DNA Research, 2017, 24(3), 289–301
doi: 10.1093/dnares/dsw063
Advance Access Publication Date: 27 January 2017
Full Paper

Full Paper

Integrated high-density consensus genetic map of *Pyrus* and anchoring of the 'Bartlett' v1.0 (*Pyrus communis*) genome

Leiting Li[1], Cecilia H. Deng[2], Mareike Knäbel[2], David Chagné[2], Satish Kumar[2], Jiangmei Sun[1], Shaoling Zhang[1], and Jun Wu[1,*]

[1]State Key Laboratory of Crop Genetics and Germplasm Enhancement, Centre of Pear Engineering Technology Research, Nanjing Agricultural University, Nanjing, China, and [2]The New Zealand Institute for Plant & Food Research Limited (Plant & Food Research), New Zealand

*To whom correspondence should be addressed. Tel. +86 25 84396485, Fax. +86 25 84396485.
Email: wujun@njau.edu.cn

Edited by Dr. Sachiko Isobe

Received 28 August 2016; Editorial decision 15 December 2016; Accepted 16 December 2016

Abstract

Genetic maps are essential tools for pear genetics and genomics research. In this study, we first constructed an integrated simple sequence repeat (SSR) and single nucleotide polymorphism (SNP)-based consensus genetic map for pear based on common SSR markers between nine published maps. A total of 5,085 markers, including 1,232 SSRs and 3,853 SNPs, were localized on a consensus map spanning 3,266.0 cM in total, with an average marker interval of 0.64 cM, which represents the highest density consensus map of pear to date. Using three sets of high-density SNP-based genetic maps with European pear genetic backgrounds, we anchored a total of 291.5 Mb of the 'Bartlett' v1.0 (*Pyrus communis* L.) genome scaffolds into 17 pseudo-chromosomes. This accounted for 50.5% of the genome assembly, which was a great improvement on the 29.7% achieved originally. Intra-genome and inter-genome synteny analyses of the new 'Bartlett' v1.1 genome assembly with the Asian pear 'Dangshansuli' (*Pyrus bretschneideri* Rehd.) and apple (*Malus* × *domestica* Borkh.) genomes uncovered four new segmental duplication regions. The integrated high-density SSR and SNP-based consensus genetic map provided new insights into the genetic structure patterns of pear and assisted in the genome assembly of 'Bartlett' through further exploration of different pear genetic maps.

Key words: pear, genetic map, genome assembly, SNP, SSR

1. Introduction

Pear (*Pyrus* spp.) is an important and popular rosaceous fruit crop cultivated in temperate regions for its economic and nutritional value. In 2014, the world production of pear was 25.8 million tonnes, from a harvest area of 1.6 million ha (http://faostat3.fao.org, 31 December 2016, date last accessed), making it the second most important fruit crop in the Maloideae, after apple. Improving pear germplasm is necessary for fruit breeders to meet consumer preferences and adapt to variable cultivation conditions. Increasingly, traditional breeding approaches are complemented by molecular breeding to take advantage of its efficiency in seedling selection and breeding for complex traits[1] using genomic

selection. Essential molecular breeding tools include molecular markers, genetic maps, and whole genome sequences. These tools have many applications in pear, including genetic diversity analyses[2–4] and quantitative trait loci (QTL) mapping for production traits and fruit quality.[5–12]

Prototypes of pear genetic maps used marker techniques, such as random amplified polymorphic DNA (RAPD),[13] amplified fragment length polymorphism (AFLP) and simple sequence repeat (SSR).[14–17] Among different types of markers, SSR markers have been widely used in constructing pear genetic maps because of their co-dominant and transferable nature,[18] thus can be used as common markers to compare and construct consensus maps. From 2007, various pear genetic maps were constructed based on SSR markers.[16,17,19–21] However, the map with the highest density[19] consisted of only 734 markers and covered 1,661.4 cM, which makes it difficult to facilitate map-based gene cloning. The construction of these maps were limited to the available molecular markers, either developed from pear expressed sequence tags (ESTs) or transferred from related species, both of which are labour- and time-consuming.

When the cost of next-generation sequencing (NGS) decreased, pear whole genome sequences were released,[22] and single nucleotide polymorphism (SNP) markers were applied rapidly for genetic map construction.[10,20,23,24] The first SNP-based pear genetic map had 857 markers,[23] but subsequently a 'Bayuehong (BYH)' × 'Dangshansuli' map with 3,143 SNP markers was developed with the restriction site-associated DNA sequencing (RADseq) technology and 98 SSR markers.[10] A 'Hosui' (*Pyrus pyrifolia*) map of 609 SNP markers and 252 SSR markers was developed from transcriptome data.[24] Compared with previous genetic maps of pear, SNP-based maps significantly increased the number of genetic markers covering the whole genome.

High-quality genetic maps are not only valuable tools for QTL mapping, but can also serve to anchor genome assembly scaffolds into pseudo-chromosomes.[25] The draft genome of 'Dangshansuli' (*Pyrus bretschneideri* Rehd.)[22] was reported as highly heterozygous, and assembled with a ×194 coverage of BAC-by-BAC (bacterial artificial chromosome) and NGS. The assembly consisted of 512.0 Mb sequences and 2,103 scaffolds, corresponding to 97.1% of the estimated genome size. From the assembly, 386.7 Mb (or 75.5%), was anchored into 17 pseudo-chromosomes corresponding to the basic chromosome number of pear ($x = 17$), using a genetic map of 2,005 SNP markers. Concurrently, the draft genome of the European pear 'Bartlett' (*Pyrus communis* L.) was assembled with ×11.4 coverage NGS data and consisted of 142,089 scaffolds covering 577.3 Mb in the assembled genome,[26] accounting for most of the expected 600 Mb genome size. However, using 2,279 genetically mapped loci (1,391 apple and 888 pear SNPs), only 171.4 Mb (29.7%) of these scaffolds were anchored to linkage groups (LGs), due to the relatively fragmented scaffold sequences and an insufficient genetic map used to anchor the scaffolds. Therefore, it is important and necessary to develop higher quality genetic maps to help improve the scaffolds anchoring ratio of the 'Bartlett' v1.0 genome assembly.

For this study, we collected nine published pear genetic maps to build an integrated SSR and SNP-based pear consensus genetic map, which was used for map comparisons. Additionally, a set of three high-quality SNP-based pear genetic maps with European pear genetic backgrounds was used for anchoring the 'Bartlett' v1.0 genome scaffolds. The new 'Bartlett' v1.1 genome assembly was further used for synteny analyses.

2. Materials and methods

2.1. Construction of a new genetic map for 'Old Home (OH)' × 'Louise Bon Jersey (LBJ)'

A new OHxLBJ genetic map was constructed using SNPs derived from genotyping-by-sequencing (GBS).[27] The two European pear cultivars 'OH' and 'LBJ' were used to create a segregation population of 297 progenies at The New Zealand Institute for Plant & Food Research (PFR).[5,23] Total DNA was extracted from young leaves of sixty F1 trees in their third year of growth after grafting using the CTAB method.[28] The GBS libraries were prepared using restriction enzyme *Bam*HI following a modified version of the protocol developed by Elshire et al.[27] with the following modifications: one microgram of DNA was used for the restriction digestion; adapter annealing was done following the protocol of Ko et al.[29]; the adapter ligation step was performed following digestion without drying out the DNA/adaptor mixture; high fidelity enzyme (AccuPrime Taq DNA polymerase High Fidelity, Life Technologies) was used for amplification; the libraries were amplified and cleaned up prior to pooling. The pooled GBS libraries were sequenced on an Illumina HiSeq2000 platform in parallel in two lanes in single-end mode with reads length of 101 bases. The average read count per library was 1.85 million and 1.86 million for the two lanes. The reads were trimmed from 3′ and 64 bases long fragments were kept after demultiplexing. The Phred score for bases in the trimmed reads were higher than 20, so no read was discarded after trimming. The sequencing quality was checked with FastQC version 0.11.2.[30] Genotyping was carried out using TASSEL/GBS v5 pipeline,[31] using the 'Bartlett' v1.0 genome assembly as a reference genome. For SNP discovery, we requested at least eight supporting reads per site. Mapping of tags (64-base long unique genomics sequences) to the reference genome was performed using Bowtie2 v2.2.5[32] in—end-to-end and—very-sensitive modes. To get the flanking sequences, 250 nucleotides at either side of an SNP site were extracted from the reference genome. Additionally, SNP markers were considered of good quality when they segregated in a Mendelian manner: ABxAA (1:1 ratio), ABxAB (1:2:1 ratio) and AAxAB (1:1 ratio).[33] All SNPs with missing data for >10 individuals were removed from the analysis. A merged genetic map for both parents was constructed using the package Onemap[34] version 2.0-4 in R-studio (R Core Team, 2016). A LOD (logarithm of odds) score of five or higher was used for grouping and the genetic maps were calculated using the Kosambi function. LG numbers were assigned by comparing markers in the new OHxLBJ with the SNP array-based OHxLBJ map published by Montanari et al.[23]

