

RESEARCH ARTICLE

Open Access

Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using Insertion-Deletion (InDel) and Simple Sequence Repeat (SSR) markers

Kun Wu¹, Minmin Yang¹, Hongyan Liu¹, Ye Tao², Ju Mei¹ and Yingzhong Zhao^{1*}

Abstract

Background: Sesame is an important and ancient oil crop in tropical and subtropical areas. China is one of the most important sesame producing countries with many germplasm accessions and excellent cultivars. Domestication and modern plant breeding have presumably narrowed the genetic basis of cultivated sesame. Several modern sesame cultivars were bred with a limited number of landrace cultivars in their pedigree. The genetic variation was subsequently reduced by genetic drift and selection. Characterization of genetic diversity of these cultivars by molecular markers is of great value to assist parental line selection and breeding strategy design.

Results: Three hundred and forty nine simple sequence repeat (SSR) and 79 insertion-deletion (InDel) markers were developed from cDNA library and reduced-representation sequencing of a sesame cultivar Zhongzhi 14, respectively. Combined with previously published SSR markers, 88 polymorphic markers were used to assess the genetic diversity, phylogenetic relationships, population structure, and allele distribution among 130 Chinese sesame accessions including 82 cultivars, 44 landraces and 4 wild germplasm accessions. A total of 325 alleles were detected, with the average gene diversity of 0.432. Model-based structure analysis revealed the presence of five subgroups belonging to two main groups, which were consistent with the results from principal coordinate analysis (PCA), phylogenetic clustering and analysis of molecular variance (AMOVA). Several missing or unique alleles were identified from particular types, subgroups or families, even though they share one or both parental/progenitor lines.

Conclusions: This report presented a by far most comprehensive characterization of the molecular and genetic diversity of sesame cultivars in China. InDels are more polymorphic than SSRs, but their ability for deciphering genetic diversity compared to the later. Improved sesame cultivars have narrower genetic basis than landraces, reflecting the effect of genetic drift or selection during breeding processes. Comparative analysis of allele distribution revealed genetic divergence between improved cultivars and landraces, as well as between cultivars released in different years. These results will be useful for assessing cultivars and for marker-assisted breeding in sesame.

Keywords: Sesame, InDel, Microsatellites, Genetic diversity, Population structure, Allele distribution

* Correspondence: zhaoyz63@163.com

¹Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Sesame Genetic Improvement Laboratory, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences (OCRI-CAAS), Wuhan, Hubei 430062, China

Full list of author information is available at the end of the article

Background

Sesame (*Sesamum indicum* L.) ($2n = 26$) is an important and ancient oilseed crop in tropical and subtropical regions of Asia, Africa and South America [1]. It is a diploid species belonging to the *Sesamum* genus of Pedaliaceae family with an estimated genome size of ~369 Mb [2]. Sesame seeds are considered to have the highest oil contents among major oilseed crops including also peanut, soybean and rapeseed [3]. It is also rich in proteins, vitamins and antioxidants such as sesamin and sesamol [4,5]. China is one of the most important sesame producing countries that contributes over 20% and consumes ~30% of the world's production, with the highest yield level around the world (2001 to 2010, UN Food and Agriculture Organization Data).

There are currently 4251 accessions in the Chinese sesame germplasm collection. More than 80 cultivars were released in the period between 1950 and 2012 [6,7]. Despite of the number of commercial cultivars, a main hindrance in sesame production is the lack of cultivars with high-yield stability and adaptability. Domestication and modern plant breeding have presumably narrowed the genetic basis of cultivated sesame, as has been in wheat, maize and other field crops [6,8,9]. These modern sesame cultivars were bred with a limited number of landrace cultivars in their pedigree. For example, more than 12 important improved cultivars including the well know Yuzhi 4, Wanzhi 2, Ezhi 6, Zhongzhi 11 and Zhongzhi 12 were developed from a common parent of Yiyangbai, directly or indirectly. Assessment of genetic variation among these modern and landrace sesame cultivars can provide breeders with insight into the need to introgress more elite germplasm into their programs to broaden genetic variation.

