

# A high-density genetic map for P genome of *Agropyron* Gaertn. based on specific-locus amplified fragment sequencing (SLAF-seq)

Yan Zhang<sup>1,2</sup> · Jinpeng Zhang<sup>1</sup> · Long Huang<sup>3</sup> · Ainong Gao<sup>1</sup> · Jing Zhang<sup>1</sup> · Xinming Yang<sup>1</sup> · Weihua Liu<sup>1</sup> · Xiuquan Li<sup>1</sup> · Lihui Li<sup>1</sup>

Received: 3 February 2015 / Accepted: 13 July 2015 / Published online: 2 August 2015  
© Springer-Verlag Berlin Heidelberg 2015

## Abstract

**Main conclusion** This genetic map for *Agropyron* Gaertn. contained 1023 markers on seven linkage groups, with a total of 907.8 cM and an average distance of 1.5 cM between adjacent loci.

Many wheat-*Agropyron cristatum* derivative lines exhibit superior agronomic traits, and part of them are valuable for future wheat breeding. To date, no high-density genetic map for *Agropyron* Gaertn. has been published. Specific-locus amplified fragment sequencing (SLAF-seq), a recently developed strategy for large scale de novo discovery and genotyping of single nucleotide polymorphisms (SNPs), was employed in this study to develop sufficient markers for a segregating *Agropyron* F<sub>1</sub> population derived from an interspecific cross between two cross-pollinated diploid collections *A. cristatum* (L.) Beauv. ‘Z1842’ and *A. mon-golicum* Keng ‘Z2098’. In total, we obtained raw data consisting of 128,932,358 pair-end reads of ~80 bp long after sequencing. Then 69,325 high-quality SLAFs were detected,

of which 26,248 SLAFs were polymorphic and 1752 of the polymorphic markers were used for the genetic map construction. The final map contained 1023 markers on the seven linkage groups (LGs), which spanned a total of 907.8 cM with an average number of 146 markers and 89 loci per LG and an average distance of 1.5 cM between adjacent loci. To our knowledge, this map is the densest genetic linkage map for *Agropyron* so far. Through BLAT alignment of *Agropyron* SLAF marker sequences with the draft genome assemblies of wheat and barley, the *Agropyron* LGs were assigned as LG1-7 according to their corresponding homoeologous chromosomal groups of wheat. Results of this study will not only provide a platform for gene/QTL fine mapping, but also serve as a reference to assist the assembling of the P genome sequence in future.

**Keywords** *Agropyron* · Genetic map · High density · SLAF marker

## Abbreviations

<i>Agropyron</i>	<i>Agropyron</i> Gaertn.
CP	Cross-pollinated
LG	Linkage group
SLAF-seq	Specific-locus amplified fragment sequencing
SNP	Single nucleotide polymorphism

Y. Zhang, J. Zhang and L. Huang contributed equally to this work.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00425-015-2372-7) contains supplementary material, which is available to authorized users.

✉ Lihui Li  
lilihui@caas.cn

- <sup>1</sup> National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China
- <sup>2</sup> Department of Crop Genomics and Genetic Improvement, Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China
- <sup>3</sup> Biomarker Technologies Corporation, Beijing 100081, China

## Introduction

*Agropyron* Gaertn. is an important perennial genus of the tribe Triticeae. Approximately 15 species exist in the genus and most of them are distributed on the temperate-frigid grassland and sand land of Eurasia (Dewey 1984). *A.*

*mongolicum*, *A. cristatum*, *A. cristatum* ssp. *imbricatum*, *A. cristatum* ssp. *pectinatum*, and *A. fragile* are five most common members of the *Agropyron* genus (Chen et al. 2013). The *Agropyron* genus exists naturally at three ploidy levels: diploid,  $2n = 2 \times = 14$ , PP; tetraploid,  $2n = 4 \times = 28$ , PPPP; and hexaploid,  $2n = 6 \times = 42$ , PPPPPP. *Agropyron* is not only a kind of pasture grass, but also an excellent wild germplasm for wheat breeding as a genus of wild relatives of wheat (Ford-Lloyd et al. 2011). Importing exogenous superior genes into wheat through wide cross is an effective approach to improve the wheat genetic diversity. *Agropyron* (P genome) exhibits many superior traits beneficial to disease resistance, abiotic stress tolerance and high yield (Dong et al. 1992). Introduction of desirable genes from *A. cristatum* into common wheat has been fulfilled through intergeneric hybridization (Li et al. 1998). Subsequently, a series of derivative lines were produced and characterized, and part of them are valuable for future wheat breeding as novel germplasms (Wu et al. 2006; Chen et al. 2012). For example, the disomic addition line 4844 shows superior numbers of florets and kernels per spike, and the alien chromosomes were designated as 6P of *A. cristatum* (Wu et al. 2006). Recently an *A. cristatum*-wheat compensating Robertsonian translocation conferring resistance to leaf rust was reported (Ochoa et al. 2015).

The detection of alien chromosome/fragment/gene in the wheat background can provide a guide to utilize the wheat-*A. cristatum* derivative lines with high efficiency, and a great deal of efforts have been made using genomic in situ hybridization (GISH), fluorescence in situ hybridization (FISH), molecular markers from *A. cristatum* (Luan et al. 2010; Wu et al. 2010; Song et al. 2013; Han et al. 2014). These work provide valuable clues for the application of wheat-*A. cristatum* translocation and introgression lines in wheat breeding. However, so far the P genome of *Agropyron* has not been sequenced due to its huge size. Construction of a high-density and high-quality genetic map will be of great benefit in germplasm study before whole genome sequencing. To date, only one genetic linkage map of *Agropyron* Gaertn. has been published, in which 152 AFLP and 23 RAPD markers were ordered in seven linkage groups (LGs; Yu et al. 2012). The progress in P genome study lags far behind the production and application of wheat-*A. cristatum* translocation and introgression lines, which limited the further elucidation of the genetic effects of alien fragments/genes on the agronomic traits of these derivative lines. Therefore, developing and mapping molecular markers to construct a high-density genetic map for P genome are very urgent and important.

High-throughput sequencing can provide new strategies for sequence-based single nucleotide polymorphism (SNP) genotyping. To reduce the costs of sequencing, reduced representation library (RRL) sequencing to reduce the

complexity of the genome has been developed within the past several years (Hyten et al. 2010). One of the simplest methods is the separation and purification of restriction fragments within a given size range. For example, restriction site-associated DNA (RAD) sequencing has been used to successfully construct high-density genetic map for *Lolium perenne* and grape (Pfender et al. 2011; Wang et al. 2012). Based on the next generation sequencing (NGS) technology, specific-locus amplified fragment sequencing (SLAF-seq) is a recently developed high-resolution strategy for large scale de novo discovery and genotyping of SNP (Sun et al. 2013). SLAF-seq is similar to RAD sequencing, but it is an enhanced RRL sequencing method. Compared with previous single-end RAD sequencing, pair-end reads obtained by SLAF-seq can somewhat increase marker specificity and accuracy. Double barcodes in SLAF reads can easily and accurately assign these reads to the individuals in a big population. Several high-density genetic maps for different species have been successfully constructed based on SLAF-seq in the last 2 years, such as sesame and soybean (Zhang et al. 2013; Qi et al. 2014).