2.2. Collection of pear genetic maps and genome resources

Nine previously published genetic maps were selected to construct an integrated pear consensus genetic (I-PCG) map using common SSR markers among maps as shown in Fig. 1A.[10,16,17,19–21,23,24] These maps were generated from four pear populations, including 'BYH' × 'Dangshansuli' (two maps), 'Bartlett' × 'Hosui' (four maps), 'Shinsei' × 282-12 ('Hosui' × 'La France') (one map), and 'Akiakari' × 'Taihaku' (two maps). 'BYH' is a descendant of European pear 'Clapp's Favorite' and Chinese pear 'Zaosuli'. 'Dangshansuli' is a native Chinese pear cultivar. The convention of SSR marker names was not exactly the same, especially for those SSRs indicating multiple loci. For instance, TsuGNH235-1 and TsuGNH235_m1 should be the same marker. Thus, we scanned all

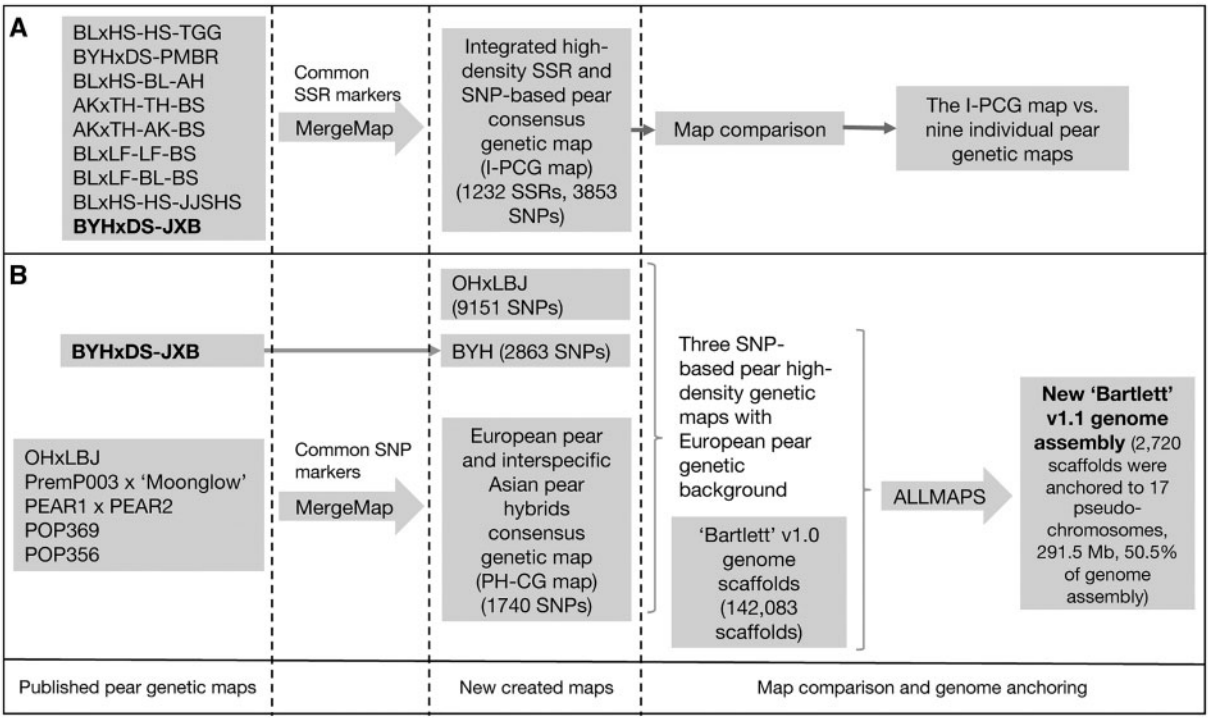

Figure 1. Workflow for the integrated SSR and SNP-based pear consensus genetic (I-PCG) map construction and 'Bartlett' v1.0 genome scaffolds anchoring.

Table 1. Pear genetic maps used to create the integrated SSR and SNP-based pear consensus genetic (I-PCG) map using common SSR markers as bridging markers

Map ID	Parents	Number of markers	Number of SSRs	Number of SNPs	Number of linkage groups	Length (cM)	Average marker interval (cM)	References code
BYHxDS-JXB[a]	'Bayuehong' × 'Dangshansuli'	3,241	98	3,143	17	2,243.4	0.70	10
HS-TGG[b]	'Bartlett' × 'Hosui'	922	252	670	22	1,341.9	1.49	24
BYHxDS-PMBR[c]	'Bayuehong' × 'Dangshansuli'	734	734	0	17	1,661.4	2.32	19
BL-AH[d]	'Bartlett' × 'Hosui'	467	400	67	17	965.0	2.14	20
TH-BS[e]	'Akiakari' × 'Taihaku'	275	275	0	17	1,039.0	4.03	21
AK-BS[f]	'Akiakari' × 'Taihaku'	208	208	0	18	799.0	4.20	21
LF-BS[g]	'Shinsei' × 282-12 ('Hosui' × 'La France')	134	134	0	17	1,156.0	9.88	16
BL-BS[h]	'Hosui' × 'Bartlett'	118	118	0	17	1,000.0	9.90	16
HS-JJSHS[i]	'Bartlett' × 'Hosui'	105	105	0	17	1,174.0	13.34	17

SSR marker names and renamed them if necessary. The map IDs were named based on their cultivar abbreviations and publication abbreviations (detail information is listed in Table 1).

Three high-density SNP-based genetic maps (Fig. 1B) with European pear genetic backgrounds were used for anchoring the 'Bartlett' v1.0 genome scaffolds, while other genetic maps with Asian pear (*P. bretschneideri* or *P. pyrifolia*) genetic backgrounds were not involved in the anchoring because of the large differences in SNPs and their low efficiency of transferability among different species. These three maps included the newly created genetic map OHxLBJ, a 'BYH' map extracted from the 'BYH' × 'Dangshansuli' consensus map (named BYHxDS-JXB),[10] and a European pear and interspecific Asian pear hybrid consensus genetic (PH-CG) map attained from the 10 genetic maps published by Montanari et al.[23] derived from five populations, including OHxLBJ, PremP003 × 'Moonglow' (T003xM), PEAR1 × PEAR2 (T042xT081),[9] POP369 (T052xT003), and POP356 (T052xT064).[8] The BYH map was a subset of BYHxDS-JXB[10] with marker names start with Pyb and Pybd, representing SNPs derived from 'BYH' and both parents, respectively.

The I-PCG map was not used for anchoring the 'Bartlett' v1.0 genome scaffolds (Fig. 1) because the consensus genetic map had a mixed genetic background of *P. bretschneideri*, *P. pyrifolia* and *P. communis*, which could cause more unpredictable conflicts owing to genome variation. The OHxLBJ and PH-CG maps were not merged into the I-PCG map (Fig. 1) because they were SNP-based maps, and markers in these maps were all unique and had no common markers with the I-PCG map for anchoring the corresponding LGs.

Pseudo-chromosome sequences of 'Dangshansuli' genome v1.0 were retrieved from the Pear Genome Project website (http://pearge nome.njau.edu.cn/default.asp?d=4&m=2, 31 December 2016, date last accessed); scaffold sequences of the 'Bartlett' v1.0 genome (https://www.rosaceae.org/species/pyrus/pyrus_communis/genome_

Table 2. Number of pairwise common SSR markers for nine pear genetic maps used for creating the integrated SSR and SNP-based pear consensus genetic (I-PCG) map

Map ID	BYHxDS-JXB	HS-TGG	BYHxDS-PMBR	BL-AH	TH-BS	AK-BS	LF-BS	BL-BS
HS-TGG	42	–	–	–	–	–	–	–
BYHxDS-PMBR	63	45	–	–	–	–	–	–
BL-AH	52	153	65	–	–	–	–	–
TH-BS	31	80	38	139	–	–	–	–
AK-BS	26	87	34	105	115	–	–	–
LF-BS	43	56	59	95	62	46	–	–
BL-BS	44	53	45	112	54	39	95	–
HS-JJSHS	36	94	36	61	39	36	55	53

v1.0, 31 December 2016, date last accessed); and pseudo-chromosome sequences of the 'Golden Delicious' genome version 1.0 primary (https://www.rosaceae.org/species/malus/malus_x_domestica/genome_v1.0p, 31 December 2016, date last accessed) were retrieved from the Genome Database of Rosaceae (https://www.rosaceae.org/).

