CrossMark

ORIGINAL ARTICLE

Construction of a high-density genetic map of *Ziziphus jujuba* Mill. using genotyping by sequencing technology

Zhong Zhang[1,2,3] · Tianjun Wei[4] · Ying Zhong[5] · Xingang Li[1,2,3] · Jian Huang[1,3]

Abstract The Chinese jujube (*Ziziphus jujuba* Mill., $2n = 2 \times = 24$), one of the most popular fruit trees in China, is widely cultivated and utilized in Asia. High-density genetic linkage maps are valuable resources for molecular breeding and functional genomics; however, they are still under-developed for the jujube. The genotyping by sequencing (GBS) strategy could be an efficient and cost-effective tool for single nucleotide polymorphism (SNP) discovery based on the sequenced jujube genome. Here, we report a new high-density genetic map constructed using GBS technology. An F1 population with 145 progenies and their parents ('Dongzao' × 'Zhongningyuanzao') were sequenced on the Illumina HiSeq 4000 platform. In total, 79.8 Gb of raw data containing 256,708,177 paired-end reads were generated. After data filtering and SNP genotyping, 40,372 polymorphic SNP markers were developed between the parents and 2540 (1756 non-redundant) markers were mapped onto the integrated genetic linkage map. The map spanned 1456.53 cM and was distributed among 12 linkage groups, which is consistent with the haploid chromosome number of the jujube. The average marker interval was 0.88 cM. The genetic map allowed us to anchor 224 Mb (63.7 %) of scaffolds from the sequenced 'Junzao' genome, containing 52 newly anchored scaffolds, which extended the genome assembly by 7 Mb. In conclusion, GBS technology was applied efficiently for SNP discovery in this study. The high-density genetic map will serve as a unique tool for molecular-assisted breeding and genomic studies, which will contribute to further research and improvement of the jujube in the near future.

Communicated by D. Grattapaglia

✉ Xingang Li
xingangle@nwsuaf.edu.cn

✉ Jian Huang
huangj@nwsuaf.edu.cn

1   College of Forestry, Northwest A&F University, Yangling, Shaanxi 712100, China

2   Research Centre for Jujube Engineering and Technology, State Forestry Administration, Yangling, Shaanxi 712100, China

3   Key Comprehensive Laboratory of Forestry, Shaanxi Province, Yangling, Shaanxi 712100, China

4   Institute of Gemplasm Resource, Ningxia Academy of Agricultural and Forestry Science, Yinchuan, Ningxia 750000, China

5   Novogene Bioinformatics Institute, Beijing 100000, China

Introduction

The Chinese jujube (*Ziziphus jujuba* Mill.) is one of the most popular fruits in Asia, known for its delicious taste, high nutritional value, and high sugar and vitamin C contents (Chen et al. 2013; Gao et al. 2013; Liu and Wang 2009). With the rapid increase in jujube cultivation, there is an urgent need for breeding improvement (Wang et al. 2007). Despite great efforts that have been made in crossbreeding, limited progress has been achieved in the past decades due to the unique biological characteristics of the jujube tree, such as early flowers dropping, small flower size, and high percentages of seedless fruits (Liu et al. 2014a; Ma et al. 2008). Consequently, jujube breeding is far behind that of other popular fruit trees (Zhang et al. 2015). In particular, breeding strategies of crops have involved molecular marker-assisted selection (MAS) in which

a molecular genetic linkage map is required as a fundamental resource; however, few high-density genetic maps have been constructed for the jujube. Therefore, an advanced breeding strategy and more genetic maps are necessary for jujube breeding as well as for identifying genes related to important genotypes.

The first molecular genetic map of the jujube was reported by Shen (2005), based on an F1 population generated from 'Dongzao' × 'Linyillizao'. It comprised 14 linkage groups (LGs) and spanned 1237.4 cM with 333 amplified fragment length polymorphism (AFLP) markers. Qi et al. (2009) improved the map to 15 LGs spanning 1309.4 cM with 388 AFLP markers and 35 random amplified polymorphic DNA (RAPD) markers, with an average marker interval of 3.1 cM. However, these two linkage maps are not sufficient for fine mapping of quantitative trait loci (QTLs) due to the long marker interval distance. Also, AFLP and RAPD markers are not compatible for genome anchoring due to the lack of sequencing information. More recently, a high-density genetic map derived from the cross of Z. jujuba 'JMS2' × Z. jujuba var. spinosa 'Xing 16' was reported (Zhao et al. 2014) that contained 2872 SNP markers with an average marker interval of 0.34 cM and comprising 12 LGs. This map provides new insight for jujube crossbreeding and genome research, and also shows great potential for the discovery and application of SNP markers with next-generation sequencing (NGS) technology.

Several strategies for genome complexity reduction have proved to be efficient for SNP discovery in various fruit species, such as the grape (Wang et al. 2012a), pear (Wu et al. 2014), peach (Bielenberg et al. 2015), and sweet cherry (Guajardo et al. 2015). The jujube has a genome size of ~360 Mb with a high level of heterozygosity (Liu et al. 2014b; Wu et al. 2013). A rapid and low cost approach had been considered for application in a large mapping population. Genotyping by sequencing (GBS) utilizes one or multiple restriction enzymes to digest genomic DNA into fragments that can be sequenced on high-throughput platforms (Elshire et al. 2011). A low cost for each sample can be achieved by multiplexing different samples and adding specific barcodes ligated to digested fragments prior to sequencing (Huang et al. 2014; Pootakham et al. 2015). GBS has been widely used in genetic linkage map construction (Bielenberg et al. 2015; Fiedler et al. 2015; Moumouni et al. 2015), scaffold anchoring, and genome assembly (Glazer et al. 2015; Mathew et al. 2014; Soto et al. 2015).

From the perspective of jujube breeding and improvement, a cross between two excellent cultivars would offer great potential for selection of superior progenies. Therefore, we performed hybridization between two excellent cultivars: 'Dongzao' and 'Zhongningyuanzao'. Based on the F1 population, a new high-density genetic map was constructed using a GBS strategy. Sequenced scaffolds from the reference genome of Z. jujuba Mill. 'Junzao' (Accession: LPXJ 00000000) were also anchored based on this map.

