Tree Genetics & Genomes (2018) 14:79
https://doi.org/10.1007/s11295-018-1290-9

ORIGINAL ARTICLE

High-density genetic map of *Populus deltoides* constructed by using specific length amplified fragment sequencing

Lecheng Fang[1] · Hailin Liu[1] · Suyun Wei[1] · Ken Keefover-Ring[2] · Tongming Yin[1]

Received: 29 March 2018 / Revised: 12 September 2018 / Accepted: 14 September 2018

Abstract

Populus deltoides is an important industrial tree species widely planted in many areas of the world, and de novo genome sequencing of this plant has been carried out by several research groups. A dense genetic map associating genome sequences is highly desirable for reconstructing the chromosome pseudomolecules using the obtained sequence scaffold assemblies. For this purpose, an intra-specific full-sib F_1 mapping pedigree was established by using the sequenced *P. deltoides* as the maternal parent. With this mapping pedigree, we constructed a high-density genetic map using 92 randomly selected progenies. Single nucleotide polymorphism (SNP) markers were discovered by using specific length amplified fragment sequencing (SLAF-seq). In total, 487,038 SLAFs were generated, of which 179,360 were polymorphic. A high-density genetic map was built using HighMap software, which included 11,680 Mendelian segregation markers distributing in 2851 marker bins. The established map consisted of 19 linkage groups (LGs) that equaled to the 19 haploid chromosomes in poplar genome, and spanned a total genetic length of 3494.66 cM, with an average distance of 1.23 cM per marker bin. The map presented here will be useful for anchoring the genome sequence assemblies along chromosomes, and for many other aspects of genetic studies on *P. deltoides* and the closely related species.

Keywords *Populus deltoides* · Specific length amplified fragment sequencing · Single nucleotide polymorphism · Linkage map

Introduction

Poplars are ecologically and economically important woody plants that are widely distributed in the northern hemisphere. Their fast growth rate makes them desirable tree species for short rotation forestry. The genus *Populus* possesses many characteristics conducive to functional genomic studies, and poplars are widely used as the model system in tree genome research (Bradshaw et al. 2000; Wullschleger 2002; Tuskan et al. 2004). The genus *Populus* consists of six subgenera, including *Abaso*, *Leuce*, *Leucoides*, *Aigeiros*, *Turanga*, and *Tacamahaca* (Eckenwalder 1996). *P. trichocarpa*, a species in the *Tacamahaca* subgenus, was the first sequenced woody plant (Tuskan et al. 2006). More recently, the genome of *P. euphratica* and *P. pruinosa*, the desert poplars belonging to *Turanga* subgenus, has also been sequenced (Ma et al. 2013; Yang et al. 2017). The availability of sequenced poplar genomes offers many possibilities to study the biological characteristics specific to perennial woody plants, which cannot be easily addressed in annual herbaceous species.

P. deltoides belongs to the *Aigeiros* subgenus in the genus *Populus*. It is a cottonwood poplar growing throughout the eastern, central, and southwestern USA, the southernmost part of eastern Canada, and northeastern Mexico (Kartesz and Meacham 1999). *P. deltoides* is one of the fastest growing hardwood trees in North America and has been introduced to many other regions of the world (Wan et al. 2015). This fast-growing tree species has long been cultivated as superior resource for timber and pulp (Eckenwalder 1977). To accelerate the domestication of *P. deltoides*, a variety of genetic and genomic resources need to be developed. Genetic maps constructed with DNA-based markers provide a key step toward unraveling genomic

Lecheng Fang and Hailin Liu contributed equally to this work.

Communicated by P. Ingvarsson

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11295-018-1290-9) contains supplementary material, which is available to authorized users.

✉ Tongming Yin
tmyin@njfu.com.cn

[1] The Southern Modern Forestry Collaborative Innovation Center, College of Forestry, Nanjing Forestry University, No. 159, Longpan Road, Nanjing 210037, China

[2] Departments of Botany and Geography, University of Wisconsin – Madison, Madison, WI 53706, USA

Published online: 03 October 2018

structure, which will lead to many promising applications in tree improvement (Fretwell 1996; Dinus et al. 2001). Previously, several linkage maps for *Populus* were constructed with anonymous markers, mainly with RAPDs and AFLPs (Bradshaw et al. 1994; Wu et al. 2000; Yin et al. 2002; Zhang et al. 2000; Gaudet et al. 2008). To reconcile genetic maps built for poplars, intensive efforts were made toward establishing genetic maps with sequence-based markers, like STSs and SSRs (Cervera et al. 2001; Yin et al. 2004, 2008). Sequence-based markers were not only desirable in comparative mapping, but also useful in mapping the sequence scaffolds produced from *Populus* genome sequencing projects along the poplar chromosomes. However, low-throughput techniques for genotyping sequence-based markers is laborious and time intensive, and thus only good for establishing a genetic map with limited density. Recent progress in next-generation sequencing technologies now allows the simultaneous generation of large amounts of sequence-based markers, which are highly desirable for genetic map construction due to their abundance, uniform distribution, and association with genome sequences (Ganal et al. 2009). Restriction site-associated DNA sequencing (Chutimanitsakun et al. 2011) is one of several high-throughput techniques that have promoted large-scale marker production. Specific length amplified fragment (SLAF) sequencing, a new marker generation technique, also has been developed as a high-resolution strategy for de novo SNP discovery with good repeatability (Sun et al. 2013). This technique has been widely adopted for genetic map construction in many plant species (e.g., Zhang et al. 2013; Qi et al. 2014; Xu et al. 2015).

