DNA Research, 2015, 22(3), 183–191
doi: 10.1093/dnares/dsv003
Advance Access Publication Date: 15 March 2015
Full Paper

Full Paper

High-density genetic map construction and identification of a locus controlling weeping trait in an ornamental woody plant (*Prunus mume* Sieb. et Zucc)

Jie Zhang[1], Qixiang Zhang[1,*], Tangren Cheng[1], Weiru Yang[1], Huitang Pan[1], Junjun Zhong[1], Long Huang[2], and Enze Liu[2]

[1]Beijing Key Laboratory of Ornamental Plants Germplasm Innovation and Molecular Breeding, National Engineering Research Center for Floriculture, Beijing Laboratory of Urban and Rural Ecological Environment, College of Landscape Architecture, Beijing Forestry University, Beijing 100083, China, and [2]Biomarker Technologies Corporation, Beijing, China

*To whom correspondence should be addressed. Tel. +86 010-62336321. Fax. +86 010-62336321. E-mail: zqxbjfu@126.com

Edited by Dr Satoshi Tabata

Received 12 December 2014; Accepted 21 February 2015

Abstract

High-density genetic map is a valuable tool for fine mapping locus controlling a specific trait especially for perennial woody plants. In this study, we firstly constructed a high-density genetic map of mei (*Prunus mume*) using SLAF markers, developed by specific locus amplified fragment sequencing (SLAF-seq). The linkage map contains 8,007 markers, with a mean marker distance of 0.195 cM, making it the densest genetic map for the genus *Prunus*. Though weeping trees are used worldwide as landscape plants, little is known about weeping controlling gene(s) (*Pl*). To test the utility of the high-density genetic map, we did fine-scale mapping of this important ornamental trait. In total, three statistic methods were performed progressively based on the result of inheritance analysis. Quantitative trait loci (QTL) analysis initially revealed that a locus on linkage group 7 was strongly responsible for weeping trait. Mutmap-like strategy and extreme linkage analysis were then applied to fine map this locus within 1.14 cM. Bioinformatics analysis of the locus identified some candidate genes. The successful localization of weeping trait strongly indicates that the high-density map constructed using SLAF markers is a worthy reference for mapping important traits for woody plants.

Key words: mei, high density, linkage mapping, SLAF-seq, weeping trait

1. Introduction

High-density linkage maps are exceptionally valuable tools in many genetic and genomic applications, especially for fine-scale mapping and map-based cloning of trait-controlled genes. A key prerequisite for constructing a high-density linkage map is the availability of plenty genetic markers, which has been feasible due to the advances in next-generation sequencing (NGS) technology for large-scale marker discovery.

Reduced representation genome sequencing (RRGS) is a rapid and cost-effective strategy for massive SNP discovery and genotyping.[1–4] Various methods of RRGS have been developed along with the spread of NGS technology.[5–8] SLAF-seq is a recently developed approach based on RRGS and high-throughput pair-end sequencing. It has been applied in many species for genetic map construction.[9–11] With barcode multiplexed sequencing, large populations with massive loci can be genotyped simultaneously. Besides, a pre-designed scheme can optimize marker efficiency, and stringent genotyping procedure can ensure the accuracy of genotypes.

Mei (Japanese apricot) is a popular ornamental plant widely cultivated in the entire East Asia and, as an important member of genus

Prunus, plays a pivotal role in systematic studies of the Rosaceae.[12] It owned many prominent ornamental features, such as colourful corollas, pleasant fragrance and weeping habit.[13] Weeping trees characterized by soft, pendant branches are used worldwide as landscape plants due to their graceful and intriguing forms. Meanwhile, being a specific tree architecture, weeping trait has received increasing attention during the past years.[14–16] In the peach, a set of random amplified polymorphic DNA (RAPD) markers were developed in an F2 population, and the *Pl* was located preliminarily on the second linkage group at comparatively large distances (11.4 and 17.2 cM) between two RAPD markers.[17] Subsequently, the simple sequence repeat (SSR) marker CPPCT029 was found to co-segregate with *Pl* in peach.[18] However, RAPD markers are difficult to transfer to other mapping population. Besides, the located region of *Pl* was just an estimate and no candidate genes had been detected.

In the current study, a high-density linkage map of mei was constructed by applying SLAF-seq method and HighMap linkage mapping strategy for the first time. To test the utility of this map, fine-scale mapping of weeping trait was conducted. Gene annotation results suggest the accuracy of the localization. It is for the first time that weeping trait was fine-mapped, which gives a typical example for further QTL fine mapping of important traits in perennial woody plants.

2. Materials and methods

2.1. Mapping population

The pedigree used for QTL mapping was generated by the female 'LiuBan' (upright type) from QingDao, China (36.20169°N, 120.4162°E) and the male 'FenTaiChuiZhi' (weeping type) from WuHan, China (30.54526°N, 114.39511°E), which were selected based on their phenotypic divergence in tree architecture, flower shape and colourful corollas (Supplementary Fig. S1). In spring 2013, a clonally replicated plantation of this entire pedigree was established in a randomized complete block design with three replicates and two tree plots at Hangzhou, China (30.566389°N, 119.879582°E).

2.2. SLAF library construction and high-throughput sequencing

An improved SLAF-seq strategy was utilized in our experiment. Firstly, reference genome of mei was used to design marker discovery experiments by simulating *in silico* the number of markers produced by different enzymes. Next, a SLAF pilot experiment was performed, and the SLAF library was conducted in accordance using the predesigned scheme. For the mei F1 population, two enzymes (HaeIII and Hpy166II, New England Biolabs, NEB, USA) were used to digest the genomic DNA. A single nucleotide (A) overhang was added subsequently to the digested fragments using Klenow Fragment ($3' \rightarrow 5'$ exo$^-$) (NEB) and dATP at 37°C. Duplex tag-labelled sequencing adapters (PAGE-purified, Life Technologies, USA) were then ligated to the A-tailed fragments using T4 DNA ligase. Polymerase chain reaction (PCR) was performed using diluted restriction-ligation DNA samples, dNTP, Q5® High-Fidelity DNA Polymerase and PCR primers (Forward primer: 5′-AATGATACGGCGACCACCGA-3′, reverse primer: 5′-CAAGCAGAAGACGGCATACG-3′) (PAGE-purified, Life Technologies). PCR products were then purified using Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK) and pooled. Pooled samples were separated by 2% agarose gel electrophoresis. Fragments ranging from 214 to 294 base pairs (with indexes and adaptors) in size were excised and purified using a QIAquick gel extraction kit (Qiagen, Hilden, Germany). Gel-purified products were then diluted. And pair-end sequencing (each end 100 bp) was performed on an Illumina HiSeq 2500 system (Illumina, Inc., San Diego, CA, USA) according to the manufacturer's recommendations.