2.3. Construction and comparison of pear consensus genetic maps

MergeMap,[35] a software that uses common markers from different maps as bridging markers to merge maps, was used to create the I-PCG map from nine published maps (Fig. 1A) and the PH-CG map (Fig. 1B). To prepare input data for Mergemap, we firstly removed conflict markers that presented in inconsistent LGs. Secondly, each individual LG in the input genetic maps for MergeMap was filtered using three criteria: number of markers in a LG not <3, LG length not <5 cM, and maximum common markers of a LG with other maps not <3. Thirdly, for the I-PCG map, each individual map was set as equal weights in the map merging process, while for the PH-CG map, individual maps were set based on the genetic background of European pear, weight = 4 if an individual map was *P. communis*, weight = 2 if one of its parents was *P. communis*, and weight = 1 if one of its grandparents was *P. communis*. Lastly, Mergemap tried to resolve conflicts in map positions among the individual maps by deleting a minimum set of marker occurrences.[35] Genetic map comparison between the I-PCG and individual maps were visualized by the Strudel[36] software.

Most maps for constructing the I-PCG map consisted of 17 LGs, which was the same as the basic chromosome number of pear ($x = 17$), except for two maps (HS-TGG and AK-BS) (Table 1). To confirm whether these maps were appropriate for merging into one consensus map, we firstly calculated the number of common SSR markers between these maps (Table 2). Each map was found to have at least 26 SSR markers in common with any other map. For instance, the number of common SSR markers for AK-BS vs. TH-BS, AK-BS vs. BL-AH and TH-BS vs. BL-AH was 115, 105 and 139, respectively.

2.4. Anchoring 'Bartlett' scaffolds and synteny analyses of intra- or inter-genomes

SSR primer sequences and SNP flanking sequences were mapped onto 'Bartlett' scaffolds using isPcr and blat programs, respectively.[37] In either condition, only the best match of each marker was kept, and markers were removed if there were two equally good best matches. The ALLMAPS software, with the strength of computing a scaffold ordering that maximizes collinearity across a collection of maps,[25] was used for anchoring the 'Bartlett' genome with the guidance of three high-density SNP-based genetic maps with European pear genetic backgrounds as mentioned above. The multiple collinearity scan toolkit (MCScanX)[38] was used to perform intra-genome synteny analyses for the 'Bartlett' genome, and inter-genome analyses for the 'Bartlett' versus 'Dangshansuli' genomes, and 'Bartlett' genome versus the genome of apple cultivar 'Golden Delicious'.

3. Results

3.1. Merging multiple genetic maps leads to a high quality and density integrated SSR and SNP-based consensus map of pear

Nine published pear genetic maps (Table 1) constructed mainly using SSR and SNP markers were integrated into a consensus genetic map I-PCG, using SSR markers in common and the software MergeMap. The number of markers for these maps ranged from 105 for HS-JJSHS[17] to 3,241 for BYHxDS-JXB.[10] Among these, BYHxDS-JXB had the most SNP markers (3,143) and BYHxDS-PMBR[19] had the most SSR markers (734). The integrated SSR and SNP-based pear consensus genetic map (I-PCG map) consisted of 17 LGs made up of 5,085 markers in total (1,232 SSRs, and 3,853 SNPs), spanning 3,266 cM, with an average marker interval of 0.64 cM (Table 3; markers listed in Supplementary Table S1). As shown in Fig. 2, a large proportion (75.8%) of the markers in the map were SNPs, from a minimum of 63.3% on LG 16 (abbreviated as LG16, same as below), to 90.2% on LG5. SSR markers were interspersed with SNP markers in almost all parts of the 17 LGs, except about 75.5 cM at the lower end of LG5, which contained only one SSR marker, and the top 56.5 cM of LG12, which contained no SNP markers. The LG length varied from the shortest 87.7 cM (LG4) to the longest 269.3 cM (LG15). Among the 17 LGs, only LG4 and LG7 (96.6cM) were shorter than 100cM, and eight LGs (LG3, LG9, LG10, LG11, LG12, LG14, LG15, and LG17) were longer than 200cM. The total genetic length of 3,266 cM makes I-PCG the longest pear consensus genetic map being constructed to date. On the other hand, the large number of markers in this genetic map resulted in the smallest average marker interval of 0.64 cM in pear. In addition, locus clusters were observed in all LGs (Table 3), which were defined as 5 cM regions with not <10 markers.[39] In total, 3,776 markers in 237 clusters accounting for 1,185 cM length were identified in the I-PCG map (Supplementary Table S1), representing 74.3% of markers and varying from 32 markers (3 clusters) on LG1 to 471 markers (27 clusters) on LG15.

Table 3. Summary of the integrated SSR and SNP-based pear consensus genetic (I-PCG) map

Linkage group	Number of markers	Number of SSRs	Number of SNPs	Number of common SSR markers	Length (cM)	Average marker interval (cM)	Length of the locus cluster	Marker number in the locus clusters
1	124	38	86	20	147.9	1.20	15	32
2	367	90	277	29	198.6	0.54	100	324
3	374	82	292	29	247.6	0.66	90	278
4	169	51	118	15	87.7	0.52	55	146
5	325	32	293	19	185.5	0.57	65	248
6	318	69	249	19	158.3	0.50	75	257
7	206	37	169	20	96.6	0.47	55	178
8	273	56	217	20	168.0	0.62	65	212
9	265	66	199	32	224.2	0.85	55	173
10	373	96	277	34	244.3	0.66	80	274
11	401	86	315	25	230.1	0.58	100	326
12	247	86	161	19	206.4	0.84	45	146
13	255	58	197	21	165.3	0.65	50	161
14	285	84	201	33	211.4	0.74	75	198
15	562	131	431	37	269.3	0.48	135	471
16	207	76	131	21	180.3	0.88	50	130
17	334	94	240	32	244.5	0.73	75	222
Total	5,085	1,232	3,853	425	3,266	0.64	1,185	3,776

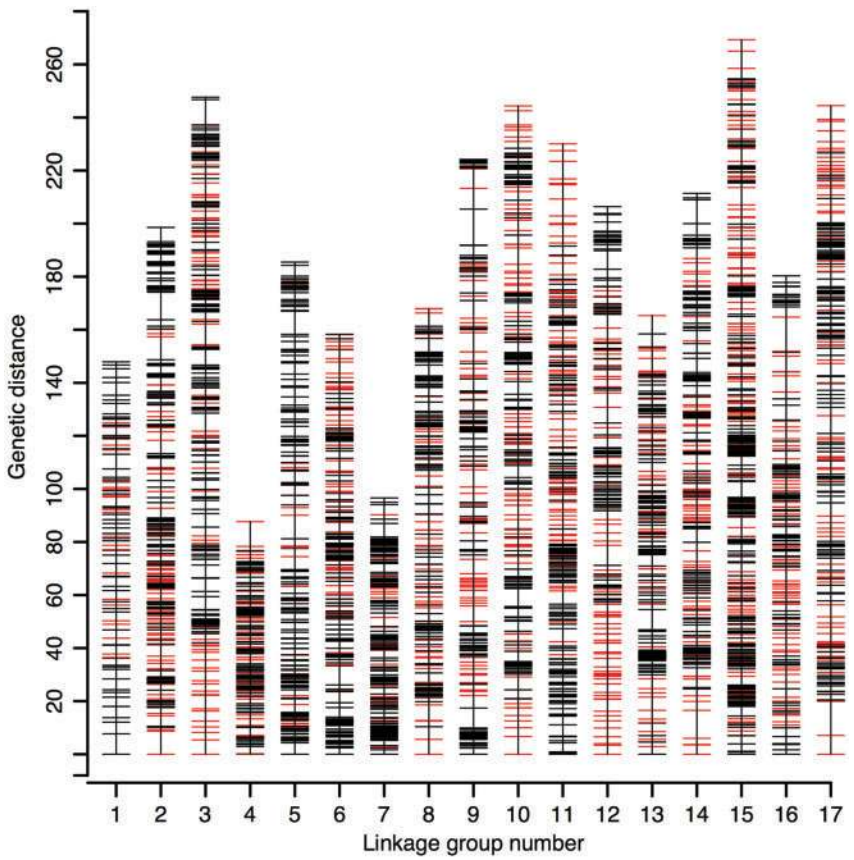

Figure 2. Distribution of SNP and SSR markers on 17 linkage groups (LGs) of the integrated SSR and SNP-based pear consensus genetic (I-PCG) map. Red bars indicate SSR markers, and black bars indicate SNP markers. *X*-axis indicates LG number, and *Y*-axis LG length in centiMorgans (cM). This figure is available in black and white in print and in colour at DNA Research online.

