grown in the open field to the maturity, when gender was recorded, were selected: 41 were female and 53 were male.

Parents and the 94 offsprings were used to produce a linkage map based on SSR and SNP markers. SSR were analyzed

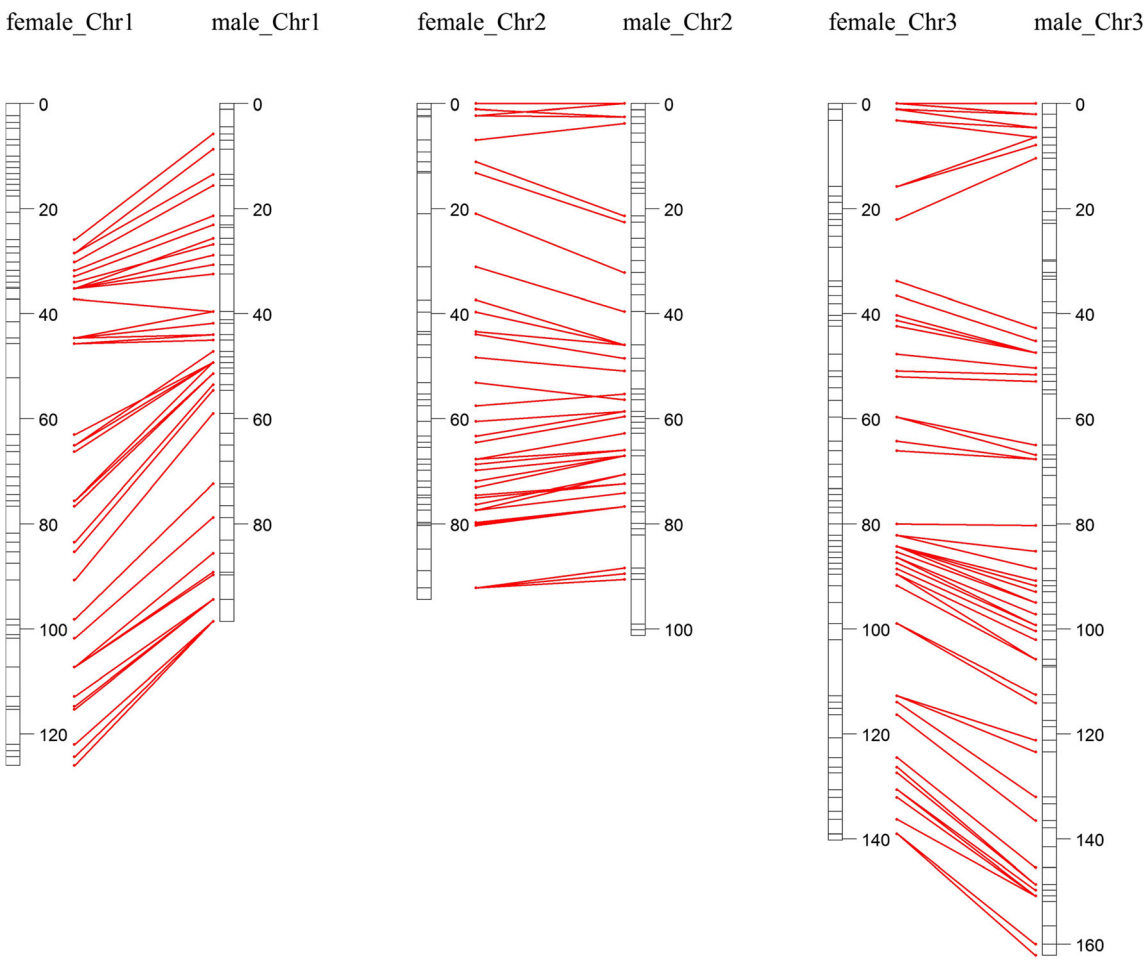

Fig. 2 Alignment of parental maps for the first three chromosomes. RAD coordinates have been rounded to 10^4 bp to generate pseudo-loci for matching

according to the procedure previously described (Huang et al. 1998), whereas SNPs were produced adopting a custom double digest restriction-associated DNA sequencing (ddRADseq) protocol based on the method proposed by Peterson et al. (2012) using the *SphI* and *MboI* restriction enzymes to produce the DNA libraries of the two parents and the offsprings. DNA was extracted following a modified CTAB-chloroform as described by Cipriani and Morgante (1993), followed by a purification with Ampure XP beads (Agencourt, Beverly, MA, USA.) with 1.8 volumes.