Besides *P. trichocarpa* (Tuskan et al. 2006), *P. euphratica* (Ma et al. 2013), and *P. pruinosa* (Yang et al. 2017), de novo genome sequencing of other poplar species, including *P. deltoides*, is being carried out by several research groups. High-density genetic maps constructed with sequence-based markers are in demand for reconstruction of the poplar chromosome structure and for comparison of poplar genomes. In this study, we report a high-density genetic map for *P. deltoides* built with SNP markers produced by SLAF-seq. To our knowledge, it is the densest genetic map for poplars established with sequence-based markers thus far. Besides application in poplar genome assembly and comparison, the mapped marker sequences also provide a useful genomic resource for quantitative trait locus (QTL) dissection and map-based cloning.

Materials and methods

Plant materials

The mapping population used in this study is an intraspecific full-sib F_1 pedigree established in 2013. The maternal parent of this pedigree is the individual selected for sequencing. Both parents of this pedigree are clones selected from germplasm collections of *P. deltoides*, which were imported from TX, USA, in 1994. In total, 1436 progenies were obtained for these mapping pedigrees and were maintained at Chenwei Forest Farm in Sihong county, Jiangsu province of China. For SLAF sequencing, we genotyped the mapping parents and 92 randomly selected progenies. In spring of 2015, young leaf tissues were collected and DNA extracted with the CTAB protocol (Doyle 1987), and then stored at − 80 °C until needed.

SLAF marker generation

Electronic enzyme digestion analysis was performed with the SLAF-predict software (Biomarker) by using the genome sequences of *P. trichocarpa* as reference (Tuskan et al. 2006). Restriction enzymes HaeIII and Hpy166II were selected for SLAF library construction, which predicted over 200 thousand SLAF tags in the reference genome. *Oryza sativa* was chosen as a control to monitor the quality of the SLAF library. Library construction was conducted following the protocol described by Sun et al. (2013) with minor modifications. First, 0.5 μg genomic DNA of each progeny was digested with 5 U HaeIII and 5 U Hpy166II at 37 °C for 6 h, then the mixture was incubated at 65 °C for 20 min with T4 DNA ligase, ATP, HaeIII-adapter, and Hpy166II-adapter, followed by incubation at 20 °C for 12 h. Subsequently, polymerase chain reaction (PCR) was performed with primer pairs each having one selective nucleotide, and products were purified with E.Z.N.A. Cycle Pure kit (Omega, Lilburn, GA, USA). Purified products were pooled and incubated at 37 °C with HaeIII, Hpy166II, T4 DNA ligase, ATP, and Solexa adapter (Illumina, San Diego, CA, USA). After incubation, the pooled samples were purified with a Quick Spin column (Qiagen Inc., Valencia, CA), then electrophoresed on 2% agarose gel. After electrophoresis, DNA fragments in the range of 500–550 bp (with indices and adaptors) were isolated with the gel extraction kit (Qiagen Inc., Valencia, CA). The isolated fragments were amplified with Phusion Master Mix (NEB, Ipswich, MA, USA) and Solexa Amplification Primer Mix (Illumina, San Diego, CA, USA) to add barcodes. Finally, SLAFs in range of 100–450 bp were isolated and diluted for paired-end sequencing on the Illumina-HiSeq 2500 platform (Illumina, San Diego, CA, USA) by Beijing Biomarker Technologies Corporation (http://www.biomarker.com.cn).

The raw data were processed as follows: reads containing 10% uncalled bases were discarded, and adapter sequences were filtered out; remaining reads were evaluated with NGS QC toolkit (Patel and Jain 2012), and if base quality scores in each read were greater than 30, sequences were retained for generating SLAF markers in the mapping pedigree based on dual-index. All clear SLAF reads were then clustered by BLAT (Kent 2002). Sequences with over 90% similarity were

grouped in one SLAF locus. Alleles in each locus were subsequently defined by using the minor allele frequency (MAF) evaluation. SLAF tags with markedly lower MAF values were corrected to the most similar genotype to improve data accuracy. According to parents' genotype, SLAF tags segregated in the mapping pedigree were classified into five segregation patterns, and these markers were further examined by chi-square test with P value < 0.05.

Map construction

For map construction, we screened the SLAFs with average depth > 10, containing less than three SNPs, and followed Mendel's law of segregation in the mapping pedigree. Linkage maps were constructed with HighMap software (Liu et al. 2014). Following a pseudo-testcross strategy (Sun et al. 2017), the markers heterozygous in one parent, but homozygous in the alternate parent were used to construct the sex-specific maps for each parent. Subsequently, the consensus map was established by integrating the parent's maps by anchoring markers that were heterozygous in both parents (Van Ooijen 2011; Guo et al. 2018). The SLAFs were assigned into different LGs based on their pairwise modified logarithm of odds (MLOD) score. The ordering module combines Gibbs sampling, spatial sampling, and simulated annealing algorithm to order markers and estimate map distances. Map distances were calculated with the Kosambi mapping function (Kosambi 1944). The error correction module identified singletons according to parental contribution of genotypes and eliminated them from the data using k-nearest neighbor algorithm. To order markers correctly, the processes of ordering and error correction were carried out iteratively.

Besides the observed genetic length, we also calculated the estimated genetic length of each LG by using the formula in Hulbert et al. (1988), based on partial linkage data: $L = m(m-1)X/K$, where m is the number of markers used for estimation; X is the map distance corresponding to the LOD threshold Z for declaring linkage; and K is the number of marker pairs having LOD values at or above Z. In this estimation, we randomly resampled 100 markers from each LG without replacement and performed the calculations for 1000 iterations. The mean value of the 1000 iterations was taken as the expected genetic length of the corresponding LG. Based on the standard deviation of the estimations, we also calculated the variation range of estimated genetic length for each LG.

The correspondence of each LG in *P. deltoides* to the chromosome in *P. trichocarpa* was determined based on alignment of mapped markers (Tuskan et al. 2006) by using BLAST. And then the Pearson correlation between linkage groups and the corresponding chromosomes was calculated using the R script. Haplotype maps and heat maps were used to evaluate the quality of the map following the description by West et al. (2006).