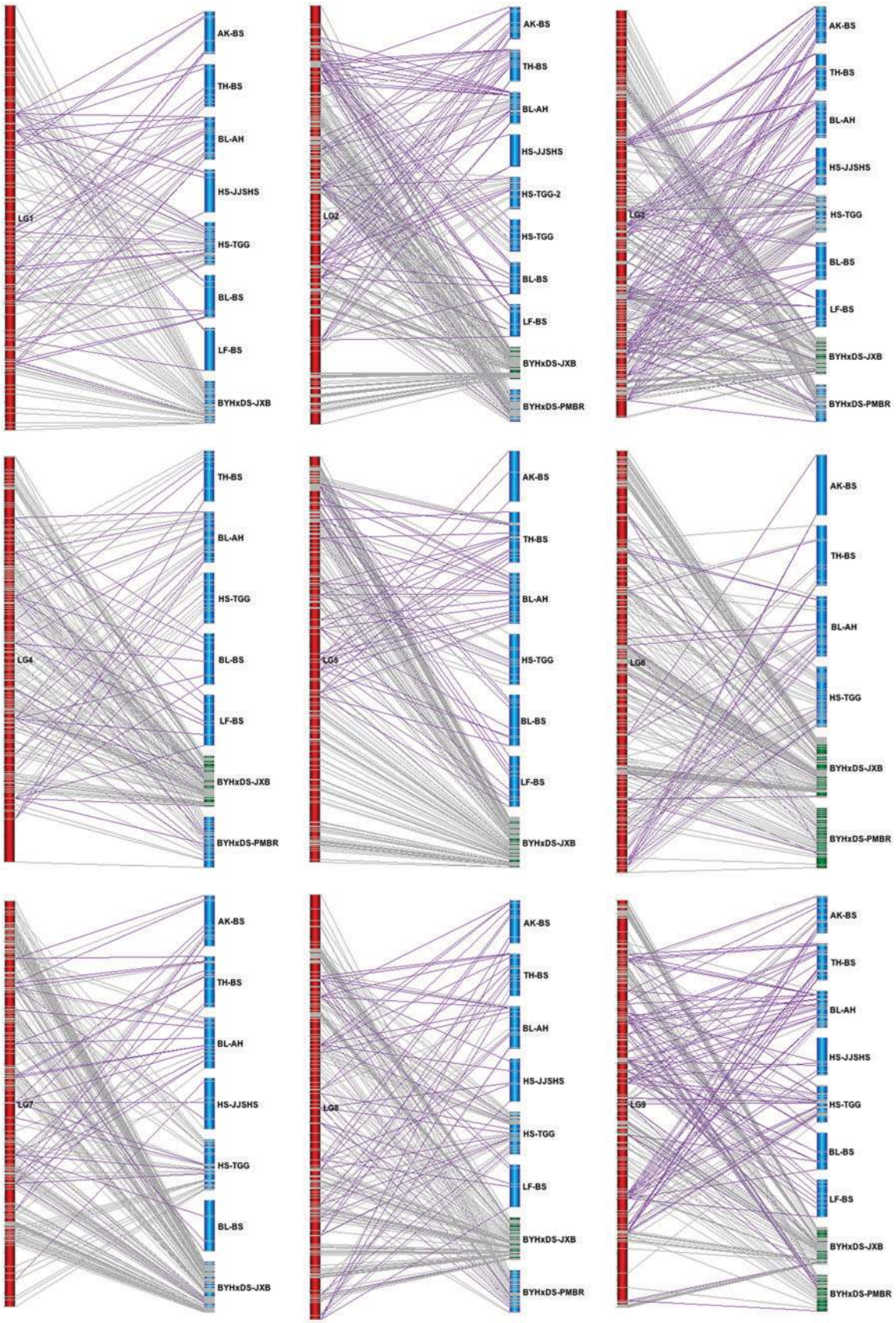

Figure 3. Alignment of the integrated SSR and SNP-based pear consensus genetic (I-PCG) map and the nine individual maps. Red bars on the left denote linkage groups (LGs) of the I-PCG map, and blue/green bars on the right represent the corresponding LGs for individual maps. Blue and green colours indicate forward and reverse orientation, respectively. Purple lines connecting the I-PCG map and individual maps indicate common markers (present in at least two individual maps), and grey lines indicate non-common markers. LG length is relative to total length. This figure is available in black and white in print and in colour at DNA Research online.

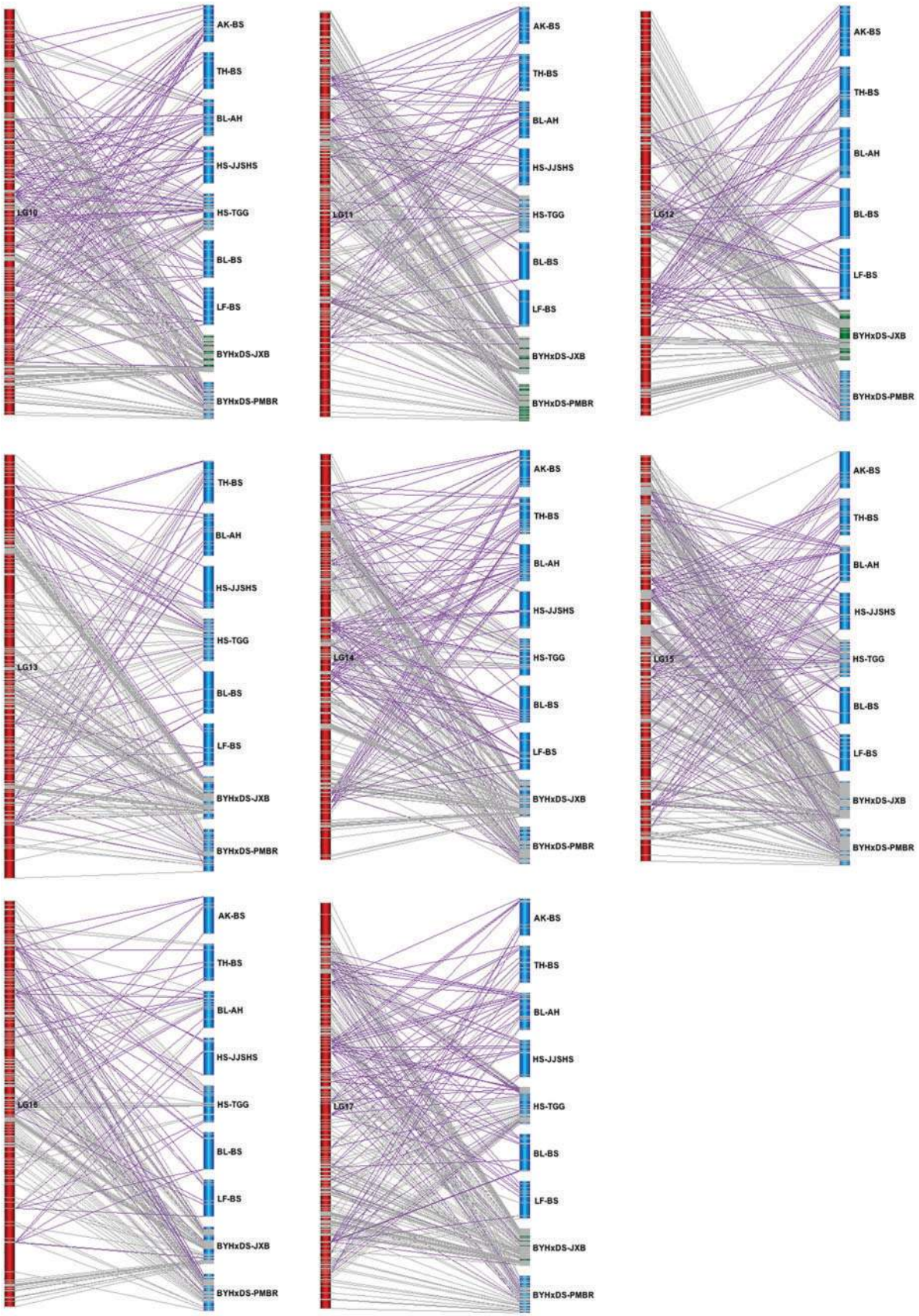

Figure 3. Continued.