It is necessary to take reliable identification of these sesame cultivars through DNA fingerprinting by molecular markers, which has been widely used for checking the identity and purity of cultivars and for assessing their genetic variability in different crops [8-13]. In sesame, the genetic diversity has been detected by universal markers such as amplified fragment length polymorphism (AFLP) [14,15], sequence-related amplified polymorphisms (SRAP) [6,7,16], random amplified polymorphic DNA (RAPD) [17-19] and inter-simple sequence repeat (ISSR) [20,21]. Applications of sequence-specific markers such as genomic simple sequence repeats (Genomic-SSR) [22-24] and expressed sequence tag-SSR (EST-SSR) [25,26] were also documented. Since most of the aforementioned studies used only limited number of improved cultivars or markers, a more comprehensive analysis of common sesame cultivars in a nationwide level is required to reach a definitive understanding of their genetic variation.

SSRs are short (1-8 bp) repeat motifs usually associated with high frequency of length polymorphism. With the

advantages of simplicity, effectiveness, abundance, reproducibility, codominant inheritance and extensive genomic coverage, SSRs have been applied to disclose genetic diversity and relationship in a number of crop species [27-32]. Few polymorphic SSR markers have been identified in sesame [22-26,33]. In addition, Insertion-Deletion (InDel) markers, which arise from insertion of transposable elements, slippage in simple sequence replication or unequal crossover events, also share these advantages for SSRs [34]. InDels have also been widely applied for genotyping, genetic diversity analysis, QTLs mapping, map-based cloning, and even marker-assisted selections in Arabidopsis, rice, wheat, turnip, sunflower, citrus, and Atlantic salmon [35-43].

In this study, we developed and characterized 349 EST-SSR markers from a cDNA library [44], and 76 InDel markers from a reduced-representation gDNA library of the same commercial sesame cultivar Zhongzhi 14. We applied these newly developed markers with 600 published EST-SSR or Genomic-SSR markers to 82 improved cultivars or inbred lines, which collectively represent virtually all the available Chinese improved sesame cultivars, and made comparison with the results from assessing 48 important landraces or wild germplasm accessions.

Results

Development and characterization of sesame SSRs and InDels

For those 1,949 non-redundant SSRs identified from uni-genes of 'Zhongzhi 14' [44], 349 primer pairs named as SBM series were successfully designed and synthesized for genetic diversity analysis in sesame (Additional file 1: Table S1). Superadded previously published sesame SSRs, a total of 815 EST-SSRs and 134 genomic-SSRs were surveyed on the genomic DNA of 'Zhongzhi 14' and 'Miaoqianzhima'. As a result, 82.52% EST-SSR and 79.85% genomic-SSR primer pairs generated reproducible and clear amplicons in two reference templates. Among these markers, 39 EST-SSRs (5.17%) and 13 genomic-SSRs (12.15%) detected polymorphisms (Table 1).

Ninety-seven InDels were identified through comparative Restriction-site Associated DNA (RAD) sequencing of the genomes of 'Zhongzhi 14' and 'Miaoqianzhima', with the GenBank accession numbers KG777470-KG777548. And 79 primer pairs were successfully designed and synthesized for genetic diversity analysis (Additional file 2: Table S2). As a result, 75 primer pairs generated single and clear bands as expected. And 36 InDels detected repeatable polymorphisms between two references (Table 1).