Results

Analysis of SLAF-seq data and SLAF markers

We obtained approximately 115 Gb of sequencing data containing 890.71 million reads with read lengths of 100 bp. The average Q30 ratio was 88.45% and the average guanine-cytosine (GC) content was 39.55% (Table 1).

After removing low-quality reads, a total of 487,038 SLAFs were recovered, varying from 299,908 to 398,708 among the progeny (Fig. 1). On average, the sequencing depth of these SLAFs is 25.63× and 12.33× for the mapping parents and for the progenies, respectively, ranging from 8.76× to 17.20× (Fig. 1). Among the identified SLAFs, 179,360 were polymorphic with a polymorphic ratio of 36.83% (Table 2).

Genetic linkage maps

To ensure the quality of the genetic map, SLAFs were selected following the aforementioned criteria described in the "Materials and methods" section. Only markers that could be ordered with high confidence were included for constructing the genetic map. Examined by chi-square test with P value < 0.05, Mendelian segregation SLAF markers were selected for map construction, which were classified into five segregation categories (ab×cd, ef×eg, hk×hk, nn×np, lm×ll) (Fig. 2, Supplementary Table S1). The mean sequencing depth of the mapped SLAFs was 29.71× (Table 1), and the number of SLAFs in each LG ranged from 415 to 833 (Table 3). The established consensus map consisted of 2851 marker bins and included 11,680 markers that covered a genetic distance of 3494.66 cM, with LG sizes ranging from 138 to 230 cM (Supplementary Table S2, S3). The average distance between marker bins was 1.23 cM (Fig. 3 and Supplementary Figure S1). For the sex-specific maps, the maternal map contained 4962 markers with a total genetic distance of 2228 cM, and the paternal map contained 4499 markers covering a total genetic length of 2599 cM (Supplementary Table S4).

Marker distribution was analyzed by counting the number of loci and all mapped markers (including bins) by sliding 10-cM

Table 1 Summary of SLAF-seq results

Total number of reads (M)	890.71
Q30 percentage (%)	88.45
GC percentage (%)	39.55
Total SLAFs	487,038
Mean raw reads sequencing depth of offsprings	20.10×
Mapped markers	11,680
Mean sequencing depth of SLAFs for the parents	25.63×
Mean sequencing depth of SLAFs for the progeny	12.33×
Mean sequencing depth of the SLAFs on the map	29.71×

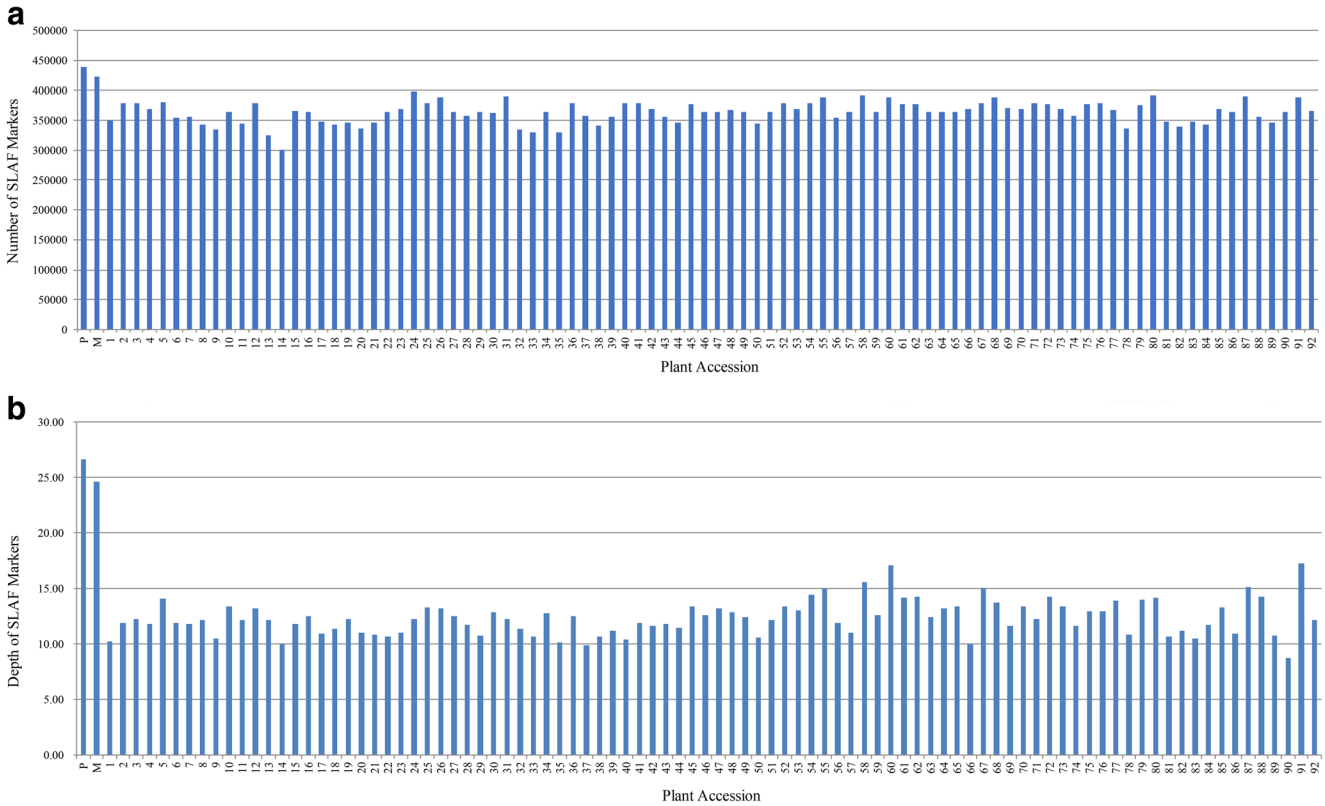

Fig. 1 Number of markers and average sequencing depth of each progeny. The *x*-axes in **a** and **b** indicate the identity of the mapping samples, and the *y*-axes indicate the number of markers (**a**) and average depth (**b**)