Briefly, for each sample, 200 ng of DNA was processed with 2 U of both enzymes (NEB, Ipswich, MA, USA) at 37 °C for 2 h, followed by inactivation at 65 °C for 20 min, and purified with 1.5 volumes of Ampure XP beads (Beckman Coulter, Brea, CA, USA). Common (P2) and barcoded (P1) adapters (see Online Resources 1) were added to a ligation reaction, 3 and 1 pmol respectively, with 200 U of T4 DNA ligase and incubated at 23 °C for 2 h and 65 °C for 20 min. Samples were then pooled in 24-plex, concentrated, and run on low-melting 1.5 % agarose gel. Fragments were selected in the range of 300–450 bp and recovered with QIAquick Gel Extraction kit (Qiagen, Venlo, Netherlands). Enrichment PCR was performed with PCR primers to incorporate Illumina hybridization and sequencing sites along with index sequences for combinatorial multiplexing (Online Resources 1). Cycling parameter were set as follows: 95 °C (5 min), 10 cycles with 95 °C (30 s), 60 °C (30 s), and 72 °C (45 s), with final extension at 72 °C (2 min). Quality, quantity, and reproducibility of libraries were assessed on a Bioanalyzer instrument (DNA High Sensitivity chip). Sequencing was carried out on Illumina HiSeq 2500 platform (Illumina Inc, San Diego, CA, USA) following the manufacturer's protocol.

RAD sequences analysis

Raw Illumina reads were processed in order to de-multiplex samples on the basis of Illumina TruSeq index and custom inline barcodes. After removal of variable-length inline barcode sequence, all reads were trimmed to 85 bp. Alignment to the reference genome (Huang et al. 2013) was carried out using Bowtie2 aligner (Langmead and Salzberg 2012) with default parameters and processed with Stacks software (Catchen et al. 2011). Alleles where retained with a minimum of three reads of depth while intra-sample polymorphism was called using a bounded SNP model with maximum error likelihood of 0.5. A catalog of candidate segregating loci was populated by comparing parental haplotypes on the basis of their genomic coordinates. Reconstructed loci for each progeny individual were then matched against the former catalog in order to score segregating haplotypes. A minimum coverage of six reads was applied to generate homozygous calls. By leveraging a priori knowledge of SNP sites expected to segregate from parents, a correction to low-coverage alleles (i.e., below the minimum of three reads depth) was used to recover heterozygous calls in progeny individuals as implemented in the stacks software, while imposing a minimum allele frequency of 0.10. Haplotype calls were retained when at least 65 individuals out of the 94 in the progeny were genotyped and used to generate the final segregation scoring matrix.

Linkage analysis

The "two-ways pseudo-test cross" analysis allowed to produce a linkage map for each cross parent by managing the