The genetic study on cross-pollinated plant is difficult because of its self-incompatibility or selfing depression, and outcrossing can result in high levels of individual heterozygosity and population heterogeneity. The mapping population for cross-pollinated plant can be a “CP” (cross-pollinated) population which consisted of  $F_1$  individuals (cross-pollinated progeny) generated from a cross between two heterogeneously heterozygous and homozygous diploid parents with no prior knowledge of linkage phases. The genetic map for tea, grape and *L. perenne* using CP population has been reported (Pfender et al. 2011; Wang et al. 2012; Hu et al. 2013).

In this study, we obtained a segregating *Agropyron*  $F_1$  population derived from an interspecific cross between two cross-pollinated diploid collections *A. cristatum* (L.) Beauv. ‘Z1842’ and *A. mongolicum* Keng ‘Z2098’. Then the  $F_1$  population was used as CP population for map construction. SLAF-seq was used to generate genotype data, which allowed the efficient development of a large number of SNP markers. Subsequently a high-density genetic map for *Agropyron* was constructed, and its synteny to wheat and barley was analyzed, and future applications of the map were also discussed.

## Materials and methods

### Plant material and DNA extraction

A segregating *Agropyron*  $F_1$  population was derived from an interspecific cross between two cross-pollinated diploid

collections ‘Z1842’ and ‘Z2098’. The male parent was Z1842 [*A. cristatum* (L.) Beauv.,  $2n = 2 \times = 14$ , PP], and the female parent was Z2098 [*A. mongolicum* Keng,  $2n = 2 \times = 14$ , PP]. In 2013, seedlings of progeny and parents were planted in the greenhouse of Chinese Academy of Agricultural Sciences, Beijing, China. Young healthy leaves were collected from the 113 individuals and their parents and genomic DNA was extracted according to the cetyltrimethylammonium bromide (CTAB) method with minor modifications (Porebski et al. 1997). The components of 100 ml CTAB buffer contained 2 g CTAB, 1.4 M NaCl, 20 mM EDTA, pH 8.0, 100 mM Tris-HCl, pH 8.0). DNA concentration and quality were estimated with a NanoDrop-2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) and by electrophoresis in 0.8 % agarose gels with a DNA marker.

### SLAF library construction and high-throughput sequencing

SLAF-seq was used to genotype a total of 113 individuals and their two parents, as previously described with minor modifications (Sun et al. 2013). 2 µg genomic DNA from each sample was digested with *NlaIII/MseI* (New England Biolabs, Beverly, MA, USA), and then a quick blunting kit (New England Biolabs) was used to convert the digested sample to 5′-phosphorylated, blunt-ended DNA. A 3′-adenine overhang was added to the resulting samples using Klenow Fragment (3′ → 5′ exo<sup>-</sup>) (New England Biolabs) and dATP at 37 °C. Duplex tag-labeled sequencing adapters (PAGE-purified, Life Technologies, Gaithersburg, MD, USA) were then ligated to the A-tailed fragments using T4 DNA ligase. The reaction products were purified using a Quick Spin column (Qiagen, Chatsworth, CA, USA) and electrophoresed in a 2 % agarose gel. Fragments of 280–330 bp (not including adapter sequence indexes and adaptors) were isolated using Gel Extraction Kits (Qiagen) and then subjected to PCR with Phusion Master Mix (New England Biolabs) and Solexa amplification primer mix. PCR products were purified for paired-end sequencing (40 bp each end) on an Illumina HiSeq 2500 sequencing platform (Illumina, San Diego, CA, USA) at Biomarker Technologies Corporation in Beijing.

### Genotyping

SLAF marker identification and genotyping were performed using procedures described by Sun et al. (2013). According to the barcode sequences, raw reads were classified to different individuals. Then low-quality reads (quality score <20) were filtered out. After barcodes were trimmed from reads, each read were 80 nucleotides long. All SLAF pair-end reads with clear index information were

clustered based on sequence similarity detected by BLAT (–tileSize = 10, –stepSize = 5) (Kent 2002). Sequences with over 95 % identity were grouped in one SLAF locus. Allelic tags were detected by parental sequences (depth >10) and those of progeny were genotyped by similarity with parental sequences. Alleles were defined in each SLAF using the minor allele frequency (MAF) evaluation. SLAF marker containing only one allelic tag was defined as non-polymorphic, while SLAF marker with 2–4 allelic tags was considered to be potential marker. In this study, the average sequence depths of SLAF markers were more than 19-fold in the two parents and 3.6-fold in 113 individuals. Polymorphic markers were classified into eight segregation patterns (paternal genotype × maternal genotype): ab × cd, ef × eg, hk × hk, lm × ll, nn × np, aa × bb, ab × cc and cc × ab. According to JoinMap 4 manual, the two characters left of the ‘×’ in these codes represent the alleles of the first parent, the two on the right represent those of the second parent; each distinct allele is represented with a different character. Among them, ab × cd, ef × eg, hk × hk, lm × ll and nn × np were appropriate for CP population and used to construct the genetic map.

### Genetic map construction and segregation distortion analysis

JoinMap 4.0 software was used to calculate recombination rates between markers, create linkage groups and analyze segregation distortion (Stam 1993). After data had been imported, a ‘CP’ type was selected for data mining. The integrities of SLAF markers and individuals are both more than 80 %. Marker segregation ratios were calculated using the chi-square test. Markers showing significant ( $P < 0.05$ ) segregation distortion and 100 % similarity of loci were excluded from the map construction. Independence LOD was used as grouping parameter with a threshold range 3–10. HighMap Strategy was used to calculate map distances, order SLAF markers and correct genotyping errors within LGs by ML (Maximum Likelihood) mapping (Liu et al. 2014).

### Analysis of syntenic relationship between *Agropyron* and wheat and barley

*Agropyron* SLAF marker sequences were aligned with the draft genome assemblies of wheat and barley ([https://urgi.versailles.inra.fr/download/iwgsc/ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public\\_data/sequences](https://urgi.versailles.inra.fr/download/iwgsc/ftp://ftpmips.helmholtz-muenchen.de/plants/barley/public_data/sequences)) by BLAT. The sequence identity should be ≥80 % and the most significant BLAT alignment was used to assign putative ortholog and identify homoeologous relationship. In addition, two strict criteria were used to screen the putative

ortholog: (1) both 40-bp end sequences had to align with the same contig; (2) the coverage on homologous contig had to fall between 200 and 500 bp. The genetic positions for these contigs of wheat and barley were retrieved from: <https://urgi.versailles.inra.fr/download/iwgsc/POPSEQ/>; [ftp://ftp.mips.helmholtz-muenchen.de/plants/barley/public\\_data/popseq\\_IPK](ftp://ftp.mips.helmholtz-muenchen.de/plants/barley/public_data/popseq_IPK). The syntenic relationships between *Agropyron* markers and wheat (or barley) contigs were described as plots of the genetic positions (cM) of *Agropyron* markers on the linkage groups against the genetic positions (cM) of the homologous contigs of wheat and barley.