2.3. Sequence data grouping and genotyping

SLAF marker identification and genotyping were performed using procedures described by Sun *et al.*[10] Briefly, low-quality reads (quality score <20e) were filtered out and then raw reads were sorted to each progeny according to duplex barcode sequences. After the barcodes and the terminal 5-bp positions were trimmed from each high-quality reads, clean reads from the same sample were mapped onto the mei genome sequence using SOAP software.[19] Sequences mapping to the same position with over 95% identity were defined as one SLAF locus.[11] Allele tags of each SLAF locus were then detected according to parental reads with sequence depth >20-fold and >30% integrity in the offspring. Single nucleotide polymorphisms (SNPs) were detected according to the allele tags and SLAFs with >3 SNPs were filtered out firstly. For diploid species, one SLAF locus can contain no more than four allele tags, so SLAF loci with more than four alleles were defined as repetitive SLAFs and discarded subsequently. Only SLAFs with two to four alleles were identified as polymorphic and considered potential markers. All polymorphism SLAFs loci were genotyped with consistency in the parental and offspring SNP loci. The marker code of the polymorphic SLAFs was analysed according to the population type CP (cross pollinators), which consist of five segregation types (ab × cd, ef × eg, hk × hk, lm × ll, nn × np). The segregation type 'abxcd' represents that the two alleles of one marker are different in both parents.

Genotype scoring was then performed using a Bayesian approach to further ensure the genotyping quality.[10] Briefly, a posteriori conditional probability was calculated using the coverage of each allele and the number of SNPs. Then, genotyping quality score translated from the probability was used to select qualified markers for subsequent analysis, which was a dynamic optimization process.[10] When the average genotype quality scores of all SLAF markers reached the cut-off value, we ceased this process. After that, high-quality SLAF markers for the genetic mapping were filtered by the following criteria. First, average sequence depths should be >7-fold in each progeny and >80-fold in the parents. Second, markers with >30% missing data were filtered. Third, the χ^2 test was performed to examine the segregation distortion. Markers with significant segregation distortion were initially excluded from the map construction and were then added later as accessory markers.

2.4. Linkage map construction

Marker loci were partitioned primarily into linkage groups (LGs) based on their locations on mei genome. Next, the modified logarithm of odds (MLOD) scores between markers were calculated to further confirm the robustness of markers for each LGs. Markers with MLOD scores <5 were filtered prior to ordering. To ensure efficient construction of the high-density and high-quality map, a newly developed HighMap strategy was utilized to order the SLAF markers and correct genotyping errors within LGs.[20] Firstly, recombinant frequencies and LOD scores were calculated by two-point analysis, which were applied to infer linkage phases. Then, enhanced gibbs sampling, spatial sampling and simulated annealing algorithms (GSS) were combined to conduct an iterative process of marker ordering.[21,22] Briefly, in the first stage of the ordering procedure, SLAF markers were selected using spatial sampling. One marker was taken randomly in a priority order of pseudo-testcross, and markers with a recombination

frequency smaller than a given sampling radius are excluded from the marker set. Subsequently, simulated annealing was applied to searching for the best map order. Summation of adjacent recombination fractions was calculated as illustrated by Liu *et al.*[20] The annealing system continued until, in a number of successive steps, the newly generated map order is rejected. Blocked Gibbs sampling was employed to estimate multipoint recombination frequencies of the parents after the optimal map order of sample markers was obtained.[20] The updated recombination frequencies were used to integrate the two parental maps, which optimize the map order in the next cycle of simulated annealing. Once a stable map order was obtained after 3–4 cycles, we turned to the next map building round. A subset of currently unmapped markers was selected and added to the previous sample with decreased sample radius. The mapping algorithm repeats until all the markers were mapped appropriately. The error correction strategy of SMOOTH was then conducted according to parental contribution of genotypes,[23] and a k-nearest neighbor algorithm was applied to impute missing genotypes.[24] Skewed markers were then added into this map by applying a multipoint method of maximum likelihood.[25]

The sex-specific maps were constructed using markers that were heterozygous in female or male parent, while the consensus map was established by integrating the parental maps through the anchor markers (markers that were heterozygous in both parents).[22] For anchor markers, map distances were calculated as the averages over the two parental distances. The remaining markers segregating in only one of the parents were placed on the consensus map by interpolation or extrapolation according to the relative position between the flanking anchor markers on the relevant parental map. Map distances were then estimated using the Kosambi mapping function.[26]

2.5. Marker-trait association analysis

Firstly, QTL analysis was performed using the R/QTL package.[27] A binary option was used to detect the weeping QTL. The significance thresholds were determined using 1,000 permutations. Results from the interval mapping analysis were used to construct the QTLs, and their positions were used in a default model.

For the Mutmap-like strategy, high-quality SLAF markers from each individual were first divided into two pools, weeping and upright, according to the phenotypic segregation of the F1 population. To examine the relative genotype contribution from each parent, the difference in SLAF allele frequencies between the weeping and upright pools was evaluated using SNP-index-like methods[28] and, hereafter, will be referred to as the SLAF-index. The difference in proportions [Δ(SLAF-index)] of one genotype in the weeping pool was compared with the upright pool. The proportion of one genotype derived from the male parent (M_g) corresponding to the two parental genotypes (N_g) was defined as the SLAF-index. If the proportion of M_g in N_g was equal in the two pools, then the Δ(SLAF-index) was equal to 0. A Δ(SLAF-index) equal to 1 suggests that one genotype almost entirely belongs to the weeping pool, indicating that the associate SLAF markers are linked closely to the weeping trait. Accordingly, the Δ(SLAF-index) was calculated for all high-quality SLAF markers between parents, which was then used as an important indicator of the relationship between marker and weeping trait.

Then, the weeping trait was considered a marker (marker0), with parents' genotype being nnxnp. Accordingly, we got the genotype of each individual based on their phenotype ('weeping' is 'nn', while 'upright' is 'np'). Recombination rates between marker 0 and markers of other genotypes were then calculated.

2.6. Candidate gene selection and annotation

We initially detected the genes in the causal region according to the functional annotation of the mei genome (http://prunusmumegenome.bjfu.edu.cn). The gene annotation list was then screened for polymorphism at the amino acid level between the resequenced 'LiuBan' genome and 'FenTaiChuiZhi' genome (unpublished data). Genes showing at least one change at the amino acid level were considered candidates. For a more detailed functional annotation, the candidates were compared with the NR protein sequences available at UniProt database using the BLASTX algorithm, with an *E*-value threshold of 10^{-01}. The associated hits were then searched for their respective Gene Ontology (GO) terms at www.geneontology.org.[29]

3. Results

3.1. SLAF sequencing and genotyping

Based on the results of SLAF pilot experiment, HaeIII and Hpy166II were used for SLAF libraries construction. In total, 472.66 M pair-end reads were obtained for both parents and 387 progeny, with each descendant sharing an average of 1.11 M reads. The raw data of SLAF-seq have all been deposited in NCBI SRA database under project accession PRJNA273338. Among them, the sequence depth of parents had more impact on alleles calling, since they were the basis for genotyping for each marker. The parents obtained 43.09 M reads, a rate of 9.31%, which was higher than average. A total of 230,584 SLAF loci were detected after reads clustering. The average sequence depths of these SLAFs were 35.15-fold for parents and 4.00-fold for each individual. The extreme depths of the locus-specific sequences suggested the accuracy of the SLAF marker discovery. Repetitive SLAFs were then disregarded, and polymorphic SLAFs accounted for 40.35% of all SLAFs (Table 1). All polymorphic SLAFs were then genotyped separately for both parents and all individuals.

