American Genetic Association

Journal of Heredity, 2020, 515–530
doi:10.1093/jhered/esaa039
Original Article
Advance Access publication September 15, 2020

OXFORD

Original Article

High-Quality SNP Linkage Maps Improved QTL Mapping and Genome Assembly in *Populus*

Chunfa Tong, Dan Yao, Hainan Wu, Yuhua Chen, Wenguo Yang, and Wei Zhao

From the Co-Innovation Center for Sustainable Forestry in Southern China, College of Forestry, Nanjing Forestry University, Nanjing 210037, China.

Address correspondence to C. Tong at the address above, or e-mail: tongchf@njfu.edu.cn.

Received June 15, 2020; First decision July 28, 2020; Accepted September 11, 2020.

Corresponding Editor: Mark Chapman

Abstract

With the advances in high-throughput sequencing technologies and the development of new software for extracting single nucleotide polymorphisms (SNPs) across a mapping population, it is possible to construct high-quality genetic maps with thousands of SNPs in outbred forest trees. Two parent-specific linkage maps were constructed with restriction site-associated DNA sequencing data from an F_1 hybrid population derived from *Populus deltoides* and *Populus simonii*, and applied in QTL mapping and genome assembly. The female *P. deltoides* map contained 4018 SNPs, which were divided into 19 linkage groups under a wide range of LOD thresholds from 7 to 55. The male *P. simonii* map showed similar characteristics, consisting of 2097 SNPs, which also belonged to 19 linkage groups under LOD thresholds of 7 to 29. The SNP order of each linkage group was optimal among different ordering results from several available software. Moreover, the linkage maps allowed the detection of 39 QTLs underlying tree height and 47 for diameter at breast height. In addition, the linkage maps improved the anchoring of 689 contigs of *P. simonii* to chromosomes. The 2 parental genetic maps of *Populus* are of high quality, especially in terms of SNP data quality, the SNP order within linkage groups, and the perfect match between the number of linkage groups and the karyotype of *Populus*, as well as the excellent performances in QTL mapping and genome assembly. Both approaches for extracting and ordering SNPs could be applied to other species for constructing high-quality genetic maps.

Subject areas: Quantitative genetics
Keywords: genetic linkage map, restriction site-associated DNA sequencing, single nucleotide polymorphism, quantitative trait locus, genome assembly, *Populus*

The genus *Populus* comprises approximately 30 species of deciduous flowering trees in the willow family (Salicaceae) and is widely distributed in the Northern Hemisphere (Boes and Strauss 1994). Three common names, poplar, aspen, and cottonwood, refer to different *Populus* species (Sannigrahi et al. 2010). According to morphological similarity and crossability, the species within this genus are grouped into 6 sections (*Abaso*, *Aigeiros*, *Leucoides*, *Populus*, *Tacamahaca*, *Turanga*), although some controversies remain unresolved (Eckenwalder 1996; Wang et al. 2014). Most of these species are of great commercial and ecological importance because they are fast-growing trees and can provide valuable raw materials within a short period of time for the production of pulp and paper, wood-based panels, biofuels, and other biobased products (Sannigrahi et al. 2010; Zhigunov et al. 2017). This genus has been chosen as a

model system for forest trees due to biological characteristics such as a small genome size (~480 Mbp), a fast growth rate, and the ease of asexual and seed reproduction (Woolbright et al. 2008). Whole-genome sequences are now available for several species of the genus, including *Populus trichocarpa* (Tuskan et al. 2006), *P. euphratica* (Ma et al. 2013), and *P. deltoides* (http://www.phytozome.net). However, the completion of these valuable genome resources relied on genetic linkage maps for resolving misassembly issues and anchoring chromosome-scale sequences (Fierst 2015). A genetic linkage map that displays the linear orders and genetic distances of groups of molecular markers has been used for over 100 years, since the first linkage map was generated for *Drosophila melanogaster* (Sturtevant 1913). Indeed, genetic linkage maps not only play an essential role in the assembly of genome scaffolds and comparative genomics (Krutovsky et al. 2004; Kakioka et al. 2013) but are also prerequisites for identifying quantitative trait loci (QTLs), which indicates the relationship between phenotype and genotype in plants and animals (Lander and Botstein 1989; Zeng 1994).

Although the precision and accuracy of a genetic map are affected by many factors, such as the population size, the number of available genetic markers, and the methods used for ordering markers within a linkage group (Tao et al. 2018), considerable efforts have been devoted to constructing genetic linkage maps in at least 10 *Populus* species for the past 3 decades (Wang et al. 2011). Most linkage maps generated in *Populus* are sparse due to the use of a small population size (<200) or a very limited number (<1000) of traditional molecular markers (Bradshaw et al. 1994; Yin et al. 2002; Zhang et al. 2004; Woolbright et al. 2008; Zhang et al. 2009; Paolucci et al. 2010; Zhou et al. 2015). The assessment of these traditional markers, including randomly amplified polymorphic DNA (RAPD), restriction fragment length polymorphisms (RFLPs), amplified fragment length polymorphisms (AFLPs), and simple sequence repeats (SSRs), is considered expensive and laborious, and these markers may be unstable due to uncontrolled experimental conditions (Davey et al. 2011; Wang et al. 2012). Du et al. (Du et al. 2016) applied 2613 traditional markers (including AFLPs, SSRs, and InDels) from 1200 individuals to produce a high-resolution linkage map that consisted of 42 linkage groups in the *Populus* section. Despite the large numbers of markers and individuals used in their work, the number of linkage groups seriously deviated from the karyotype of *Populus* (2n = 38), possibly because the majority of the markers were less effective dominant markers (Gerber et al. 2000). With the advances in the next-generation sequencing (NGS) technologies, an increasing number of linkage maps have been built on the basis of single nucleotide polymorphism (SNP) markers generated from restriction site-associated sequencing (RADseq) data (Baird et al. 2008; Chutimanitsakun et al. 2011; Kakioka et al. 2013). Recently, Tong et al. (2016) and Zhigunov et al. (2017) applied RADseq data to construct linkage maps in F_1 hybrid populations of *Populus deltoides* × *Populus simonii* and *Populus tremula* × *Populus tremula*, respectively, based on the reference genome of *P. trichocarpa* (Tuskan et al. 2006). In contrast, Mousavi et al. (2016) proposed a strategy for the *de novo* generation of SNPs with RADseq data across 150 progeny in an F_1 hybrid population of *P. deltoides* × *P. simoniii*, by incorporating whole-genome sequencing data from the 2 parents. This strategy successfully identified 3442 high-quality SNPs and produced the 2 parental linkage groups under a wide range of LOD (logarithm of odds) thresholds, which perfectly matched the karyotype of *Populus*. Their work demonstrated that genetic mapping with RADseq in *Populus* presents a significant advantage over traditional PCR-based methods.

To perform genetic linkage mapping and QTL analysis in *Populus*, we established an F_1 hybrid population by crossing *P. deltoides* and *P. simoniii* over 3 years from 2009 to 2011. The female parent, *P. deltoides*, and the male parent, *P. simoniii*, exhibit substantial differences in their growth rates and stress responses, thus producing permanent material for genetic mapping (Mousavi et al. 2016). In our previous studies (Mousavi et al. 2016; Tong et al. 2016; Yao et al. 2020), different strategies for performing genetic mapping with RADseq data from the F_1 population were investigated, ultimately leading to the development of the novel software gmRAD for *de novo* extraction of high-quality SNP genotypes. In the present study, we incorporated more individuals from the F_1 hybrid population to construct 2 parental high-quality genetic maps, and gmRAD was utilized for the generation of high-quality SNP genotypes solely from RADseq data. Additionally, a new strategy was adopted for obtaining the optimal order of SNPs within a linkage group, which was highly beneficial to the quality of the 2 linkage maps. Based on the newly constructed parental linkage maps, we further performed QTL analysis of tree height (H) and diameter at breast height (DBH) with phenotypic data measured over 5 years. The candidate genes of the detected QTLs and their relationship with QTLs previously identified in other studies were also investigated. In addition, the linkage maps significantly improved the anchoring of a draft genome sequence of the male parent *P. simonii* to the chromosomes.

Materials and Methods

Mapping Population and Measurement of Growth Traits

The mapping population was established by crossing *P. deltoides* Marsh cv. "Lux" and *P. simonii* Carr. in the springs from 2009 to 2011 (Mousavi et al. 2016; Tong et al. 2016). The female *P. deltoides* parent came from the "I-69" cultivar, which originated from a natural population in Illinois, United States and was selected in Italy in the 1950s and introduced to China in 1972 (Zhang et al. 2009). The male *P. simonii* "L-3" was chosen from a forestland in Luoning County, Henan Province, China; the voucher specimen number of this parent is 35353, which was identified by S. Y. Wang and deposited at the Henan Normal University Biological Specimen Depository (HENU), China. All the 3-year interspecific crosses were conducted on the Siyang Forest Farm (SFF) in Siyang County, Jiangsu Province, China. Approximately 200 seedlings that were generated in 2009 and 2010 were first planted in Qiaolin Town, Nanjing, China, and the cuttings of these plants were collected to plant at SFF in 2011 for acclimatization. In the spring of 2012, a total of 527 individuals belonging to 145 clones were transplanted with a 2 × 2 m spacing and were cut close to the soil surface thereafter, on Xiashu Forest Farm (XFF) of Nanjing Forestry University, Jurong County, Jiangsu Province, China. Additionally, 384 seedlings were derived in 2011 and transplanted to XFF in 2012. These seedlings were planted with a tight spacing due to the limitation of land resources. Therefore, we obtained an F_1 hybrid population of *P. deltoides* and *P. simonii* for linkage mapping in 2012, including 145 clones each having 3–5 individuals and 384 seedlings. The seedlings were reserved just for genotypic resources but not expected for QTL mapping because their growth was seriously affected over the years by the tight spacing each other. The clones were designed as a permanent material for QTL mapping. The H and DBH traits of each individual in the clonal experiment were continuously measured for 5 years, from 2013 to 2017. The trait mean of each clone per year was applied for QTL mapping.