Table 4. High-density SNP-based pear genetic maps with European pear genetic background used for anchoring the 'Bartlett' v1.0 genome scaffolds

Map ID	Number of SNPs	Number of linkage groups	Length (cM)	Average marker interval (cM)	References
OHxLBJ[a]	9,151	17	1,664.6	0.18	This study
BYH[b]	2,863	17	2,080.0	0.73	10
PH-CG[c]	1,740	17	2,773.3	1.61	23

Population: [a]'Old Home' × 'Louise Bon Jersey'.
[b]'Bayuehong' × 'Dangshansuli'.
[c]Consensus genetic map derived from five populations: 'Old Home' × 'Louise Bon Jersey', PremP003 × 'Moonglow' (T003xM), PEAR1 × PEAR2 (T042xT081),[9] POP369 (T052xT003), and POP356 (T052xT064).[8]

Table 5. Summary of the new Bartlett v1.1 genome assembly constructed using the ALLMAPS software

Pseudo-chromosome	Length of pseudo-chromosome (bp)	Number of scaffolds	Length of scaffolds (bp)	Number of markers	Markers from BYH	Markers from PH-CG	Markers from OHxLBJ
Chr1	10,263,956	110	10,253,056	408	2	65	341
Chr2	17,147,276	141	17,133,276	618	212	107	299
Chr3	18,593,297	184	18,574,997	996	188	89	719
Chr4	14,885,393	156	14,869,893	703	95	60	548
Chr5	20,478,923	210	20,458,023	847	241	97	509
Chr6	15,384,309	167	15,367,709	784	174	79	531
Chr7	16,091,674	154	16,076,374	684	110	71	503
Chr8	16,073,589	130	16,060,689	650	150	58	442
Chr9	14,040,893	127	14,028,293	353	127	82	144
Chr10	20,488,728	212	20,467,628	1,019	196	105	718
Chr11	19,300,883	200	19,280,983	859	209	102	548
Chr12	16,230,967	154	16,215,667	766	144	80	542
Chr13	15,861,480	113	15,850,280	619	126	87	406
Chr14	14,746,025	131	14,733,025	553	146	53	354
Chr15	28,200,744	212	28,179,644	1,075	316	127	632
Chr16	15,009,554	146	14,995,054	554	82	61	411
Chr17	19,000,797	173	18,983,597	765	150	81	534
Sum	291,798,488	2,720	291,528,188 (50.5%)[a]	12,253	2,668	1,404	8,181

[a]The total size of 'Bartlett' v1.0 (*Pyrus communis*) scaffold sequences was 577,335,413 bp.

In total, 425 common SSR markers were detected on the nine individual maps, as mentioned above, to construct the I-PCG map (Table 3, Fig. 3). Of these, 200 (47.1%) were derived from apple, 148 (34.8%) from pear genomic sequences, and 77 (18.1%) from pear EST sequences. There was a large proportion of markers from apple because few SSR markers were developed from pear before the whole genome sequences were released. Common markers on the 17 LGs varied from 15 on LG4 to 37 on LG15. Of these markers, 70% were common between two or three maps, and five markers were common on all nine individual maps: NB113a (LG3), CH03g12-2 (LG3), CH02c11 (LG10), CH02b03b (LG10), and CH04h02 (LG11).

3.2. Construction of a new GBS SNP map for the 'OH' × 'LBJ' population

The new OHxLBJ map, constructed using SNPs derived from the GBS approach, consists of 9,151 markers and spans 1,664.6 cM with an average of 0.18 cM distance between markers (Supplementary Table S4 for a summary and Supplementary Table S5 for a detailed list of markers). The SNPs are located on 1,719 'Bartlett' assembly scaffolds. The information on the LG assignment from the OHxLBJ SNP array map[23] enabled the assignment of LGs on the new GBS-based OHxLBJ pear genetic map in this study. The lengths of the LGs showed a wide range, from 12.3 cM (LG8) to 255.0 cM (LG17). In total, five LGs were >100 cM, and another three LGs were <50 cM, indicating that some LGs were saturated, while other LGs were partially covered. In addition, we noticed high-level marker co-segregation in this map; for instance, 403 SNP markers on LG1 were located in 24 unique genetic positions, suggesting 16.8 markers per unique genetic position for LG1. Overall, there were 428 unique genetic positions on this GBS-based OHxLBJ map, and on average 21.4 markers per unique genetic position, with an average interval of 4.05 cM.

3.3. Improvement of 'Bartlett' genome assembly anchoring by multiple high-density maps

Since genetic maps were commonly used for anchoring scaffolds into pseudo-chromosomes, we evaluated the potential application of the I-PCG map and three high-density SNP-based pear genetic maps with European pear background to improve the 'Bartlett' v1.0 genome assembly. The PH-CG map[23] consisted of 1,740 markers and spanned 2,773.3 cM (Supplementary Table S2 for a summary,

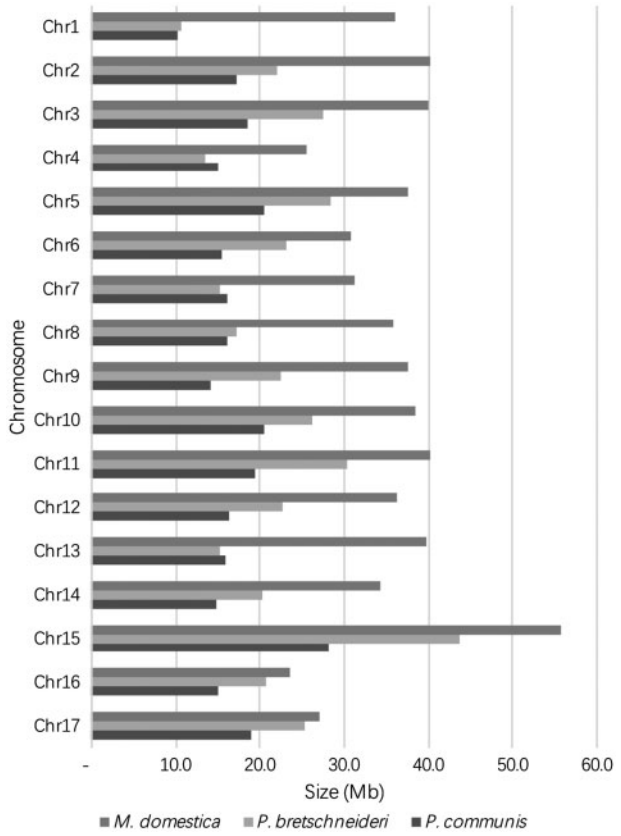

Figure 4. Comparison of pseudo-chromosome length of 'Bartlett' (*P. communis*), 'Dangshansuli' (*P. bretschneideri*), and 'Golden Delicious' (*Malus* × *domestica*) genome assemblies. *X*-axis means pseudo-chromosome size, million bases (Mb) as unit, and *Y*-axis means chromosome number.

Supplementary Table S3 for a detailed list), while the new OHxLBJ consisted of 9,151 markers and spanned 1,664.6 cM (Supplementary Table S4 for a summary and Supplementary Table S5 for a detailed list). A brief overview of the maps is shown in Table 4. Using these maps, a total of 291.5 Mb 'Bartlett' v1.1 genome assembly was obtained in 17 pseudo-chromosomes, increasing the assembly ratio to 50.5% (Table 5 for a summary and Supplementary Table S6 for a list of scaffold ordering). Pseudo-chromosome 1 (abbreviated as Chr1, same as below) was the shortest (10.3 Mb) and Chr15 was the longest (28.2 Mb), which were the same with the 'Dangshansuli' genome: the shortest of 10.7 Mb for Chr1 and the longest of 43.6 Mb for Chr15. In apple, the longest pseudo-chromosome was also Chr15 (55.8 Mb), but the shortest one was Chr16 (23.5 Mb) (Fig. 4).

Synteny analyses for the new 'Bartlett' v1.1 genome assembly against 'Dangshansuli' and 'Golden Delicious' genomes were performed (Fig. 5). Intra-genome synteny analyses of 'Bartlett' genome (Fig. 5A) identified homologous pseudo-chromosome pairs as per previous reports[22,40]: 1–7, 2–7, 2–15, 3–11, 4–12, 5–10, 6–14, 8–15, 9–17, 12–14 and 13–16. Additionally, we identified four segmental duplication regions for pseudo-chromosome pairs 6–13, 9–13, 14–16 and 14–17, which were also confirmed by inter-genome synteny analyses between the 'Bartlett' pear and 'Golden Delicious' apple genomes. Inter-genome synteny analyses between both pear genomes of 'Bartlett' and 'Dangshansuli' (Fig. 5B) showed good collinearity. The near-linear arrangement of homologous regions for all 17 corresponding pseudo-chromosomes was observed. Nonetheless, the arrangements of Chr1, Chr2, Chr3 and Chr5 were not as well aligned as other pseudo-chromosomes. Interestingly, inter-genome synteny analyses between 'Bartlett' and the apple genome of 'Golden Delicious' (Fig. 5C), both in the subfamily Maloideae, showed not only good collinearity, but also homologous relationships, similar to the 'Bartlett' intra-genome synteny analyses. This phenomenon was also observed when inter-genome synteny analysis was performed between 'Dangshansuli' and 'Golden Delicious' (data not shown), indicating the unique genome structure for apple.