## Results

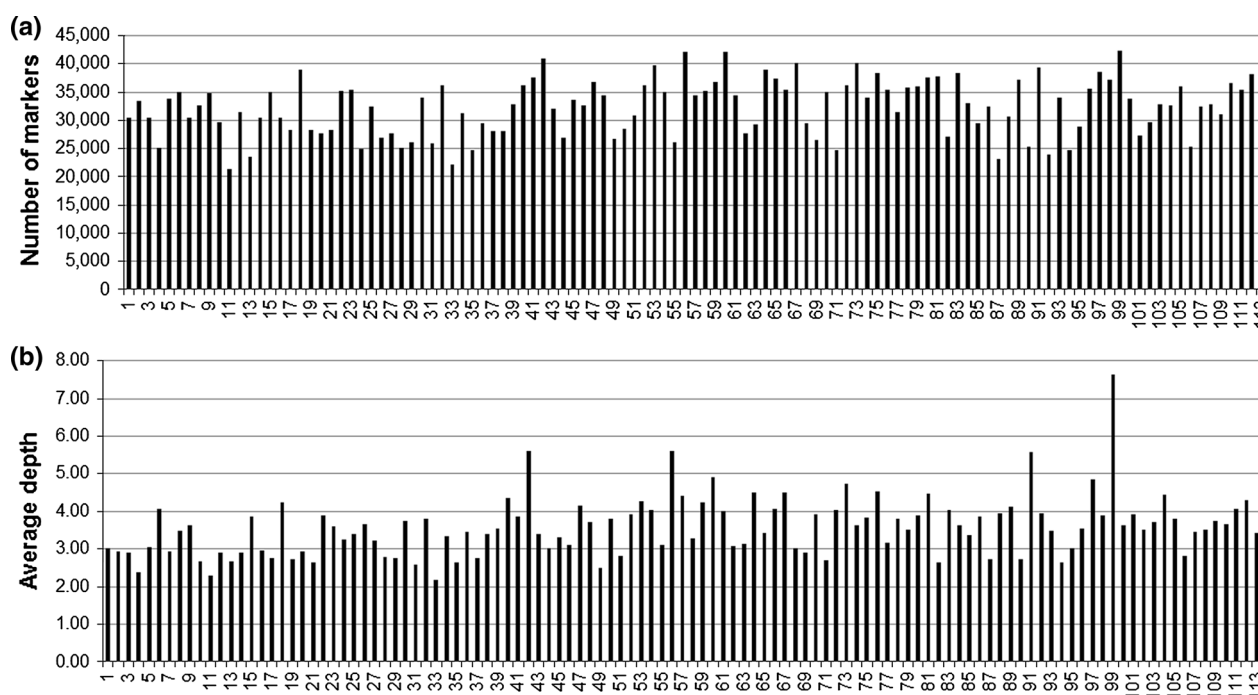
### Analysis of SLAF-seq data and SLAF markers

After SLAF library construction and high-throughput sequencing, we obtained 19.88 Gb of raw data consisting of 128,932,358 pair-end reads of ~80 bp long. Among them, 9,447,505 and 7,802,086 reads were from the male and female parent, respectively. Read numbers of the 113 individuals in the CP population ranged from 466,226 to 2,045,138 with an average of 917,133.

The numbers of SLAFs in the male and female parents were 44,866 and 41,344, respectively. The read numbers of SLAFs were 1,229,975 and 810,857 in the male and female

parents, respectively. The average coverage for each SLAF was 27.4-fold in the male parent and 19.6-fold in the female parent. In the F<sub>1</sub> population, the numbers of SLAF markers in each individual ranged from 21,297 to 42,314 with an average of 32,301 (Fig. 1a). The read numbers of SLAFs ranged from 48,436 to 323,024 with an average of 118,565, and the average depth ranged from 2.2-fold to 7.6-fold with an average of 3.6-fold in the F<sub>1</sub> population (Fig. 1b).

In total, 69,325 high-quality SLAFs were detected, of which 26,248 were polymorphic with a polymorphism rate of 37.9 % (Supplemental Table S1). Of the 26,248 polymorphic SLAFs, 14,174 were classified into eight segregation patterns (Supplemental Table S2). Five segregation patterns (ab × cd, ef × eg, hk × hk, lm × ll and nn × np) were appropriate for CP population. In total, 4231 markers fell into this five types, and the examples for this five segregation types were shown in Supplemental Fig. S1. Among these 4231 markers, 1752 markers with more than 40-fold of parental sequence depth and more than fourfold of individual sequence depth were used for the genetic map construction (Supplemental Table S3), and their genotype data in the 113 F<sub>1</sub> individuals was shown in Supplemental Table S4. All 113 individuals contained more than 95 % of the 1752 SLAF markers. Then markers showing significant ( $P < 0.05$ ) segregation distortion, less than 80 % integrity of individual segregation pattern, and 100 % similarity of loci were excluded from the map construction.



**Fig. 1** Number of markers (a) and average depth (b) for each F<sub>1</sub> individual. The x-axes indicate the 113 F<sub>1</sub> individuals, the y-axes indicate the number of SLAF markers (a) and average depth (b)

**Table 1** Summary of seven linkage groups

Linkage group	Number of markers			Genetic sizes (cM)		
	Female (Z2098)	Male (Z1842)	Integrated map	Female (Z2098)	Male (Z1842)	Integrated map
1	130	157	175	127.5	243.3	186.1
2	161	200	225	101.2	267.9	185.5
3	86	98	107	59.2	119.6	89.4
4	86	107	117	108.1	222.5	165.3
5	121	142	152	46.9	157.3	102.1
6	91	96	115	58.8	85.3	72.1
7	118	121	132	53.7	160.8	107.3
Total	793	921	1023	555.4	1256.7	907.8

### Basic characteristics of the genetic map

After linkage analysis, 921 markers for Z1842 (male), 793 markers for Z2098 (female), and 1023 markers for the integration were mapped onto the genetic maps and assigned to 7 LGs (Table 1; Supplemental Figs. S2–S8). Taking into account the size of all LGs, marker coverage amounted to 1256.7 cM for Z1842 (male), 555.4 cM for Z2098 (female), and 907.8 cM for the integrated map. The final integrated map contained 622 loci with an average distance of 1.5 cM between adjacent loci (Figs. 2, 3, 4). Positions of the SLAF markers on these LGs are shown in Supplemental Table S5.