Materials and methods

Plant material and DNA extraction

Two jujube cultivars, 'Dongzao' (8 years old) and 'Zhongningyuanzao' (8 years old), were selected as the hybrid parents. They exhibited obvious differences in several phenotypes, such as fruit shape, fruit quality, and fruit maturity time. The hybridization was performed in Zhongning County (37.48° N, 105.70° E), Ningxia, China, in 2012. In total, 385 seeds were harvested and 294 were grown into seedlings. Among them, 145 individuals were selected randomly for genetic map construction. All progenies and the two parents were maintained in the experimental field of the Ningxia Academy of Agricultural and Forestry Science (Zhongning County, Ningxia).

Young leaves from the parents and progenies were collected at 3 weeks after sprouting in May 2015. Collected leaves were immediately stored in liquid nitrogen, then transferred to the laboratory and stored at −80 °C. Genomic DNA was isolated using the MiniBEST Plant Genomic DNA Extraction Kit 9768 (Takara, Dalian, China) following the manufacturer's protocol. DNA purity and concentration were determined using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, USA), and finally checked on 1 % agarose gels.

GBS protocol and high-throughput sequencing

We used a GBS strategy in this study to develop SNP markers. First, we performed a GBS pre-design for restriction enzyme selection. In this step, the enzyme combination and sizes of digested fragments were predicted and evaluated according to the reference genome of 'Junzao' (Accession: LPXJ 00000000). Generally, the following characteristics were required for the selection of appropriate restriction enzymes: (1) The digested tags should be evenly distributed throughout the sequences to be examined; (2) Repeated tags should be avoided; (3) The paired-end length of each tag should be suitable for the coverage of the Illumina HiSeq 4000 sequencing platform; (4) The tag numbers must be sufficient for the subsequent steps. To maintain sequence depth uniformity of different fragments, a tight length range of ~50 base pairs (bp) was selected.

Next, we constructed the GBS library according to the predesigned scheme. The pipeline diagram was shown in Fig. 1. Here, three restriction enzymes were selected for DNA digestion and 0.1–1 μg of genomic DNA was incubated at 37 °C with MseI (New England Biolabs; NEB), T4 DNA ligase

Fig. 1 GBS pipeline diagram. The genomic DNA were digested with *Mse*I and ligated with Y adapter N containing a barcode for each individual. Then two additional restriction enzymes *Hae*III and *Eco*RI were simultaneously added into the *Mse*I ligations for further digestion. Then PCR was applied for removing the fragments lacking *Mse*I cutting sites in both pair-end and adding the index as well as completing i5 and i7 sequences on the Y adapters. Then PCR products were purified and pooled, and run on a 2 % agarose gel. Finally, fragment length varied from 400 to 450 bp (with indexes and adaptors) were cut from the gel and purified for sequencing
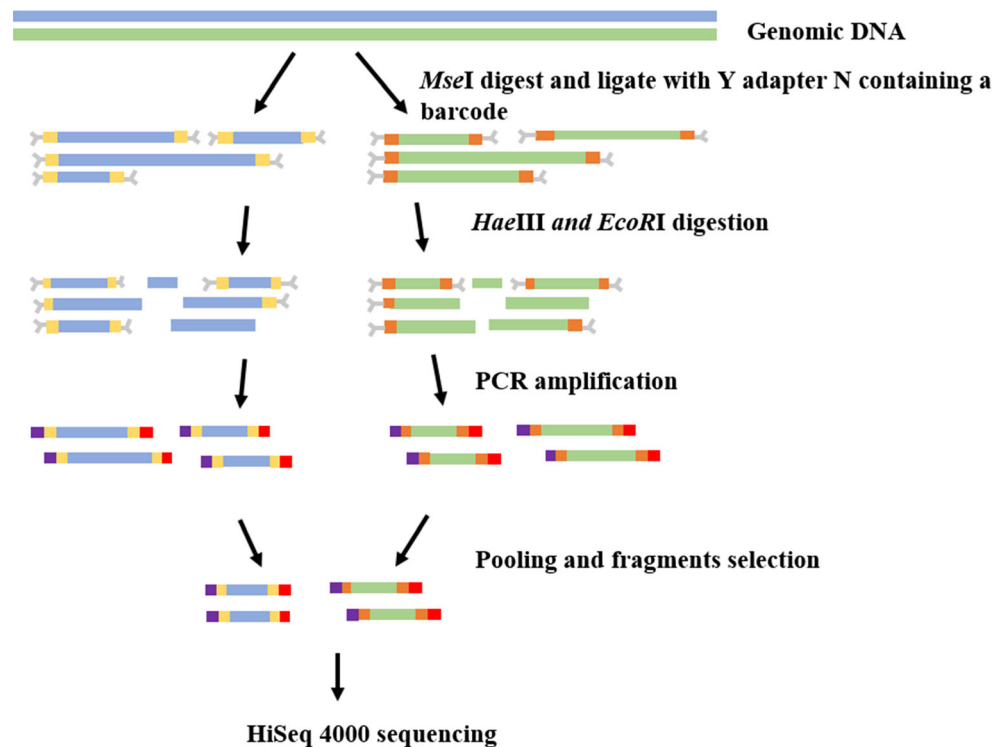
(NEB), ATP (NEB), and *Mse*I Y adapter N containing barcodes, and then heat-inactivated at 65 °C. To further decrease the complexity and increase the sequencing depth and genome coverage, two additional enzymes, *Hae*III and *Eco*RI (NEB), were simultaneously added into the *Mse*I digestions to further digest the fragments at 37 °C. Then, the digested fragments with ligations were purified with Agencourt AMPure XP (Beckman) and subjected to PCR amplification using Phusion Master Mix (NEB) with universal primers as well as index primers to add index and complete i5 and i7 sequences. The PCR products were purified using Agencourt AMPure XP (Beckman) and pooled, then run on a 2 % agarose gel. Fragments of 400–450 bp (with indexes and adaptors) were cut from the gel and purified with a gel extraction kit (QIAGEN). These purified products were then further cleaned with Agencourt AMPure XP (Beckman) prior to sequencing. Then, paired-end 150-bp sequencing was performed on the selected tags on the Illumina HiSeq 4000 platform.