3.4. Comparison of the I-PCG map with genome assemblies of pear and apple revealed both collinearity and misalignments

To further evaluate the collinearity between the pear genetic and physical maps, we compared the I-PCG map with the new 'Bartlett' v 1.1 (Fig. 6A), 'Dangshansuli' and apple genome assemblies (Fig. 6B). In total, 4,036, 4,157 and 3,707 markers were mapped onto the 'Bartlett' 'Dangshansuli' and 'Golden Delicious' genome assemblies, and among these, 3,622, 4,069 and 3,707 markers were mapped on pseudo-chromosomes, respectively (Supplementary Table S1). As shown in Fig. 6A, the I-PCG map and the new 'Bartlett' v1.1 genome assembly displayed good collinearity. The uniformity of the genetic map with the physical maps could also be observed in the 'Dangshansuli' and 'Golden Delicious' (Fig. 6B), but with more misalignments (markers were not located on corresponding pseudo-chromosomes). For 'Bartlett', 3,427 markers (94.6% of markers mapped on pseudo-chromosomes) were located in the corresponding pseudo-chromosomes, 3,649 markers (89.7%) for 'Dangshansuli', and 2,117 markers (57.1%) for 'Golden Delicious'. Among these, 3,414 markers were located in corresponding pseudo-chromosomes in at least two of the above three genome assemblies (Supplementary Table S1). Furthermore, misaligned markers between the I-PCG map and 'Bartlett' genome assembly were enriched in five LG/Chr pairs (at least five markers for each LG/Chr pair): 1–7, 9–17, 3–11, 13–16 and 6–14. For the 'Dangshansuli' genome, they were enriched in six LG/Chr pairs (at least 17 markers for each LG/Chr pair): 1–7, 11–15, 8–15, 3–11, 5–11 and 13–16. For 'Golden Delicious', they were enriched in nine LG/Chr pairs (at least 41 markers for each LG/Chr pair): 5–10, 3–11, 1–7, 9–17, 13–16, 8–15, 2–15, 6–14 and 1–15. Overall, comparing the I-PCG map with the 'Dangshansuli', 'Bartlett' and 'Golden Delicious' assemblies confirmed the reliability of the consensus genetic map.

4. Discussion

Common methods to build high-density genetic maps with NGS usually start with genotyping individuals in a hybrid population through reduced-genome complexity technologies, such as RADseq,[41] GBS,[27] and SLAF-Seq.[42] In this study, we merged published pear genetic maps by treating common SSR markers presented in maps from multiple sources as bridging markers and constructed an integrated SSR and SNP-based pear consensus genetic map (the I-PCG map). The advantages of merging genetic maps include integrating multiple population maps, obtaining larger genetic maps, discovering the genome and genetic variation, and resolving conflicts between different maps. Nevertheless, the length of the consensus genetic map could be inflated (>3,000 cM in this study) due to the method used to develop the consensus map, i.e. artifact of the merging process, not true recombination. Thus, the consensus genetic map may be more relevant

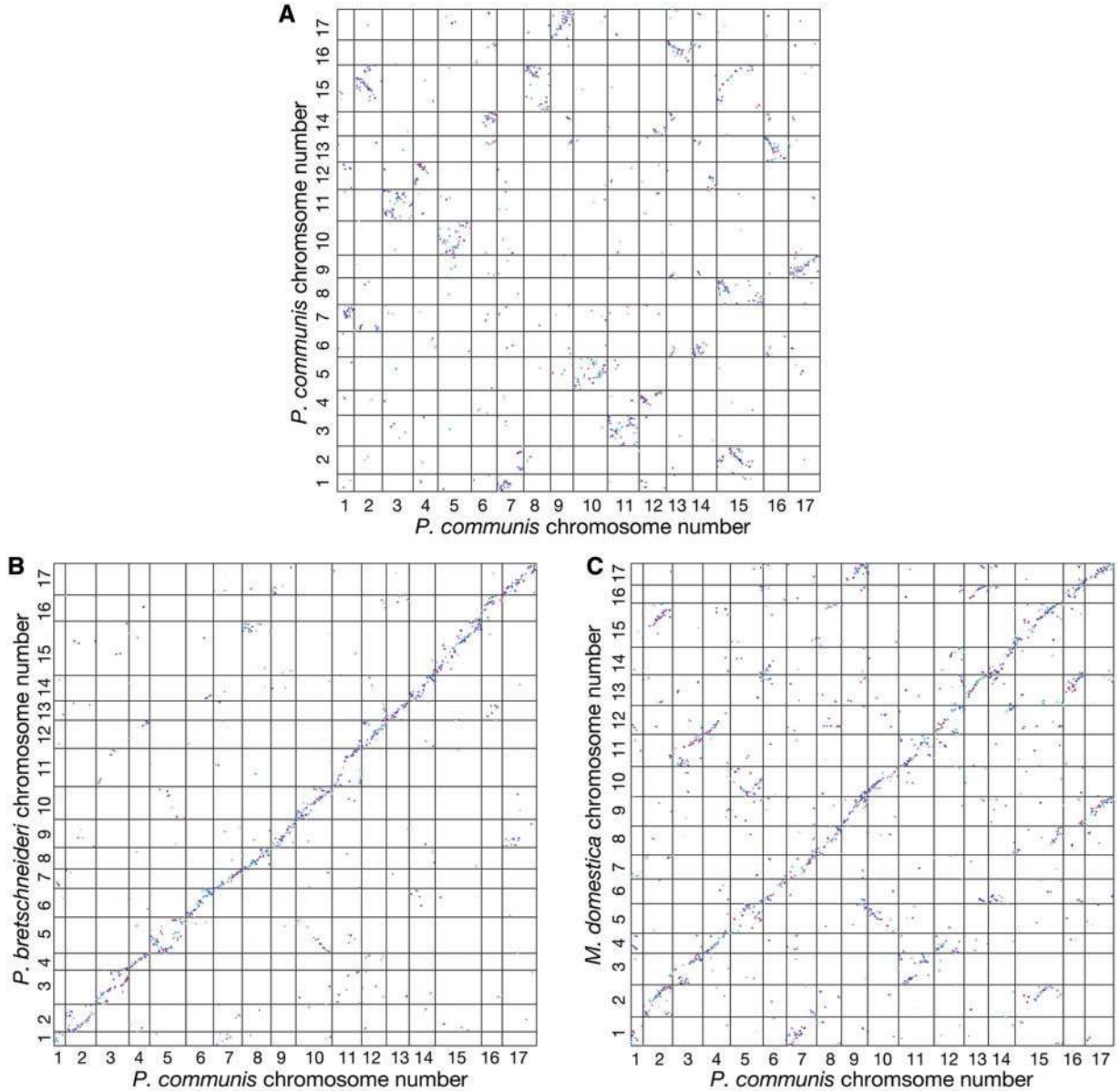

Figure 5. (A) Intra-genome synteny for the 'Bartlett' (*P. communis*) v1.1 pseudo-chromosomes; (B) Inter-genome synteny for the 'Bartlett' v1.1 and 'Dangshansuli' (*P. bretschneideri*) pseudo-chromosomes; (C) Inter-genome synteny for the 'Bartlett' v1,1 and 'Golden Delicious' (*Malus* × *domestica*) pseudo-chromosomes. This figure is available in black and white in print and in colour at DNA Research online.

to the relative positioning of markers than the absolute distances between them.[43]