DNA Extraction and RAD Sequencing

In the springs of 2013 and 2016, fresh leaf tissue was collected from the 2 parents and a total of 418 progeny. The samples were stored in a −80°C freezer, and DNA was extracted according to the manufacturer's protocol of the Plant Genomic DNA Kit (Tiangen, Tiangen Biotech Co. Ltd., Beijing, China) or following the CTAB protocol (Doyle and Doyle 1987). The RAD library for each individual was constructed basically following the protocol described by Baird et al. (Baird et al. 2008). Briefly, genomic DNA from each sample was digested with the restriction enzyme *EcoRI* (NEB, USA), and a P_1 adaptor containing a MID (molecular identifier) was then ligated to the products of the restriction reaction. Thereafter, the DNA samples were pooled and randomly sheared by sonication before the addition of the paired-end P_2 adaptor. Libraries with an insert size ranging from 300 to 500 bp were prepared and sequenced from both ends (paired-end) with 82–150 bp read lengths. The RAD sequencing of all the individuals was performed at 5 different times in 4 sequencing companies. In 2013, the 2 parents and 150 progeny were sequenced in 8 lanes on the Illumina HiSeq 2000 platform at Novogene Bioinformatics Institute, Beijing, China (NBI), while the same 2 parents and another 150 progeny were sequenced in 7 lanes on the Illumina HiSeq 2000 platform at Beijing Genomics Institute, Shenzhen, China (BGI) (Tong et al. 2016). In 2016, we also sequenced 39 progeny at Majorbio Company, Shanghai, China (MBC), 40 at Genepioneer Biotechnologies, Nanjing, China (GPB), and 47 at BGI on the Illumina HiSeq 4000 platforms. In addition to the 2 parents, there were 8 progeny that were each sequenced separately in 2 different sequencing companies (Supplementary Tables S1–S5). The same individual reads data from different companies can be used to validate the accuracy of sequencing and subsequent SNP calling.

SNP Discovery and Genotyping

The raw data from each sequencing company were processed to generate so-called clear data via the following procedures: 1) each sample reads were segregated by its unique nucleotide MID to generate the first and second read files in fastq format; 2) if the paired reads contained primer/adaptor sequences, they were removed from the fastq files; 3) if 10% of bases were uncalled in a single read, the pair was deleted from the data files; and 4) if a single read contained more than half low-quality bases (Phred score ≤5), the pair was also removed from the dataset. To obtain high-quality (HQ) reads, the NGS QC toolkit (Patel and Jain 2012) was applied to further filter out those paired reads in which a single read contained more than 30% bases that had a Phred score less than 20.

We used the software gmRAD for SNP discovery and genotype calling in the HQ read data for all individuals across the F_1 mapping population (Yao et al. 2020). The main parameters of the software used for analyzing the whole dataset were set as follows: 1) the hamming distance for grouping the first reads of each parent was set to 5; 2) the edit distance for filtering mapped reads was set as 5; 3) the coverage depth of an allele of a heterozygote was required to be at least 3; 4) the percent of genome repeat regions was set to ~70% based on the genome sequence of *P. trichocarpa* (Tuskan et al. 2006); 5) the minimum score of an SNP genotype was set as 30; 6) the minimum percent of non-missing genotypes at an SNP was set equal to 80%; and 7) the minimum *P*-value from the chi-squared test for the segregation ratio at an SNP was set as 0.05.

Genetic Map Construction

Owing to the biological characteristics of the F_1 hybrid population, the majority of SNPs segregated in the types of *aa*×*ab* and *ab*×*aa* (Mousavi et al. 2016; Tong et al. 2016). We used these 2 SNP segregation datasets to construct the linkage maps of female *P. deltoides* and male *P. simonii* separately. First, each dataset was divided into linkage groups under an LOD threshold after performing 2-point linkage analysis with the software FsLinkageMap (Tong et al. 2010). Second, the SNPs in each linkage group were ordered with different types of available software that can be used for genetic mapping in an F_1 hybrid population of outbred species. We used JoinMap (Van Ooijen 2006), OneMap (Margarido et al. 2007), and Lep-MAP (Rastas et al. 2013) to order the SNPs within a linkage group 5, 1, and 3 times, respectively. Among the ordering results from the 3 software programs, we chose the optimal result for the final map based on the ordering-objective function of the minimum sum of adjacent recombination fractions (SARF) (Falk 1989). Third, to match the linkage groups with chromosomes, we extracted a flanking sequence of 41 bp for each SNP and aligned it with BLAST to the reference genome of *P. trichocarpa*. Therefore, the linkage group number was determined to be the chromosome number if the majority of its SNPs were mapped to the reference chromosome. Finally, linkage maps were first plotted in WMF format with Kosambi's genetic mapping function (Kosambi 1944) by FsLinkageMap, and then further modified into PDF format with the software Mayura Draw (http://www.mayura.com).

QTL Mapping for Longitudinal Growth Traits

We performed QTL analysis of the 2 growth traits (H and DBH) with a phenotypic dataset obtained from the F_1 hybrid population based on the parental linkage maps constructed in this study. The dataset included 5-year H and DBH data measured from 332 individuals belonging to 109 clones. We applied the means of the growth traits for each clone per year to conduct QTL analysis using the software *mvqtlcim* (Liu et al. 2017), which is designed to implement composite interval mapping (Zeng 1994) of longitudinal or multiple traits in an outbred full-sib family. Because all the SNP markers in the linkage maps segregate in a 1:1 ratio, the QTL segregation patterns were assumed to be *Qq*×*QQ* in the maternal linkage map and *QQ*×*Qq* in the paternal linkage map. When performing the CIM calculations, the number of background markers was iterated from 11 to 49 with a step size of 2 and the window size was set to 10, 15, or 20 cM. For each parameter combination of the background marker number (BMN) and window size (WS), the LR was calculated at each map position in steps of 1 cM, and the LR threshold for declaring a significant QTL was determined from 1000 permutation tests (Doerge and Churchill 1996). To obtain the optimal mapping result, we fitted a multivariate linear regression model for each trait on the genotype effects of all detected QTLs over the 5 time points for each parameter combination. The regression model can be expressed for *N* detected QTLs as follows:

$$y_{it} = \mu_t + \sum_{j=1}^{N} (p_{ij1} - p_{ij2}) a_{jt} + e_{it},\ i = 1,\ 2,\ \cdots,\ n;\ t = 1,\ 2,\ \cdots,\ 5 \quad (1)$$

where y_{it} is the trait value of the ith clone at time t; μ_t is the overall mean at time t; a_{jt} is the homozygous genotype (QQ) effect of the jth detected QTL at time t, while $-a_{jt}$ is the heterozygous genotype (Qq) effect; p_{ij1} is the conditional probability of the jth QTL genotype QQ on its flanking marker genotype and p_{ij2} is the conditional probability for Qq (Liu et al. 2017); and e_{it} is the random error at time t for the ith individual. The coefficient of determination (R^2) of model (1) for each time point can be regarded as the proportion of the phenotypic

variance at that time explained by all detected QTLs. Therefore, the optimal mapping result was chosen as the result that corresponded to the maximum mean R^2 values over the time points among all the parameter combinations. For the optimal mapping result, the effects of each QTL over different times were re-estimated with model (1), and the corresponding heritabilities were calculated as the differences in the R^2 values when it was removed from the model.

Investigation of QTL Candidate Genes

To investigate the candidate genes for the identified QTLs, we first obtained the coding sequences (CDS) of those genes within the physical regions of QTLs from the gene annotation database of *P. trichocarpa* v3.1 at Phytozome (https://phytozome.jgi.doe.gov). Then, the genes were re-annotated with Blast2GO by subjecting their CDSs to Blast searches against the protein database and then mapping the hits to Gene Ontology (GO) terms. With the gene annotations, we identified candidate genes for growth and development as well as response to stresses.

Chromosomal Assembly With Genetic Maps

As a genome assembly application, we used the 2 parental linkage maps constructed here to re-anchor the contigs of the male parent *P. simonii* to chromosomes. The draft genome assembly contained 686 contigs with accession numbers from VJNQ02000001.1 to VJNQ02000686.1 at the NCBI, with a total length of 441.38 Mb and a contig N50 of 1.94 Mb, which were assembled from single-molecule long reads generated by the PacBio platform in our previous study (Wu et al. 2020). First, the flanking sequence of each SNP in the 2 linkage maps was subjected to Blast searches against the contigs by setting the query coverage value equal to 90%. For an SNP, the best blast hit with a maximum query coverage and a minimum expected value was remained, letting it connect to a unique contig. Next, 2 comma-delimited (csv) files for each parental linkage map were prepared by recording the contig and the position to which an SNP sequence was aligned as well as the linkage number and position of the SNP in the linkage map. Thereafter, we performed anchoring using the software ALLMAPS (Tang et al. 2015) with equal weights for the 2 linkage maps.

Results

RAD Sequencing Data

A total of 1486.2 Gb RADseq data were obtained from the 2 parents and 418 progeny in the F_1 hybrid population of *P. deltoides* and *P. simonii*. These data consisted of paired-end (PE) reads with different lengths and came from the companies NBI, BGI, GPB, and MBC, and were considered clean data because they were gently filtered by the sequencing companies (as described in Materials and Methods). All the clean data are available under accession numbers in the NCBI SRA database (http://www.ncbi.nlm.nih.gov/Traces/sra), which were listed in Supplementary Tables S1–S5. After further filtering with the NGS QC toolkit (Patel and Jain 2012), we obtained 1385.0 Gb of high-quality (HQ) reads data for genetic mapping. Table 1 summarizes the clean and HQ data to provide the average information regarding reads number and base numbers for each of the 5 batches of sequencing conducted at the 4 companies. More details of each sample data are presented in Supplementary Tables S1–S5, which includes the PE reads number, the number of bases, the first and second read lengths, and the accession number prefixed with SRR in the NCBI SRA database.

SNP Discovery and Quality Assessment

We analyzed the HQ reads with the software package gmRAD to extract SNP genotype data for the 2 parents and 418 progeny in the F_1 hybrid population. As expected, the majority of SNPs segregated in the types of *aa*×*ab* and *ab*×*aa*, which amounted to 2101 and 4021, respectively. We counted the number of non-missing genotypes at each of these SNPs and found that the average number was 388 (92.82%) and that the minimum number was 335. Additionally, we found that the missing genotype rate in an individual ranged from 0.05% to 39.42% with an average value of 7.18%. To evaluate the accuracy of the genotypes at these SNPs, 8 progeny that were sequenced independently at 2 different companies were used to confirm the consistency of the genotypes. For each individual, the ratio of the number of SNP sites that were genotyped consistently with independent RADseq data was calculated. The consistency rate was shown to be very high, in the range of 98.61% to 99.98%, and the overall rate was up to 99.59% (Supplementary Table S6).

Table 1. Summary of the average reads number of the RADSeq data with the number of bases in brackets for the 5 batches of sequencing in the 4 companies

Batch	Company	Sample type	Sample number	Raw reads number (Gb)	HQ reads number (Gb)
1	NBI[a]	Male parent	1	14 024 713 (2.80)	12 617 155 (2.52)
		Female parent	1	12 949 974 (2.59)	11 500 364 (2.30)
		Progeny	150	4 555 004 (0.91)	4 084 337 (0.82)
2	BGI[b]	Male parent	1	57 159 139 (9.91)	55 789 694 (9.68)
		Female parent	1	21 619 787 (3.72)	20 956 868 (3.60)
		Progeny	150	6 701 881 (1.16)	6 500 976 (1.13)
3	GPB[c]	Progeny	40	9 880 332 (2.91)	9 869 660 (2.90)
4	MBC[d]	Progeny	39	11 962 418 (3.68)	11 121 430 (3.29)
5	BGI	Progeny	47	129 945 928 (19.08)	120 126 084 (17.63)
Total		Male parent	1	71 183 852 (12.71)	68 406 849 (12.20)
		Female parent	1	34 569 761 (6.31)	32 457 232 (5.90)
		Progeny	418[e]	20 712 294 (3.51)	19 287 668 (3.27)

[a]NBI: Novogene Bioinformatics Institute, Beijing, China.
[b]BGI: Beijing Genomics Institute, Shenzhen, China.
[c]GPB: Genepioneer Biotechnologies, Nanjing, China.
[d]MBC: Majorbio Company, Shanghai, China.
[e]In the progeny, there are 8 individuals each sequenced in 2 batches indipendently.