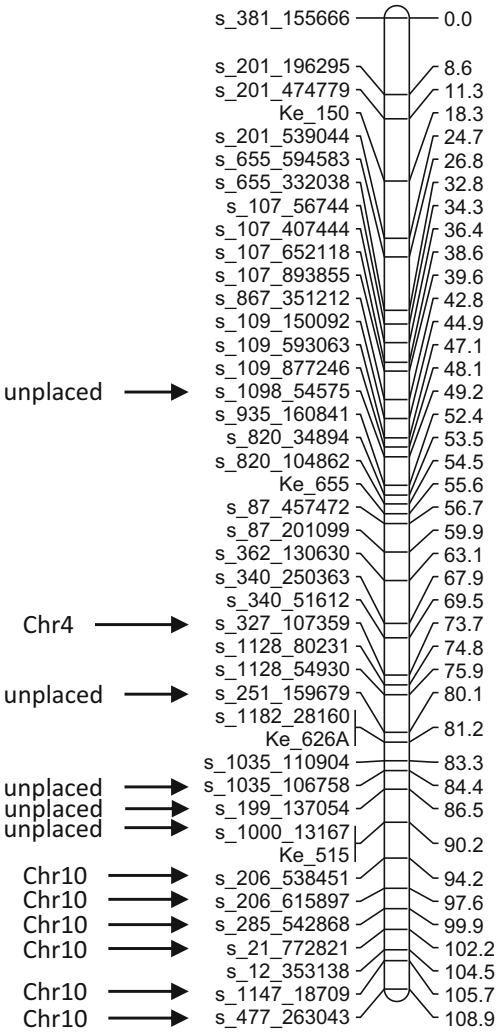

Fig. 3 Major miss-assemblies were found in pseudomolecule "Chr 16." The picture depicts how RAD-based linkage map provided evidences of miss-anchored scaffolds in the current genome assembly as well as placement for un-positioned ones. On the left side of the LG the comments is highlighted the unplaced or miss-anchored scaffolds by means of a representative markers for each genetic bin. For miss-anchored scaffolds, the chromosomal pseudomolecule to which they were wrongly assigned is reported

Table 2 Genome anchoring improvements for each chromosome

LG name as from Huang et al. 2013	Detected scaffold (#) with misplacement	Previously unmapped sequence (bp)	Previously unmapped scaffolds
Chr1		2,184,434	20
Chr2		2,218,279	12
Chr3		4,233,834	30
Chr4		1,604,561	9
Chr5		3,555,725	29
Chr6	*From Chr23*: 212 (p)	2,660,972	20
Chr7		1,797,529	19
Chr8		4,028,411	15
Chr9		4,938,913	17
Chr10		10,836,296	36
Chr11		942,358	7
Chr12		5,605,708	22
Chr13		792,000	6
Chr14		6,156,229	20
Chr15		358,128	10
Chr16	*From Chr10*: 206, 438, 477, 803, 21, 1147, 252, 285	8,665,036	39
Chr17		3,323,865	21
Chr18	*From Chr11*: 54	5,978,979	38
Chr19	*From Chr6*: 418, 307	11,323,227	37
Chr20	*From Chr11*: 22 (p)	4,296,753	17
Chr21	*From Chr5*: 36 (p)	1,571,490	13
Chr22		7,419,277	20
Chr23		1,002,392	9
Chr24		688,414	6
Chr25		5,369,140	27
Chr26		3,133,093	25
Chr27		7,474,929	29
Chr28		1,260,193	10
Chr29		5,732,178	29
		119,152,343	592

Scaffolds showing evidences of misplacement only for some RAD loci are reported with "p" as "partial" (i.e., possibly caused by erroneous scaffolding)

two parental meioses separately. A first analysis was conducted to obtain linkage groups for each parent and data cleanup. RAD markers were grouped using R/QTL package with limit of detection (LOD) score >10 and rec <0.20. Heat maps of LOD scores along with recombination fractions were manually inspected to determine *trans* states across groups and to switch markers accordingly, while singlets due to repetitive sequences or miss-alignments were removed. Markers ordering and correction of genotyping errors were achieved with two iterations of MSTmap software (Wu et al. 2008) with default parameters (using Kosambi function for distance estimation) followed by a Perl implementation of the SMOOTH program as described in Van Os et al. (2005). For each genetic bin, a single RAD marker was selected as representative and integrated to the SSR dataset. A final ordering state was calculated with the JoinMap (v4.0) program (Van Ooijen 2006; Stam 1993) using the maximum likelihood (ML) method. This last step was taken to reduce the conflicts in assigning the distances between markers due to the small size of the mapping population.

Analysis of the gender-controlling region and identification of new markers for the MAS

The 16 scaffolds of the pseudomolecule number 25 of the genome assembly (Huang et al. 2013), whose RAD markers co-segregated with the sex determinant in the mapping population (Online Resources 2), were analyzed, and many SSR-containing sequences were extracted using a modified version of the software Sputnik (Morgante et al. 2001). Primers were designed on the flanking regions using Primer3 (Rozen and Skaletsky 2000). Markers were preliminarily amplified in the parents and a few offsprings, and four of them that provided adequate amplificates and clear bands were then tested in several controlled crosses of the *A. chinensis*/*Actinidia deliciosa* complex, with different ploidy levels (2*n*, 4*n*, and 6*n*) in the attempt of identifying gender-linked alleles suitable for the marker-assisted selection.