windows along each LG (Fig. 4). The region with the highest marker density, on LG14, contained 162 SLAFs within a 10-cM interval. According to the 10-cM sliding window analysis, the average number of marker bins ranged from 2 to 14, with the average number of SLAFs varied from 4 to 162. The mapped markers were distributed relatively uniformly along each LG, and in our linkage map, "Gap ≤ 5 cM" ranging from 97.82 to 100% with an average value of 99.12%. Twenty-three gaps larger than 5 cM were detected on 15 LGs, of which the largest gap spanned a genetic length of 8.85 cM on LG1.

We estimated the genetic length of each LG by the method-of-moment estimator under $Z = 9$ (Hulbert et al. 1988), which is the LOD support to assign markers into different LGs in map construction. The mean total length of the estimations was $L = 4140$ cM (Table 4), which was 645 cM larger than the total observed length. For each LG, the observed genetic length was in the variation range of the estimated length.

Table 2 Classification of SLAF markers

Type	Polymorphic SLAFs	Non-polymorphic SLAFs	Total SLAFs
Number	179,360	307,678	487,038
Percentage (%)	36.83	63.17	100

The Pearson correlation was used to evaluate the marker synteny between the linkage groups of *P. deltoides* and the chromosomes of *P. trichocarpa*. The Pearson correlation coefficients for the corresponding linkage groups and chromosomes were in range of 0.93 to 1.00 (Fig. 5, Table 5, Supplementary Table S5). In conclusion, high collinearity was revealed for the mapped markers on the established genetic map and in the poplar genome assemblies.

Evaluation of the genetic map

Double crossovers are closely associated with genotyping errors (Qi et al. 2014) and could be revealed by a haplotype map within short genetic distances. In this study, haplotype maps of the genetic map were constructed for each or the 92 progeny with SLAF markers (Supplementary Figure S2). Recombination events of each individual could be clearly visualized with the haplotype maps. These maps indicated that the ratio of double crossovers was about 0.48%, while the percentage of missing data on each LG was less than 0.15% (Table 6).

Heat mapping is an effective way to evaluate the quality of a genetic map and is derived from the pairwise recombination values of the mapped SLAFs (Supplementary Figure S3). Among the generated heat maps, an obvious trend was observed: the pairwise linkage becomes looser with an increase

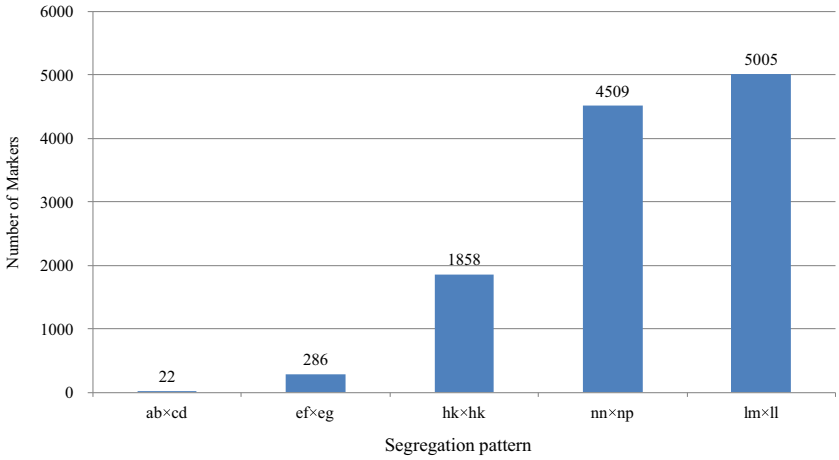

Fig. 2 Number of markers for five segregation patterns

in genetic distance between mapped markers. This pattern indicated that the markers in each LG were well ordered. In conclusion, we obtained a high-quality linkage map constructed with dense sequence-based markers for *P. deltoides*.

Discussion

The genus *Populus* is a member of the Salicaceae family that has a wide natural distribution in the Northern Hemisphere and a small representation in tropical Africa (Bradshaw et al. 2000). *Populus* is widely used as a model system for characterizing the unique biological process of perennial woody plants due to its modest genome size, fast growth, and ease of vegetative propagation (Wullschleger 2002; Jansson and Douglas 2007). With the most recent classifications, 29 poplar species in six separate subgenera are recognized (Eckenwalder 1996). Analyzing the genome sequence of *P. trichocarpa* suggests that *Populus* originates from a common paleotetraploid ancestor (Tuskan et al. 2006), which provides a desirable system to study the genomic evolution after the paleopolyploidization event among closely related species.

Table 3 Basic information of consensus linkage map

Linkage group ID	Total marker (bin)	Total distance (cM)	Average distance (cM)	Max gap (cM)	Gap ≤ 5 cM (%)
LG1	833 (166)	226.63	0.27	8.85	98.8
LG2	482 (152)	199.19	0.41	6.40	99.35
LG3	537 (162)	179.32	0.33	3.99	100
LG4	415 (105)	141.97	0.34	6.13	99.06
LG5	426 (116)	155.83	0.37	5.75	98.29
LG6	531 (151)	202.95	0.38	8.79	98.68
LG7	498 (137)	177.91	0.36	5.66	97.82
LG8	771 (156)	230.23	0.30	6.99	98.09
LG9	691 (179)	190.88	0.28	7.19	99.44
LG10	637 (127)	178.99	0.28	8.90	99.22
LG11	725 (190)	199.12	0.28	3.72	100
LG12	815 (164)	191.88	0.24	4.68	100
LG13	578 (177)	198.56	0.34	6.84	98.88
LG14	602 (121)	158.83	0.26	7.24	99.18
LG15	720 (164)	187.03	0.26	6.30	99.39
LG16	724 (177)	194.52	0.27	5.29	99.43
LG17	642 (115)	138.72	0.22	8.67	99.14
LG18	523 (159)	172.95	0.33	4.68	100
LG19	530 (133)	169.15	0.32	8.33	98.5
Total	11,680 (2851)	3494.66	1.23	8.85	–