Genetic Linkage Maps

After performing 2-point linkage analysis with FsLinkageMap, 4018 SNPs segregating in *ab*×*aa* were grouped into 19 linkage groups in the female parent under a wide range of LOD thresholds from 7 to 55, while 2097 SNPs segregating in *aa*×*ab* in the male parent also constituted 19 linkage groups, with LOD thresholds ranging from 7 to 29. For each linkage group, the optimal order of markers was chosen among the 9 ordering results from JoinMap, OneMap, and Lep-MAP (Supplementary Files 2 and 3), and the map was drawn in this order with the genetic distances converted from the recombination fractions of adjacent markers by Kosambi's function. Each linkage group was numbered according to the reference chromosome to which most of its SNPs were mapped (Supplementary Files 4 and 5). The maternal linkage map spanned a total genetic distance of 7838.48 cM of the genome with an average distance of 1.96 cM between adjacent SNPs, and the paternal linkage map covered 5506.35 cM of the genome with an average distance of 2.65 cM (Figures 1 and 2).

More detailed information about the 2 parental linkage maps is presented in Supplementary Table S7 and Supplementary Files 4 and 5, including SNP interval distances, cumulative distances, linkage phases, SNP flanking sequences, and their mapped positions in the reference genome of *P. trichocarpa*. With these results, we revealed that there remained strong relationships between the genetic linkage maps and the physical map of the reference genome. It was not only the SNP number but also the length of the linkage groups that was highly correlated with chromosome size in the reference genome, with correlation coefficients greater than 0.83 being obtained for both parental linkage maps (Supplementary Table S8). Figure 3 shows the scatter plots of the SNP positions on linkage map against their mapped positions in the reference genome of *P. trichocarpa* for all the linkage groups of the female (DLGs) and male (SLGs) parents. Overall, there was a high level of collinearity between the genetic maps and the physical map. Additionally, we observed that a few short dotted lines in DLGs 2, 5, 6, 9, 18, and 19 and SLGs 2, 4, 7, 8, 9, 12, and 18 were not consistent with the 45° line. These inconsistencies may be due to some complex reasons such as the existence of inverse regions between the 2 parental genomes, the incorrect orders of local SNPs in linkage maps, and the imperfection of the reference genome itself (Tong et al. 2016).

QTL Mapping of Growth Traits

We performed QTL mapping based on the 2 parental linkage maps with the mean H and DHB data for each of the 109 clones using the software *mvqtlcim* and considering different parameter values of the background marker number (BMN) and window size (WS). For tree height (H), 20 significant QTLs were identified in the female linkage map and 19 in the male linkage map (Table 2), explaining 67.60% of the observed phenotypic variance in total. These results were obtained from the optimal runs based on a BMN of 45 and WS of 10.0 for the female linkage map and a BMN of 49 and WS of 10.0 for the male linkage map. Figure 4 displays scatter plots of the log-likelihoods (LR) against the positions of the female and male linkage maps, with dashed lines representing the LR thresholds for declaring significant QTLs. It can be seen that each chromosome presented at least one detected QTL, with Chromosome 1 exhibiting the maximum number of 4. For the trait of diameter at breast height (DBH), we detected 24 QTLs in the female linkage map and 23 in the male linkage map, accounting for 62.58% of the phenotypic variance in total (Table 3). These results corresponded to the run with parameter values of BMN 49 and WS 10.0 for both parental linkage maps. Figure 5 shows the LR profiles of the 2 linkage maps, where the LR thresholds are indicated by dashed lines. We found that these 47 QTLs for DBH were distributed on all chromosomes except Chromosome 19.

More detailed information about these QTLs for H and DBH is listed in Tables 2 and 3, including the QTL IDs, their positions in the linkage maps and chromosomes, LR statistics, and the average heritability over 5 years re-estimated with model (1). The IDs of these QTLs were selected according to the trait, chromosome number, order on a chromosome and one of the 2 parental linkage maps. For example, HQ3D1 represents that the QTL is for tree height (H) and is the third QTL in linkage group 1 of the female *P. deltoides* (D) map; DQ2S6 represents that the QTL is for the diameter (D) at breast height and is the second QTL in linkage group 6 of the male *P. simonii* (S) map. It was observed that on average over the 5 years, each QTL explained 0.07%–6.11% of the phenotypic variance for H and 0.04%–4.69% for DBH.

QTL Candidate Gene Investigation

To explore the candidate genes of the 86 QTLs identified above, we used the software Blast2GO (https://www.blast2go.com) to re-annotate the candidate genes within the physical region corresponding to the marker interval of each QTL (Tables 2 and 3). We found that each QTL contained 3–193 (37 on average) candidate genes for H and 2–117 (29 on average) candidate genes for DBH, 99.3% of which were annotated with descriptions from the Blast hits, GO terms, Enzyme codes, and KEGG maps (Supplementary Files 6 and 7). With the annotation results, we searched key words related to tree growth and development, stress responses, and disease resistance. Consequently, 48 QTLs (26 for H and 22 for DBH) presented at least one candidate gene related to these key words were found (Figure 6, Supplementary Table S9). Among these QTLs, 6 QTLs exhibited candidate genes related to the growth and development of leaves, 18 to roots, 5 to flowers, 8 to seeds, and 7 to the xylem. For the responses to stress or disease resistance, 10 QTLs presented candidate genes for salt stress, 7 for heat stress, 5 for cold stress, and 5 for water deprivation. Additionally, 13 QTLs exhibited candidate genes for disease resistance. In addition, candidate genes related to brassinosteroids were identified in 3 QTLs, auxin responses in 13 QTLs, and photosynthesis in 7 QTLs. It is worth noting that brassinosteroids regulate plant growth, development, and immunity and exert strong effects on plant height (Dubouzet et al. 2013). Further analysis showed that QTLs HQ1S6 and DQ2S6 presented 43 common candidate genes, while HQ2S17 and DQ2S17 presented one common candidate gene. It is likely that each of the 2 pairs belongs to a single QTL that controls the growth traits of H and DBH simultaneously.

Comparison of QTLs With Previous Studies

Compared with previous studies mapping growth traits in *Populus*, we were able to identify more QTLs with a greater total PVE, some of which were related to those detected in previous studies. We detected 39 and 47 QTLs with a total PVE of 67.60% and 62.58% for the growth traits of H and DBH, respectively, while previous studies detected limited numbers of QTLs, typically not exceeding 12 (Bradshaw and Stettler 1995; Wu 1998; Monclus et al. 2012; Du et al. 2016; Liu et al. 2017). Because there is no available position information in the physical map for the QTLs identified in 2 early studies (Bradshaw and Stettler 1995; Wu 1998), we tried to identify the relationship

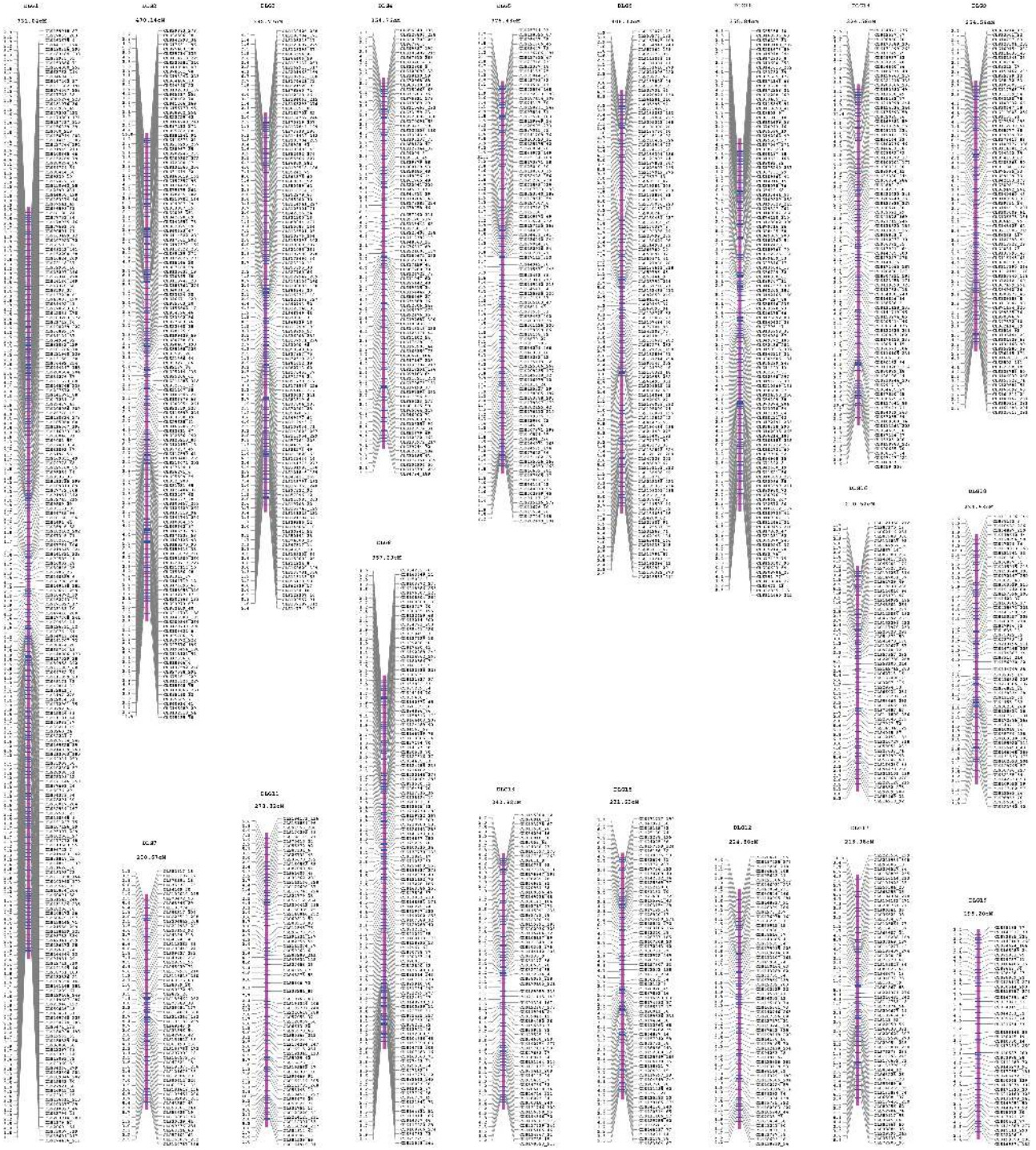

Figure 1. The genetic map of linkage groups DLG1-DLG19 of the maternal *Populus deltoides*. The length of each linkage group is shown under its name. The SNPs are named after the cluster number of the first reads from the female parent and its position on it, prefixed with string "CLS."

between the QTLs found in the current study and 3 recent studies (Monclus et al. 2012; Du et al. 2016; Liu et al. 2017). For tree height, we found that all the QTLs identified in the 3 studies except for a few that were not mapped to the physical map of *P. trichocarpa* were either positioned not far from or their confidence intervals contained a QTL that was identified in this study on the same chromosome (Supplementary Table S10). The distance between the related QTLs was less than 5.0 Mb in most cases and greater than 5.0 Mb but less than 15.0 Mb in a few cases. For the trait of DBH or circumference, we found a similar relationship between the QTLs detected in the current study and 2 studies by Monclus *et al.* (Monclus et al. 2012) and Du et al. (Du et al. 2016), without considering the study of Liu *et al.* (Liu et al. 2017) because they did not perform QTL analysis of such a trait (Supplementary Table S11).