Basic characteristics of the 7 LGs of the integrated map were shown in Table 2. To our knowledge, this map is the densest genetic linkage map for *Agropyron* so far. On average, one LG contained 146 markers and 89 loci which spanned an average of 129.7 cM. The genetic distance of the LGs ranged from 72.1 cM (LG6) to 186.1 cM (LG1). According to total distance, the largest LG was LG1 with 175 markers, a length of 186.1 cM, and an average distance of 1.8 cM between adjacent loci. The smallest LG was LG6, with 115 markers, a length of 72.1 cM, and an average distance of 1.0 cM between adjacent loci. The degree of linkage between markers was reflected by ‘Gap  $\leq 5$ ’ ranging from 83.1 to 100 % with an average value of 95.4 %. The largest gap on this map was 18.4 cM located in LG1.

### Segregation distortion markers on the map

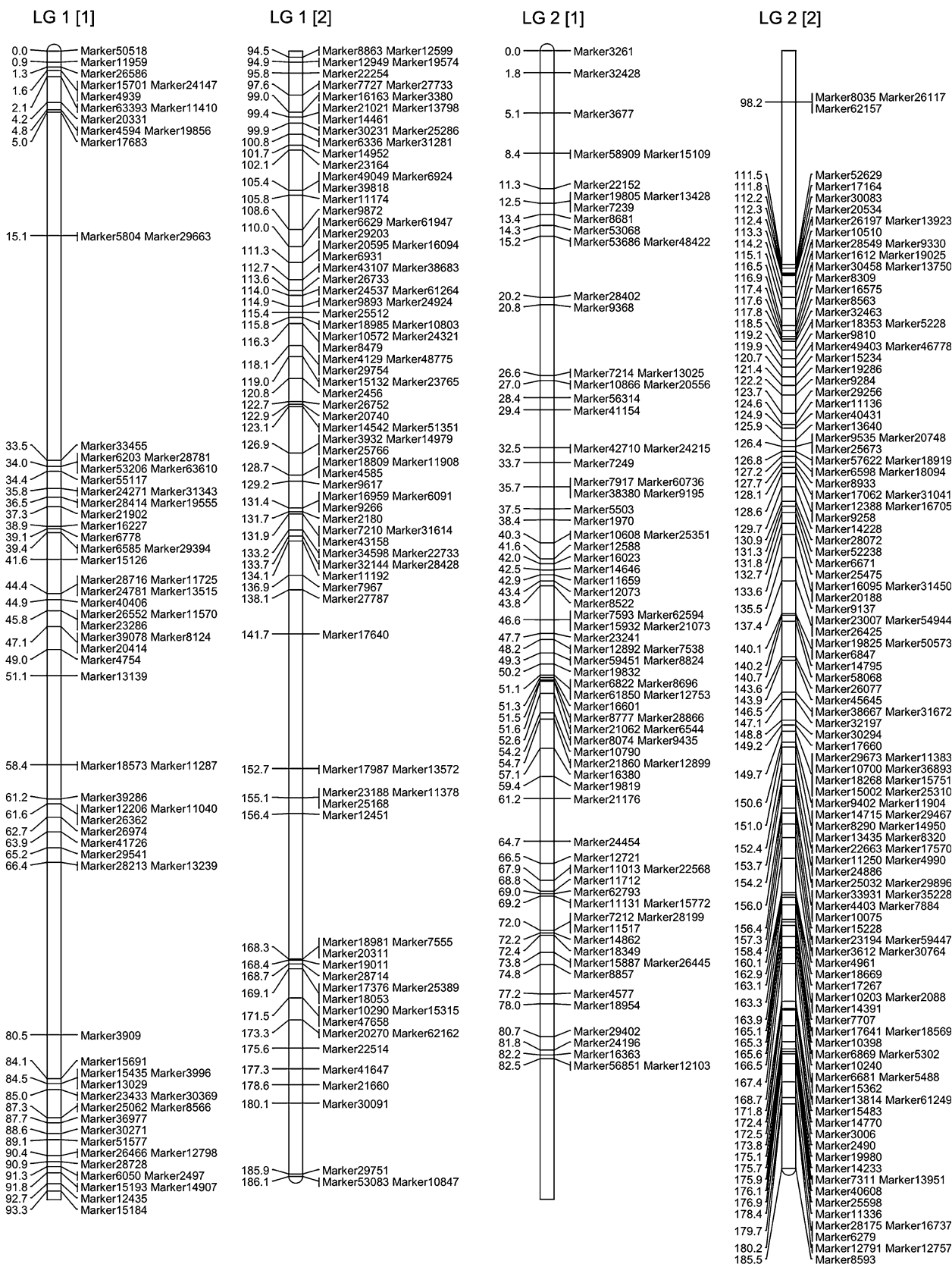
Among the 1752 SLAF markers used for the map construction, 293 markers showed segregation distortion, which genotype frequencies in the  $F_1$  individuals were shown in Supplemental Table S6. It can be found that, for the two segregation types  $\langle ab \times cd \rangle$  and  $\langle ef \times eg \rangle$  in Supplemental Table S6, in most cases, there is one most predominant allele in the genotypes of  $F_1$  individuals, accompanying with two predominant genotypes. For

example, the genotypes of Marker26833 (segregation type  $\langle ab \times cd \rangle$ ) was 60 ac, 1 ad, 45 bc and 7 bd. For Marker26833, *c* was the most predominant allele, and both ac and bc were predominant genotypes in 113  $F_1$  individuals. The allele *c* came from the second parent’s first parent. For two-allele  $\langle lm \times ll \rangle$  and  $\langle nn \times np \rangle$ , both heterozygous and homozygous genotypes are likely to be predominant.

Segregation distortion markers at mild significance level ( $0.05 \leq P \leq 0.1$ ) were kept and 126 markers were mapped onto the map (Supplemental Table S5). The distribution of segregation distortion markers was not similar to that of all markers except for LG5 (Table 3). For example, LG7 had the highest percentage (35.7 %) of segregation distortion markers, whereas it only had 12.9 % of all markers. The highest frequency of segregation distortion markers was also on LG7, which was 34.1 %, while the lowest frequency (2.6 %) was on LG4. Furthermore, 7 of 16 segregation distortion regions (SDR) were detected on LG7, while none was detected on LG3 and LG4.