Then, 88 primer pairs, including 39 EST-SSRs, 13 genomic-SSRs and 36 InDels, that amplified reproducible and polymorphic bands were used to genotype 130 sesame cultivars, landraces or wild germplasm. A total of 223 and 102 alleles were detected using SSR and

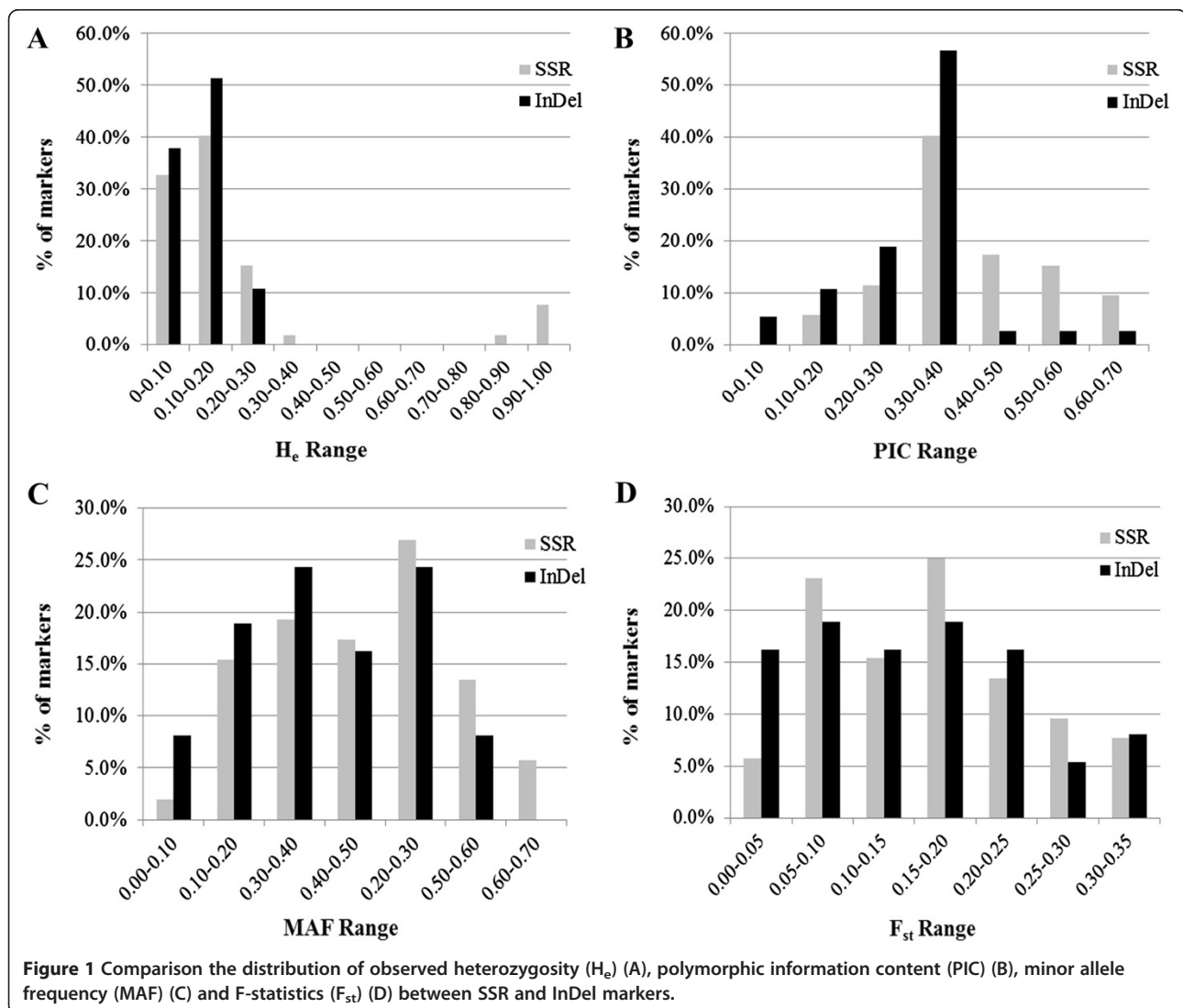
Table 1 Types of markers surveyed and the polymorphism detection rates between 'Zhongzhi 14' and 'Miaoqianzhima'