Sequence data analysis

The sequence data from each individual were sorted according to the barcodes in the raw reads. To ensure that reads were reliable and without artificial bias (such as low quality paired reads, which resulted mainly from base-calling duplicates and adapter contamination) in the following analysis, raw data (raw reads) in FASTQ format were first processed through a series of quality control (QC) procedures with in-house C scripts. QC standards were as follows: (1) Reads with ≥10 % unidentified nucleotides (Ns) were removed, (2) Reads with >50 % of bases having a Phred quality <5 were removed, (3) Reads with >10 nt aligned to the adapter were removed, allowing ≤10 % mismatches, and (4) Reads containing the *Hae*III or *Eco*RI sequences were removed. Then, the Burrows-Wheeler Aligner (BWA) software (Li and Durbin 2009) was used to align the clean reads from each individual against the reference genome (settings: mem –t − 4 –k 32 –M -R). Alignment files were converted to bam files using the SAMtools software (Li et al. 2009) (settings: –bS –t). If multiple read pairs had identical external coordinates, only the pair with the highest mapping quality was retained.

SNP calling and genotyping

SNP calling was performed for parents and progenies using the SAMtools software (Li et al. 2009). The number of SNPs and types of transitions or transversions were counted. Then, a Perl script was used to filter the SNPs that had more than two genotypes. Polymorphic markers between the two parents were detected and classified into eight segregation patterns (*ab × cd, ef × eg, hk × hk, lm × ll, nn × np, aa × bb, ab × cc*, and *cc × ab*) according to the CP model in JoinMap 4.0 software (Van Ooijen 2011).

Linkage map construction and evaluation

Markers showing significantly distorted segregation ($P < 0.001$), integrity (> 65 %), or containing abnormal bases were filtered by JoinMap 4.0. The segregation patterns $hk \times hk$ and $nn \times np$ were used for the male parent map construction while the patterns $lm \times ll$ and $hk \times hk$ were used for the female parent map using JoinMap 4.0. The regression algorithm, three times circulation sequence, and Kosambi (1943) mapping function were used in marker distance calculation. The LOD value was set to 2.0–20. The integrated map for both the male and female parents was computed using the combine group for map integration function in MergeMap software (Wu et al. 2008). A Perl script SVG was used to visualize exported maps, and heat maps were constructed to evaluate the maps.

Anchoring sequenced scaffolds to the genetic map

Markers localized on the genetic map were used to anchor sequenced scaffolds of 'Junzao' using a Perl script. The visualized map of LGs that correlated with anchored scaffolds was then constructed and exported.

Results

Restriction enzyme selection for library construction

Before construction of the sequencing library, restriction enzymes were evaluated based on the predicted number of tags, length of fragments, and distribution across the reference genome. After scanning the entire 'Junzao' genome, a combination of three restriction enzymes, *Mse*I (TTAA), *Hae*III (GGCC), and *Eco*RI (GAATTC), was selected for GBS library construction. According to their restriction sites, the predicted sizes of the DNA fragments ranged from 400 to 450 bp and the number of tags was 99,050 with only 3.4 % repetitive tags. These tags covered 332,817,121 bp (94.74 %) of the genome and were distributed across 2745 scaffolds, accounting for 7.6 % of total scaffolds (Online Resource 1 ESM_1).

Analysis of GBS data

In total, 79.8 Gb of raw data containing 256,708,177 paired-end reads were generated by sequencing the parents and 145 progenies. After data filtering, 92 % of reads were high quality, with an average Q20 ratio of 93.79 % and a GC content of 35.31 %. The parents were sequenced at a higher level to enhance the chances of detecting more SNP markers. Finally, clean data covering 1,480,832,064 (99.99 %) and 1,840,679,712 (99.99 %) bp were obtained for the female and male parents, respectively. For each individual, the clean

data ranged from 189,546,336 to 700,333,632 bp, with an average of 486,968,574 bp (Online Resource 2 ESM_2).

High-quality clean reads were aligned against the 'Junzao' genome. Consequently, 9,876,597 clean reads were obtained for the female parent and 7,802,479 clean reads for the male parent. For F1 individuals, an average of 2,789,866 clean reads was aligned to the reference genome. From all of these data, 98 % could be aligned to unique positions on the genome, while the remaining 2 % were aligned to either multiple or zero positions (Online Resource 3 ESM_3). Only reads aligned to unique positions on the reference genome were retained for the following SNP calling and genotyping.

SNP discovery and genotyping

SNP calling of the two parents and F1 individuals was performed with SAMtools (Li et al. 2009). In total, 136,178 and 109,275 SNPs were detected in the female and male parents, respectively. For F1 individuals, an average of 68,921 SNPs was discovered for an individual progeny. Among them, the parents exhibited a higher heterozygous rate (27.89 %) of SNPs than the F1 individuals (18.51 %) (Online Resource 4 ESM_4).