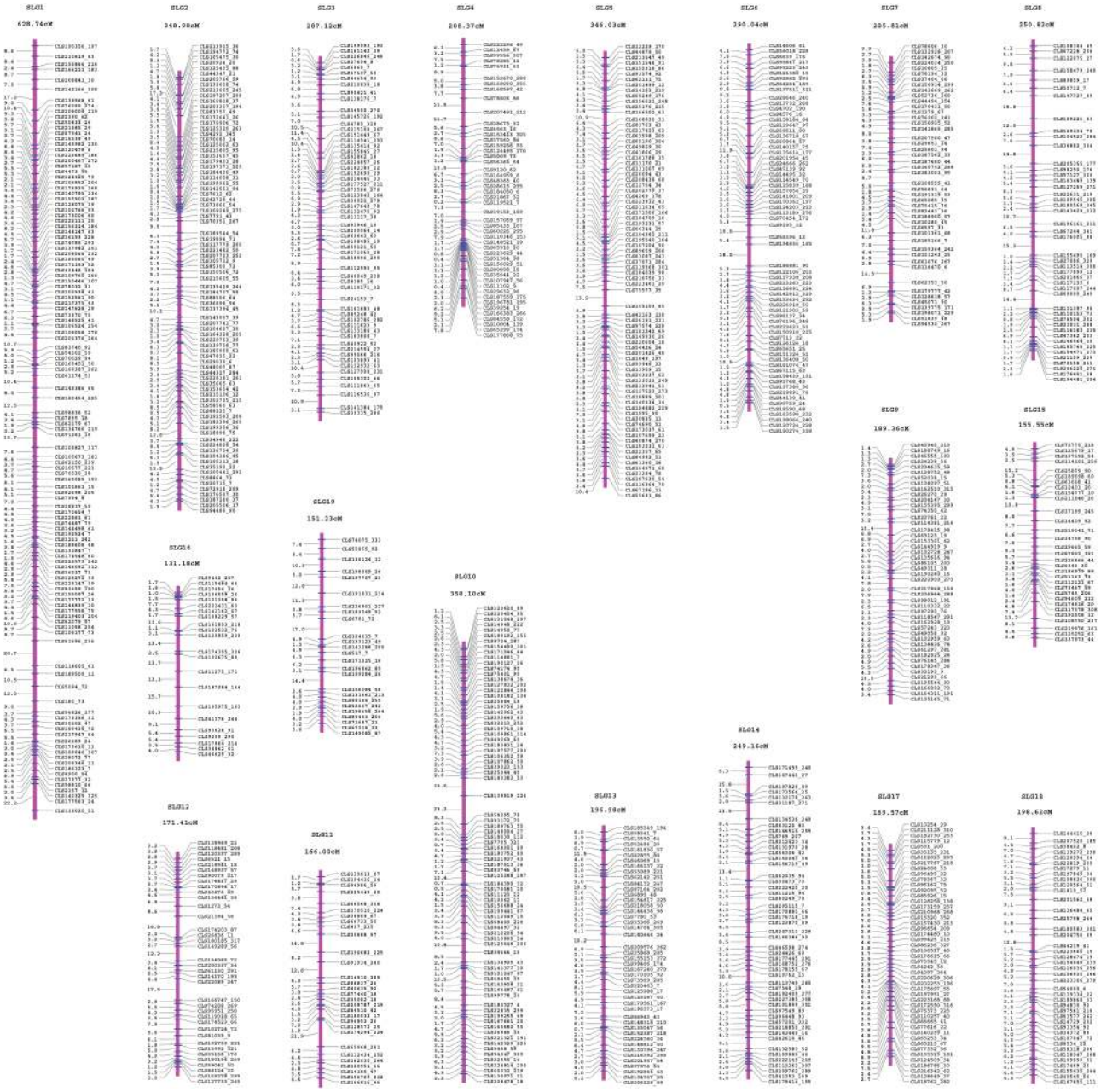

Figure 2. The genetic map of linkage groups SLG1-SLG19 of the paternal *Populus simonii*. The length of each linkage group is shown under its name. The SNPs are named after the cluster number of the first reads from the female parent and its position on it, prefixed with string "CLS."

Genome Assembly of *P. simonii*

With the 2 parental linkage maps constructed above, we first performed the primary assembly of the 686 contigs into chromosomes using the software ALLMAPS (Tang et al. 2015), which were assembled with single-molecule long reads sequenced by the PacBio platform in our previous study (Wu et al. 2020). The result showed that there were 3 contigs each corresponding to 2 separated regions in the same or different linkage groups, indicating that possible errors existed in the assembly of these contigs (Supplementary Figures S1–S3). We split each of the 3 contigs into 2 parts as sub-contigs according to the intersection of the separated regions on the linkage groups, leading to a total of 869 contigs. With these contigs, we were able to successfully anchor a total of 394.08 Mb of sequences to the 19 chromosomes of the male parent *P. simonii*, representing approximately 89.30% of the total bases (Table 4). The assembly data can be accessed in Supplementary File 8 at the Figshare database (https://doi.org/10.6084/m9.figshare.9918515). Overall, 322 contigs were assembled into chromosomes while the remaining 367 contigs were not placed because they contained no SNPs in the genetic maps. However, these unplaced contigs were relatively small fragments, with an average length of 128.87 kb. In particular, all the N50 contigs were positioned in the genome. Regarding the SNPs, 99.64% (6093 vs. 6115) of all the SNPs in the 2 linkage maps were aligned to the contigs, among which 98.15% (5980 vs. 6093) were anchored to the chromosomes. In addition, Figure 7A shows the side-by-side connections between the positions on chromosome 1

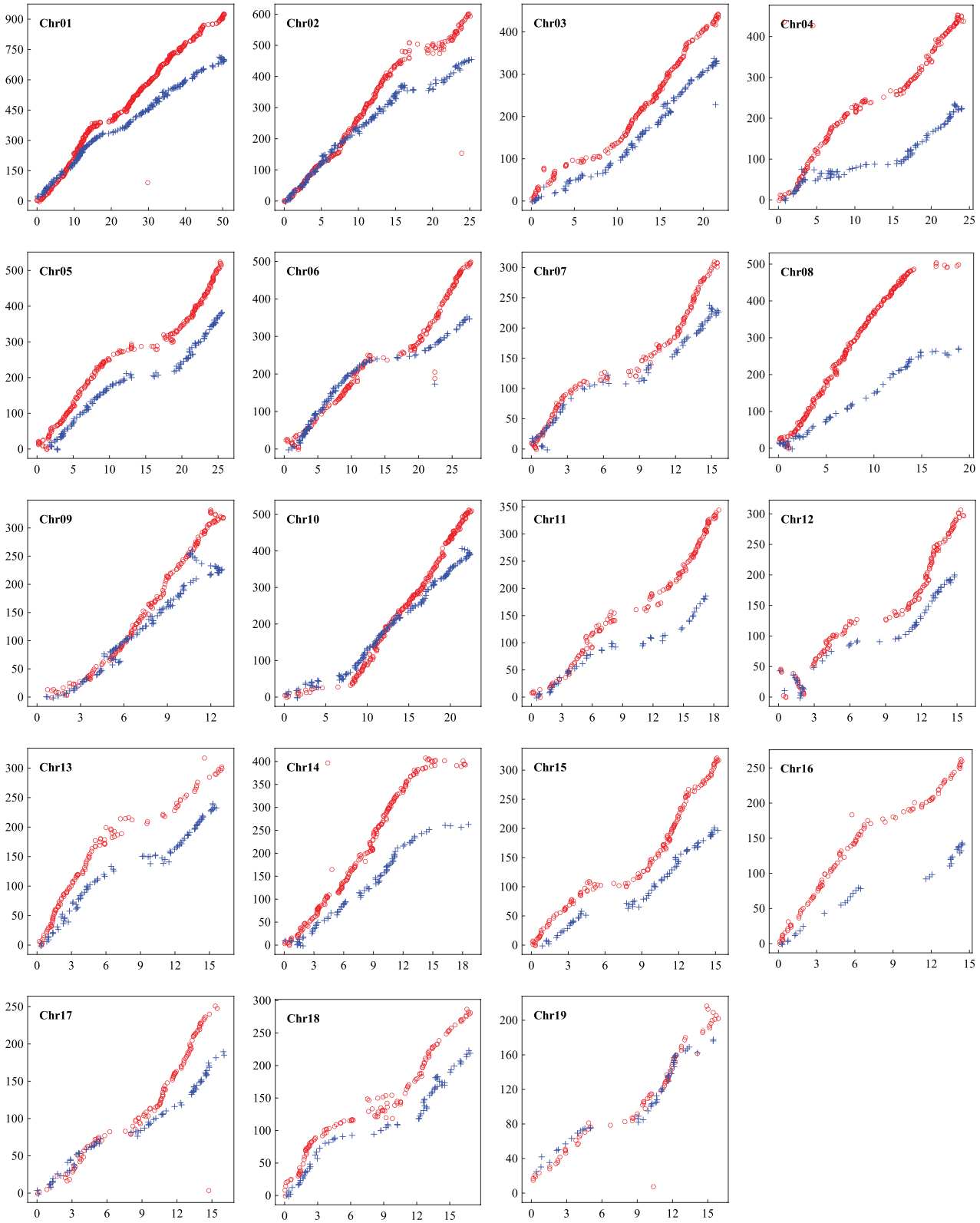

Figure 3. Collinear comparison of the 2 parental linkage maps with the reference genome of *Populus trichocarpa*. The horizontal axis indicates the reference genome position with the unit of Mb; The vertical axis indicates the linkage map position with the unit of cM. The circle and cross points indicate SNP positions on the maternal linkage map of *Populus deltoides* and the paternal linkage map of *Populus simonii* against the positions of the reference genome, respectively.