Type	Series code	Number	Number (%) of markers with		Source
			Clear bands	Detecting polymorphism ^a	
Genomic-SSR	GBssr, Saesam	23	22 (95.7)	4 (18.2)	Dixit et al. [22]; Cho et al. [24]
	No. (named 'GSSR')	111	85 (76.6)	9 (10.6)	Spandana et al. [45]
EST-SSR	ZHY	25	24 (96.0)	2 (8.3)	Wei et al. [33]
	HS	342	316 (93.0)	14 (4.4)	Zhang et al. [25]; Yue et al. [46]
	ZM	99	99 (100.0)	12 (12.1)	Wei et al.[59]; Wang et al. [26]
	SBM	349	315 (90.3)	11 (3.5)	Authors' laboratory
InDel	SBI	79	75 (94.9)	36 (48.0)	Authors' laboratory

^a% Number of markers detecting polymorphism VS number of markers producing clear bands.

InDel markers, respectively. Allele number per locus for SSR and InDel markers ranged from 2 to 9 and from 2 to 6 (with average number of 4.29 and 2.76), H_e average was 20.7% and 12.0%, gene diversity average was 0.47 and 0.39, PIC average was 0.40 and 0.32, average minor

allele frequency (MAF) was 35.58% and 28.78%, average F_{st} was 0.16 and 0.15, respectively. And the average alleles number per locus, gene diversity and PIC of SSR markers were significant higher than InDel markers ($P < 0.01$). The distribution of H_e , MAF and F_{st} among the whole



population confirmed that InDel markers are less polymorphic than SSR markers but showed similar differentiation between sesame accessions (Figure 1). The observed H_e was obviously lower for InDel than SSR markers. The MAF and F_{st} values were similar between InDel and SSR markers. So these InDel and SSR markers showed comparable ability in deciphering genetic diversity of sesame in this study.

Genetic diversity

Genotyping of 130 individuals including white seeded improved cultivars or inbred lines [WIC(L)], white seeded landraces (WLR), black seeded improved cultivars (BIC), black seeded landraces (BLR) and wild germplasm accessions revealed a total of 325 alleles. The average allele number per locus for the five different subsets varied from 2.3034 to 2.9213, with the highest number in wild germplasm accessions. Four wild germplasm accessions showed higher MAF, gene diversity, heterozygosity and PIC than the rest four subsets. Seventy WIC(L) accessions had the significantly lowest MAF, gene diversity and PIC values ($P < 0.01$) (Table 2). Compared to the WLR or BLR, respectively, WIC(L) and BIC had significantly higher level of gene diversity and PIC (Figure 2A, B). Furthermore, these improved cultivars (including both white and black seeded) were also compared for genetic diversity among subsets by releasing period (Table 2). Compared to landraces, the five subsets including Y1970s, Y1980s, Y1990s,

Y2000s and Y2010s cultivars had lower MAF, gene diversity and PIC values. Landraces and Y1990s cultivars had similarly higher heterozygosity level than other three subsets. The MAF, gene diversity and PIC of Y2010s cultivars were significantly lower than those of all other subsets ($P < 0.05$) (Table 2). For gene diversity, Y1990s cultivars had the largest variation, followed by Y2000s and Y1980s (Figure 2C). The variations of PIC within Y1970s, Y1990s and Y2000s were similarly higher than those in Y1970s and Y2010s (Figure 2D).