Sequencing depth alone cannot ensure the genotyping quality; thus, genotyping quality scores (details see methods) were used to select qualified markers. The worst low-quality marker was discarded during each cycle. This dynamic process was repeated until the average genotype quality score of all SLAF markers reached the cut-off value of 30. Finally, 9,412 SLAFs were defined as high-quality SLAF

Table 1. SLAF-seq data summary for mei F1 population

Total reads	
No. of reads	472,655,544.00
Reads in high-quality SLAFs	183,454,384.00
Reads in repeat SLAFs	13,619,794.00
Reads in low depth SLAFs	8,778,576.00
High-quality SLAFs	
No. of SLAFs	230,584.00
Average SLAF depth	795.61
Average depth in parents	35.13
Average depth in individuals	4.00
Polymorphic SLAFs	
No. of polymorphic SLAFs	93,031.00
Average depth in parents	127.31
Average depth in individuals	7.85
No. of SNPs	14,388.00
SNP ratio per kb	6.02
High-quality SLAF markers	
No. of high-quality SLAF markers	9,412

No: number; SLAF: specific length amplified fragment.

markers with over 80-fold parental sequence depth, over 7-fold progeny sequence depth and >70% integrity among the F1 population.

3.2. High-density linkage map construction

The 9,412 high-quality markers were distributed into eight LGs according to their locations on mei genome and the MLOD scores with other markers (at least one MLOD score >5). A total of 8,007 markers containing skewed markers (segregation and genotype sequence data are listed in Supplementary Tables S1 and S2) were used to construct the final linkage map (Fig. 1). The information of all markers on the map is organized in Supplementary Table S3, which includes marker names, LGs, genetic position and physical location in mei genome. Following linkage analysis, coverage of the markers was 155.77-fold in the male parent, 98.21-fold in the female parent and 8.18-fold in each F1 individual (on average). The integrity of each marker among the 387 F1 individuals was also a key parameter to control map quality. All markers on the map demonstrated 96% integrity (on average). The final map was 1,550.62 cM in length with an average distance of 0.195 cM between adjacent markers. As shown in Table 2, the largest LG was LG2, with 1,722 SLAF markers and a 0.15-cM marker interval, while LG6 was the smallest with 698 SLAF markers and a 0.20-cM marker interval (Supplementary Fig. S2). The uneven length of LGs is identical with the karyotypic parameters of mei chromosomes,[30] and the distribution of SLAF markers on LGs is shown in Fig. 2. The largest gap was 24.13 cM in LG1, followed by 12.79 cM in LG6. 3.50% skewed markers with $P < 0.05$ and 4.98% with $P < 0.01$ have been identified on the final map. Here, we defined an area containing four skewed markers as segregation distortion regions (SDRs), and one SDR was detected in LG2, while three other SDRs were located in LG6. These SDRs may be related with preferential selection and will not affect the accuracy of the genetic map.[31]

3.3. Comparative analysis of the high-density linkage map

To evaluate the quality of the genetic map, we firstly mapped the SLAF markers to the mei genome. An excellent collinearity between physical distances and genetic distances of all SLAF markers in eight LGs is presented in Fig. 3. All consecutive curves generated from the eight LGs indicate two facts: the mei genome was sufficiently covered with SLAF markers, and SLAF markers were placed accurately within each LG.

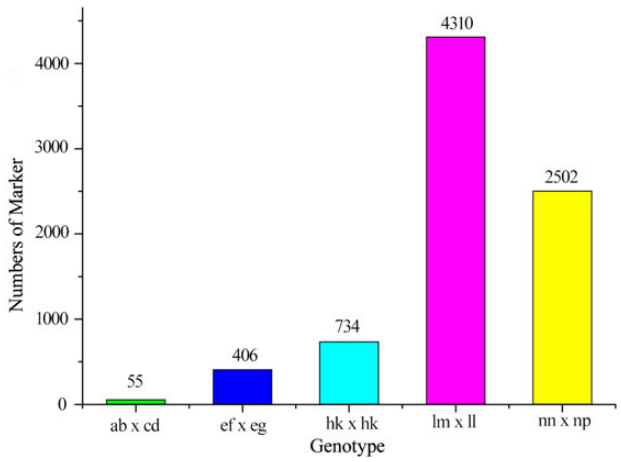

Figure 1. Numbers of each marker segregation type on the linkage map of mei. This figure is available in black and white in print and in colour at *DNA Research* online.

Most parts of these curves represent a falling trend, suggesting that their genetic and physical positions follow an identical order. Meanwhile, two possible recombination hotspots can be observed intuitively on pseudo-chromosome 1 and pseudo-chromosome 4, and the relative positions in each LG can be seen in Supplementary Fig. S2.

Macro-collinearity was revealed by aligning the SLAF markers on the mei high-density map to the peach genome sequence. Using 90% identity as the standard, 6,837 SLAF markers were mapped onto the peach genome (Supplementary Fig. S3). LG1 and LG6 appeared to have adverse orientations with the peach chromosomes (Ch6 and Ch8), which were nominal since, there was no stringent direction of the markers during linkage mapping. LG5 was syntenic to two peach chromosomes, Ch2 and Ch4, suggesting that fission or translocation events have occurred between the two *Prunus* species.

3.4. Inheritance mode of weeping trait of mei

As a prerequisite to mapping the loci responsible for weeping, the inheritance of the weeping trait was evaluated. In this study, two F1 segregation populations were obtained, of which the parents differed remarkably in tree architecture (upright cultivar × weeping cultivar). Of the 320 genotypes in the 'LiuBanDan' × 'ShuangBi' population, 143 were classified as weeping and 177 as upright. A χ^2 test indicated that the growth habit segregation fit the expected ratio of 1 : 1. In the 'LiuBan' × 'FenTaiChuiZhi' mapping population (387 individuals), the growth habits of 165 and 222 trees were ranked as weeping (W) and upright (U), respectively (Table 3). The growth habit segregation in this population was slightly distorted from the expected 1(W) : 1(U) ratio due to an excess of uprights. Meanwhile, diverse branching habit that determined different tree architecture of weeping progeny was observed. Combined with former molecular basis of weeping trait, we suggested that weeping trait of mei may be controlled by a major gene together with one or more modifiers to account for a deficiency of weeping seedlings in the 'LiuBan' × 'FenTaiChuiZhi' population.