and the linkage groups, while Figure 7B provides scatter plots of the SNP positions on the chromosome against the positions in the 2 linkage maps. For the other chromosomes, the same visualizations can be found in Supplementary File 9. With these visualizations, we found that each anchored chromosome presented a very high level of collinearity with each of the 2 parental linkage maps, exhibiting

Table 2. Summary of the identified QTLs for the tree height (H) about their positions, LR statistics, effects of genotype *QQ* at each year, and average heritabilities over the 5 years on the 2 parental linkage maps of *Populus deltoides* and *Populus simonii*

QTL ID	Chr/LG[a]	Marker Interval	Map Position (cM)	Genome Position[b] (Mb)	Region Length[b] (kb)	LR	Average Heritability (%)
HQ1D1	1	45	100.61	5.58	271.02	60.08	0.60
HQ2D1	1	108	241.06	10.36	185.41	64.24	0.28
HQ3D1	1	144	315.67	12.41	150.46	110.56	3.17
HQ4D1	1	413	773.85	39.76	99.43	71.00	0.07
HQ1D2	2	100	156.61	7.32	304.62	77.92	2.53
HQ2D2	2	144	264.52	9.80	123.05	82.95	1.27
HQD6	6	178	381.03	23.31	293.61	62.01	1.80
HQD7	7	160	304.08	15.35	228.92	121.66	2.85
HQD8	8	213	381.96	10.42	131.74	102.56	3.94
HQD9	9	107	175.21	8.42	171.32	85.25	0.72
HQ1D10	10	17	26.47	4.23	683.00	72.47	0.12
HQ2D10	10	94	189.47	12.38	178.02	87.26	0.66
HQ1D11	11	81	183.13	11.71	65.63	60.96	0.09
HQ2D11	11	104	229.76	14.65	205.43	78.09	0.83
HQ1D12	12	39	75.99	3.83	13.75	63.90	0.48
HQ2D12	12	75	136.55	10.10	577.68	83.79	1.21
HQD14	14	30	52.44	2.61	64.44	91.51	4.68
HQD16	16	19	39.76	1.67	111.06	61.57	0.22
HQD17	17	118	220.02	13.94	266.68	89.95	4.46
HQD18	18	2	13.85	0.06	59.68	59.08	0.36
HQS3	3	82	209.47	15.79	384.29	67.67	2.98
HQS4	4	98	207.04	22.64	513.66	115.85	1.81
HQ1S5	5	8	20.57	2.19	453.76	98.8	0.53
HQ2S5	5	119	300.33	22.3	35.84	73.23	1.69
HQ1S6	6	11	29.66	2.53	402.61	94.41	1.46
HQ2S6	6	140	350.21	27.38	330.2	80.77	0.11
HQ1S8	8	40	127.17	8.44	1442.00	81.67	6.11
HQ2S8	8	71	269.74	18.35	1044.70	112.58	4.27
HQ1S9	9	12	19.08	2.89	305.16	66.3	1.01
HQ2S9	9	124	254.53	10.65	194.51	74.46	0.70
HQS10	10	110	258.89	16.53	74.56	63.02	1.78
HQS12	12	30	108.82	11.10	250.35	71.38	0.53
HQ1S13	13	9	24.55	1.62	471.81	71.08	2.60
HQ2S13	13	72	182.20	12.97	286.94	72.31	0.37
HQS14	14	53	154.23	9.95	345.44	62.11	1.35
HQS15	15	71	187.40	14.60	53.80	142.94	2.83
HQ1S17	17	53	116.40	11.43	726.85	63.59	0.40
HQ2S17	17	77	166.48	14.67	184.50	103.34	2.44
HQS19	19	14	30.79	0.62	393.11	102.82	3.64

[a]Chr, chromosome; LG, linkage group.
[b]Esitmated with the flanking SNPs mapped to the reference sequence of *Populus trichocarpa* v3.0.

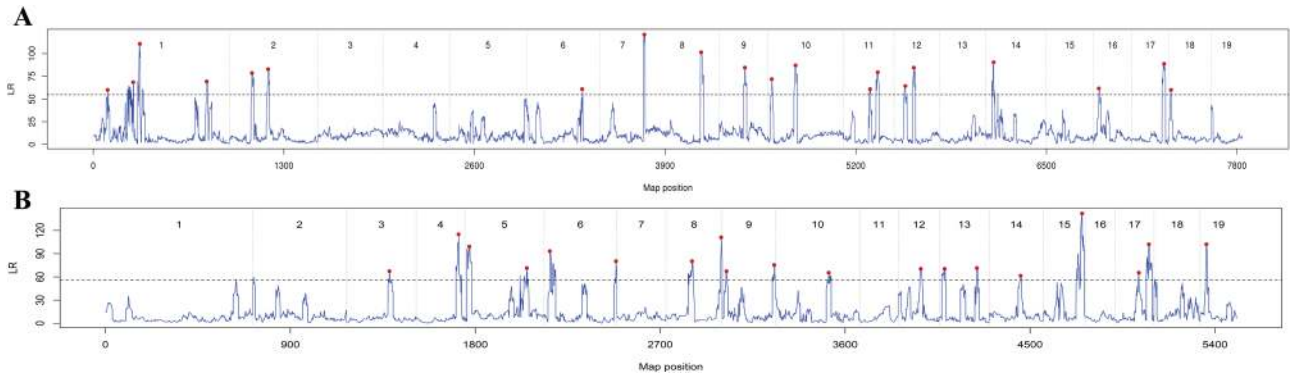

Figure 4. The profile of the log-likelihood ratios (LR) for identifying QTLs controlling the tree height. The profile was based on **(A)** the maternal linkage map of *Populus deltoides* and **(B)** the paternal linkage map of *Populus simonii*. The LR threshold values for declaring existence of a QTL at the significant level of 0.05 are 54.76 on the maternal map and 56.12 on the paternal map, indicated as horizontal dashed lines that were determined by 1000 permutation tests. The vertical dashed lines separate the linkage groups. Each peak labeled with a dot is the highest one within a window size of 20.0 cM, representing a significant QTL.

Table 3. Summary of the identified QTLs for the diameter at breast height (DBH) about their positions, LR statistics, effects of genotype *QQ* at each year and average heritabilities over the 5 years on the 2 parental linkage maps of *Populus deltoides* and *Populus simonii*

QTL ID	Chr/LG	Marker interval	Map position (cM)	Genome position (Mb)	Region length (kb)	LR	Average heritability (%)
DQ1D2	2	107	181.51	7.60	50.36	123.84	0.31
DQ2D2	2	128	232.70	8.97	309.48	82.03	0.71
DQ1D3	3	42	90.93	4.92	1707.45	85.23	0.06
DQ2D3	3	200	380.19	18.72	36.85	130.17	1.14
DQD4	4	9	18.93	1.46	252.54	108.7	0.73
DQ1D5	5	34	64.14	2.63	314.63	128.43	2.82
DQ2D5	5	122	247.08	9.49	117.26	106.47	0.78
DQ1D6	6	50	134.05	8.12	87.96	129.80	0.55
DQ2D6	6	95	206.41	10.83	114.94	77.09	0.34
DQ3D6	6	142	290.08	20.45	74.06	79.01	1.18
DQD7	7	48	85.28	2.62	126.55	115.90	0.45
DQD9	9	68	109.61	6.48	35.06	148.71	1.34
DQ1D10	10	26	45.90	8.47	37.58	79.20	3.96
DQ2D10	10	36	65.47	9.01	154.81	93.69	3.40
DQ1D11	11	26	56.88	3.87	81.90	125.15	3.01
DQ2D11	11	122	263.86	16.02	19.65	78.20	0.06
DQD13	13	8	19.19	0.70	242.51	118.53	0.61
DQ1D14	14	116	212.48	8.81	49.55	116.63	4.69
DQ2D14	14	154	296.56	10.44	93.16	91.53	0.12
DQ3D14	14	176	330.97	11.80	394.18	80.37	0.32
DQD15	15	116	218.82	11.98	109.32	136.02	0.05
DQ1D16	16	21	47.21	1.79	117.64	80.60	0.18
DQ2D16	16	31	75.85	3.12	66.39	128.90	0.75
DQD17	17	101	187.18	12.96	51.61	113.94	4.21
DQ1S1	1	34	104.05	4.48	320.22	72.20	0.50
DQ2S1	1	137	420.99	26.6	248.66	69.44	0.77
DQ3S1	1	159	472.96	31.18	110.45	87.53	0.82
DQ1S2	2	134	327.97	14.71	719.84	77.93	0.22
DQ2S2	2	150	359.59	17.39	121.75	116.12	0.17
DQS4	4	74	132.70	17.95	38.51	87.35	0.16
DQ1S5	5	34	81.52	5.10	18.11	152.66	1.14
DQ2S5	5	112	286.42	21.64	552.69	97.92	7.70
DQ1S6	6	2	5.98	1.31	314.44	91.37	1.43
DQ2S6	6	11	27.66	2.53	402.61	78.62	0.36
DQ3S6	6	51	104.6	5.53	245.6	88.53	0.39
DQ1S7	7	29	70.79	2.55	295.29	100.06	0.08
DQ2S7	7	58	160.70	12.08	56.68	135.04	1.01
DQ3S7	7	71	197.11	14.02	159.72	145.95	2.19
DQ1S8	8	11	17.82	0.74	536.00	68.32	0.23
DQ2S8	8	48	202.92	12.00	153.37	106.63	0.67
DQS9	9	3	2.13	1.05	823.06	88.92	0.46
DQS12	12	46	163.53	13.03	258.08	92.61	0.04
DQS16	16	2	1.18	0.50	457.10	104.01	0.17
DQ1S17	17	64	143.43	13.56	262.33	126.79	0.15
DQ2S17	17	76	163.39	14.51	132.45	130.43	0.09
DQ1S18	18	3	3.24	0.46	126.75	138.38	0.18
DQ2S18	18	52	159.64	13.38	76.61	78.28	0.14

[a]Chr, chromosome; LG, linkage group.
[b]Esitmated with the flanking SNPs mapped to the reference sequence of *Populus trichocarpa* v3.0.

a correlation coefficient in the range of 0.978 to 1.000, with a mean of 0.996.

Discussion

We successfully extracted a large number of high-quality SNPs from very large amounts of RADseq data from the 2 parents and their 418 progeny in the F_1 hybrid population of *P. deltoides* and *P. simonii*. With these SNPs, a high-density genetic linkage map for each parent was constructed, with the number of linkage groups perfectly matching the karyotype of *Populus*. We applied the linkage maps to perform QTL mapping, resulting in the identification of dozens of QTLs dominating the growth traits of H and DBH. We also applied the linkage maps to anchor the contigs of *P. simonii* to chromosomes, significantly improving the assembly quality over the previous result (Wu et al. 2020). Compared with previous similar studies in *Populus*, the linkage maps constructed in this study and the downstream applications have some salient characteristics worth emphasizing.