By excluding the missing information, 40,372 (54.99 %) polymorphic SNPs were detected between the two parents. These SNPs could be classified into eight segregation types according to the CP model in JoinMap 4.0. As shown in Fig. 2, seven patterns were detected. Among them, four major patterns including $lm \times ll$, $nn \times np$, $hk \times hk$, and $aa \times bb$ accounted for nearly 99.8 %, while the other three patterns, $ab \times cc$, $cc \times ab$, and $ef \times eg$, only accounted for 0.2 %. The type $aa \times bb$ was excluded, as it was homozygous in both parents and resulted in no segregation in F1 individuals. Only segregation types $lm \times ll$, $nn \times np$, and $hk \times hk$ were selected for genotyping in F1 individuals. Among the 35,733 selected
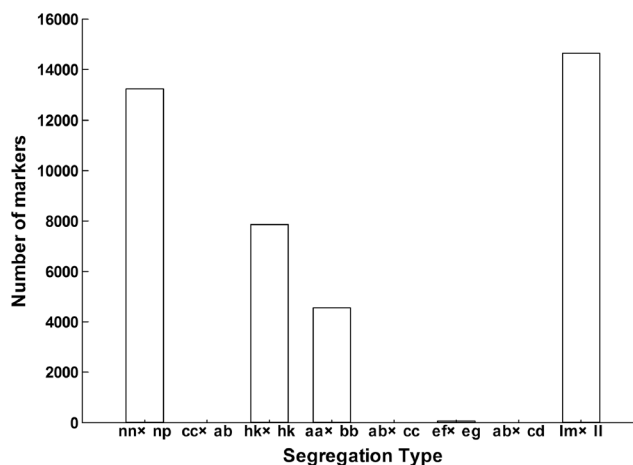
Fig. 2 Segregation types of polymorphic SNP markers. The *x-axis* indicates the eight segregation types; the *y-axis* indicates the corresponding number of markers

polymorphic SNPs, the ratios of transitions and transversions were 56.49 and 43.51 %, respectively. Each type of transition (A/G and C/T) and transversion (A/C, A/T, C/G, and G/T) was nearly balanced (Online Resource 5 ESM_5).

Genetic linkage map construction and evaluation

Available markers were filtered for <65 % integrity in each individual and chi-square tests with $P < 0.001$. Finally, 694 markers with $hk \times hk$, 1755 markers with $lm \times ll$, and 1656 markers with $nn \times np$ segregation types could be used for map construction (Online Resource 6 ESM_6). After data preparation, 2350 markers with types $nn \times np$ and $hk \times hk$ and 2449 markers with types $lm \times ll$ and $hk \times hk$ were used for the male and female map constructions, respectively. On the female map, 1409 (971 non-redundant) markers fell into 12 LGs and the genetic length was 1250.564 cM with an average marker interval of 1.38 cM (Online Resource 7 ESM_7). On the male map, 1214 (843 non-redundant) markers fell into 12 LGs; the genetic length was 1140.657 cM with an average marker interval distance of 1.49 cM (Online Resource 8 ESM_8). Heat maps reflect the linkage relationship between each marker in a single linkage group (Online Resource 9 ESM_9). Most of the LGs in both parent maps performed well in visualization.

Then the two parent maps were merged; the integrated map spanned 1456.53 cM with 2540 (1756 non-redundant) markers and fell into 12 LGs, which is consistent with the chromosome number of the jujube (Fig. 3, Online Resource 10 ESM_10). Among the 12 LGs, LG01 was the largest group, with a genetic distance of 183.8 cM and 292 markers. LG02 was the shortest group with 90 markers and spanning 74.27 cM. The average marker interval ranged from 0.54 to 1.20 cM, with an average distance of 0.88 cM (Table 1). Between the markers, 1745 gaps were detected. Among them, 1724 gaps (98.8 %) were less than 5 cM, 19 gaps were between 5 and 10 cM, and only two gaps were larger than 10 cM, which were on LG05 and LG07, respectively.

## Anchoring scaffolds of the sequenced jujube genome to the genetic map

The sequenced jujube genome ('Junzao') consisted of 36,147 scaffolds covering 351 Mb of sequences. The 1756 non-redundant markers were used as anchors to orient these scaffolds. In total, 316 unique scaffolds representing 224 Mb (63.7 %) of the total genome were localized on the 12 LGs (Fig. 4). LG01 anchored the highest number of scaffolds with a physical length of 32.8 Mb. LG12 anchored the lowest number of scaffolds with a length of 17.2 Mb. The correlation between genetic and physical length was 153.81 Kb/cM on average (Table 2).

Based on the scaffold orientation information of the draft genome of 'Junzao', we compared the orders and directions of scaffolds anchored by the current map. As shown in Table 3, our map shared 264 anchored scaffolds with the anchored genome. Of them, 180 (68.2 %) were in accordance with the placement orders. Most of the LGs showed a good collinear relationship between the different anchorings, such as LG04, 05, 06, and 11 (ESM_11a and b). In addition, 52 scaffolds covering 7,056,409 bp of data were recruited with the current map. These scaffolds were distributed on 12 LGs, and the sizes ranged from 406 to 1,210,059 bp. LG11 was enriched with the highest number (14) of new scaffolds, while LG02 and LG03 had the lowest number with only one new scaffold each. In this way, the linkage map of the 'Junzao' jujube genome was extended by 7 Mb with new anchored scaffolds (Online Resource 12 ESM_12).

Discussion

GBS is a rapid, efficient, and cost-effective strategy for SNP development, genetic linkage map construction, marker-based complex trait selection, and draft genome assembly in many species with or without reference genomes (Elshire et al. 2011; He et al. 2014; International Barley Genome Sequencing Consortium et al. 2012; Poland et al. 2012; Ward et al. 2013). Here, we used the combination of three restriction enzymes for GBS library construction. *Mse*I has a digestion site with a 4-bp sequence (TTAA) and has a higher distribution frequency on the jujube genome. The additional two enzymes *Hae*III and *Eco*RI were added to further digest fragments after *Mse*I. The combination of three restriction enzymes improved the efficiency of GBS by increasing the tag number, sequencing depth, and genome coverage, and also supplied more chances to detect suitable regions for targeted fragments. In total, 79.8 Gb of raw sequencing data with 92 % clean data were generated, and finally 98 % of the clean data were mapped to unique positions on the reference genome. Thus, these data are indicative of good results in the library construction. Moreover, 73,417 SNPs and 40,372 (54.99 %) polymorphic SNPs were detected between the two parents. A previous study revealed that 'Dongzao' and 'Zhongningyuanzao' had a distant genetic relationship (Huang et al. 2015), which could increase the chances of detecting more polymorphic markers. These results demonstrated that GBS is a low cost, high efficiency, and rapid approach for SNP development and map construction.