Population structure and genetic clustering

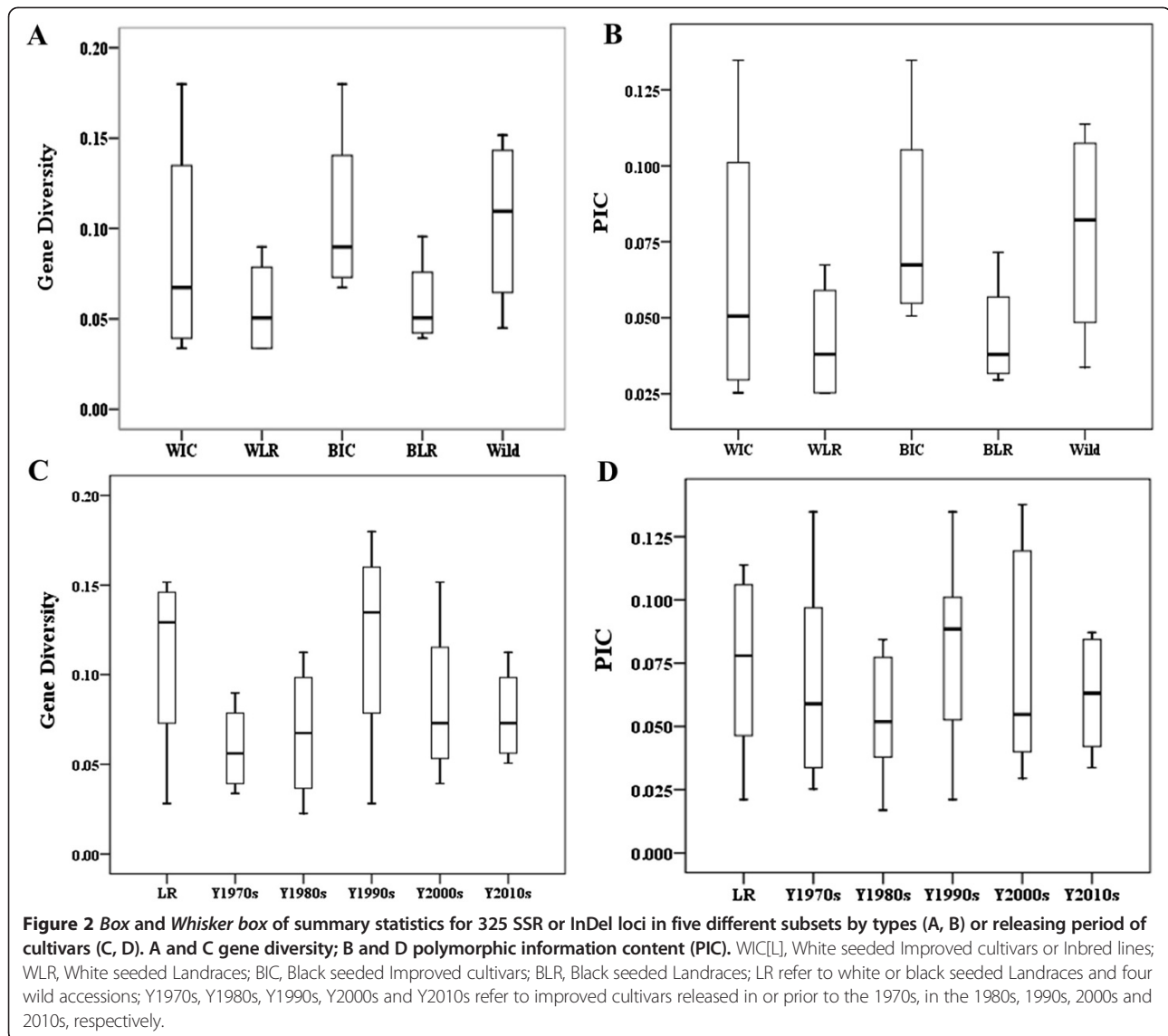
To examine the relatedness among these 130 lines, the genotypic data for 52 SSRs and 36 InDels were analyzed using a model-based approach implemented in STRUCTURE. Fifty datasets were obtained by setting the number of possible clusters (k) from 1 to 10 with five replications each. The $\text{LnP}(D)$ for each given k increased with the increase of k and the most significant change was observed when k increased from 1 to 2. In addition, a sharp peak of the second-order likelihood, Δk , appeared at $k = 2$ (Figure 3A). Accordingly, the total panel could be divided into two main groups, designated as G1 and G2, respectively (Figure 3D, Additional file 3: Table S3). The G1 group contained 98 lines, most of which are white seeded. The G2 group contained 21 lines, mostly black seeded (Additional file 3: Table S3). The remaining 11 lines each had a membership