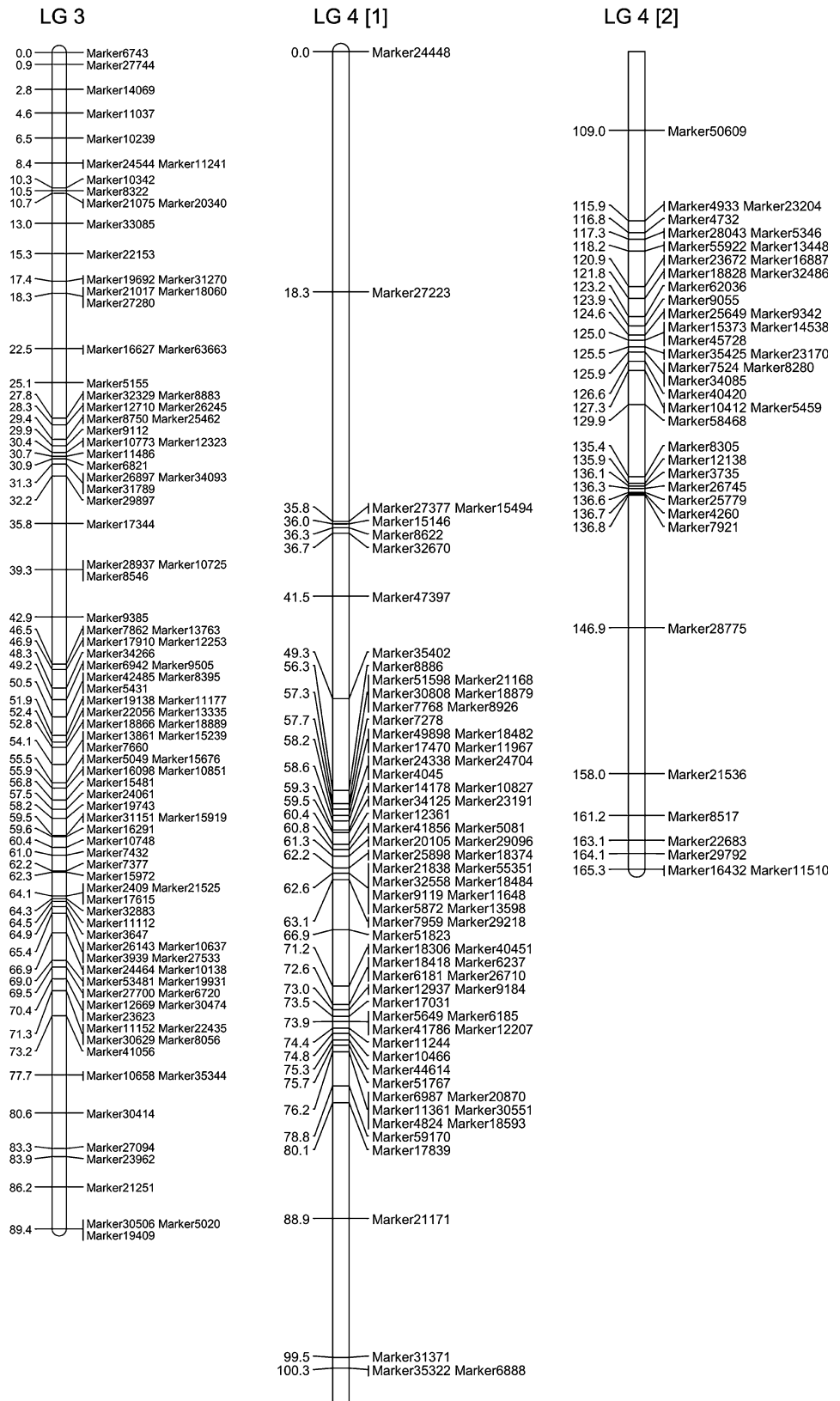
### Homoeologous relationship between *Agropyron* linkage group and chromosomal group of wheat and barley

To evaluate the accuracy of the genetic map and analyze the homoeologous relationship between *Agropyron* and wheat and barley, we aligned *Agropyron* SLAF marker sequences with the draft genome assemblies of wheat and barley. 161 and 46 of 1023 mapped *Agropyron* markers showed high sequence identity ( $\geq 80$  %) with wheat and barley contigs, respectively (Supplemental Table S7 and 8). The serial number of *Agropyron* linkage group was assigned according to the corresponding homoeologous chromosomal group of wheat, which was defined by the highest hit number of contigs. 15 of 22 markers on LG5 were homologous to 5A/B/D contigs, and 11 of 17 markers on LG6 were homologous to 6A/B/D contigs, showing substantial homoeologous relationship between LG5 and