The integrated genetic linkage map comprised 2540 (1756 non-redundant) SNP markers and spanned 1456.33 cM, with an average marker interval of 0.88 cM. The map was divided into 12 LGs and was consistent with the haploid chromosome number. Compared with previously reported linkage maps (Qi et al. 2009; Shen 2005), a large number of markers with
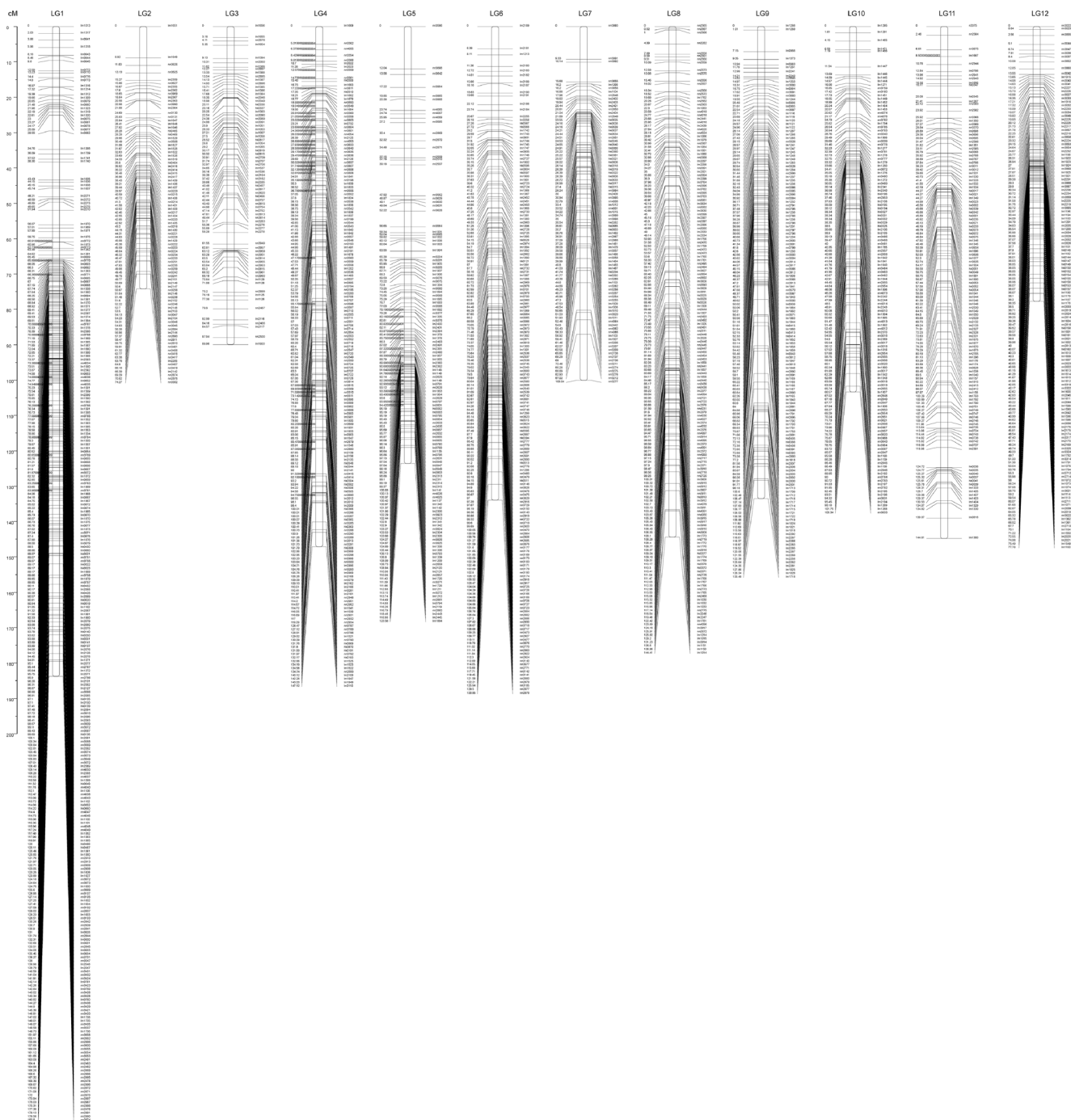
Fig. 3 Linkage groups 01–12 of the integrated map. The *x-axis* indicates the numbers of linkage groups; the *y-axis* indicates the genetic length (cM)

sequencing information as well as a higher map density with a shorter marker interval distance were achieved in the current study. However, this map is still lower density than the 'interspecific' map reported by Zhao et al. (2014), which is attributed to the different sequencing strategies and cross combinations. The RAD strategy could identify more SNP alleles with its higher genome coverage (Baird et al. 2008; Davey et al. 2011). Notably, interspecific hybridization used in genetic

map construction would contribute to more recombination and polymorphic markers (Wang et al. 2012a). However, the new high-density map developed by crossing two excellent jujube cultivars will provide more breeding resources for further research.