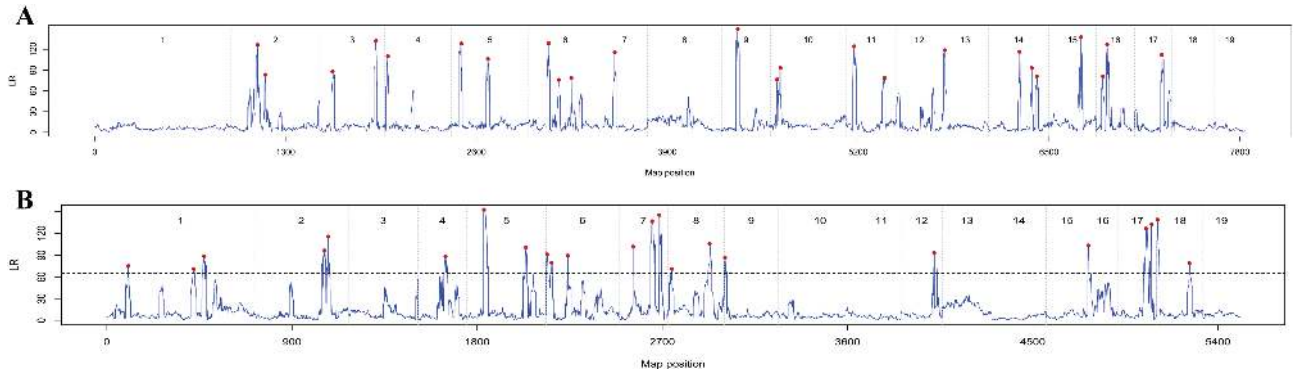

Figure 5. The profile of the log-likelihood ratios (LR) for identifying QTLs controlling the diameter at breast height (DBH). The profile was based on **(A)** the maternal linkage map of *Populus deltoides* and **(B)** the paternal linkage map of *Populus simonii*. The LR threshold values for declaring existence of a QTL at the significant level of 0.05 are 64.91 on the maternal map and 65.27 on the paternal map, indicated as horizontal dashed lines that were determined by 1000 permutation tests. The vertical dashed lines separate the linkage groups. Each peak labeled with a dot is the highest one within a window size of 20.0 cM, representing a significant QTL.

Adequate High-Quality SNP Genotype Data for Genetic Mapping

Currently, RADseq data can be obtained in a fast and low-cost way across many individuals in a mapping population. However, the extraction of enough high-quality SNPs for genetic mapping from massive RADseq data is a challenging prerequisite. Therefore, we provided a powerful software pipeline, gmRAD, for performing this task in a recent work (Yao et al. 2020). We used this software to analyze the RADseq data from the F_1 population and obtained a total of 6122 SNPs that segregated in the type of *aa*×*ab* or *ab*×*aa*, leading to the 2 parental linkage maps. These SNP genotype data could be assessed as more accurate from several aspects. First, the number of SNPs generated here was greater than that extracted with the software Stacks as described in Yao et al. (Yao et al. 2020) because gmRAD handles not only PE reads but also different lengths of reads within or between samples. Second, the SNP genotype data exhibit a high level of accuracy. This was confirmed by the RADseq data from the 8 progeny that were each sequenced separately in 2 different companies, with an overall consistency rate of 99.59% (Supplementary Table S6). Third, it was most surprising that the number of linkage groups for each parental linkage map perfectly matched the karyotype of *Populus* under a wide range of LOD threshold values, from 7 to 55 for the female map and 7 to 29 for the male map. Such strong consistency between linkage group and chromosome numbers has rarely been found in previous genetic mapping studies in *Populus* (Bradshaw et al. 1994; Yin et al. 2002; Zhang et al. 2004; Zhang et al. 2009; Paolucci et al. 2010). This result largely reflects that the identified SNPs are uniformly distributed in the genome and that the genotypes of all individuals at all SNP sites are more reliable. Finally, among all the 6115 SNPs in the 2 parental linkage maps, 96.7% (5916) were well mapped to the reference genome of *P. trichocarpa* with their flanking sequences (Supplementary Files 4 and 5). This finding showed that the large majority of SNPs are universal and are therefore reliable, at least across the species *P. deltoides*, *P. simonii*, and *P. trichocarpa*.

The missing SNP genotype rate for an individual or for an SNP site may cause concerns about influence on linkage mapping. However, the concerns are unnecessary if we clearly understand the statistical details in the analysis of genetic mapping. In fact, most genetic linkage maps constructed with current available software were based on 2-point linkage analysis as the first step followed by the steps of linkage grouping and maker ordering. Therefore, for any 2 markers, the most important thing is whether there are sufficient number of segregation genotypes available to support an accurate estimate of the recombination fraction between them. To evaluate the accuracy of the 2-point linkage analysis, the LOD value was the most important statistical index. In our current study, each SNP has an average number of 388 (92.82%) genotypes with the minimum number of 335 (80%), which provided much more information for any 2-SNPs linkage analysis. We illustrated in a previous study (Mousavi et al. 2016) that a moderate sample size of about 150 or more individuals could be used for constructing parent-specific genetic linkage maps in an F_1 hybrid population of forest trees. In contrast, we used double more sample size at each SNP for linkage mapping, exhibiting high-quality linkage maps constructed with highly significant linked SNPs within linkage groups classified under LOD thresholds up to 55 or 29.

Compared with our previous work on linkage mapping using the same population (Tong et al. 2016), the current linkage maps were improved in 2 main ways, in addition to the strategy for ordering SNPs within linkage groups. First, the number of SNPs extracted to construct the 2 parental linkage maps was increased ~141%, from 2541 to 6115, with an increase in the estimated SNP consistency rate from 98.20% to 99.59% (Tong et al. 2016; Supplementary Table S6). Second, the SNPs in each previous parental map were classified into 20 linkage groups, leaving one group ambiguously assigned to the chromosomes. In contrast, the current linkage maps contained 19 linkage groups, which were supported under a wide range of LOD thresholds, presenting a perfect one-to-one correspondence to the chromosomes. This improvement may be attributed to additional individuals incorporated and the application of a new strategy for extracting SNP genotypes.

Optimal Ordering of SNPs Within Linkage Groups

In genetic mapping, it is crucially important for markers to be ordered correctly within a linkage group. Ordering hundreds or even thousands of markers belongs to a hard scientific problem, known as the traveling salesman problem (TSP) (Wu et al. 2008; Monroe et al. 2017). Linkage maps with erroneous marker orders will negatively affect downstream applications such as QTL mapping and genome scaffold or contig assembly. Therefore, we utilized as many software as possible for ordering the SNPs within linkage groups and then

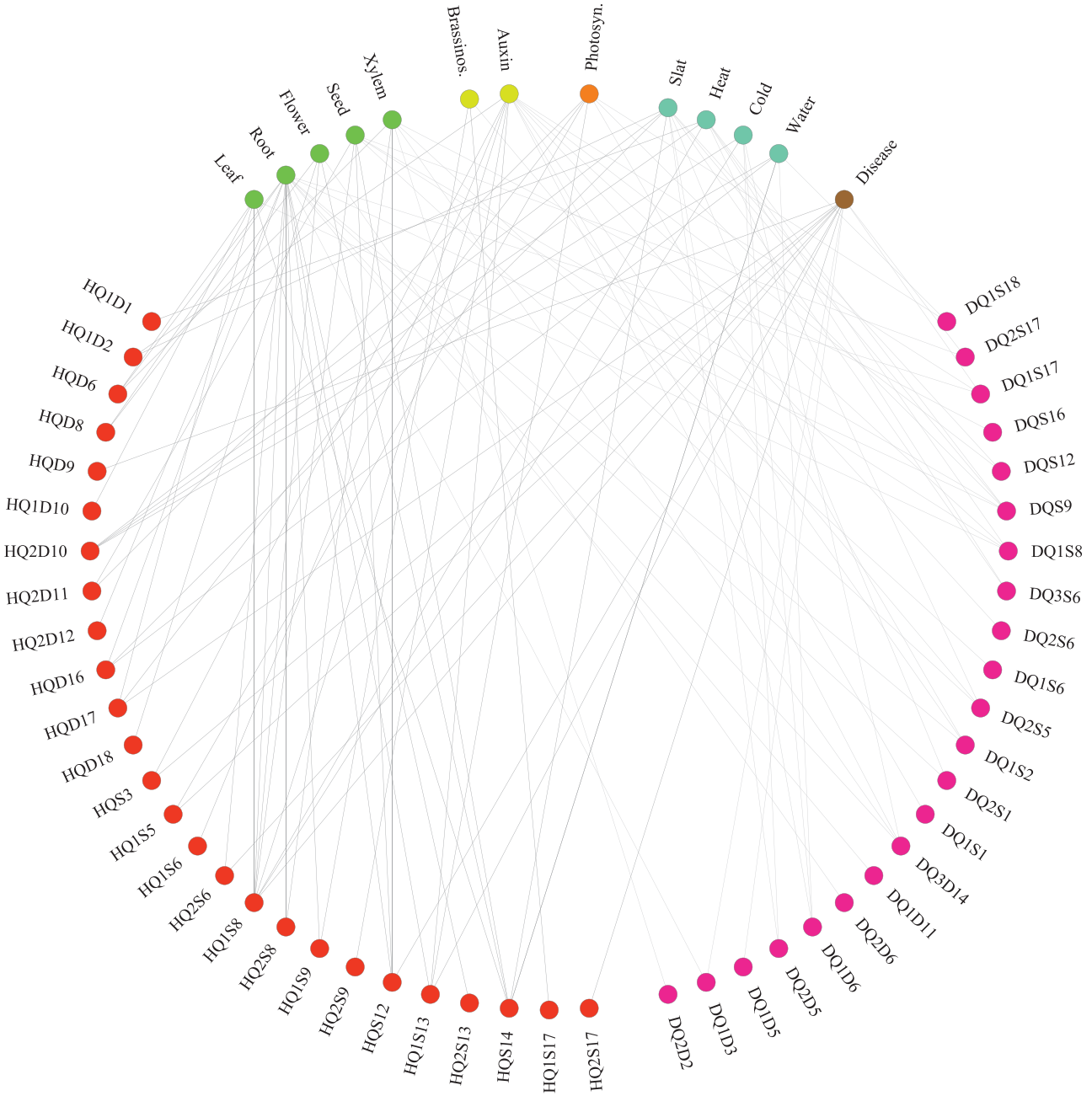

Figure 6. Potential candidate genes of QTLs related to biological functions and processes. The 26 QTLs prefixed with "H" for tree height (H) and 22 QTLs prefixed with "D" for diameter at breast height (DBH) are related to the tree growth and development of leaf, root, flower, seed, and xylem, to stress responses of salt, heat, cold, and water deprivation, to disease resistance, or involved in brassinosteroid, auxin, and photosynthesis.