Table 2 Statistical summary of the genetic diversity of five different sesame subsets

Subsets	Accession Number	MAF	Genotype number	Allele number	Gene Diversity	Heterozygosity	PIC
WIC(L)	70	0.2714**	3.5169	2.4157	0.3635**	0.1535	0.3047**
WLR	31	0.3337	3.6067	2.5843	0.4306	0.1769	0.3588
BIC	12	0.3172	2.8315*	2.3034**	0.4066	0.1464	0.3346
BLR	13	0.3124	3.0674	2.4719	0.4098	0.1858	0.3425
Wild	4	0.4232	2.6854**	2.9213	0.5221	0.3801	0.4564
Y1970s	8	0.2667	2.3820*	2.0674*	0.3524	0.1701	0.2905
Y1980s	10	0.2891	2.6180	2.2135	0.3788	0.1236**	0.3151
Y1990s	10	0.2612	2.8090	2.2247	0.3502	0.1989	0.2925
Y2000s	36	0.2773	3.4607	2.4045	0.3724	0.1450*	0.3135
Y2010s	12	0.2525*	2.6180	2.1348	0.3358*	0.1490	0.2781*
LR	48	0.3742	4.9663	3.6067	0.4791	0.1911	0.4056
P1	28	0.2424	3.0337	2.2921	0.3313	0.1409	0.2795
P2	23	0.1975	2.6180**	2.0337**	0.2596**	0.1251*	0.2152**
P3	14	0.2582	2.7978	2.2022	0.3429	0.2064	0.2850
P4	13	0.2736	2.9101	2.3371	0.3628	0.1495	0.3049
P5	4	0.4232**	2.6854*	2.9213	0.5221	0.3801	0.4564
Overall		0.3265	5.1798	3.6517	0.4323	0.1686	0.3650