Table 3 SSR markers extracted from the scaffolds 514 and 776 of the chromosome 25 of kiwifruit genome sequence (Huang et al. 2013), whose RAD markers co-segregated with the sex determinant

Marker [a]	Core repeat	Primers	Primer first base in the scaffold	*T* anneal (°C)
s_514T	(GAA)13	F-ctggatcagcttctggact R-ggcaaaagatgaaaagagtg	65,435	55
s_514B	(GTC)7	F-tctttgcgatcctattttgt R- tgcaaacagaaaaacaatga	331,926	55
s_776T	(CTT)11	F-caatttgacccaagtaccac R-atggcaatcaatcactcaat	15,320	55
s_776B	(AAT)7	F-tgatttgcttgctttatgaat R-tcggttttgtctgttttag	210,402	55

[a] s-xxx stay for scaffold number, T=means toward the top, B=means towards the bottom of the scaffold

Table 4 Analysis of gender association of SSR markers of Table 3 analyzed in different kiwifruit cross populations

Marker	Species	Cross parents [a]	Parent allelic profile (female parent first)	Female progeny allelic profiles	Male progeny allelic profiles	Gender-linked allele	Recombinants/no. offsprings
s_514T	*A. chinensis (2n)*	C8 x A54.19	220/220 x 212/217	212/220	217/220	217	0/94
	A. chinensis (4n)	C3 x male pool [b]	203/211/214 x [b]	203/211/214, 203/214, 211/214	203/209/214, 203/209/211, 209/211/214	209	0/30
	A. deliciosa (6n)	Koryoku x Matua	203/206/208 x 203/206/208/229	203/206/208, 203/206, 203/208	203/206/208/229, 203/206/229, 206/208/229	229	0/31
	A. deliciosa (6n)	Koryoku x C10	203/206/208 x 203/208/217/220	203/206/208/217/220, 203/208/217/220, 206/208/217/220, 203/208/220,	203/206/8/217/220, 203/208/217/220, 203/206/208/220, 203/208/217, 203/208/220	null [i]	- [i]
	A. deliciosa (6n)	Katiuscia x Matua	203/206/220 x 203/206/208/229	203/206/208, 203/206, 203/206/220	203/206/208/229, 203/206/229, 203/206/208/220/229	229	0/16
s_514B	*A. chinensis (2n)*	C8 x A54.19	167/188 x 217/229	167/229, 188/229	167/217, 188/217	217	0/94
	A. chinensis (4n)	C3 x male pool [c]	192/217/223 x [c]	Not scorable	Not scorable	223? [j]	- [j]
	A. deliciosa (6n)	Koryoku x Matua	210/226 x 210/214/226	210/226	210/214/226, 214/226	214	0/30
	A. deliciosa (6n)	Koryoku x C10	210/226 x 217/223/226	210/217/223/226, 210/217/226, 217/223/226, 217/226	210/217/223/226, 210/217/226, 210/217/223, 210/217, 217/226	null [i]	- [i]
	A. deliciosa (6n)	Katiuscia x Matua	210 x 210/214	210	210/214	214	0/16
s_776T	*A. chinensis (2n)*	C8 x A54.19	203/203 x 203/206	203/206	203/203	203 [f]	0/94
	A. chinensis (4n)	C3 x male pool [d]	188/196/199/202 x [d]	many	many	205	0/30
	A. deliciosa (6n)	Koryoku x Matua	182/188/194 x 176/188/194	182/188/194, 182/188	176/182/188/194, 176/182/188, 176/182	176	0/27
	A. deliciosa (6n)	Koryoku x C10	182/188/194 x 176/188/194	176/182/188, 176/188/194, 182/188/194, 182/188	176/182/188/194, 176/182/194, 182/188/194	null [i]	- [i]
	A. deliciosa (6n)	Katiuscia x Matua	182/194 x 176/182/194	182/194	176/182/194	176	0/16
s_776B	*A. chinensis (2n)*	C8 x A54.19	233/233 x 231/null	231/233	233/null	null [g]	0/94
	A. chinensis (4n)	C3 x male pool [e]	228/234 x [e]	not scorable	not scorable	-	-