**Fig. 2** Linkage group 1 and 2 of high-density linkage map for *Agropyron*

**Fig. 3** Linkage group 3 and 4 of high-density linkage map for *Agropyron*



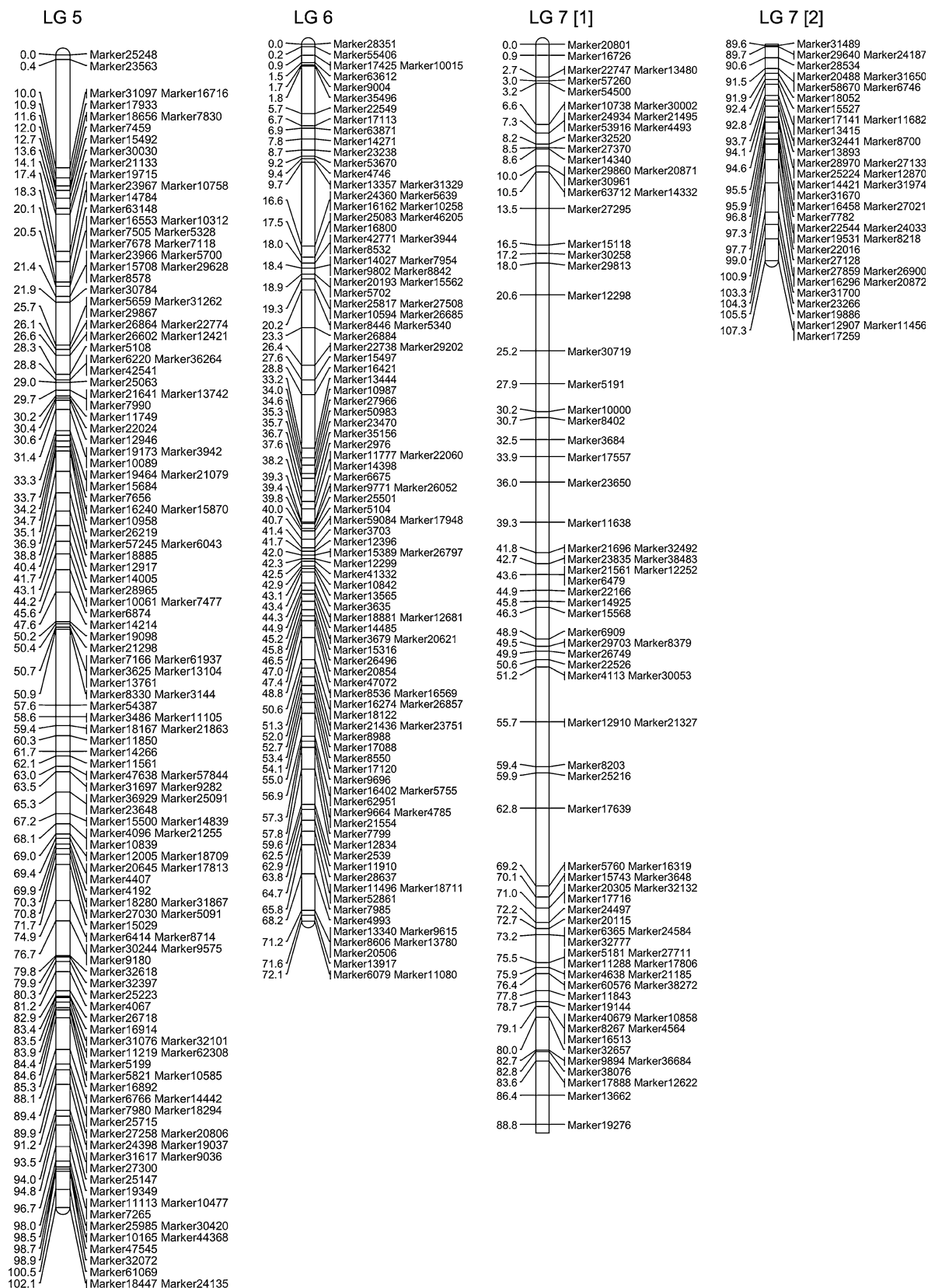


Fig. 4 Linkage group 5 to 7 of high-density linkage map for *Agropyron*



**Table 2** Basic characteristics of the seven linkage groups of the integrated map

Linkage group	Number of markers	Number of loci	Total distance (cM)	Average distance (cM)	Gaps ≤5 (%)	Largest gap (cM)
1	175	105	186.1	1.8	93.3	18.4
2	225	144	185.5	1.3	96.5	15.7
3	107	64	89.4	1.4	100	4.2
4	117	66	165.3	2.5	83.1	18.3
5	152	90	102.1	1.1	97.8	9.6
6	115	74	72.1	1.0	98.6	6.9
7	132	79	107.3	1.4	98.7	6.4
Maximum	225	144	186.1	2.5	100	18.4
Minimum	107	64	72.1	1.0	83.1	4.2
Total	1023	622	907.8	–	–	–
Average	146	89	129.7	1.5	95.4	11.4

**Table 3** Distribution of segregation distortion markers

Linkage group	All markers		Segregation distortion marker		Frequency of segregation distortion marker (%)	SDR number
	Number	Percentage (%)	Number	Percentage (%)		
1	175	17.1	29	23.0	16.6	4
2	225	22.0	19	15.1	8.4	3
3	107	10.5	5	4.0	4.7	0
4	117	11.4	3	2.4	2.6	0
5	152	14.9	18	14.3	11.8	1
6	115	11.2	7	5.6	6.1	1
7	132	12.9	45	35.7	34.1	7
Total	1023	–	126	–	–	16

The significance level of segregation distortion is  $0.05 \leq P \leq 0.1$

SDR means segregation distortion region, containing ≥3 adjacent segregation distortion loci

**Table 4** Homoeologous relationship between *Agropyron* linkage group and chromosomal group of wheat and barley

Linkage group	Wheat			Barley		
	Number of contigs	Homoeologous group		Number of contigs	Homoeologous group	
		Chromosome	Number of contigs		Chromosome	Number of contigs
1	31	1A/B/D	13	11	1H	4
2	39	2A/B/D	17	10	2H	7
3	17	3A/B/D	9	3	–	–
4	19	4A/B/D	7	3	4H	2
5	22	5A/B/D	15	5	5H	5
6	17	6A/B/D	11	5	–	–
7	16	7A/B/D	5	3	7H	2
Total	161	–	77	40	–	–

5A/B/D, as well as LG6 and 6A/B/D (Table 4). To barley, 7 of 10 markers on LG2 were homologous to 2H contigs, and five markers on LG5 were all homologous to 5H contigs (Table 4). Furthermore, using the genetic positions

of *Agropyron* markers and their homologous contigs of wheat and barley, different degrees of syntenic relationships between *Agropyron* and wheat (or barley) were discovered and shown in a dot plot (Fig. 5). Among the seven

linkage groups, highly conserved syntenic relationships between LG6 and 6A/B/D, LG2 and 2H, LG5 and 5H were shown in Fig. 5 and Supplemental Table S8.

## Discussion

### Utilization of the segregation patterns in the F<sub>1</sub> population of cross-pollinated *Agropyron*

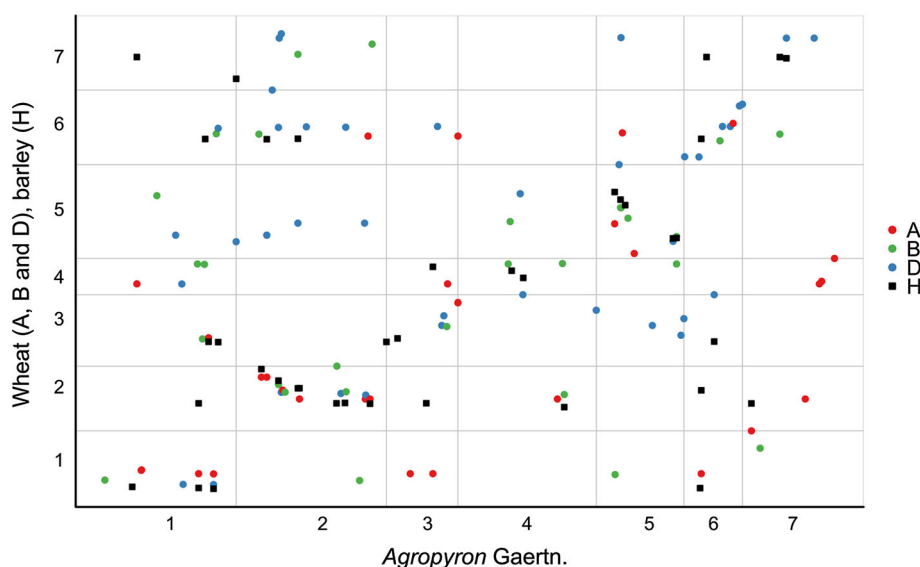
In this study, 14,174 polymorphic SLAFs were successfully classified into eight segregation patterns (Supplemental Table S2). Five segregation patterns (ab × cd, ef × eg, hk × hk, lm × ll and nn × np) were appropriate for CP population, which accounted for approximately 30 % of 14,174 SLAFs. It is noteworthy that the homozygous genotype, aa × bb, accounted for 45 % of 14,174 SLAFs, the highest proportion of eight segregation patterns. Based on these homozygous loci, the diploid parents in this study can also be used as homozygous parents. An F<sub>2</sub> population will be obtained by selfing one F<sub>1</sub> individual. Consequently, the SLAF markers which segregation pattern is aa × bb can be used for genetic map construction when F<sub>2</sub> population is used as mapping population. It is clear that the F<sub>2</sub> population size depends on the seed set of selfing. It has been reported that a genetic linkage map of *Agropyron* Gaertn. was constructed using a F<sub>2</sub> population (Yu et al. 2012). It is evident that using F<sub>1</sub> population for map construction can save considerable time

in comparison with using F<sub>2</sub> population. The efficient strategy for the construction of *Agropyron* genetic map in this study provides a valuable reference for those wild plants which are heterogeneously heterozygous and homozygous.