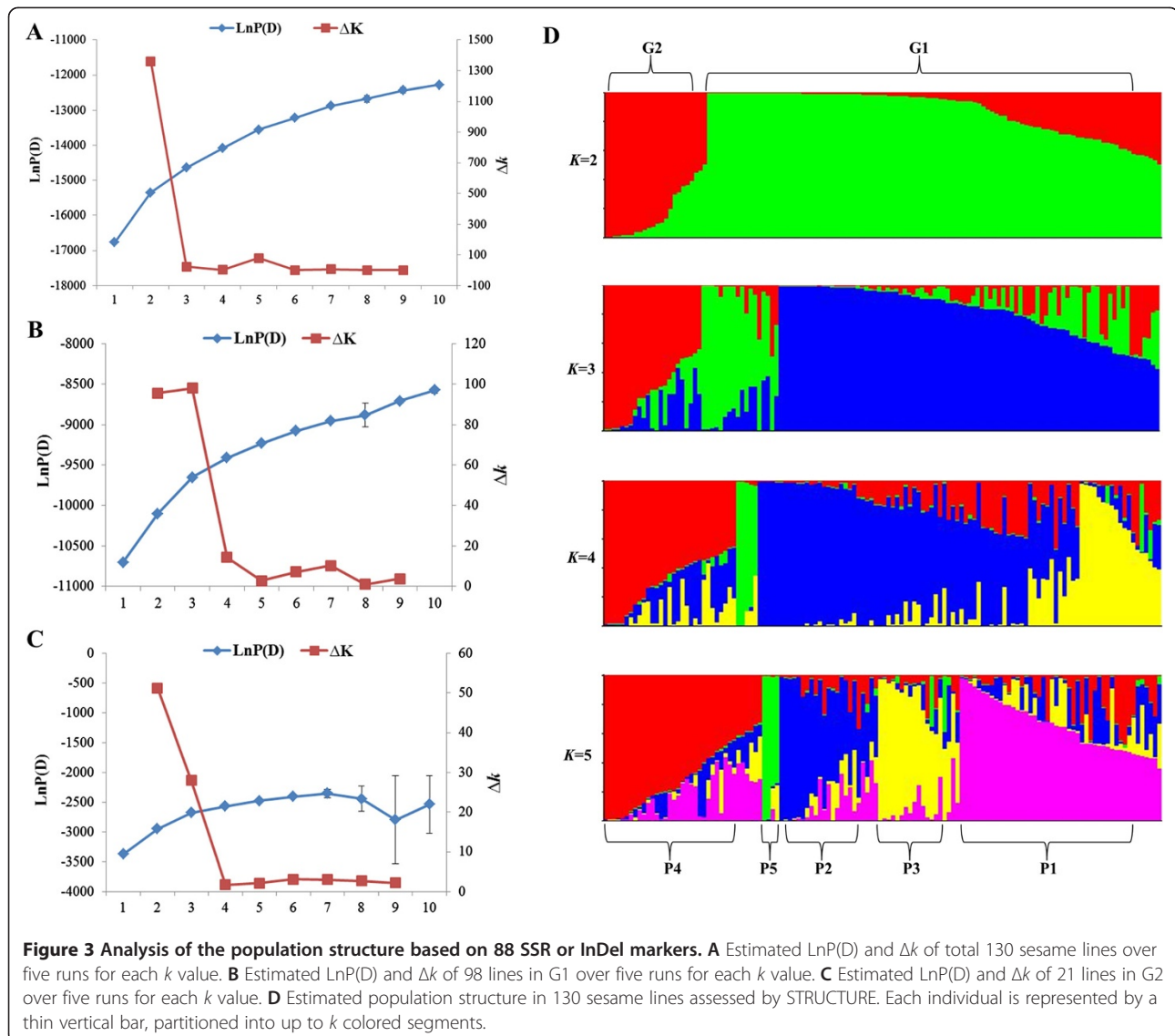
Mean values are represented in the table, * $P \leq 0.05$ and ** $P \leq 0.01$ were assessed by Z-test.

Subsets of WIC(L), WLR, BIC, BLR and Wild includes 130 accessions; Subsets of Y1970s, Y1980s, Y1990s, Y2000s, Y2010s and LR (WLR, BLR or wild accessions) includes 124 accessions except 6 white seeded inbred lines; Subsets of subgroup P1, P2, P3, P4 and P5 includes 82 accessions except 48 accessions that classified into a mixed subgroup.



probability lower than 0.60 in any given group and were thus classified into a mixed group (named Gmix). The main groups were further subdivided into P1, P2, P3 and P4, P5 subpopulations, respectively, as suggested by the STRUCTURE analysis (Figure 3). The P1 subgroup included 21 WIC(L)s and 7 WLRs (53.6% from Hubei Province). The P2 subgroup included 21 WIC(L)s, one BIC and one WLR (56.5% from Henan Province). The P3 subgroup included 5 WICs, 8 WLRs, and one BIC. The P4 subgroup contained 5 BICs (all from Jiangxi Province), 7 BLRs (such as Wuninghei, most from south China or Asia) and one WLR (C-50, from India). The P5 subgroup included only four wild germplasm accessions from India or Africa. The remaining 48 lines were classified into a mixed subgroup (named Pmix) as they had membership probabilities lower than 0.60 in any given subgroup (Additional file 3: Table S3).

Moreover, we also constructed a neighbor-joining tree and conducted PCA to examine genetic population structure and genetic clustering of these sesame accessions. The NJ phylogenetic tree based on Nei's genetic distances (1972) displayed a similar clustering pattern of relationship to that of STRUCTURE (Figure 4A). The tree had five clear branches with the "mixed" lines (Pmix, in black) distributed in each branch. PCA based on Nei's genetic distances showed a similar, five-cluster distribution pattern, with the mixed subgroup being in the middle of these five defined subgroups (Figure 4B). The top two principal components clearly separated these subgroups, but only partially distinguished P1 and P2. It appeared that P3, P4 and P5 were relatively distant from P1 and P2, which were close to each other. P3 and P4 were distant from each other. More interestingly, Wild 1 and Wild 2 from P5 were genetically far away from the rest four subgroups, while



other two wild germplasm accessions were comparatively closer to P4 and P3.