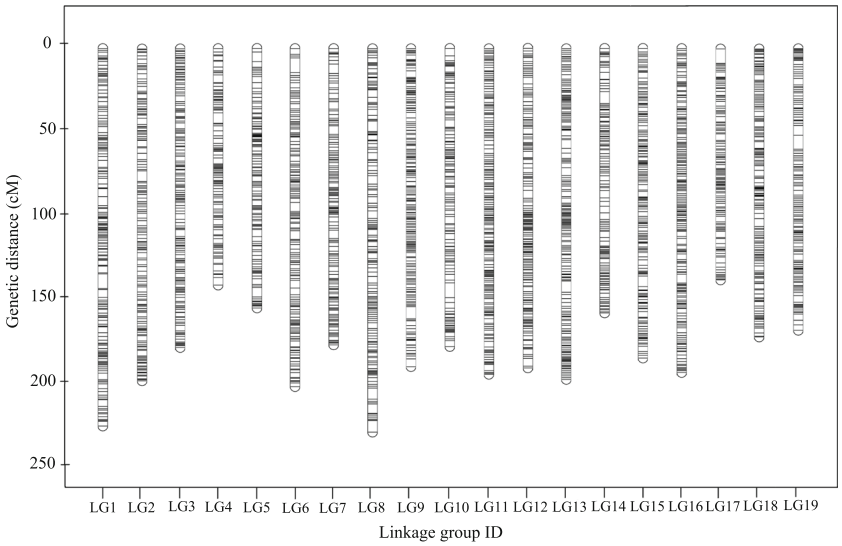

Fig. 3 Distribution of SLAFs along each of the LG in linkage map of *Populus deltoides*. Note: A black line indicates a SLAF marker

Thus far, the genomes of *P. trichocarpa* (species in *Tacamahaca* subgenus, Tuskan et al. 2006), *P. euphratica*, and *P. pruinosa* (species in *Turanga* subgenera, Ma et al. 2013; Yang et al. 2017) have been sequenced, but only the sequence scaffolds of *P. trichocarpa* have been mapped along different chromosomes of the poplar genome. Syntenic relationships can provide unambiguous identification of orthologs and extend genetic maps established in one species to related species (Stirling et al. 2003; Brunner et al. 2004), which is expected to facilitate the identification and utilization of beneficial genes, alleles, and QTL across related taxa (Yin et al. 2001). Although genetic maps have revealed high levels of marker synteny among different poplar species, detailed genomic variation among them needs to be explored at the sequence level. *P. deltoides* and its hybrids are the most important commercial poplars planted in artificial plantations worldwide, and it is a member in *Aigeiros* subgenus. No genome has been sequenced for poplar species in this subgenus, thus far.

Fig. 4 Distribution of marker bins and SLAFs in the 10-cM sliding windows among each LG. The number of marker bins and the number of mapped SLAFs were indicated with light and dark blue bars, respectively

Table 4 Genetic length estimated by the adjusted Hulbert et al.'s method-of-moment estimator

LG	Mean estimated genetic length (cM)	Coverage	SD (cM)	Range (cM)	Observed genetic length (cM)
LG1	250.32	1	18.69	206.32–289.55	226.63
LG2	240.19	0.999987	20.35	174.53–265.07	199.19
LG3	226.09	0.999998	19.21	171.73–253.73	179.32
LG4	185.95	1	16.41	140.29–203.82	141.97
LG5	204.93	1	20.61	139.97–225.39	155.83
LG6	258.45	0.999982	24.73	187.39–289.38	202.95
LG7	229.47	1	15.91	150.78–219.25	177.91
LG8	268.95	0.999995	21.30	210.76–302.28	230.23
LG9	193.87	1	11.55	167.21–220.16	190.88
LG10	212.11	1	19.10	157.61–253.41	178.99
LG11	200.61	1	19.44	150.13–255.52	199.12
LG12	235.05	1	16.34	166.09–245.55	191.88
LG13	236.92	1	17.58	192.89–273.80	198.56
LG14	198.32	1	16.53	156.37–248.12	158.83
LG15	215.69	1	20.86	167.02–261.14	187.03
LG16	237.92	1	17.68	178.72–268.00	194.52
LG17	167.70	1	14.98	125.37–222.06	138.72
LG18	186.74	1	19.75	123.47–235.62	172.95
LG19	190.99	1	14.95	146.56–228.30	169.15
Total	4140.26	–	–	3113.21–4760.15	3494.66

Extensive collaboration between geneticists, physiologists, and pathologists has set a solid scientific foundation for joint efforts on poplar studies in the future. Sequencing and anchoring the sequence scaffolds along chromosomes of *P. deltoides* will enrich the international collection of *Populus* genomic resources, and be useful for comparative and functional genomic studies for the *Populus* research community. Establishment of high-density genetic map with sequence-based markers provides a basic platform for reconstructing the pseudomolecule of poplar genome and for reconciling the genetic information obtained by different research groups.