Table 4 (continued)

Marker	Species	Cross parents [a]	Parent allelic profile (female parent first)	Female progeny allelic profiles	Male progeny allelic profiles	Gender-linked allele	Recombinants/no. offsprings
	A. deliciosa (6n)	Koryoku x Matua	230 [h] x 230/234	230/234, 230	230/234, 230	null [i]	- [i]
	A. deliciosa (6n)	Koryoku x C10	230 [h] x 119	119/230	119/230	null [i]	- [i]
	A. deliciosa (6n)	Katiuscia x Matua	230/234 x 230/234	230/234, 230, 234	230/234	null [i]	- [i]

[a] In crosses between polyploids (4*n* and 6*n*), multiple copies of alleles cannot be detected and therefore those occurrences are not marked in the profiles

[b] A pool of pollen of male individuals from the following controlled cross (C3 x A125.93). The parent allelic profiles were 203/211/214 x 203/209

[c] A pool of pollen of male individuals from the following controlled cross (C3 x A125.93). The parent allelic profiles were 192/217/223 x 223/226

[d] A pool of pollen of male individuals from the following controlled cross (C3 x A125.93). The parent allelic profiles were 188/196/199/202 x 182/205/208

[e] A pool of pollen of male individuals from the following controlled cross (C3 x A125.93). The parent allelic profiles were 228/234 x 228/234

[f] The presence of the allele 203 must be associated with the absence of the alternative allele 206 segregating from the male parent

[g] The presence of the null allele must be associated with the absence of the alternative allele 231 segregating from the male parent

[h] A further allele of 119 bp is present and has been considered belonging to a second locus

[i] Likely presence of a null not scorable allele

[j] The allele 223 likely linked to the sex in the male parent was present also in the female parent

Results and discussion

RAD and SSR genotyping

A total of 540 million reads were produced on Illumina HiSeq2500 corresponding to some 270 M pairs. About 92 % of raw reads contained the expected restriction site overhang on both pairs (*SphI* on forward read and *MboI* on reverse reads) along with discriminating inline barcodes. The average number of successfully de-multiplexed reads per sample was 2.8 M, with a standard deviation of 1.2 M, excluding parents. These latter ones obtained 7.8 and 16.3 M reads for the female and the male parent, respectively, by increasing their load to guarantee the coverage saturation of loci (Fig. 1). Reads were all trimmed to a final length of 85 bp and aligned to the reference genome of *A. chinensis* recently published (Huang et al. 2013). As much as 93.9 and 95.7 % reads from female and male parent, respectively, were successfully mapped to the reference genome. An in silico prediction of double-digested fragments in the range of 220–370 bp on the available assembly was used to simulate a gel selection of 300–450 bp (having some 80 bp introduced by pre-PCR adapters ligation). This exercise led to the estimation of some 30,000 fragments, which would result in 60,000 RAD loci given that single-end read-stacks are analyzed independently. Figure 1 depicts an excess of loci collected on each individual sample, with parental lines counting up to 123,618 and 113,398. This phenomenon finds explanation in the fact that each individual can stochastically yield loci that are out of the target. This effect is also visible in progeny samples, where the yield of more reads directly reflects in the counting of more loci. However, by simply considering loci that are matching across a minimum set of 65 samples (i.e., they are truly enriched in the given size selection), we obtained some 56,700, which agrees with our prediction. Moreover, by analyzing the coverage distribution of the over-sequenced female sample, we calculate a weighted median coverage of 150× while all genomic coordinates with a coverage above 30× and 75× totalized for the 92.6 % and the 81.4 % of all the available reads. This indicated that our complexity reduction strategy was effective in sampling a well-defined set of genomic loci ensuring cost-effectiveness of the sequencing output and providing the possibility of accurate heterozygous versus homozygous calling.