Anchoring scaffold sequences into pseudo-chromosomes requires both high-quality scaffold sequences and high-quality genetic maps.[25,44] In this study, we selected three high-quality maps with European pear genetic backgrounds for re-anchoring the 'Bartlett' scaffolds. The constructed pseudo-chromosomes captured 291.5 Mb of the 'Bartlett' genome sequence and dramatically increased the anchored 'Bartlett' genome from 29.7% in the original assembly to 50.5% in version 1.1. The improvement of a larger ratio for 'Bartlett' genome assembly was due to the higher density genetic maps used compared with the original assembly, which contained 1,391 *Malus* SNPs, and 888 *P. communis* SNPs.[26] In this study, we used a 5.4-fold number of SNP markers, including 2,668 from the BYH map, 1,404 from the PH-CG map, and 8,181 from the OHxLBJ map (Table 5), ensuring the capture of more 'Bartlett' scaffolds. Synteny analyses of the new 'Bartlett' v1.1 assembly indicated an extensive conserved genome collinearity with *P. bretschneideri* 'Dangshansuli' and *Malus* × *domestica* 'Golden Delicious', indicating that the new 'Bartlett' v1.1 is consistent with its closely related species. Furthermore, the homologous pseudo-chromosome pairing relationships demonstrated by intra-genome synteny analyses were similar with previous studies in 'Dangshansuli'[22] and 'Golden Delicious',[40] indicating that the genome structure of the new 'Bartlett' v1.1 genome assembly was correct. However, we noticed that four pseudo-chromosomes of 'Bartlett' did not have such good collinearity with 'Dangshansuli' and 'Golden Delicious' as other pseudo-chromosomes, namely Chr1, Chr2, Chr3, and Chr5, and

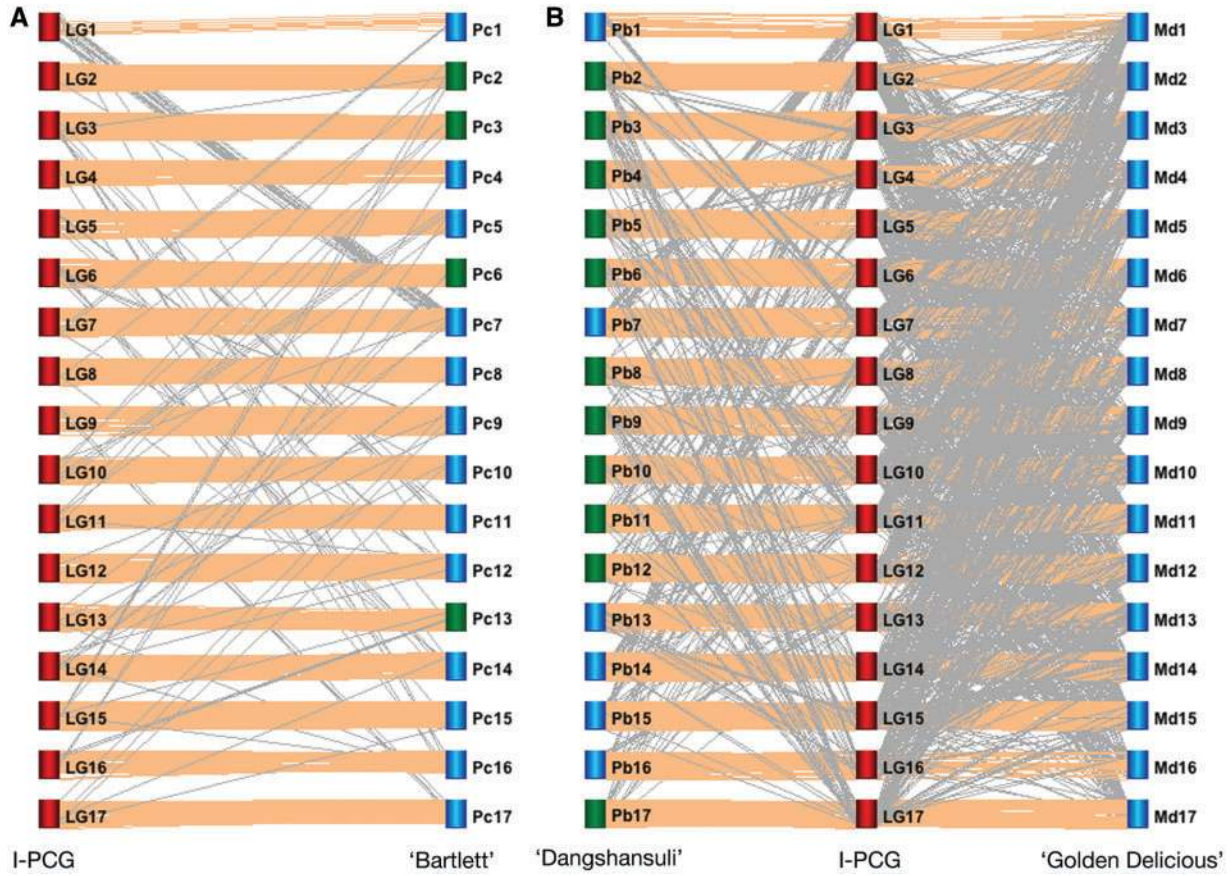

Figure 6. Alignment of the integrated SSR and SNP-based pear consensus genetic (I-PCG) map with (A) 'Bartlett' pseudo-chromosomes (Pc), (B) 'Dangshansuli' (Pb) and apple (Md) pseudo-chromosomes. Red bars indicate linkage groups (LGs) of the I-PCG map. Blue and green bars indicate forward and reverse orientation LGs, respectively. LG length is relative to total length. The yellow lines indicate markers located in the same linkage group/pseudo-chromosome number and the grey lines indicate markers in different linkage group/pseudo-chromosome number. This figure is available in black and white in print and in colour at DNA Research online.

that there were a number of dispersed collinear regions. These were possibly due to differences in the genome structure between 'Bartlett' and the other two genomes. On the other hand, this assembly ratio was still lower than that for 'Dangshansuli', which accounted for 75.5% of the genome. This may be the result of the genetic maps we used not capturing more scaffold sequences, although we used 5.4 times more SNP markers than for the original 'Bartlett' genome assembly. Also, the scaffold N50 length of the 'Bartlett' genome assembly was 88 kb compared with 540.8 kb of the 'Dangshansuli' one, which made it difficult to capture more scaffold sequences.[25] Overall, we significantly improved the 'Bartlett' genome assembly, but more work is required for further improvement, especially a better quality scaffold assembly, which is the target of a recent initiative by an international European pear genome consortium.

The integrated I-PCG map was compared with the improved 'Bartlett' v1.1, 'Dangshansuli'[22] and 'Golden Delicious'[40] genome assemblies. Both collinearity and misalignment of markers were observed for this comparison, indicating that the I-PCG map we constructed was reliable, but both the pear genetic and physical map are far from being complete. The reason for the misalignment may be repeat content or genome-wide duplication in the pear and apple genomes.[45] The results demonstrated that most misaligned markers were coordinated with homologous pairs identified in the pear and apple genomes. Furthermore, there were 42.9% misaligned markers for the apple genome of 'Golden Delicious', much higher than the 5.4% and 10.3% for the 'Bartlett' and 'Dangshansuli' genomes, respectively, indicating the much difference between pear and apple. In addition, similar homologous pseudo-chromosome pairing relationships were revealed by both comparisons of the consensus map and synteny analyses of the genome assembly of pear, which was the first step for further integration of genetic and physical maps in the future.

In conclusion, we demonstrated that a set of different pear genetic maps is a valuable resource for map integration and genetic comparison, which has been exploited in anchoring genome scaffolds of European pear 'Bartlett' (*P. communis*). It was proved that the abundant common markers could be merged into a consensus map, resulting in an integrated SSR and SNP-based pear consensus genetic map with the highest marker density to date. It consists of 5,085 markers and spans 3,266.0 cM, with an average marker interval of 0.64 cM. Meanwhile, the multiple high-density SNP-based maps enabled a significant improvement (291.5 Mb) in scaffold anchoring to pseudo-chromosomes, which accounted for 50.5% of the 'Bartlett' genome. The genetic and physical map comparison further revealed synteny and variation between European pear and Chinese white pear or apple.

Accession number

PRJNA349767.

Data availability

The sequencing data for the OHxLBJ GBS libraries were deposited at NCBI Sequence Read Archive (SRA) database under BioProject accession number of PRJNA349767. The pseudo-chromosome sequences for the European pear (*P. communis*) 'Bartlett' v1.1 genome were deposited at the Pear Genome Project website (http://pearge nome.njau.edu.cn/default.asp?d=4&m=2).

Conflict of interest

None declared.

Supplementary data

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

Funding

This work was financially supported by the National High-tech R&D Program (863) of China (2013AA102606-02), the National Natural Science Foundation of China (31372045), the Earmarked Fund for China Agriculture Research System (CARS-29), the Science Foundation for Distinguished Young Scientists in Jiangsu Province (BK20150025) and the Six Talent Peaks Project in Jiangsu Province (2014-NY-025).