3.5. Fine mapping of loci conferring weeping trait

First, QTL analysis was performed using R/qtl based on the high-density genetic map and phenotypic separation among progeny.[27] Threshold of significance ($P = 0.05$) for each marker after 1,000 permutation was set to 4.35, which resulted in the identification of a single locus at 71.88 cM on pseudo-chromosome 7 (Fig. 4). On the linkage map, three markers (marker353041, marker437413 and marker321918) located at 71.88 cM. The LOD scores on pseudo-chromosome 7 were much higher than those on other pseudo-chromosomes, with the LOD score of the strongest QTL peak (peak LOD score = 87.65) being the highest. This phenomenon suggested that the locus on pseudo-chromosome 7 was strongly associated with weeping trait. However, the 95% confidence intervals for the QTL position were 4.80–87.65 cM, almost covered the whole pseudo-chromosome 7. To conduct finer mapping of the loci controlling weeping trait, a Mutmap-like strategy was initially applied to identify polymorphisms of the SLAF markers. Mutmap was developed on the basis of bulked segregation analysis (BSA) by calculating a parameter (SNP-index) to judge the contribution of an allele to the mutant phenotype.[28] Here, we considered weeping as the mutant trait and calculated Δ(SLAF-index) for each SLAF marker on the genome. The relationship between Δ(SLAF-index) and SLAF position on the genome was graphically represented (Fig. 5). As expected, Δ (SLAF-index) values were distributed randomly around 0 for most of the genome, with the exception of pseudo-chromosome 7. An obvious peak was observed when the Δ(SLAF-index) threshold was set

Table 2. Basic characteristics of mei linkage groups

LGID	Total marker	Total distance (cM)	Average distance (cM)	Max gap	Segregation SLAF ($P < 0.05$)	Segregation SLAF ($P < 0.01$)	SDR number
LG1	1,139	238.14	0.21	24.13	10	49	0
LG2	1,722	263.84	0.15	2.95	156	181	1
LG3	1,131	216.97	0.19	2.86	3	9	0
LG4	917	205.79	0.22	5.53	1	1	0
LG5	972	231.69	0.24	3.4	32	39	0
LG6	698	142.48	0.2	12.79	78	105	3
LG7	704	129.8	0.18	6.93	0	15	0
LG8	724	121.9	0.17	3.13	0	0	0
Max group	1,722	263.84	0.24	24.13	156	181	3
Min group	698	121.9	0.15	2.86	0	0	0
Total	8,007	1,550.62	1.56	61.72	280	399	4

LG: linkage group; SDR: segregation distortion region; cM: centiMorgan.

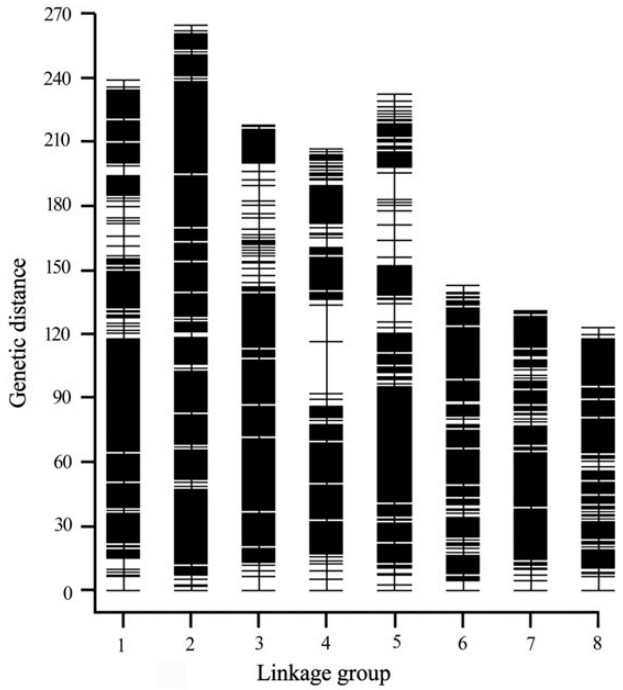

Figure 2. Distribution of SLAF markers on eight linkage groups of mei. A black bar indicates a SLAF marker. The *x*-axis represents linkage group number and the *y*-axis indicates genetic distance (centiMorgan as unit).

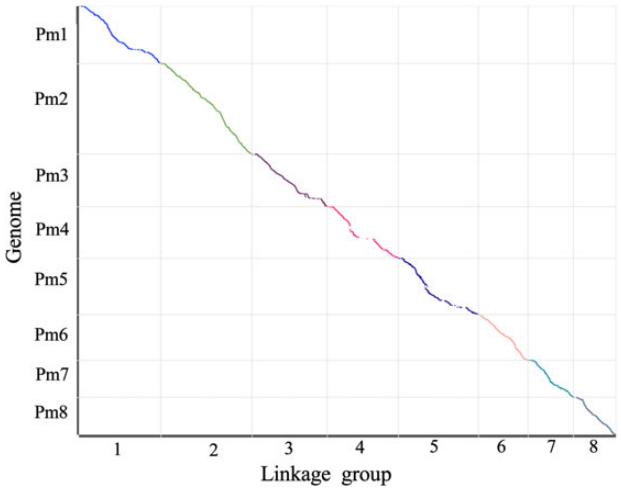

Figure 3. Collinearity analysis of all mei linkage groups with mei genome sequence. The *x*-axis scales genetic distance of mei LGs accordingly, while the *y*-axis represents the position of mei peseudo-chromosomes. SLAF markers in these LGs are plotted as dots on the Figure. This figure is available in black and white in print and in colour at *DNA Research* online.

to 0.85. Five markers (marker334902, marker430976, marker446598, marker431969 and marker348136) above the threshold value with an average Δ(SLAF-index) equal to 0.9 were considered to be tightly linked markers to the weeping trait, and ranged from 69.63 to 75.52 cM on LG7. The above results strongly suggested that a major locus controlling the weeping trait was located on pseudo-chromosome 7 of mei. Thus, we assumed that weeping trait was a completely monogenetic trait to further confirm this localization. On this basis, we considered weeping trait as a marker (marker0) and calculated the recombination rate of marker0 with other markers. Eventually, the locus conferring weeping trait was determined to be located between two markers, whose recombination rate with marker0 was the lowest in LG7. The two flanking markers were positioned at 76.83 cM (marker315769) and 82.99 cM (marker418421), respectively.

To avoid loss of candidate genes, we integrated the results of the analyses. In total, 10 markers were considered to be tightly linked to weeping trait. Based on the positions of the SLAF markers on the fully sequence genome of mei, the candidate genomic region for weeping trait was located at 10.54–11.68 Mb on pseudo-chromosome 7 of mei (Fig. 6A).

3.6. Annotated genes within the candidate region

The interval of the candidate region is 1.14 Mb in length and contains 159 predicted protein encoding genes. Genes within the candidate region were then detected based on annotation of the mei genome (http://prunusmumegenome.bjfu.edu.cn./). All 159 predicted genes were first screened for predicted amino acid polymorphisms between the 'LiuBan' and 'FenTaiChuiZhi' alleles according to their whole-genome resequencing data. This selection resulted in 69 genes. Among them, 19 could not be annotated using the known annotation databases. A complete list of the remaining 50 genes is provided in Supplementary Table S4.