A total of 73,993 unfiltered SNP sites were obtained with the analysis of parental genotypes over a set of 152,966 candidate genomic loci with 33,573 and 28,069 SNPs being identified in the female and male parents, respectively (Table 1). This was achieved by adopting filtering criteria such as minimum coverage and bounded SNP probability model (see "Materials and methods" section). A catalog of candidate segregating RAD loci was generated for the two parents by means of called haplotypes (rather than each single SNP site), given the possible presence of more than one SNP site per RAD site.

By matching offspring haplotypes against the populated catalog of polymorphic RAD sites and applying a minimum threshold of 65 successfully genotyped individuals, 12,586 RAD loci were found to be polymorphic and to segregate in a test cross fashion. Along with a set of 167 genotyped SSR markers, RAD loci were split according to the two-way pseudo-test cross design, collecting 6347 and 6470 informative markers for the female parent and the male one, respectively.

Linkage map

Grouping of RAD markers with the R/QTL package provided 29 linkage groups for the female parent, totaling 6112 markers; by side, the male map included 6262 markers, with the same number of groups, corresponding to the haploid series of chromosomes of the diploid *A. chinensis*. The ordering procedure conducted with MSTmap provided a first framework to apply the correction/imputation routine on the large data frame. After correction and final ordering of bins along SSR markers (see "Materials and methods" section), the female map consisted of 1300 genetic bins, comprising 6112 RAD loci, and 132 SSRs; the male map had 1364 genetic bins, containing 6262 RAD loci and 109 SSR markers (Table 1). SSR markers segregating from both parents along with shared polymorphic RAD loci were able to align the 29 linkage groups of the two parents.

Total map length was 3614 cM for the female parent and 3276 cM for the male one, with a mean distance between genetic bins of 2.78 and 2.40 cM for the female and male maps, while raw map densities of 0.58 and 0.51 cM, were reached, respectively (Table 1 and Online Resources 2). As expected, the map length exceeded by far the length of previously published ones, which contained a more limited number of markers. For instance, the female and male maps published in 2009 by Fraser and coworkers (2009) contained 464 and 365 markers and had a length of 2266 and 2078 cM, respectively. Authors reported a theoretical estimate of 3090 and 2782 cM, by means of the method 3 as proposed by Chakravarti et al. (1991). Shorter linkage maps as compared to those reported in this paper are likely due to irregular marker distribution and large gaps which lead to underestimate the genetic distances. Vice versa, the map inflation at the increasing of markers number is known being mainly due to missing values and/or genotyping errors that reduce the proportion of correctly ordered markers and provide a less precise estimation of recombination distances (Hackett and Broadfoot 2003; Van Os et al. 2005). The map length inflation could be exacerbated if markers with no recombination are not removed from the set and when a consensus map is preferred to the individual parental maps (Ronin et al. 2012). These are likely the reasons for which the linkage map used to anchor scaffolds in the paper of kiwifruit genome assembly was 5504 cM long (Huang et al. 2013). Obviously, in the last map, the correct calculation of genetic distances was sacrificed in favor of a better scaffold ordering.

Female and male maps, aligned through common SSR markers and scaffold coordinates of RAD markers, showed similar marker saturation with very small arrangements in the marker order. An exception, of which an explanation has not been found, is represented by chromosome 10, where the male linkage group is by far shorter than the female one with only 11 markers and a total length of 42.0 cM against the 52 markers and the 124.3 cM of the female linkage group. Skewed segregations were observed for a number of markers, but they were not analyzed in details, considering that they little affect the accuracy of maps.

SSR marker sequences together with scaffold coordinates of RAD markers allowed anchoring of all linkage groups to the chromosomal pseudomolecoles of the assembly by Huang et al. (2013). This information was exploited as a proxy to assess reliability of ordering across the two parental maps by reciprocal alignment (Fig. 2 and Online Resources 3).