A high-density genetic linkage map can facilitate genome assembly and has been one of the fundamental components of genome sequencing (Gaur et al. 2012). Assisted by the high-

Table 1 Characteristics of the integrated genetic map derived from the F1 population ('Dongzao' × 'Zhongningyuanzao')

Linkage group	Number of markers		Genetic length (cM)	Average marker interval (cM)	Gap <5 cM	Maximum gap (cM)
	All SNPs	Non-redundant SNPs				
LG01	432	292	183.8	0.63	289	5.21
LG02	137	90	74.27	0.83	88	8.92
LG03	108	75	89.96	1.20	74	3.78
LG04	259	181	147.12	0.81	177	7.18
LG05	197	132	123.56	0.94	129	12.04
LG06	236	176	133.95	0.76	174	6.39
LG07	138	89	100.04	1.12	84	13.02
LG08	232	172	144.41	0.84	169	7.45
LG09	228	150	133.49	0.89	145	6.14
LG10	185	131	103.34	0.79	131	4.97
LG11	179	124	144.81	1.17	121	5.74
LG12	209	144	77.78	0.54	143	2.54
Average	212	146	121.38	0.88	144	–
Total	2540	1756	1456.53	–	1724	–

density genetic map, we anchored 63.7 % of assembled scaffolds (224 Mb) of the 'Junzao' genome. Comparison of our map to the previously anchored genome revealed a highly collinear relationship between the genetic and physical maps. Also, 52 newly anchored scaffolds extended the genome by 7 Mb, demonstrating that the high-density genetic map was useful and efficient for genome assembly improvement. Additionally, the inconsistent scaffold placement order may

be attributed to the different cultivars sequenced for the genetic map and genome sequencing. Rearrangements, translocations, gains or losses of DNA segments, and copy number variations (CNVs) have been widely observed in different genotypes of the same species (Swanson-Wagner et al. 2010; Zmienko et al. 2014). On the other hand, the markers on different genetic maps could influence the anchoring results. Certainly, improper mapping or errors present in the
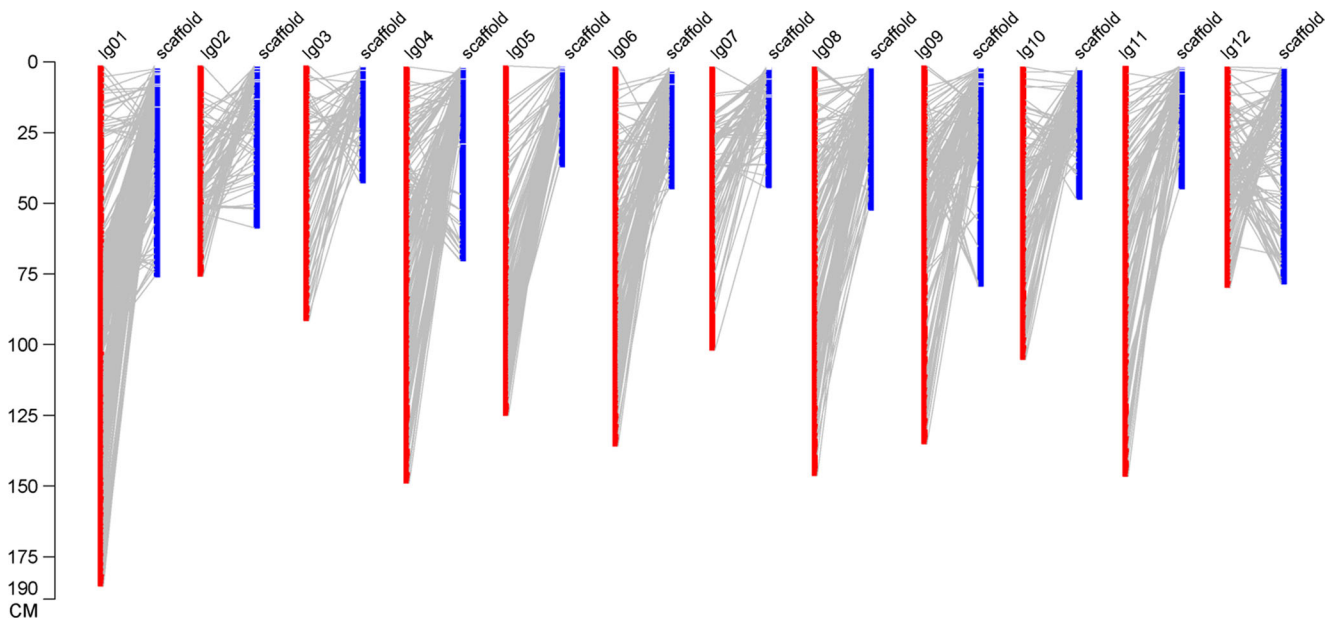
Fig. 4 Anchoring of sequenced scaffolds of 'Junzao' (Accession number: LPXJ00000000) to the 12 linkage groups. The *x-axis* indicates linkage groups (LG01–12) and scaffolds, the *y-axis* indicates the genetic length (cM)

Table 2 Anchored sequenced genome scaffolds of *Z. jujuba* 'Junzao' with SNP markers

Linkage group	Number of scaffolds	Number of anchored markers	Genetic length (cM)	Physical length (Kb)	Relationship of genetic map to physical length (Kb/cM)
LG01	41	292	183.80	32,821.96	178.57
LG02	19	90	74.27	15,006.77	202.06
LG03	16	75	89.96	13,610.75	151.30
LG04	32	181	147.12	22,721.20	154.44
LG05	29	132	123.56	15,294.72	123.78
LG06	29	176	133.95	21,279.97	158.87
LG07	21	89	100.04	13,122.14	131.17
LG08	29	172	144.41	20,740.73	143.62
LG09	24	150	133.49	17,379.20	130.19
LG10	25	131	103.34	17,923.40	173.44
LG11	37	124	144.81	16,928.61	116.90
LG12	14	144	77.78	17,196.32	221.09
Average	26	146	121.38	18,668.81	153.81
Total	316	1756	1456.53	224,025.78	–

genome assembly would also contribute to inconsistent placement orders (Soto et al. 2015). We calculated the relationships between the genetic and physical maps in the present study. The ratio of physical/genetic distance was 153.81 Kb/cM on average, and this information will be useful for further research on gene cloning and genome structure analysis.

Segregation distortion is regarded as the driving force of evolution (Cai et al. 2015; Taylor and Ingvarsson 2003). Segregation distortion is widely present in many species and is affected by population types and environmental factors (Wang et al. 2012b; Zhou et al. 2015; Zhou et al. 2014). It is perhaps caused by gametophyte and/or zygotic selection and chromosomal rearrangements (Faure et al. 1993; Wang et al. 2012b). Considering the self-incompatible behavior of the jujube (Asatryan and Tel-Zur 2013a, b), unsurprisingly, 2337 (36.28 %) distorted markers were found in the mapping process. A similar phenomenon with the female parent 'Dongzao' was also reported by Shen (2005). Although these markers could increase the map density and might not significantly influence QTL mapping (Xu 2008; Zhang et al. 2010; Zhang et al. 2013), these markers were filtered and not used in this study.