chose the optimal result for the final maps among multiple ordering results. The software tools currently available for genetic mapping include MapMaker (Lander et al. 1987), JoinMap (Van Ooijen 2006), OneMap (Margarido et al. 2007), MSTmap (Wu et al. 2008), FsLinkageMap (Tong et al. 2010), Lep-Map (Rastas et al. 2013), and TSPmap (Monroe et al. 2017). However, only the JoinMap, OneMap, FsLinkageMap, and Lep-Map software can be applied to the F_1 hybrid population as described in this study, but FsLinkageMap cannot handle a large number of markers. Although JoinMap provides 2 methods for ordering markers (i.e., the regression and maximum likelihood (ML) methods), the regression method takes an intolerably long time to complete for a large dataset (Monroe et al. 2017). Thus, we reasonably used the ML method in JoinMap and the OneMap and Lep-Map software to order SNPs. Because the ML method in JoinMap and the Lep-Map software generate different ordering results for different running times, we ran JoinMap 5 times and Lep-Map 3 times, obtaining 5 and 3 ordering results for each linkage group, respectively. Under the ordering criterion of the SARF (Falk 1989), we found that the optimal ordering results for all the linkage groups came from JoinMap (Supplementary Files 2 and 3). Other criteria such as the sum of adjacent LOD scores (SALOD) and the likelihood for an order of markers can certainly also be applied to ordering markers (Weeks and Lange 1987; Lander and Botstein 1989; Tong et al. 2010). However, different ordering criteria may

Table 4. Summary statistics for anchoring the contigs of *Populus simonii* into chromosomes using the 2 parental linkage maps of *Populus deltoides* and *Populus simonii*

	P. deltoides	*P. simonii*	Anchored	Oriented	Unplaced
Linkage Groups	19	19	19		
SNPs	4004	2089	5980	5740	102
SNPs per Mb	10.1	5.5	15.2	15.7	2.2
N50 Contigs	72	72	72	71	0
Contigs	337	281	322	245	367
Contigs with 1 SNP	75	70	41	0	27
Contigs with 2 SNPs	42	35	39	25	7
Contigs with 3 SNPs	21	23	18	15	6
Contigs with ≥4 SNPs	199	153	224	205	5
Total bases (Percent of genome)	397 177 400 (90.0%)	382 067 890 (86.6%)	394 078 642 (89.3%)	365 757 506 (82.9%)	47 296 709 (10.7%)

As described in Tang et al. (2015), contigs with no SNPs, or ambiguous placements, are separately counted ("Unplaced"). The SNP density for the anchored and unplaced contigs represent the sum of unique SNPs from all input datasets. N50 contigs refer to those equal to or longer than contig N50.

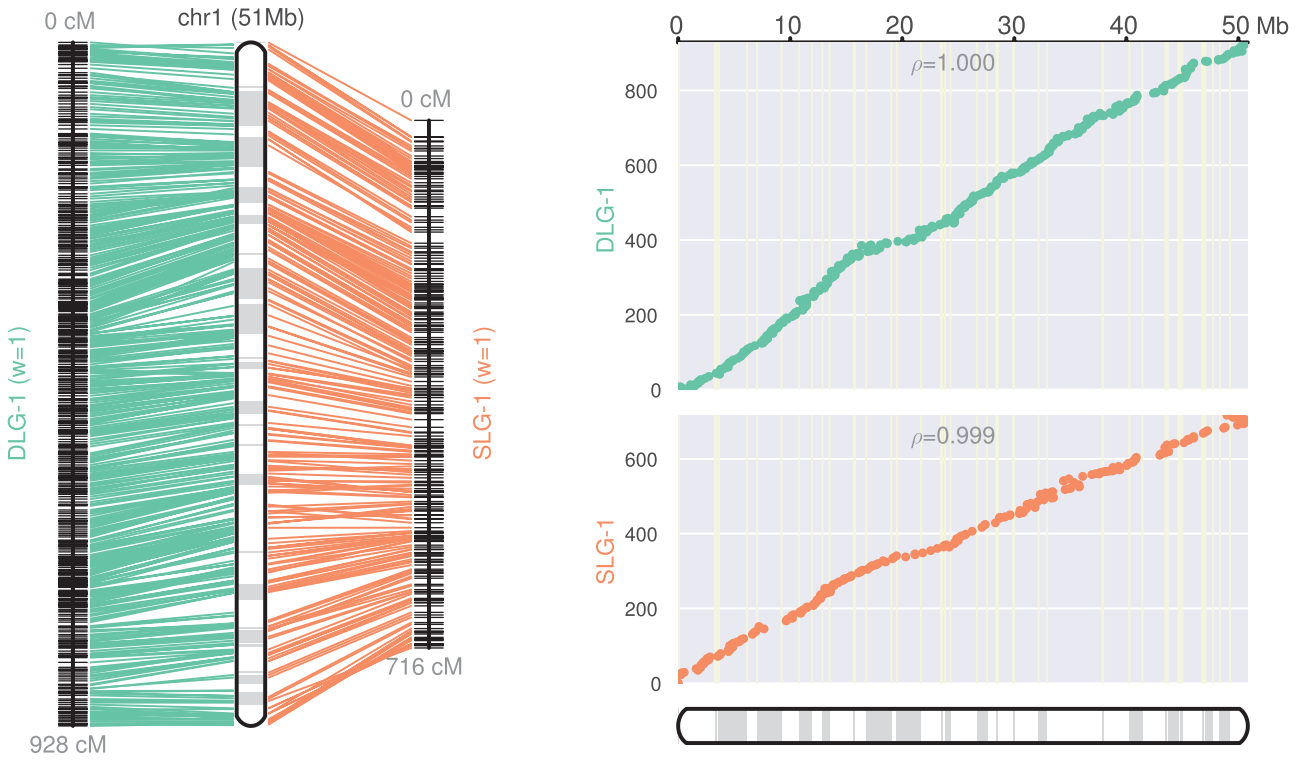

Figure 7. The assembly of chromosome 1 of *Populus simonii* genome. The chromosome was assembled with the 689 contigs and the 2 parental linkage maps of *Populus deltoides* and *P. simonii*. **(A)** Connections between the physical positions on the assembled chromosome and the linkage map positions. **(B)** Scatter plots of the physical position on the chromosome against the genetic position on the 2 linkage maps. Boxes of alternating shades represent the contigs within the assembled chromosome. The ρ-value on each plot is the Pearson correlation coefficient between the SNP positions on the physical and linkage maps.

produce different so-called optimal results, which is truly a frustrated issue worth of further study.

Choice of Statistical Models in QTL Mapping

We performed QTL mapping of the growth traits using the 2 parental linkage maps constructed in this study. One of the main characteristics of the linkage maps was that the linkage phase between adjacent SNPs was not fixed in either coupling or repulsion, unlike genetic maps derived from the traditional backcross (BC) and F_2 populations. The phenotype data were measured from 2013 to 2017, representing time series or longitudinal data. Therefore, the software *mvqtlcim* was applied for QTL mapping, satisfying the characteristics of both the linkage phase and phenotype data in particular. The software was designed to implement the composite mapping (CIM) method (Zeng 1994) for mapping multivariate traits in an F_1 hybrid population, in which the main parameters of WS and BMN need to be specified in practical computing. We applied the strategy of performing QTL analysis by choosing the optimal result from results obtained with different parameter values. However, what is the criterion for an optimal result? In a previous study (Liu et al. 2017), the criterion was defined as the maximum proportion of the variance explained (PVE) by all the detected QTLs. However, the PVE of each

QTL was simply added, and the total was compared over all the mapping results obtained with different parameters. This approach very likely led to the total PVE beyond 100%. In contrast, we estimated the total PVE of all the detected QTLs with model (1) in the present study. This approach is more reasonable because the effective of each QTL was re-estimated under the same statistical model, and the total PVE would never be greater than 100%.

According to a modified version of the infinitesimal model, a quantitative trait is controlled by a few genes with large effects and many genes with small effects (Hu et al. 2012). It is reasonable that we detected 39 QTLs for H and 47 for DBH, of which many have small effects. This demonstrated that our method is more powerful in identifying QTLs compared with previous studies where only a small number of QTLs were detected (Supplementary Tables S10 and S11). Although some QTLs have higher LR values but with smaller PVEs (Tables 2 and 3), this inconsistency can be explained by the fact that the CIM model for detecting a potential QTL at a specific position varies along the genome because the background markers incorporated in the model may be different. This characteristic was inherited from the traditional CIM method (Zeng 1994) and the kind of inconsistency between LRs (or LODs) and PVEs can be found in previous studies (Monclus et al. 2012; Du et al. 2016). Moreover, the PVEs listed in Tables 2 and 3 were not directly resulted from the CIM method with the software *mvqtlcim* but were simultaneously estimated from model (1) as described in Methods.

Improvement of Chromosome-Level Assembly With High-Quality Genetic Maps

Although the 686 contigs were already used to generate chromosome-level sequences in Wu et al. (2020) with the help of the linkage maps constructed in Yao et al. (2020), we re-anchored these contigs into chromosomes with the new high-quality linkage maps constructed here, expecting an improvement in chromosome-level assembly. It can be seen that each of the new anchored chromosome sequences showed a high level of collinearity with each parental linkage map, presenting a correlation coefficient greater than to 0.970 (Supplementary File 8). In contrast, the anchored sequences of chromosomes 4, 5, 6, 8, 9, 13, and 15 in the previous study showed much lower collinearity with the corresponding linkage groups, having absolute values of correlation coefficient less than 0.90 ranging from 0.704 to 0.893 (Supplementary Table S12). Overall, the coefficients in this study were consistently greater than their corresponding values in the previous study, except the 2 pairs highlighted in Supplementary Table S12. Furthermore, the *t*-test showed that significant difference existed between the means of the 2 group correlation coefficient values with a *P*-value of 0.00067. This demonstrated that the 2 newly constructed parental linkage maps were of much higher quality than the maps used in the previous study, especially regarding SNP order within the linkage groups. Compared with the 19 chromosome-sized scaffolds of *P. trichocarpa*, which represented ~92% of the genome assembly, the sequences anchored to the chromosomes exhibited a little lower representation rate (89.3%) in the *P. simonii* genome. However, to improve the representation rate, efforts can be made in 2 regards. The first approach is to increase the density of the linkage maps, which could be achieved by obtaining deeper RADseq data from each individual and then generating more SNP genotype data. On the other hand, novel methods should be proposed or applied for precisely ordering a large number of molecular markers within a linkage group.

Conclusions

We constructed 2 high-quality, high-density parental genetic linkage maps using only RADseq data from the 2 parents and a large number of progeny in the F_1 hybrid population of *P. deltoides* and *P. simonii*. The SNPs in the maps generated with the newly developed software gmRAD were confirmed to be of high quality so that the number of distinguished linkage groups perfectly matched the karyotype of *Populus* under a wide range of LOD thresholds. Moreover, the SNP order within each linkage group was chosen as the optimal order among multiple ordering results from different linkage mapping software platforms. In addition, the linkage maps facilitated the identification of more QTLs underlying growth traits than has been achieved in previous studies, and significantly improved the anchoring of contigs in a draft genome assembly of *P. simonii* to chromosomes. To a large extent, these excellent characteristics indicated that the 2 parental linkage maps of *Populus* are of high quality. The approaches for both extracting and ordering SNPs could be applied to other plant species, especially to other forest trees, for constructing high-density, high-quality genetic maps.