Some of the genes are conspicuous. Nine were obviously associated with tension wood (TW) formation and development, which are

Table 3. Segregation ratios of progenies with weeping and upright shape in mei F1 populations

Population	Total	Weeping	Upright	Expected ration	χ^2	Significance	$\chi^2_{0.05}$
LB × FT	387	165	222	1 : 1	4.20	Significance	3.87
LBD × SB	320	143	177	1 : 1	3.61	No significance	3.87

Cross combinations: LB × FT: 'LiuBan' × 'FenTaiChuiZhi'; LBD × SB: 'LiuBanDan' × 'ShuangBi'.

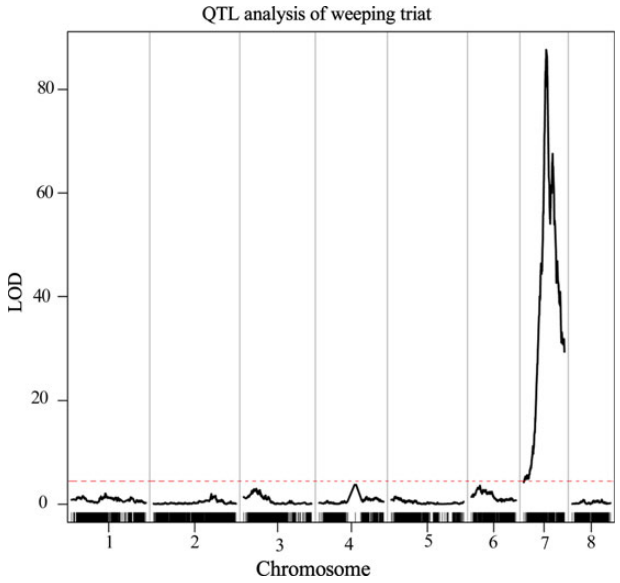

Figure 4. R/QTL analysis of weeping trait. The *x*-axis scales genetic distance of mei LGs accordingly, while the *y*-axis represents the LOD scores. The dashed line represents the significant threshold value of 4.35 ($P = 0.05$). This figure is available in black and white in print and in colour at *DNA Research* online.

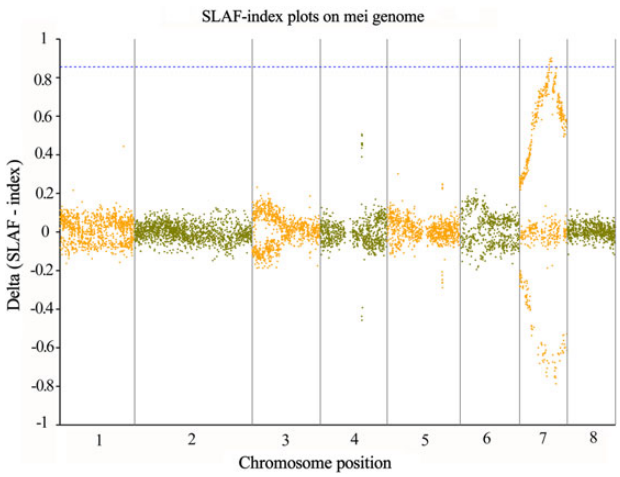

Figure 5. Mutmap-like analysis of weeping trait in mei. The *x*-axis represents the position of pseudo-chromosomes, whereas the *y*-axis shows the value of △(SLAF-index). SLAF markers unique to weeping (i.e. not present in 'LiuBan' and upright individuals) are shown in grey, whereas those unique to upright (i.e. not present in 'FenTaiChuiZhi' and weeping individuals) are in black. The dashed line represents the threshold value of 0.85. This figure is available in black and white in print and in colour at *DNA Research* online.

related with weeping trait. Two members of the cellulose synthase-like (CSL) family, Pm024150 and Pm024152, are important for cell wall development and growth.[32] Pm024254 that encodes endo-beta-1,4-

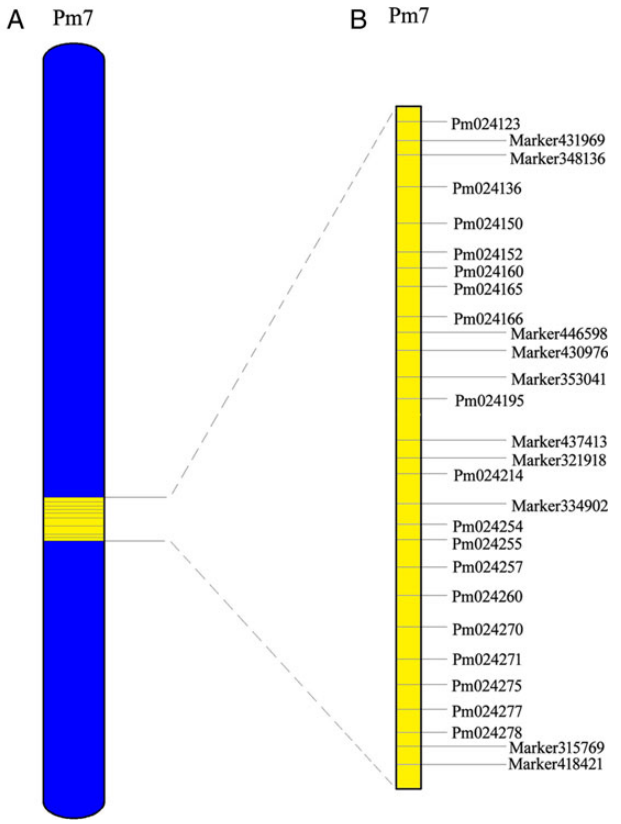

Figure 6. Distribution of tightly linked markers and plausible candidate genes of weeping on Pseudo-chromosome 7 of mei. (A) Candidate genomic region on pseudo-chromosome 7. (B) Schematic diagram of the order of the 10 tightly linked markers and the 18 plausible candidate genes of weeping based on gene annotation. This figure is available in black and white in print and in colour at *DNA Research* online.

glucanase is believed to function in cell wall changes associated with diverse growth processes and cell elongation.[33,34] Two predicted genes (Pm024195 and Pm024255) have been suggested to be involved in cell wall formation and assembly.[35,36] Auxin-induced protein 5NG4-like (Pm024277), a positive regulator of auxin metabolism, may be involved in secondary cell wall biogenesis.[37] Also, genes involved in lignin biosynthesis that affect fibre development may result in TW abnormalities.[38] Pm024278 is a probable cinnamyl alcohol dehydrogenase 1, which catalyzes the final specific step for lignin monomer production,[39] while Pm024136 is predicted to be involved in xylan metabolism. NAC domain-containing protein 43 (Pm024277) may regulate the secondary cell wall lignification of tissues.[40]