References

1. Ashraf, M. and Foolad, M. R. 2013, Crop breeding for salt tolerance in the era of molecular markers and marker-assisted selection. *Plant Breed*., **132**, 10–20.
2. Liu, Q. W., Song, Y., Liu, L., et al. 2015, Genetic diversity and population structure of pear (*Pyrus* spp.) collections revealed by a set of core genome-wide SSR markers. *Tree Genet. Genomes*, **11**, 128.
3. Song, Y., Fan, L., Chen, H., et al. 2014, Identifying genetic diversity and a preliminary core collection of *Pyrus pyrifolia* cultivars by a genome-wide set of SSR markers. *Sci. Horticult*., **167**, 5–16.
4. Zhang, M. Y., Fan, L., Liu, Q. Z., et al. 2014, A novel set of EST-derived SSR markers for pear and cross-species transferability in Rosaceae. *Plant Mol. Biol. Rep*., **32**, 290–302.
5. Knäbel, M., Friend, A. P., Palmer, J. W., et al. 2015, Genetic control of pear rootstock-induced dwarfing and precocity is linked to a chromosomal region syntenic to the apple *Dw1* loci. *BMC Plant Biol*., **15**, 1.
6. Montanari, S., Guérif, P., Ravon, E., et al. 2015, Genetic mapping of *Cacopsylla pyri* resistance in an interspecific pear (*Pyrus* spp.) population. *Tree Genet. Genomes*, **11**, 1–14.
7. Montanari, S., Brewer, L., Lamberts, R., et al. 2016, Genome mapping of postzygotic hybrid necrosis in an interspecific pear population. *Hortic. Res*., **3**, 15064.
8. Saeed, M., Brewer, L., Johnston, J., et al. 2014, Genetic, metabolite and developmental determinism of fruit friction discolouration in pear. *BMC Plant Biol*., **14**, 241.
9. Won, K., Bastiaanse, H., Kim, Y. K., et al. 2014, Genetic mapping of polygenic scab (*Venturia pirina*) resistance in an interspecific pear family. *Mol. Breed*., **34**, 2179–2189.
10. Wu, J., Li, L. T., Li, M., et al. 2014, High-density genetic linkage map construction and identification of fruit-related QTLs in pear using SNP and SSR markers. *J Exp. Bot*., **65**, 5771–5781.
11. Zhang, R.-P., Wu, J., Li, X.-G., et al. 2013, An AFLP, SRAP, and SSR genetic linkage map and identification of QTLs for fruit traits in pear (*Pyrus* L.). *Plant Mol. Biol. Rep*., **31**, 678–687.
12. Terakami, S., Moriya, S., Adachi, Y., et al. 2016, Fine mapping of the gene for susceptibility to black spot disease in Japanese pear (*Pyrus pyrifolia* Nakai). *Breed. Sci*., **66**, 271.
13. Iketani, H., Abe, K., Yamamoto, T., et al. 2001, Mapping of disease-related genes in japanese pear using a molecular linkage map with RAPD markers. *Breed. Sci*., **51**, 179–184.
14. Yamamoto, T., Kimura, T., Shoda, M., et al. 2002, Genetic linkage maps constructed by using an interspecific cross between Japanese and European pears. *Theor. Appl. Genet*., **106**, 9–18.
15. Dondini, L., Pierantoni, L., Gaiotti, F., et al. 2005, Identifying QTLs for fire-blight resistance via a European pear (*Pyrus communis* L.) genetic linkage map. *Mol. Breed*., **14**, 407–418.
16. Yamamoto, T., Kimura, T., Terakami, S., et al. 2007, Integrated reference genetic linkage maps of pear based on SSR and AFLP markers. *Breed. Sci*., **57**, 321–329.
17. Terakami, S., Kimura, T., Nishitani, C., et al. 2009, Genetic linkage map of the Japanese pear 'Housui' identifying three homozygous genomic regions. *J. Jpn. Soc. Hortic. Sci*., **78**, 417–424.
18. Fan, L., Zhang, M. Y., Liu, Q. Z., et al. 2013, Transferability of newly developed pear SSR markers to other Rosaceae species. *Plant Mol. Biol. Rep*., **31**, 1271–1282.
19. Chen, H., Song, Y., Li, L. T., et al. 2015, Construction of a high-density simple sequence repeat consensus genetic map for pear (*Pyrus* spp.). *Plant Mol. Biol. Rep*., **33**, 316–325.
20. Yamamoto, T., Terakami, S., Moriya, S., et al. 2013, DNA markers developed from genome sequencing analysis in Japanese pear (*Pyrus pyrifolia*). *Acta Hortic*., **976**, 477–483.
21. Yamamoto, T., Terakami, S., Takada, N., et al. 2014, Identification of QTLs controlling harvest time and fruit skin color in Japanese pear (*Pyrus pyrifolia* Nakai). *Breed. Sci*., **64**, 351–361.
22. Wu, J., Wang, Z., Shi, Z., et al. 2013, The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res*., **23**, 396–408.
23. Montanari, S., Saeed, M., Knabel, M., et al. 2013, Identification of Pyrus single nucleotide polymorphisms (SNPs) and evaluation for genetic mapping in European pear and interspecific Pyrus hybrids. *PLoS ONE*, **8**, e77022.
24. Terakami, S., Nishitani, C., Kunihisa, M., et al. 2014, Transcriptome-based single nucleotide polymorphism markers for genome mapping in Japanese pear (*Pyrus pyrifolia* Nakai). *Tree Genet. Genomes*, **10**, 853–863.
25. Tang, H., Zhang, X., Miao, C., et al. 2015, ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biol*., **16**, 3.
26. Chagné, D., Crowhurst, R. N., Pindo, M., et al. 2014, The draft genome sequence of European pear (*Pyrus communis* L. 'Bartlett'). *PLoS ONE*, **9**, e92644.
27. Elshire, R. J., Glaubitz, J. C., Sun, Q., et al. 2011, A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.
28. Doyle J. and Doyle J. L. 1987, Genomic plant DNA preparation from fresh tissue-CTAB method. *Phytochem. Bull*., **19**, 11–15.
29. Ko J. H., Chow K. S. and Han K. H. 2003, Transcriptome analysis reveals novel features of the molecular events occurring in the laticifers of *Hevea brasiliensis* (para rubber tree). *Plant Mol. Biol*., 2003, **53**, 479–492.
30. Andrews, S. 2010, FastQC, Available: http://sourceforge.net/projects/picard/files/picard-tools/, 31 December 2016, date last accessed.
31. Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. and Buckler, E. S. 2007, TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.
32. Langmead, B. and Salzberg, S. L. 2012, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
33. Wu R., Ma C. X., Painter I. and Zeng Z. B. 2002, Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. *Theor. Popul. Biol*., **61**: 349–363.

34. Margarido G. R. A., Souza A. P. and Garcia A. A. F. 2007, OneMap: software for genetic mapping in outcrossing species. *Hereditas*, **144**, 78–79.
35. Wu, Y. H., Close, T. J. and Lonardi, S. 2011, Accurate construction of consensus genetic maps via integer linear programming. *IEEE/ACM Trans. Comput. Biol.*, **8**, 381–394.
36. Bayer, M., Milne, I., Stephen, G., et al. 2011, Comparative visualization of genetic and physical maps with Strudel. *Bioinformatics*, **27**, 1307–1308.
37. Kent, W. J. 2002, BLAT – the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
38. Wang, Y., Tang, H., Debarry, J. D., et al. 2012, MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.*, **40**, e49.
39. Shirasawa, K., Bertioli, D. J., Varshney, R. K., et al. 2013, Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of Arachis and divergence of the legume genomes. *DNA Res.*, **20**, 173–184.
40. Velasco, R., Zharkikh, A., Affourtit, J., et al. 2010, The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nat. Genet.*, **42**, 833–839.
41. Baird, N. A., Etter, P. D., Atwood, T. S., et al. 2008, Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.
42. Sun, X., Liu, D., Zhang, X., et al. 2013, SLAF-seq: an efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE*, **8**, e58700.
43. Khan, M. A., Han, Y. P., Zhao, Y. F., Troggio, M. and Korban, S. S. 2012, A multi-population consensus genetic map reveals inconsistent marker order among maps likely attributed to structural variations in the apple genome. *PLoS ONE*, **7**, e47864.
44. Pop, M., Kosack, D. S. and Salzberg, S. L. 2004, Hierarchical scaffolding with Bambus. *Genome Res.*, **14**, 149–159.
45. Michael, T. P. and VanBuren, R. 2015, Progress, challenges and the future of crop genomes. *Curr. Opin. Plant Biol.*, **24**, 71–81.