Table 3 Comparison of anchored scaffold information with 'Junzao' genome assembly

Linkage group	Number of anchored scaffolds	Anchored scaffolds in assembled genome	Common scaffolds	Colinear scaffolds	New anchored scaffolds	Size of new anchored scaffolds (bp)
LG01	41	65	37	19	4	402,810
LG02	19	45	18	12	1	2372
LG03	16	50	15	6	1	540
LG04	32	65	28	25	4	2,024,518
LG05	29	67	24	23	5	687,027
LG06	29	37	26	22	3	437,662
LG07	21	41	17	12	4	983,501
LG08	29	39	24	17	5	946,575
LG09	24	44	21	16	3	975,074
LG10	25	50	20	6	5	1,643,691
LG11	37	62	23	18	14	1,426,783
LG12	14	31	11	4	3	1,231,051
Total	316	596	264	180	52	7,056,409

Conclusions

We used GBS technology to develop SNP markers based on an F1 population derived from the cross of 'Dongzao' × 'Zhongningyuanzao'. Using this map, we anchored 63.7 % (224 Mb) of assembled genome scaffolds, containing 52 new scaffolds, which extended the draft genome assembly by 7 Mb. These results have enriched marker resources with sequencing information of the jujube. The newly constructed high-density genetic map will serve as an efficient tool for molecular-assisted breeding and genomic studies. This work will contribute to the further research and improvement of the jujube in the near future.

Data archiving statement The supplementary materials mentioned in the manuscript can be found in Online Resources ESM_1 through ESM_12. Detailed sequencing information of the SNP markers mapped onto the two parent maps can be found in Online Resources 7 and 8 (ESM_7 and ESM_8). The raw sequencing data reported here have been submitted to the Short Read Archive (SRA) of the National Center for Biotechnology Information (NCBI); related accession numbers are listed in Online Resource 13 (ESM_13).

References

Asatryan A, Tel-Zur N (2013a) Intraspecific and interspecific crossability in three *Ziziphus* species (Rhamnaceae). Genet Resour Crop Evol 61:215–233. doi:10.1007/s10722-013-0027-8

Asatryan A, Tel-Zur N (2013b) Pollen tube growth and self-incompatibility in three *Ziziphus* species (Rhamnaceae). Flora 208: 390–399. doi:10.1016/j.flora.2013.04.010

Baird NA et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376. doi:10.1371/journal.pone.0003376

Bielenberg DG et al. (2015) Genotyping by sequencing for SNP-based linkage map construction and QTL analysis of chilling requirement and bloom date in peach [*Prunus persica* (L.) Batsch]. PLoS One 10:e0139406. doi:10.1371/journal.pone.0139406

Cai C, Cheng FY, Wu J, Zhong Y, Liu G (2015) The first high-density genetic map construction in tree peony (*Paeonia* Sect. Moutan) using genotyping by specific-locus amplified fragment sequencing. PLoS One 10:e0128584. doi:10.1371/journal.pone.0128584

Chen J et al. (2013) Chemical and biological assessment of *Ziziphus jujuba* fruits from China: different geographical sources and developmental stages. J Agric Food Chem 61:7315–7324. doi:10.1021/jf402379u

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510. doi:10.1038/nrg3012

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379. doi:10.1371/journal.pone.0019379

Faure S, Noyer JL, Horry JP, Bakry F, Lanaud C, Gonzalez de Leon D (1993) A molecular marker-based linkage map of diploid bananas (*Musa acuminata*). Theor Appl Genet 87:517–526. doi:10.1007/BF00215098

Fiedler JD, Lanzatella C, Okada M, Jenkins J, Schmutz J, Tobias CM (2015) High-density single nucleotide polymorphism linkage maps of lowland switchgrass using genotyping-by-sequencing. Plant Genome 8. doi:10.3835/plantgenome2014.10.0065

Gao QH, CS W, Wang M (2013) The jujube (*Ziziphus jujuba* Mill.) fruit: a review of current knowledge of fruit composition and health benefits. J Agric Food Chem 61:3351–3363. doi:10.1021/jf4007032

Gaur R et al. (2012) High-throughput SNP discovery and genotyping for constructing a saturated linkage map of chickpea (*Cicer arietinum* L.). DNA Res 19:357–373. doi:10.1093/dnares/dss018

Glazer AM, Killingbeck EE, Mitros T, Rokhsar DS, Miller CT (2015) Genome assembly improvement and mapping convergently evolved skeletal traits in sticklebacks with genotyping-by-sequencing. G3 (Bethesda) 5:1463–1472. doi:10.1534/g3.115.017905

Guajardo V, Solis S, Sagredo B, Gainza F, Munoz C, Gasic K, Hinrichsen P (2015) Construction of high density sweet cherry (*Prunus avium* L.) linkage maps using microsatellite markers and SNPs detected by genotyping-by-sequencing (GBS). PLoS One 10:e0127750. doi:10.1371/journal.pone.0127750

He J, Zhao X, Laroche A, ZX L, Liu H, Li Z (2014) Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front Plant Sci 5:484. doi:10.3389/fpls.2014.00484

Huang J, Yang X, Zhang C, Yin X, Liu S, Li X (2015) Development of chloroplast microsatellite markers and analysis of chloroplast diversity in Chinese jujube (*Ziziphus jujuba* Mill.) and wild jujube (*Ziziphus acidojujuba* Mill.). PLoS One 10:e0134519. doi:10.1371/journal.pone.0134519

Huang YF, Poland JA, Wight CP, Jackson EW, Tinker NA (2014) Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat. PLoS One 9:e102448. doi:10.1371/journal.pone.0102448