Weeping trait may be controlled by a regulatory gene causing differential expression of the downstream network.[41] Our annotation analysis revealed nine genes that may regulate gene expression at the transcriptional level or post-transcriptional level. 26S proteasome (Pm024160) deserves close attention, since it may play an important

role in balancing cell expansion with cell proliferation during shoot development.[42] Twenty-one Protein NLP6 (Pm024165) is an uncharacterized transcription factor, while 22 Protein NLP7 (Pm024166) is a transcription factor involved in the regulation of nitrate assimilation and nitrate signal transduction.[43] BHLH155 (Pm024214) is a transcription factor that may regulate root development, while growth-regulating factor 8 (Pm024257) may act as a transcriptional activator that plays a role in the regulation of cell expansion in meristem tissues.[44] Four genes (Pm024270, Pm024271, Pm024275 and Pm024123) with significant function variation in the 'LiuBan' and 'FenTaiChuiZhi' genome were predicted to be DNA-directed RNA polymerase II (Pol II) subunit RPB1, a structure to promote translocation of Pol II, which serves as a platform for the assembly of factors that regulate transcription initiation, elongation, termination and mRNA processing.[45]

In addition, some genes are related to enzymatic activities. Among these genes, Pm024200 corresponds to the tightly linked marker437413 and is suggested to participate in coenzyme transport and metabolism. It has been suggested that ethylene promotes cell elongation in cotton.[38] Pm024219, which encodes serine/threonine-protein kinase WNK8, and Pm024247, which encodes subtilisin-like protease (precursor), may both contribute to cell elongation, since a conformational change in the serine/threonine protein kinase activates the downstream signalling of ethylene.[38] N6-adenosine-methyltransferase MT-A70-like (Pm024237), which methylates adenosine residues of some mRNAs, may also play a role in the efficiency of mRNA splicing, transport or translation.

4. Discussion

4.1. Characteristics of SLAF-seq strategy for large-scale marker development

The SLAF-seq strategy, a combination of locus-specific amplification and high-throughput sequencing, is an effective method for large-scale SNP discovery and genotyping, which has been applied successfully in various species.[9–11]

Compared with conventional methods of marker development, the SLAF-seq method owned some unique superiority. First, since copy number uniformity among fragments can reduce bias sequencing result, a pre-designed scheme was conducted to ensure the density, uniformity and efficiency of marker development. The well-assembled reference genome of mei was analysed beforehand, considering the information on genomic GC content, repeat conditions and genetic characteristics. After that, a pilot experiment was performed to check the reduced representation library features in targeted length range (214–294 bp), which was supposed to include fragments with similar amplification features on the gel. When non-specific amplified bands appeared, the pre-design step would be repeated. SLAF markers were designed such that an even distribution throughout the genome was established and repeats were avoided. Second, sequence depth together with the genotyping quality scores enhanced the genotyping accuracy. Since lower coverage per locus always lead to a lower confidence in each genotype call and each newly discovered marker,[46] sequencing depth is an important consideration in sequencing-based genotyping. It is suggested that 6-fold is a minimal sequencing depth for each individual when SLAF-seq strategy is used.[10] A sequencing depth of over 80-fold for parents and over 7-fold for progenies of our study can provide sufficiently high genotyping accuracy. Sequencing depth alone cannot ensure the genotyping quality; thus, genotyping quality scores were used to detect suspicious markers. In this study, suspicious marker was discarded during each cycle of the dynamic process, until the average genotype quality score of all SLAF markers reached the cut-off value of 30. Third, successive stringent principles were used for high-quality SLAF marker selection. All these features mentioned above make the SLAF-seq strategy an effective new tool for large-scale genotyping and rapidly markers development.

This study provides the first marker development on large scale for mei; in total, 9,412 high-quality SLAF markers were developed, and 8,007 polymorphic markers were finally identified for genetic linkage map construction. The integrity and accuracy of the SLAF markers were high, while the quality and quantity of them also met the requirements for construction of a high-quality and high-density genetic map. It is for the first time that SLAF-seq strategy was successfully applied in developing markers in ornamental woody plants.

4.2. The densest genetic map of mei

The genetic map in this study was the first mei genetic map constructed using over 8,000 SLAF markers developed from SLAF-seq strategy. The earliest genetic map was of the Chinese mei cultivars 'Fenban' and 'Kouzi', using 144 SSR markers.[47] Later, by applying the RAD-seq strategy, 779 SNP markers were developed for the same mapping population, which were then used in anchoring and orienting scaffolds for mei genome assembly.[48] However, the resolutions of these maps and the sizes of the mapping populations were not enough for fine-scare QTL localization. In this study, the linkage map contained 8,007 markers, a marker density that has, to our knowledge, never been reached for mei and even any other species of *Prunus*.[12,49] This linkage map covered nearly the whole genome with a resolution of 0.195 cM.

To increase the quantity of markers and genomic coverage on the final genetic map, skewed markers were not discarded. In total, 8.5% skewed markers were inserted in the genetic map at the last step. Marker distortion may be caused by preferential selection or gametic/zygotic selection,[31,50,51] which always resulted in higher genetic variance.[52] It is reported that skewed markers can be applied for QTL mapping without detrimental effect and can be beneficial if used properly, as areas containing skewed markers may be recombination hotspots for specific traits.[25,53]

Besides, a HighMap strategy (the development of HighMap software is under way) was utilized to construct the high-density genetic map in this study. By exploiting an iterative ordering and error correction strategy as well as optimized GSS algorithm, HighMap was efficient for constructing linkage maps using markers developed from NGS. First, the marker order and the map distance were relatively accurate for data with genotyping errors, since an error correction strategy was performed. Second, HighMap greatly curbed the genetic map distance, while the iterative ordering strategy ensured an accurate marker order.[20] The excellent collinearity between physical distance and genetic distance of all SLAF markers in eight LGs strongly indicates the high quality of our map. Meanwhile, high macro-collinearity is also observed among the SLAF markers of our map and the peach genome, which is identical with former researches.[47,48]

The mapping population should segregate for flower shape, corolla colour and cold tolerance, since the parents of the F1 population show obvious differences in these traits (Supplementary Fig. S1). As the mei whole-genome sequence has been released,[48] it can be used in combination with our map for fine-scale mapping of genes related to important ornamental traits, since the sequences of each marker is available and genes around the marker are easy to obtain. As the species of *Prunus* are closely related, this high-quality and high-density genetic map will also accelerate the molecular breeding for perennial woody plants in *Prunus*.

4.3. Location of locus conferring weeping trait utilizing the high-density map

Previous studies have suggested that the weeping trait was controlled by a single recessive gene both in peach and mulberry.[14,16,54–57] However, Tu *et al.* suggested that weeping trait in *Castanea mollissima* was controlled by a major recessive gene together with some modifier genes. Although self-cross of a weeping parent yielded 1 (W) : 3(U) ratio of the progeny, different branching drooping degrees were observed among the weeping offspring.[58] In terms of our mapping population, the segregation ratio was slightly distorted from the expected ratio, while no skewed markers had been detected around the flanking regions of the weeping locus. Besides, the branching drooping degree of the progeny was also diverse. The inheritance pattern of mei was similar to *C. mollissima*. Thus, we supposed that weeping trait of mei was controlled by a major recessive gene together with some modifier genes account for an excess of upright seedlings.