In this study, we establish a high-density genetic map for *P. deltoides* that comprises 19 LGs, equal to the haploid chromosome number of *Populus*. Comparing to the previous

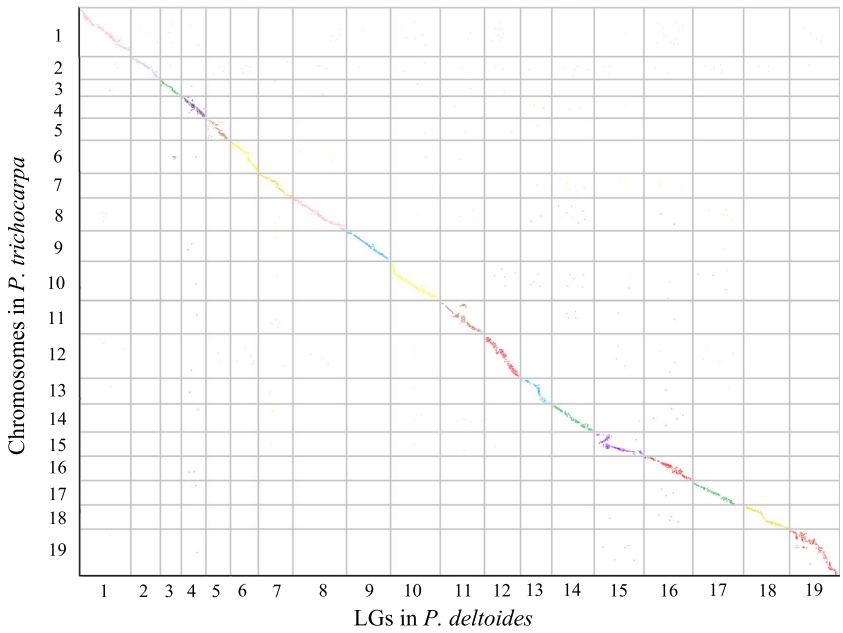

Fig. 5 Syntenic analysis for the mapped markers between the linkage groups of *P. deltoides* and the chromosomes of *P. trichocarpa*. Note: Synteny of the mapped markers on different LGs to the corresponding chromosome were separately showed in different colors

Table 5 The marker synteny between the linkage groups of *P. deltoides* and the chromosomes of *P. trichocarpa*

Linkage group of *P. deltoides*	Chromosome of *P. trichocarpa*	Pearson correlation
LG1	chr1	0.97
LG2	chr2	0.99
LG3	chr3	0.99
LG4	chr4	0.97
LG5	chr5	0.93
LG6	chr6	0.99
LG7	chr7	0.98
LG8	chr8	0.99
LG9	chr9	0.99
LG10	chr10	0.95
LG11	chr11	0.98
LG12	chr12	0.96
LG13	chr13	0.99
LG14	chr14	0.97
LG15	chr15	0.99
LG16	chr16	0.98
LG17	chr17	0.99
LG18	chr18	1.00
LG19	chr19	0.97

publications, it was the same with Mousavi et al. (2016), but better than Tong et al. (2016) that contained 20 linkage groups.

Table 6 Percentages of double crossovers and missing data in each LG

LG ID	Percentage of double crossovers (%)	Percentage of missing data (%)
LG1	0.00	0.00
LG2	0.20	0.00
LG3	0.00	0.00
LG4	0.24	0.00
LG5	0.00	0.00
LG6	0.38	0.00
LG7	0.20	0.09
LG8	0.00	0.00
LG9	0.14	0.04
LG10	0.31	0.00
LG11	0.00	0.02
LG12	0.24	0.00
LG13	0.17	0.15
LG14	0.00	0.00
LG15	0.14	0.03
LG16	0.00	0.02
LG17	0.00	0.00
LG18	0.00	0.03
LG19	0.57	0.04

The total observed genetic length was 3494.66 cM, which was shorter than the other two studies (Mousavi et al. 2016; Tong et al. 2016), in contrast to the mean estimated genetic length of 4140 cM. The resampled markers achieved nearly complete coverage of each LG (Table 4). Therefore, the big discrepancy between these two parameters should not be due to the difference in map coverage. For a particular mapping pedigree and a particular set of progeny, the genetic distance should be a constant when markers fully cover the genome. Genetic length should not increase as more markers are added in the map. However, it is a common scenario that genetic length would expand when more markers are added, even after full-length coverage has been achieved. Overestimate of the genetic length could be largely caused by genotyping errors in the marker data. It has been shown, for instance, that a 3% error rate in genotyping data can double the genetic map length (Brzustowicz et al. 1993). Double crossovers detected within short genetic distances are commonly caused by genotyping errors. In our mapping data, we detected a low percentage of double crossovers on the haplotype map of each progeny. Although this percentage is relatively low (Table 5), it could have a significant effect on the observed map length. In HighMap, there is an error correction module to eliminate the putative genotyping errors from the data. Error elimination might be the main underlying cause for the smaller observed map length than the mean estimated genetic length.

Error rates vary among different markers, therefore, the estimated genetic length would vary when different subsets of markers are resampled. In this study, the estimated genetic lengths vary from 3113.21 to 4760.15 cM (Table 4). In previous mapping studies of poplars, the map lengths of genetic maps that achieved nearly complete coverage varied from 1500 to 2800 cM (e.g., Bradshaw and Stettler 1993; Gaudet et al. 2008; Pakull et al. 2009), and these maps commonly contain less than 500 markers. The discrepancies in the observed genetic lengths can be partially interpreted by different error rates. Besides, the choice of map function, the linkage criteria, and the differences in recombination events in the mapping pedigree could also be underlying causes (Plomion and O'Malley 1996; Echt and Nelson 1997; Remington et al. 1999).