Supplementary Material

Supplementary material is available at *Journal of Heredity* online.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 31870654 and 31270706) awarded to C.T. and the Priority Academic Program Development of Jiangsu Higher Education Institutions. The funding bodies were not involved in the design of the study, collection, analysis, and interpretation of data, and in writing the manuscript.

Conflict of Interest

The authors declare that they have no conflict of interests.

Acknowledgments

We thank Prof. Huogen Li (College of Forestry, Nanjing Forestry University, China) for his great technical assistance in the crossing experiments. We thank Huzhi Xu (Luoning Forest Bureau of Henan Provicne, China), Jiangtao Zhang (Forest Science Research Institute of Henan Province, China), Xiangjin Yan (Siyang Agroforestry Center of Jiangsu Province, China), and Jinhai Yang (Nanjing Qiaolin Forestry Science and Technology Co. Ltd., Nanjing, China) for their great help in collecting flowering branches of *P. simonii* or in our crossing experiments and seedling cultivation.

Author Contributions

C.T. designed the research and wrote the manuscript. D.Y., H.W., Y.C. and W.Z. performed experiments. C.T., D.Y., and W.Y. analyzed data. All authors read and approved the final manuscript.

Data Availability

Supplementary files are available at the Figshare database (https://doi.org/10.6084/m9.figshare.9918515). The RADseq data for all individuals are available under accession numbers SRP052929 and SRP174603 in the NCBI SRA database (http://www.ncbi.nlm.nih.

gov/Traces/sra). The 686 contigs of *P. simonii* genome can be accessed with numbers from VJNQ02000001.1 to VJNQ02000686.1 at the NCBI.

References

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*. 3:e3376.

Boes TK, Strauss SH. 1994. Floral phenology and morphology of black cottonwood, *Populus trichocarpa* (Salicaceae). *Am J Bot*. 81:562–567.

Bradshaw HD Jr, Stettler RF. 1995. Molecular genetics of growth and development in *Populus*. IV. Mapping QTLs with large effects on growth, form, and phenology traits in a forest tree. *Genetics*. 139:963–973.

Bradshaw HD Jr, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF. 1994. Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. *Theor Appl Genet*. 89:167–178.

Chutimanitsakun Y, Nipper RW, Cuesta-Marcos A, Cistué L, Corey A, Filichkina T, Johnson EA, Hayes PM. 2011. Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley. *BMC Genomics*. 12:4.

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.

Doerge RW, Churchill GA. 1996. Permutation tests for multiple loci affecting a quantitative character. *Genetics*. 142:285–294.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochemical Bulletin*. 19:11–15.

Du Q, Gong C, Wang Q, Zhou D, Yang H, Pan W, Li B, Zhang D. 2016. Genetic architecture of growth traits in *Populus* revealed by integrated quantitative trait locus (QTL) analysis and association studies. *New Phytol*. 209:1067–1082.

Dubouzet JG, Strabala TJ, Wagner A. 2013. Potential transgenic routes to increase tree biomass. *Plant Sci*. 212:72–101.

Eckenwalder JE. 1996. Systematics and evolution of *Populus*. In: Stettler RF, Bradshaw HD, Heilman PE, Hinckley TM, editors. *Biology of Populus and its implications for management and conservation*. Ottawa (Canada): NRC Research Press, National Council of Canada, Ottawa. pp. 7–32.

Falk CT. 1989 A simple scheme for preliminary ordering of multiple loci: application to 45 CF families. In: Elston RC, Spence MA, Hodge SE, MacCluer JW, editors. *Multipoint mapping and linkage based upon affected pedigree members, Genetic Workshop 6*. New York: Liss. pp. 17–22.

Fierst JL. 2015. Using linkage maps to correct and scaffold *de novo* genome assemblies: methods, challenges, and computational tools. *Front Genet*. 6:220.

Gerber S, Mariette S, Streiff R, Bodénès C, Kremer A. 2000. Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. *Mol Ecol*. 9:1037–1048.

Hu Z, Wang Z, Xu S. 2012. An infinitesimal model for quantitative trait genomic value prediction. *PLoS One*. 7:e41336.

Kakioka R, Kokita T, Kumada H, Watanabe K, Okuda N. 2013. A RAD-based linkage map and comparative genomics in the gudgeons (genus Gnathopogon, Cyprinidae). *BMC Genomics*. 14:32.

Kosambi DD. 1944. The estimation of map distance from recombination values. *Annals of Eugenics*. 12:172–175.

Krutovsky KV, Troggio M, Brown GR, Jermstad KD, Neale DB. 2004. Comparative mapping in the Pinaceae. *Genetics*. 168:447–461.

Lander ES, Botstein D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*. 121:185–199.

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

Liu F, Tong C, Tao S, Wu J, Chen Y, Yao D, Li H, Shi J. 2017. MVQTLCIM: composite interval mapping of multivariate traits in a hybrid F1 population of outbred species. *BMC Bioinf*. 18:515.

Ma T, Wang J, Zhou G, Yue Z, Hu Q, Chen Y, Liu B, Qiu Q, Wang Z, Zhang J, *et al*. 2013. Genomic insights into salt adaptation in a desert poplar. *Nat Commun*. 4:2797. doi:10.1038/ncomms3797.

Margarido GR, Souza AP, Garcia AA. 2007. OneMap: software for genetic mapping in outcrossing species. *Hereditas*. 144:78–79.

Monclus R, Leplé JC, Bastien C, Bert PF, Villar M, Marron N, Brignolas F, Jorge V. 2012. Integrating genome annotation and QTL position to identify candidate genes for productivity, architecture and water-use efficiency in *Populus* spp. *BMC Plant Biol*. 12:173.

Monroe JG, Allen ZA, Tanger P, Mullen JL, Lovell JT, Moyers BT, Whitley D, McKay JK. 2017. TSPmap, a tool making use of traveling salesperson problem solvers in the efficient and accurate construction of high-density genetic linkage maps. *BioData Min*. 10:38.

Mousavi M, Tong C, Liu F, Tao S, Wu J, Li H, Shi J. 2016. De novo SNP discovery and genetic linkage mapping in poplar using restriction site associated DNA and whole-genome sequencing technologies. *BMC Genomics*. 17:656.

Paolucci I, Gaudet M, Jorge V, Beritognolo I, Terzoli S, Kuzminsky E, Muleo R, Scarascia Mugnozza G, Sabatti M. 2010. Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. *Tree Genet Genomes*. 6: 863–875.

Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One*. 7:e30619.

Rastas P, Paulin L, Hanski I, Lehtonen R, Auvinen P. 2013. Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinformatics*. 29:3128–3134.

Sannigrahi P, Ragauskas AJ, Tuskan GA. 2010. Poplar as a feedstock for biofuels: a review of compositional characteristics. *Biofuels Bioprod Biorefin*. 4:209–226.

Sturtevant AH. 1913. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J Exp Zool*. 14:43–59.

Tang H, Zhang X, Miao C, Zhang J, Ming R, Schnable JC, Schnable PS, Lyons E, Lu J. 2015. ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biol*. 16:3.

Tao S, Wu J, Yao D, Chen Y, Yang W, Tong C. 2018. Identification of recombination events in outbred species with next-generation sequencing data. *BMC Genomics*. 19:398.

Tong C, Li H, Wang Y, Li X, Ou J, Wang D, Xu H, Ma C, Lang X, Liu G, *et al*. 2016. Construction of high-density linkage maps of *Populus deltoides* × *P. simonii* using restriction-site associated DNA sequencing. *PLoS One*. 11:e0150692.

Tong C, Zhang B, Shi J. 2010. A hidden Markov model approach to multilocus linkage analysis in a full-sib family. *Tree Genet Genomes*. 6: 651–662.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, *et al*. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. 313:1596–1604.

Van Ooijen JW. 2006. *JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations*. Wageningen (Netherlands): Kyazma B. V.

Wang Z, Du S, Dayanandan S, Wang D, Zeng Y, Zhang J. 2014. Phylogeny reconstruction and hybrid analysis of *Populus* (Salicaceae) based on nucleotide sequences of multiple single-copy nuclear genes and plastid fragments. *PLoS One*. 9:e103645.

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

Wang Y, Zhang B, Sun X, Tan B, Xu LA, Huang M, Wang M. 2011. Comparative genome mapping among *Populus adenopoda*, *P. alba*, *P. deltoides*, *P. euramericana* and *P. trichocarpa*. *Genes Genet Syst*. 86:257–268.

Weeks DE, Lange K. 1987. Preliminary ranking procedures for multilocus ordering. *Genomics*. 1:236–242.

Woolbright SA, Difazio SP, Yin T, Martinsen GD, Zhang X, Allan GJ, Whitham TG, Keim P. 2008. A dense linkage map of hybrid cottonwood (*Populus fremontii* x *P. angustifolia*) contributes to long-term ecological research and comparison mapping in a model forest tree. *Heredity (Edinb)*. 100:59–70.

Wu RL. 1998. Genetic mapping of QTLs affecting tree growth and architecture in *Populus*: implication for ideotype breeding. *Theor Appl Genet*. 96:447–457.

Wu Y, Bhat PR, Close TJ, Lonardi S. 2008. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet*. 4:e1000212.

Wu H, Yao D, Chen Y, Yang W, Zhao W, Gao H, Tong C. 2020. De novo genome assembly of *Populus simonii* further supports that *Populus simonii* and *Populus trichocarpa* belong to different sections. *Genes Genomes Genetics*. 10:455–466.

Yao D, Wu H, Chen Y, Yang W, Gao H, Tong C. 2020. gmRAD: an integrated SNP calling pipeline for genetic mapping with RADseq across a hybrid population. *Brief. Bioinformatics*. 21:329–337. doi: 10.1093/bib/bby114.

Yin T, Zhang X, Huang M, Wang M, Zhuge Q, Tu S, Zhu LH, Wu R. 2002. Molecular linkage maps of the *Populus* genome. *Genome*. 45:541–555.

Zeng ZB. 1994. Precision mapping of quantitative trait loci. *Genetics*. 136:1457–1468.

Zhang B, Tong C, Yin T, Zhang X, Zhuge Q, Huang M, Wang M, Wu R. 2009. Detection of quantitative trait loci influencing growth trajectories of adventitious roots in *Populus* using functional mapping. *Tree Genet Genomes*. 5:539–552.

Zhang D, Zhang Z, Yang K, Li B. 2004. Genetic mapping in (*Populus tomentosa* x *Populus bolleana*) and *P. tomentosa* Carr. using AFLP markers. *Theor Appl Genet*. 108:657–662.

Zhigunov AV, Ulianich PS, Lebedeva MV, Chang PL, Nuzhdin SV, Potokina EK. 2017. Development of F1 hybrid population and the high-density linkage map for European aspen (*Populus tremula* L.) using RADseq technology. *BMC Plant Biol*. 17:180.

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.