Because of the complexity of tree architecture, it is difficult to define a methodology for describing branching habit in an accurate quantitative way on tree descendants. However, the distinction between the weeping and upright forms of the mei progeny was sharply clear according to their branching directions.[13] Thus, the progeny were scored quantitatively according to their phenotypes, and QTL analysis was then performed to map loci conferring weeping trait. According to the confidence intervals analysis, the candidate region almost covered the whole pseudo-chromosome 7. Then a Mutmap-like strategy was performed to fine map the weeping loci. The results of the two methods both suggested that the major locus controlling weeping trait was located on mei pseudo-chromosome 7. To confirm this conclusion, an extreme linkage analysis was conducted considering weeping trait as a complete recessive monogenetic trait. And the supposed weeping gene was also located on mei pseudo-chromosome 7. The results of these three methods can be mutual authentication and supplement, which could increase the reliability of the candidate locus of weeping trait.

Since the use of more highly annotated genomic databases could narrow down the selection of plausible candidate genes, it was worthwhile to determine the function of the candidate genes in the causal region. Combining the whole-genome resequencing data of the parents with the highly annotated genomic databases, we successfully narrowed down the number of plausible candidates to 69. However, due to the poor genetic background of weeping trait, it was difficult to determine the weeping controlling gene directly. Nine candidate structural genes and enzymes were predicted to be involved in cell wall and cellulose synthesis/degradation. These genes were supposed to be linked with weeping trait, since TW contains higher cellulose content and lower lignin, which may cause the drooping phenomenon.[38,59] Besides, nine genes were predicted to be related to transcriptional regulation, which were preferred for further experimental verification. In total, 18 candidates were considered as the most plausible genes of weeping in the casual region (Fig. 6B). The annotation results also confirm that the localization of weeping locus is precise which suggested the high quality of our genetic map.

In summary, the present study demonstrates that SLAF-seq strategy is a powerful method for marker discovery and high-density linkage map construction, and our study resulted in the densest linkage map for mei. Comparative analysis and fine-scale mapping of the weeping trait suggest the high quality of this genetic map. A locus conferring weeping trait was detected on a 1.14-Mb interval on pseudo-chromosome 7 of mei for the first time, which is important for a better understanding of genetic mechanism of weeping trait. The precise localization of weeping trait strongly indicates that the high-density genetic map should contribute to the breeding of important ornamental traits in mei, along with the genus *prunus*.

Acknowledgements

We really appreciate Dr Satoshi Tabata and the reviewers for the precious comments and suggestions in improving our manuscript.

Conflict of Interest statement

The authors declared that they have no competing interests.

Supplementary Data

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

This work was supported by the Fundamental Research Funds for the Ministry of Science and Technology (Grant No. 2011AA100207), the Fundamental Research Funds for Central Universities (No. BLYJ201407) and Special Fund for Beijing Common Construction Project. Funding to pay the Open Access publication charges for this article was provided by Fundamental Research Funds for the Ministry of Science and Technology (Grant No. 2011AA100207).

References

1. Altshuler, D., Pollara, V.J. and Cowles, C.R., et al. 2000, An SNP map of the human genome generated by reduced representation shotgun sequencing, *Nature*, **407**, 513–6.
2. Hyten, D.L., Cannon, S.B. and Song, Q., et al. 2010, High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence, *BMC Genomics*, **11**, 38.
3. Lucito, R., Nakimura, M. and West, J.A., et al. 1998, Genetic analysis using genomic representations, *Proc. Natl Acad. Sci.*, **95**, 4487–92.
4. Van Tassell, C.P., Smith, T.P. and Matukumalli, L.K., et al. 2008, SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries, *Nat. Methods*, **5**, 247–52.
5. Baxter, S.W., Davey, J.W. and Johnston, J.S., et al. 2011, Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism, *PloS ONE*, **6**, e19315.
6. Chutimanitsakun, Y., Nipper, R.W. and Cuesta-Marcos, A., et al. 2011, Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley, *BMC Genomics*, **12**, 4.
7. Pfender, W., Saha, M., Johnson, E. and Slabaugh, M. 2011, Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in Lolium perenne, *Theor. Appl. Genet.*, **122**, 1467–80.
8. Poland, J.A., Brown, P.J., Sorrells, M.E. and Jannink, J.-L. 2012, Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach, *PloS ONE*, **7**, e32253.
9. Chen, S., Huang, Z. and Dai, Y., et al. 2013, The development of 7E chromosome-specific molecular markers for Thinopyrum elongatum based on SLAF-seq technology, *PloS ONE*, **8**, e65122.
10. Sun, X., Liu, D. and Zhang, X., et al. 2013, SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing, *PloS ONE*, **8**, e58700.
11. Zhang, Y., Wang, L. and Xin, H., 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.