Compared to the expansion in genetic length, reliability of marker order is more important in anchoring the sequence scaffolds along poplar chromosomes. Examining the double crossovers and values based on the heat maps showed that we obtained a high-quality linkage map of *P. deltoides*. The relationship for the number of observed chiasma (μ) between loci in a haploid cell, recombination fraction (r), and genetic distance (d) can be derived based on the following formula: $\mu = -\ln(1\text{-}2r)$, and $d = -\ln(1\text{-}2r)/2$ (Haldane 1919). Based on the estimated genetic lengths, the expected chiasma in a haploid cell per meiosis of the mapping parents would be from 31.03 to 47.52. Compared to chiasma per meiosis in conifers (Remington et al. 1999; Yin et al. 2003; Lind et al. 2014), recombination occurs much more frequently in poplars. The

size of the poplar genome is about 485 ± 10 Mb (Tuskan et al. 2006), thus 1-cM genetic length contains less than 200-kb genome sequences on average. With the development of third-generation sequencing technologies and scaffolding-enhancing techniques, like 10X Genomics (Eisenstein 2015), Hi-C (Burton et al. 2013), and BioNano (Pendleton et al. 2015), scaffold N50 of de novo sequence assemblies can easily achieve up to several megabases. Therefore, the resolution of the established map is sufficient in anchoring the sequence assemblies of *P. deltoides* in practice. Meanwhile, with the availability of more reference genomes and the increasing resources of mapped sequence-based markers, the cloning of genes underlying important traits based on mapping studies can be greatly accelerated in poplars.

Conclusion

We established a mapping pedigree by using a sequenced *P. deltoides* as the maternal parent. To facilitate the de novo genome sequence assembly of *P. deltoides*, a dense genetic map was built with this mapping pedigree using sequence-based markers. The resulting map contains 19 LGs that correspond to the number of haploid chromosomes in poplars. Evaluations based on examination of double crossovers and heat maps visualization indicate the obtained map is of high quality. This map provides a robust platform to facilitate building the pseudomolecules of poplar chromosomes, and also will be useful for comparative and functional genomic studies on *P. deltoides* and its closely related species.

Data archiving statement The raw data of SLAF-seq have been submitted to NCBI SRA database under the BioProject ID number PRJNA432808.

Funding information The work is supported by the National Key Research and Development Plan of China (2016YFD0600101) and the National Natural Science Foundation of China (31500533 and 31570662). It was also made possible by research funded by the Doctorate Fellowship Foundation and the PAPD (Priority Academic Program Development) program at Nanjing Forestry University and the Doctorate Fellowship Foundation of Nanjing Forestry University.

References

Bradshaw HD, Stettler RF (1993) Molecular genetics of growth and development in *Populus*. I. Triploidy in hybrid poplars. Theor Appl Genet 86:301–307

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. TAG Theor Appl Genet 89:167

Bradshaw HD, Ceulemans R, Davis J, Stettler R (2000) Emerging model systems in plant biology: poplar (*Populus*) as a model forest tree. J Plant Growth Regul 19:306–313

Brunner AM, Busov VB, Strauss SH (2004) Poplar genome sequence: functional genomics in an ecologically dominant plant species. Trends Plant Sci 9(1):49–56

Brzustowicz LM, Merette C, Xie X et al (1993) Molecular and statistical approaches to the detection and correction of errors in genotype databases. Am J Hum Gent 53:1137

Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J (2013) Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol 31(12):1119–1125

Cervera MT, Storme V, Ivens B, Gusmão J, Liu BH, Hostyn V et al (2001) Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. Genetics 158:787

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

Dinus RJ, Payne P, Sewell MM, Chiang VL, Tuskan GA (2001) Genetic modification of short rotation poplar wood: properties for ethanol fuel and fiber productions. Crit Rev Plant Sci 20:51–69

Doyle JJ (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Echt CS, Nelson CD (1997) Linkage mapping and genome length in eastern white pine (*Pinus strobus* L.). TAG Theor Appl Genet 94(8):1031–1037

Eckenwalder JE (1977) North American cottonwoods (*Populus*, salicaceae) of sections Abaso and Aigeiros. J Arnold Arboretum 58:193–208

Eckenwalder JE (1996) Systematics and evolution of *Populus*. In: Stettler RF, Bradshaw HD Jr, Heilman PE et al (eds) Biology of *Populus* and its implications for management and conservation, vol 7. NRC Research Press, National Research Council of Canada, Ottawa, p 30

Eisenstein M (2015) Startups use short-read data to expand long-read sequencing market. Nat Biotechnol 33(5):433–435

Fretwell N (1996) A search for human Y-chromosomes specific minisatellites and microsatellites. Dissertation, University of Leicester

Ganal MW, Altmann T, Röder MS (2009) SNP identification in crop plants. Curr Opin Plant Biol 12:211–217

Gaudet M, Jorge V, Paolucci I et al (2008a) Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. Tree Genet Genomes 4:25–36

Guo J, Li C, Teng T, Shen F, Chen Y, Wang Y, Pan C, Ling Q (2018) Construction of the first high-density genetic linkage map of pikeperch (*Sander lucioperca*) using specific length amplified fragment (SLAF) sequencing and QTL analysis of growth-related traits. Aquaculture 497:299–305

Haldane JBS (1919) The combination of linkage values, and the calculation of distances between the loci of linked factors. J Genet 8:299–309

Hulbert SH, Ilott TW, Legg EJ, Lincoln SE, Lander ES, Michelmore RW (1988) Genetic analysis of the fungus, *Bremia lactucae*, using restriction fragment length polymorphisms. Genetics 120:947–958

Jansson S, Douglas CJ (2007) *Populus*: a model system for plant biology. Annu Rev Plant Biol 58(1):435–458

Kartesz JT, Meacham CA (1999) Synthesis of the north American flora. North Carolina Botanical Garden, University of North Carolina at Chapel Hill

Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12:656–664

Kosambi DD (1944) The estimation of map distances from recombination values. Ann Eugenics 12:172–175