Of the 1023 SNP markers on the integrated map, 332 were homozygous for one parent and heterozygous for the other (102 for nn × np and 230 for lm × ll; Fig. 6). In this study, in segregation type (nn × np), locus is heterozygous in the second (female) parent, while locus is heterozygous in the first (male) parent in (lm × ll). The result might indicate the larger heterozygosity of Z1842 (male) than Z2098 (female), as well as the higher heterozygosity of *A. cristatum* than *A. mongolicum*, which could be the important reason for the larger genetic sizes of the LGs for Z1842 (male) than those for Z2098 (female). The other three segregation types (ab × cd, ef × eg, and hk × hk) could be mapped on both female and male linkage maps. Thus the remaining 691 markers could be used as shared markers for the integration of the two parents' maps into one (Supplemental Figs. S2–S8).

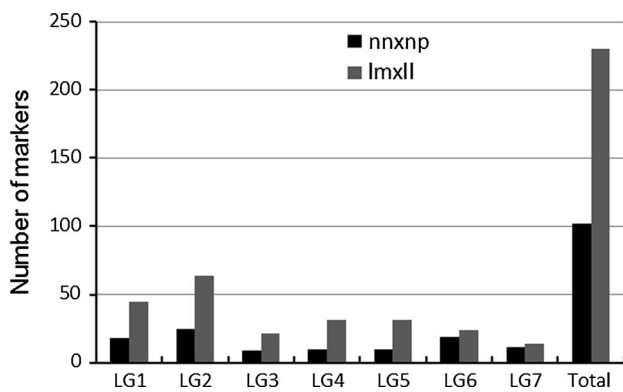
### The advantages of SLAF sequencing for developing markers

Conventional methods of developing markers usually are time-consuming, laborious and inefficient. The developed markers are sometimes with low accuracy due to the instability and complexity of DNA bands, such as RAPD



**Fig. 5** Analysis of syntenic relationship between *Agropyron* and wheat (round dot) and barley (square box). The X axis indicates the genetic position (cM) of each *Agropyron* marker, 1–7 are the serial numbers of seven linkage groups; the Y axis indicates the genetic position of each contig of wheat and barley, 1–7 are the serial numbers of seven

chromosomal groups. The genetic positions for these contigs of wheat and barley were retrieved from 140318\_wheat\_CSS\_POPSEQ\_positions (<https://urgi.versailles.inra.fr/download/iwgsc/POPSEQ/>) and 130905\_MxB\_iSelect + OWB\_GBS\_maps\_combined\_anchoring ([ftp://ftp.mips.helmholtz-muenchen.de/plants/barley/public\\_data/popseq\\_IPK](ftp://ftp.mips.helmholtz-muenchen.de/plants/barley/public_data/popseq_IPK))



**Fig. 6** Number of markers with segregation type -  $nn \times np$ ) and  $(lm \times ll)$  on the genetic map

and SSR markers. The new method was required to ensure density, uniformity, efficiency and accuracy of marker development. SLAF-seq is a recently developed high-resolution sequencing strategy for simplified genome. SLAFs are specific loci associated with certain restriction site, thus the high coverage of each marker is easier to achieve with less sequencing than random sequencing, which will help ensure marker density and accuracy (Sun et al. 2013). Meanwhile, the distribution of the selected restriction site in genome determines marker uniformity. This study provides the first large scale marker development for *Agropyron* with high efficiency. In total, 69,325 SLAF markers based on high-throughput sequencing were developed, and eventually 1023 polymorphic markers with high integrity and coverage were mapped onto the genetic map, which were accomplished after 4 months. Each polymorphic SLAF marker corresponds to 2–4 sequences with high homology and SNPs, which makes genotyping easier and more accurate than DNA electrophoretic bands. These developed SLAFs also accurately reflect the genetic and polymorphic characteristics of P genome, which expand our understanding on the P genome of *Agropyron*.

#### Different degrees of synteny between different linkage groups of *Agropyron* and their homoeologous chromosomal groups of wheat and barley

As shown in Table 4, Supplemental Table S8 and Fig. 5, 17 of 39 markers on LG2 were homologous to 2A/B/D contigs, showing homoeologous relationship between LG2 and 2A/B/D and assigning LG2 to *Agropyron* chromosome 2P. Likewise, seven linkage groups, from LG1 to LG7, were assigned to 1P-7P accordingly. However, there were differences in the synteny between different linkage groups of *Agropyron* and their homoeologous chromosomal groups of wheat and barley. Several groups performed well

in synteny, such as LG6 to 6A/B/D, LG2 to 2H, and LG5 to 5H. It has been identified that three wheat–*A. cristatum* 6P addition lines carried genes conferring high numbers of kernels per spike and resistance to powdery mildew, the important traits for wheat improvement (Han et al. 2014). The deep and comprehensive comparative analysis of 6P with 6/B/D may provide valuable cues for promoting the wheat molecular chromosome engineering. Thus the conserved synteny between LG6 and 6A/B/D discovered in this study can be an aid for this. Several groups performed moderately in synteny, such as LG1 to 1A/B/D, and LG2 to 2A/B/D. Although 1A/B/D had the highest hit number of contigs homologous to LG1 markers, over half of LG1 markers were homologous to contigs on other wheat chromosomal groups except for 2A/B/D (Supplemental Table S8; Fig. 5), implying active interchromosomal translocation on 1P. Meanwhile, the comparison of marker order implied the intrachromosomal rearrangement on 1P. Together, whether interchromosomal translocation and intrachromosomal rearrangement on 1P were the two of the reasons for the higher frequency (16.6 %) of segregation distortion marker of LG1 remains further clarification in future. Likewise, although 2A/B/D had the highest hit number of contigs homologous to LG2 markers, over half of LG2 markers were homologous to other wheat chromosomal groups except 3A/B/D (Supplemental Table S8; Fig. 5), implying active interchromosomal translocation on 2P. For LG3 to 3A/B/D, LG4 to 4A/B/D, and LG7 to 7ABD, obtaining more information about genetic positions of wheat contigs will be helpful for further syntenic analysis. Recently, de novo transcriptome sequencing of *A. cristatum* has been completed (Zhang et al. 2015). Through the alignment of the sequences of 1023 mapped SLAF markers with the transcriptome sequencing data of *Agropyron*, it has been found that only about 4 % of SLAF markers were hit. Thus in this study SLAF-seq mostly targeted non-coding regions which are less well conserved between species than genes, which can partly explain the relatively low degree of synteny between P and A/B/D/H genomes shown in Fig. 5.