Genetic placement of scaffolds sequences across the two parental maps was able to anchor some 90 % of the current genome assembly as produced by Huang et al. (2013), corresponding to 557 Mbp. This resulted in the ability to anchor some 120 Mbp of previously unplaced scaffold sequence, raising the total amount of mapped sequence to 571 Mbp (92.6 % of the current assembly) from the 333.6 Mbp of the current release. Moreover, independent segregation analysis of the two parents ensured unbiased detection of false scaffold joining in the current version of the genome pseudomolecules. These misplaced scaffolds appear in our maps conflicting with their position in the genome assembly by means of several independent RAD segregations. Figure 3 depicts the male linkage group 16, corresponding to the pseudomecule chr_16 in Huang et al. (2013) where most of misplacements were found: The discrepancy that we recorded is that of several scaffolds, currently attributed to Chr10 unambiguously mapped to Chr16, for a total of 4.5 Mbp. Other events of scaffold misplacement were detected in Chr6, Chr16, Chr18, Chr19, Chr20, and Chr21 (Table 2).

Mapping the gender determinant and development of associated molecular markers

Gender was recorded as heterozygous in male parent (XY) and homozygous in the female parent (XX), considering the model of genetic control of sex reported in the literature that is similar to that of *Silene*, with the male the heterogametic gender (Testolin et al. 1995). This Mendelian control of sex is maintained at any ploidy level (Testolin et al. 1995). Gender mapped in the linkage group 25 of the male parent, 2.166 cM from the top (Online Resources 2).

There are four markers suggested in the literature being associated to the gender, SmX, Ke225, UDK096, and SmY1, that span 4 cM in the region where the male phenotype

maps (Fraser et al. 2009; McNeilage et al. 2012). We tested SmX, SmY1, and UDK096, and they were monomorphic or did not produce PCR products, although different annealing temperatures were assayed. Ke255 was not tested because primers were not published.

The analysis of the four SSR markers isolated from the scaffolds 514 and 776, whose RAD markers co-segregated with the sex determinant in the mapping population (Table 3), revealed a rather complex scenario (Table 4). Alleles in coupling with the sex determinant were identified for all four markers in the mapping population (C8 x A54.19). No recombinant was scored and, therefore, all four markers appeared suitable to screen the gender in the progeny, even in the cases in which the male-linked allele was a null allele (e.g., marker s_776B) or the same allele was present also in the female parent, e.g., marker s_776T (Table 4). In the latter case, the heterozygosity of the marker in the male parent helped to discriminate between male and female profiles in the progeny.

The analysis of marker segregation in crosses of higher ploidy level and in crosses of species that are different from which markers had been isolated increased the difficulty in scoring the profiles. Nevertheless, the alleles linked to the sex determinant were easily identified on the male parents and offsprings, except in the cases in which the sex-linked allele was not scorable because of the lack of amplification or in the case in which the allele likely coupled with the sex in the male parent had the same size of an allele carried by the female parent (Table 4).

Conclusions

The ddRADseq protocol for genotyping by sequencing, as described by Peterson et al. (2012), has shown to be a robust method for the identification and mapping of SNP markers in outcross species such as *A. chinensis*. The ability to obtain a controlled complexity reduction by means of the choice of two restriction enzymes coupled with size selection of fragments allowed to maintain adequate coverage across most RAD loci and thus yielding accurate heterozygous versus homozygous haplotype calling. Moreover, our results show that even in the presence of a reference genome, the genetic maps produced in this work can help to improve the genome assembly. Our maps can place about 120 Mbp of previously unanchored sequence of the genome assembly. Furthermore, its potential to correct some false joining that were originated in the first anchoring procedure of the genome assembly (Huang et al. 2013) was assessed.

The saturated map produced by merging ddRAD haplotypes and SSRs, with quite regularly spaced markers, easily helps to find genetic loci that control Mendelian traits like the gender and will help finding loci controlling quantitative traits (QTLs) in the extended population. The latter work is in progress.

The new four SSR markers, identified in the scaffolds whose RAD markers co-segregated with the sex determinant in the mapping population, were able to discriminate between male and female progeny in most analyzed crosses, whatever the ploidy level of cross parents and whatever the species essayed. In all cases, the identification of the allele carried by the male parent was necessary for the analysis of the alleles associated to the gender in the progeny. Because the sex-linked alleles were not always of the same length in the different crosses, what cannot be done with these markers is to screen germplasm collections where the parentage of accessions is not known in advance.