Lind M, Källman T, Chen J, Ma XF, Bousquet J, Morgante M, Zaina G, Karlsson B, Elfstrand M, Lascoux M, Stenlid J (2014) A *Picea abies* linkage map based on SNP markers identifies QTLs for four aspects

of resistance to *Heterobasidion parviporum* infection. PLoS One 9: e101049

Liu D, Ma C, Hong W, Huang L, Liu M, Liu H, Zeng H, Deng D, Xin H, Song J, Xu C, Sun X, Hou X, Wang X, Zheng H (2014) Construction and analysis of high-density linkage map using high-throughput sequencing data. PLoS One 9:e98855. https://doi.org/10.1371/journal.pone.0098855

Ma T, Wang J, Zhou G, Yue Z, Hu Q, Chen Y, Liu B, Qiu Q, Wang Z, Zhang J, Wang K, Jiang D, Gou C, Yu L, Zhan D, Zhou R, Luo W, Ma H, Yang Y, Pan S, Fang D, Luo Y, Wang X, Wang G, Wang J, Wang Q, Lu X, Chen Z, Liu J, Lu Y, Yin Y, Yang H, Abbott RJ, Wu Y, Wan D, Li J, Yin T, Lascoux M, DiFazio SP, Tuskan GA, Wang J, Liu J (2013) Genomic insights into salt adaptation in a desert poplar. Nat Commun 4:2797

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

Pakull B, Groppe K, Meyer M, Markussen T, Fladung M (2009) Genetic linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.). Tree Genet Genomes 5:505–515

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

Pendleton M, Sebra R, Pang AWC, Ummat A, Franzen O, Rausch T, Stütz AM, Stedman W, Anantharaman T, Hastie A, Dai H, Fritz MHY, Cao H, Cohain A, Deikus G, Durrett RE, Blanchard SC, Altman R, Chin CS, Guo Y, Paxinos EE, Korbel JO, Darnell RB, McCombie WR, Kwok PY, Mason CE, Schadt EE, Bashir A (2015) Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nat Methods 12:780–786

Plomion C, O'Malley DM (1996) Recombination rate differences for pollen parents and seed parents in *Pinus pinaster*. Heredity 77(4): 341–350

Qi Z, Huang L, Zhu R, Xin D, Liu C, Xue H, Jiang H, Hong W, Hu G, Zheng H, Chen Q, Zhang J-S (2014) A high-density genetic map for soybean based on specific length amplified fragment sequencing. PLoS ONE 9(8):e104871

Remington DL, Whetten RW, Liu BH, O'malley DM (1999) Construction of an AFLP genetic map with nearly complete genome coverage in *Pinus taeda*. Theor Appl Genet 98:1279–1292

Stirling B, Yang ZK, Gunter LE, Tuskan GA, Bradshaw HD Jr (2003) Comparative sequence analysis between orthologous regions of the *Arabidopsis* and *Populus* genomes reveals substantial synteny and microcollinearity. Can J For Res 33(11):2245–2251

Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H, Xu C, Song J, Huang L, Wang C, Shi J, Wang R, Zheng X, Lu C, Wang X, Zheng H (2013) SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. PLoS One 8:e58700

Sun C, Niu Y, Ye X, Dong J, Hu W, Zeng Q, Chen Z, Tian Y, Zhang J, Lu M (2017) Construction of a high-density linkage map and mapping of sex determination and growth-related loci in the mandarin fish (*Siniperca chuatsi*). BMC Genomics 18:446

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

Tuskan GA, DiFazio SP, Teichmann T (2004) Poplar genomics is getting popular: the impact of the poplar genome project on tree research. Plant Biol 6:2

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

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

Wan Z, Li Y, Liu M, Chen Y, Yin T (2015) Natural infectious behavior of the urediniospores of *Melampsora laricipopulina* on poplar leaves. J For Res 26:225–231

West MA, van Leeuwen H, Kozik A, Kliebenstein DJ, Doerge R, Clair DAS et al (2006) High-density haplotyping with microarray-based expression and single feature polymorphism markers in Arabidopsis. Genome Res 16:787–795

Wu RL, Yin TM, Huang MR et al (2000) The application of marker-assisted selection to tree breeding. Sci Silvae Sin 36:103–113

Wullschleger SJSTG (2002) Genomics and forest biology: *Populus* emerges as the perennial favorite. Plant Cell 14:2651–2655

Xu X, Xu R, Zhu B, Yu T, Qu W, Lu L et al (2015) A high-density genetic map of cucumber derived from specific length amplified fragment sequencing (SLAF-SEQ). Front Plant Sci 5:768

Yang W, Wang K, Zhang J, Ma J, Liu J, Ma T (2017) The draft genome sequence of a desert tree *Populus pruinosa*. GigaScience 6:1–7

Yin T, Huang M, Wang M, Zhu L-H, Zeng Z-B, Wu R (2001) Preliminary interspecific genetic maps of the *Populus* genome constructed from RAPD markers. Genome 44(4):602–609

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

Yin TM, Wang XR, Andersson B, Lerceteau-Köhler E (2003) Nearly complete genetic maps of *Pinus sylvestris* L. (Scots pine) constructed by AFLP marker analysis in a full-sib family. Theor Appl Genet 106:1075–1083

Yin T, Difazio SP, Gunter LE, Riemenschneider D, Tuskan GA (2004) Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. Theor Appl Genet 109:451–463

Yin T, Difazio SP, Gunter LE, Zhang X, Sewell MM, Woolbright SA, Allan GJ, Kelleher CT, Douglas CJ, Wang M, Tuskan GA (2008) Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. Genome Res 18:422–430

Zhang X, Yin T, Qiang Z, Huang M, Zhu L, Zhai W et al (2000) Rapd linkage mapping in a *Populus deltoides*×*Populus euramericana* f (1) family. Hereditas 22:209–213

Zhang Y, Wang L, Xin H, Li D, Ma C, Xia D 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