#### Future applications for the genetic map

To our knowledge, this map is the densest genetic linkage map for *Agropyron* so far, though it is not yet saturated, because several big gaps exist in the 7 LGs ranging from 4.2 to 18.4 cM. The integrated map spanned 907.8 cM with an average number of 146 markers and 89 loci per LG and an average distance of 1.5 cM between adjacent loci. The average distance in this map is much less than the 2.47 cM previously reported (Yu et al. 2012). Future applications for the genetic map will be mainly in four aspects. The first, the map is helpful for the identification of homoeologous

chromosomal group of P genome in the wheat-*A. cristatum* addition lines. The wheat-*A. cristatum* 1P, 4P and 6P translocation lines were generated and identified, respectively (Liu et al. 2010; Luan et al. 2010), but many other addition lines are not yet clear. The SLAF-seq of addition lines can be performed using the similar strategy to the mapping population. Through SLAF sequence alignment, the correspondences of SLAFs between addition line and the genetic map can provide valuable clues on the correspondences between addition line and LG. The second, the map provides data useful for QTL fine mapping. Abundant variations of important agronomical traits were observed in F<sub>1</sub> individuals, such as those in spike (Supplemental Fig. S9). It is optimistic that mapping with SLAF markers can identify some QTLs in contribution to important agronomical traits. The third, the result in this study is also beneficial for molecular marker assisted breeding of superior wheat-*A. cristatum* germplasm and the cloning of superior *Agropyron* genes. After obtaining the SLAFs showing strong linkage with target trait, we can design primers based on the SLAF sequences (Supplemental Table S9) and then use them to perform PCR to screen target trait. In future the alignment of SLAF marker sequences with transcriptome and genome sequencing data of *Agropyron* will be beneficial for the identification of *Agropyron* gene resource for the improvement of wheat and the cloning of superior *Agropyron* genes. Finally, because these high-density LGs were constructed based on molecular markers developed at the whole genome level, they will serve as a reference for positioning sequence scaffolds on the physical map to assist the assembling of the P genome in future.

**Author contribution statement** L. L and Y. Z designed the research. Y. Z, J. P. Z, L. H and J. Z performed the research. Y. Z wrote the paper. A. G, X. Y, W. L and X. L participated in the preparation of the reagents and materials in this study.

**Acknowledgments** This work was supported by the grants from the National Basic Research Program of China (973 Grant No. 2011CB100104), the National Natural Science Foundation of China (Grant No. 31471493), and the National High Technology Research and Development Program of China (863 Grant No. 2011AA100101).

## References

- Chen D, Zhang J, Wang J, Yang X, Liu W, Gao A, Li X, Li L (2012) Inheritance and availability of high grain number per spike in two wheat germplasm lines. *J Integr Agr* 11: 1409–1416
- Chen S, Ma X, Zhang X, Huang L, Zhou J (2013) Genetic diversity and relationships among accessions of five crested wheatgrass species (Poaceae: *Agropyron*) based on gliadin analysis. *Genet Mol Res* 12:5704
- Dewey DR (1984) The genomic system of classification as a guide to intergeneric hybridization with the perennial Triticeae. In: Gustafson JP (ed) Gene manipulation in plant improvement, 16th Stadler Genetics Symposium. Springer, US, New York, pp 209–279
- Dong Y, Zhou R, Xu S, Li L, Cauderon Y, Wang R (1992) Desirable characteristics in perennial Triticeae collected in China for wheat improvement. *Hereditas* 116:175–178
- Ford-Lloyd BV, Schmidt M, Armstrong SJ, Barazani O, Engels J, Hadas R, Hammer K, Kell SP, Kang D, Khoshbakht K (2011) Crop wild relatives—undervalued, underutilized and under threat? *Bioscience* 61:559–565
- Han H, Bai L, Su J, Zhang J, Song L, Gao A, Yang X, Li X, Liu W, Li L (2014) Genetic rearrangements of six wheat-*Agropyron cristatum* 6P addition lines revealed by molecular markers. *PLoS ONE* 9:e91066
- Hu C-Y, Lee T-C, Tsai H-T, Tsai Y-Z, Lin S-F (2013) Construction of an integrated genetic map based on maternal and paternal lineages of tea (*Camellia sinensis*). *Euphytica* 191:141–152
- Hyten D, Cannon S, Song Q, Weeks N, Fickus E, Shoemaker R, Specht J, Farmer A, May G, Cregan P (2010) Highthroughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genom* 11:38
- Kent WJ (2002) BLAT—the BLAST-like alignment tool. *Genome Res* 12:656–664
- Li L, Yang X, Li X, Dong Y, Chen X (1998) Introduction of desirable genes from *Agropyron cristatum* into common wheat by intergeneric hybridization. *Sci Agric Sinica* 31:1–6
- Liu W, Luan Y, Wang J, Wang X, Su J, Zhang J, Yang X, Gao A, Li L (2010) Production and identification of wheat-*Agropyron cristatum* (1.4P) alien translocation lines. *Genome* 53:472–481
- Liu D, Ma C, Hong W, Huang L, Liu M, Liu H, Zeng H, Deng D, Xin H, Song J (2014) Construction and analysis of high-density linkage map using high-throughput sequencing data. *PLoS ONE* 9:e98855
- Luan Y, Wang X, Liu W, Li C, Zhang J, Gao A, Wang Y, Yang X, Li L (2010) Production and identification of wheat-*Agropyron cristatum* 6P translocation lines. *Planta* 232:501–510
- Ochoa V, Madrid E, Said M, Rubiales D, Cabrera A (2015) Molecular and cytogenetic characterization of a common wheat-*Agropyron cristatum* chromosome translocation conferring resistance to leaf rust. *Euphytica* 201:89–95
- Pfender W, Saha M, Johnson E, Slabaugh M (2011) Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theor Appl Genet* 122:1467–1480
- Porebski S, Bailey LG, Baum BR (1997) Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol Biol Rep* 15:8–15
- Qi Z, Huang L, Zhu R, Xin D, Liu C, Han X, Jiang H, Hong W, Hu G, Zheng H (2014) A high-density genetic map for soybean based on specific length amplified fragment sequencing. *PLoS ONE* 9:e104871
- Song L, Jiang L, Han H, Gao A, Yang X, Li L, Liu W (2013) Efficient induction of wheat-*Agropyron cristatum* 6P translocation lines and GISH detection. *PLoS ONE* 8:e69501
- Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: Join Map. *Plant J* 3:739–744
- Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, Jiang C, Guan N, Ma C, Zeng H (2013) SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE* 8:e58700

- 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
- Wu J, Yang X, Wang H, Li H, Li L, Li X, Liu W (2006) The introgression of chromosome 6P specifying for increased numbers of florets and kernels from *Agropyron cristatum* into wheat. *Theor Appl Genet* 114:13–20
- Wu M, Zhang J, Wang J, Yang X, Gao A, Zhang X, Liu W, Li L (2010) Cloning and characterization of repetitive sequences and development of SCAR markers specific for the P genome of *Agropyron cristatum*. *Euphytica* 172:363–372
- Yu X, Li X, Ma Y, Yu Z, Li Z, Gulick P (2012) A genetic linkage map of crested wheatgrass based on AFLP and RAPD markers. *Genome* 55:327–335
- Zhang Y, Wang L, Xin H, Li D, Ma C, Ding X, Hong W, Zhang X (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
- Zhang J, Liu W, Han H, Song L, Bai L, Gao Z, Zhang Y, Yang X, Li X, Gao A, Li L (2015) *De novo* transcriptome sequencing of *Agropyron cristatum* to identify available gene resources for the enhancement of wheat. *Genomics*. doi:10.1016/j.ygeno.2015.04.003