International Barley Genome Sequencing Consortium et al. (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491:711–716. doi:10.1038/nature11543

Kosambi DD (1943) The estimation of map distance from recombination values. Ann Eugenics 12:172–175. doi:10.1111/j.1469-1809.1943.tb02321.x

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. doi:10.1093/bioinformatics/btp324

Li H et al. (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352

Liu MJ, Wang JR, Liu P, Lin MJ, Xiao J, Liu ZG, Sun XC (2014a) Design and practice of emasculation-free cross breeding in Chinese jujube. Acta Hortic Sin 41:1495–1502

Liu MJ, Wang M (2009) Germplasm resource of Chinese jujube. China Forestry Publishing House, Beijing

Liu MJ et al. (2014b) The complex jujube genome provides insights into fruit biology. Nat Commun 5:5315. doi:10.1038/ncomms6315

Ma QH, JR X, Wang GX, Yao LX (2008) Progress on hybrid breeding of Chinese jujube (*Ziziphus jujuba* Mill.). Chin Agric Sci Bull 24:174–178

Mathew LS et al. (2014) A first genetic map of date palm (*Phoenix dactylifera*) reveals long-range genome structure conservation in the palms. BMC Genomics 15:285. doi:10.1186/1471-2164-15-285

Moumouni KH, Kountche BA, Jean M, Hash CT, Vigouroux Y, Haussmann BIG, Belzile F (2015) Construction of a genetic map for pearl millet, *Pennisetum glaucum* (L.) R. Br., using a genotyping-by-sequencing (GBS) approach. Mol Breed 35:5. doi:10.1007/s11032-015-0212-x

Poland J et al. (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5:103–113. doi:10.3835/plantgenome2012.06.0006

Pootakham W et al. (2015) Construction of a high-density integrated genetic linkage map of rubber tree (*Hevea brasiliensis*) using genotyping-by-sequencing (GBS). Front Plant Sci 6:367. doi:10.3389/fpls.2015.00367

Qi J, Dong Z, Mao YM, Shen LY, Zhang YX, Liu J, Wang XL (2009) Construction of a dense genetic linkage map and QTL analysis of trunk diameter in Chinese jujube. Sci Silvae Sin 45:44–49

Shen LY (2005) Construction of genetic linkage map and mapping QTLs for some traits in Chinese jujube (*Ziziphus jujuba* Mill.). Agricultural University of Hebei

Soto JC et al. (2015) A genetic map of cassava (*Manihot esculenta* Crantz) with integrated physical mapping of immunity-related genes. BMC Genomics 16:190. doi:10.1186/s12864-015-1397-4

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, Ware D, NM S (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res 20:1689–1699. doi:10.1101/gr.109165.110

Taylor DR, Ingvarsson PK (2003) Common features of segregation distortion in plants and animals. Genetica 117:27–35. doi:10.1023/A:1022308414864

Van Ooijen JW (2011) Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. Genet Res (Camb) 93:343–349. doi:10.1017/S0016672311000279

Wang CZ, Gao JC, Gao WH, Cao DY, Wang HX, Li XG (2007) Advances in research on improvement of Chinese jujube (*Ziziphus jujuba* Mill.). J Fruit Sci 24:673–678

Wang N, Fang L, Xin H, Wang L, Li S (2012a) Construction of a high-density genetic map for grape using next generation restriction-site associated DNA sequencing. BMC Plant Biol 12:148. doi:10.1186/1471-2229-12-148

Wang W et al. (2012b) Construction and analysis of a high-density genetic linkage map in cabbage (*Brassica oleracea* L. var. capitata). BMC Genomics 13:523. doi:10.1186/1471-2164-13-523

Ward JA et al. (2013) Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. BMC Genomics 14:2. doi:10.1186/1471-2164-14-2

Wu J et al. (2014) High-density genetic linkage map construction and identification of fruit-related QTLs in pear using SNP and SSR markers. J Exp Bot 65:5771–5781. doi:10.1093/jxb/eru311

Wu L, Tang Y, Li Y, Yin D, Pang X (2013) Estimation of genome size of *Ziziphus jujuba* and *Z. acdiojujuba*. J Beijing For Univ 35:77–83 doi:10.13332/j.1000-1522.2013.03.012

Wu Y, Close TJ, Lonardi S (2008) On the accurate construction of consensus genetic maps. Comput Syst Bioinformatics Conf 7:285–296

Xu S (2008) Quantitative trait locus mapping can benefit from segregation distortion. Genetics 180:2201–2208. doi:10.1534/genetics.108.090688

Zhang L et al. (2010) Effects of missing marker and segregation distortion on QTL mapping in F2 populations. Theor Appl Genet 121:1071–1082. doi:10.1007/s00122-010-1372-z

Zhang Y et al. (2013) Construction of a high-density genetic map for sesame based on large scale marker development by specific length amplified fragment (SLAF) sequencing. BMC Plant Biol 13:141. doi:10.1186/1471-2229-13-141

Zhang Z, Huan J, Yin X, Li XG, Liu TJ (2015) Fruit characterization of hybrid progenies derived from natural pollination cross between *Ziziphus jujuba* 'Dongzao'and 'Linyilizao'. J Fruit Sci 32:57–62

Zhao J et al. (2014) Rapid SNP discovery and a RAD-based high-density linkage map in jujube (*Ziziphus* Mill.). PLoS One 9:e109850. doi:10.1371/journal.pone.0109850

Zhou W, Tang Z, Hou J, Hu N, Yin T (2015) Genetic map construction and detection of genetic loci underlying segregation distortion in an intraspecific cross of *Populus deltoides*. PLoS One 10:e0126077. doi:10.1371/journal.pone.0126077

Zhou X et al. (2014) Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). BMC Genomics 15:351. doi:10.1186/1471-2164-15-351

Zmienko A, Samelak A, Kozlowski P, Figlerowicz M (2014) Copy number polymorphism in plant genomes. Theor Appl Genet 127:1–18. doi:10.1007/s00122-013-2177-7