12. Sun, L., Wang, Y. and Yan, X., et al. 2014, Genetic control of juvenile growth and botanical architecture in an ornamental woody plant,. Prunus mume Sieb. et Zucc. as revealed by a high-density linkage map, *BMC Genet.*, **15**, S1.
13. Chen, J. 1996, *Chinese mei flowers*, Haikou, China: Hainan Publishing HouseHaikou, China, pp.51–2.
14. Yamanouchi, H., Koyama, A., Machii, H., Takyu, T. and Muramatsu, N. 2009, Inheritance of a weeping character and the low frequency of rooting from cuttings of the mulberry variety 'Shidareguwa', *Plant Breed.*, **128**, 321–3.
15. Yamaguchi, S. 2008, Gibberellin metabolism and its regulation, *Annu. Rev. Plant Biol.*, **59**, 225–51.
16. Werner, D.J. and Chaparro, J.X. 2005, Genetic interactions of pillar and weeping peach genotypes, *Hort Sci.*, **40**, 18–20.
17. Dirlewanger, E. and Bodo, C. 1994, Molecular genetic mapping of peach, *Euphytica*, **77**, 101–3.
18. Ymeng, L. 2006, *Physiological characteristics of weeping phenomenon and its simple sequence repeats analysis in weeping peach* (*Prunus persica* Var. Pendula), Taian: College of Horticulture Science and Engineering, Shandong Agricultural UniversityTaian, pp.54.
19. Li, R., Li, Y., Kristiansen, K. and Wang, J. 2008, SOAP: short oligonucleotide alignment program, *Bioinformatics*, **24**, 713–4.
20. Liu, D., Ma, C. and Hong, W., et al. 2014, Construction and analysis of high-density linkage map using high-throughput sequencing data, *PloS ONE*, **9**, e98855.
21. Jansen, J., De Jong, A. and Van Ooijen, J. 2001, Constructing dense genetic linkage maps, *Theor. Appl. Genet.*, **102**, 1113–22.
22. Van Ooijen, J. 2011, Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species, *Genet. Res.*, **93**, 343–9.
23. van Os, H., Stam, P., Visser, R.G. and van Eck, H.J. 2005, SMOOTH: a statistical method for successful removal of genotyping errors from high-density genetic linkage data, *Theor. Appl. Genet.*, **112**, 187–94.
24. Huang, X., Zhao, Y. and Wei, X., et al. 2012, Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm, *Nat. Genet.*, **44**, 32–9.
25. Xu, S. and Hu, Z. 2010, Mapping quantitative trait loci using distorted markers, *Int. J. Plant Genom*, **2009**, 2–10.
26. Kosambi, D. 1943, The estimation of map distances from recombination values, *Ann. Eugenics*, **12**, 172–5.
27. Broman, K.W. and Sen, S. 2009, *A guide to QTL mapping with R/qtl*. Springer, New York.
28. Abe, A., Kosugi, S. and Yoshida, K., et al. 2012, Genome sequencing reveals agronomically important loci in rice using MutMap, *Nat. Biotechnol.*, **30**, 174–8.
29. Ashburner, M., Ball, C.A. and Blake, J.A., et al. 2000, Gene Ontology: tool for the unification of biology, *Nat. Genet.*, **25**, 25–9.
30. Chen, J., Yang, B., Yang, W., Hao, R., Zhang, J. and Zhang, Q. 2013, Chromosome sectioning of Prunus mume Siet. et Zucc all the year around, *Acta Agrticulturae Boreali-Occidentalis Sinica*, **22**, 150–4.
31. Weber, J.L. 1990, Informativeness of human (dC-dA) n·(dG-dT) n polymorphisms, *Genomics*, **7**, 524–30.
32. Liepman, A.H., Wilkerson, C.G. and Keegstra, K. 2005, Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases, *Proc. Natl Acad. Sci. USA*, **102**, 2221–6.
33. Bujang, N., Harrison, N. and Su, N.-Y. 2014, A phylogenetic study of endo-beta-1, 4-glucanase in higher termites, *Insectes Sociaux*, **61**, 29–40.
34. Lashbrook, C.C., Gonzalez-Bosch, C. and Bennett, A.B. 1994, Two divergent endo-beta-1, 4-glucanase genes exhibit overlapping expression in ripening fruit and abscising flowers, *Plant Cell Online*, **6**, 1485–93.
35. Adair, W.S. and Apt, K.E. 1990, Cell wall regeneration in Chlamydomonas: accumulation of mRNAs encoding cell wall hydroxyproline-rich glycoproteins, *Proc. Natl Acad. Sci.*, **87**, 7355–9.
36. Baumberger, N., Doesseger, B. and Guyot, R., et al. 2003, Whole-genome comparison of leucine-rich repeat extensins in Arabidopsis and rice. A conserved family of cell wall proteins form a vegetative and a reproductive clade, *Plant Physiol.*, **131**, 1313–26.
37. Busov, V.B., Johannes, E. and Whetten, R.W., et al. 2004, An auxin-inducible gene from loblolly pine (Pinus taeda L.) is differentially expressed in mature and juvenile-phase shoots and encodes a putative transmembrane protein, *Planta*, **218**, 916–27.
38. Andersson Gunnerås, S. 2005, *Wood formation and transcript analysis with focus on tension wood and ethylene biology*, Umeå: Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural SciencesUmeå, pp.3.
39. Barakat, A., Bagniewska-Zadworna, A. and Choi, A., et al. 2009, The cinnamyl alcohol dehydrogenase gene family in Populus: phylogeny, organization, and expression, *BMC Plant Biol.*, **9**, 26.
40. Mitsuda, N., Iwase, A. and Yamamoto, H., et al. 2007, NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of Arabidopsis, *Plant Cell Online*, **19**, 270–80.
41. Sugano, M., Nakagawa, Y., Nyunoya, H. and Nakamura, T. 2004, Expression of gibberellin 3. BETA.-hydroxylase gene in a gravi-response mutant, weeping Japanese flowering cherry, *Biol. Sci. Space*, **18**, 261–6.
42. Kurepa, J., Wang, S., Li, Y., Zaitlin, D., Pierce, A.J. and Smalle, J.A. 2009, Loss of 26S proteasome function leads to increased cell size and decreased cell number in Arabidopsis shoot organs, *Plant Physiol.*, **150**, 178–89.
43. Castaings, L., Camargo, A. and Pocholle, D., et al. 2009, The nodule inception-like protein 7 modulates nitrate sensing and metabolism in Arabidopsis, *Plant J.*, **57**, 426–35.
44. Heim, M.A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B. and Bailey, P.C. 2003, The basic helix-loop-helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity, *Mol. Biol. Evol.*, **20**, 735–47.
45. Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J.-Y. and Chen, X. 2009, Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis, *Genes Dev.*, **23**, 2850–60.
46. Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. and Blaxter, M.L. 2011, Genome-wide genetic marker discovery and genotyping using next-generation sequencing, *Nat. Rev. Genet.*, **12**, 499–510.
47. Sun, L., Yang, W. and Zhang, Q., et al. 2013, Genome-wide characterization and linkage mapping of simple sequence repeats in Mei (Prunus mume Sieb. et Zucc.), *PloS ONE*, **8**, e59562.
48. Zhang, Q.X., Chen, W.B. and Sun, L.D., et al. 2012, The genome of Prunus mume, *Nat. Commun.*, **3**, 1318.
49. Arús, P., Verde, I., Sosinski, B., Zhebentyayeva, T. and Abbott, A.G. 2012, The peach genome, *Tree Genet. Genom.*, **8**, 531–47.
50. Faris, J., Laddomada, B. and Gill, B. 1998, Molecular mapping of segregation distortion loci in Aegilops tauschii, *Genetics*, **149**, 319–27.
51. Wang, C., Zhu, C., Zhai, H. and Wan, J. 2005, Mapping segregation distortion loci and quantitative trait loci for spikelet sterility in rice (Oryza sativa L.), *Genet. Res.*, **86**, 97–106.
52. Zhang, L., Wang, S. and Li, H., et al. 2010, Effects of missing marker and segregation distortion on QTL mapping in F2 populations, *Theor. Appl. Genet.*, **121**, 1071–82.
53. Xu, S. 2008, Quantitative trait locus mapping can benefit from segregation distortion, *Genetics*, **180**, 2201–8.
54. Bassi, D., Rizzo, M., Geibel, M., Fischer, M. and Fischer, C. 2000, Peach breeding for growth habit, *Acta Hortic.*, **1**, 411–4.
55. Monet, R., Bastard, Y. and Gibault, B. 1988, Genetic study of the weeping habit in peach, *Agronomie*, **8**, 127–32.
56. Scorza, R., Bassi, D. and Liverani, A. 2002, Genetic interactions of pillar (columnar), compact, and dwarf peach tree genotypes, *J. Am. Soc. Hortic. Sci.*, **127**, 254–61.
57. Yamazaki, K., Okabe, M. and Takahashi, E. 1987, Inheritance of some characteristics and breeding of new hybrids in flowering peaches. *Bull. Kanagawa Hort. Exp. Sta. (Japan)*, **34**, 46–53.
58. Tu, S. and Qiu, F. 2007, The character display of *Castanea mollissima* var. pendula Sexual Progenies, *Acta Hirticuoturae Sinica*, **34**, 760–2.
59. Coutand, C., Jeronimidis, G., Chanson, B. and Loup, C. 2004, Comparison of mechanical properties of tension and opposite wood in Populus, *Wood Sci. Technol.*, **38**, 11–24.