Acknowledgments This research was developed within the program Chile-Italia D09I1136 "Mejoramiento Genético del Kiwi apoyado en la Selección Asistida por Marcadores" supported by the Chilean Government- FONDEF. Authors are grateful to Dr. Ross Ferguson for the manuscript revision.

References

Catchen J, Amores A, Hohenlohe P, Cresko W, Postlethwait J (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. G3: Genes, Genomes, Genetics 1:171–182

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

Cipriani G, Morgante M (1993) Evidence of chloroplast variation DNA variation in the genus Actinidia revealed by restriction analysis of PCR - amplified fragments. J Genet Breeding 47:319–326

Crowhurst RN, Gleave AP, MacRae EA et al (2008) Analysis of expressed sequence tags from *Actinidia*: applications of a cross species EST database for gene discovery in the areas of flavor, health, color and ripening. BMC Genomics 9:351

Fraser LG, Tsang GK, Datson PM, De Silva HN, Harvey CF, Gill GP, Crowhurst RN, McNeilage MA (2009) A gene-rich linkage map in the dioecious species *Actinidia chinensis* (kiwifruit) reveals putative X/Y sex-determining chromosomes. BMC Genomics 10:102. doi:10.1186/1471-2164-10-102

Hackett CA, Broadfoot LB (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Heredity 90:33–38. doi:10.1038/sj.hdy.6800173

Huang WG, Cipriani G, Morgante M, Testolin R (1998) Microsatellite DNA in *Actinidia chinensis*: isolation, characterisation, and homology in related species. Theor Appl Genet 97:1269–1278

Huang S-X, Ding J, Deng D et al (2013) Draft genome of the kiwifruit *Actinidia chinensis*. Nat Communications 4:2640. doi:10.1038/ncomms3640

Kumar S, Chagné D, Bink MCAM, Volz RK, Whitworth C, Carlisle C (2012) Genomic selection for fruit quality traits in apple (*Malus* x *domestica* Borkh.). PLoS ONE 7(5):e36674. doi:10.1371/ journal.pone.0036674

Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

McNeilage MA, Fraser LG, Tsang GK, Datson PM, De Silva HN, Crowhurst RN, Ferguson AR (2012) Molecular genetics and genomics and kiwifruit breeding. Acta Horticulturae 913:63–70

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Morgante M, Hanafey M, Powell W (2001) Microsatellites are preferentially associated with non repetitive DNA in plant genomes. Nat Genet 30:194–200

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7(5), e37135. doi:10.1371/journal.pone.003

Ronin Y, Mester D, Minkov D, Belotserkovski R, Jackson BN, Schnable PS, Aluru S, Korol A (2012) Two-phase analysis in consensus genetic mapping. G3 Genes Genomes Genetics 2(5):537–549. doi:10.1534/g3.112.002428

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Methods Mol Biol, vol 132. Humana Press, Totowa, pp 365–386

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J 3:739–744

Testolin R (2012) Breeding the future: what fruit breeders can learn from breeders of cows and chickens. Chronica Horticulturae 52(2):6–8

Testolin R (2013) Kiwifruit breeding: from the phenotypic analysis of parents to the genomic estimation of their breeding value (GEBV). Acta Horticulturae 913:123–130

Testolin R, Cipriani G, Costa G (1995) Sex segregation ratio and gender expression in the genus *Actinidia*. Sexual Plant Reprod 8:129–132

Van Ooijen JW (2006) JoinMap® 4, Software for the mapping of quantitative trait loci in experimental populations. Kyazma B.V, Wageningen

Van Os H, Stam P, Visser RGF, Eck HJ (2005) SMOOTH: a statistical method for successful removal of genotyping errors from high-density genetic linkage data. Theor Appl Genet 112:187–194

Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. Plos ONE. doi:10.1371/ journal.pgen.1000212

Data Archiving Statement

Raw Illumina reads used for this study are publicly available at NCBI SRA under the accession SRA293705 (project PRJNA294589, "Actinidia chinensis Map") as run SRR2229880, SRR2239887, SRR2239888, and SRR2239889.