Population differentiation and diversity

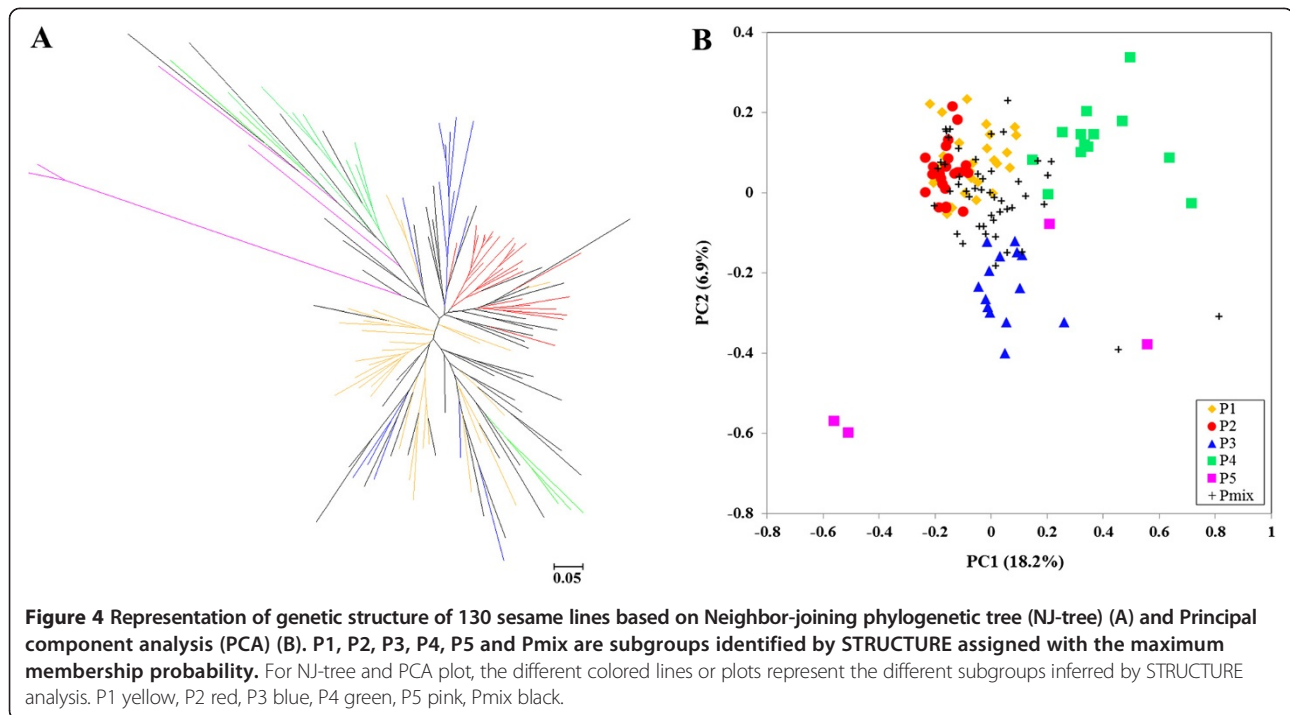
AMOVA was performed and F_{st} was calculated to investigate population differentiation and diversity. AMOVA results indicated that only 10.23% ($P < 0.001$) of the total molecular variation was partitioned among groups, 20.23% ($P < 0.001$) was attributed to differentiation among subgroups and 69.54% ($P < 0.001$) within subgroups. Pairwise F_{st} of the two inferred groups was 0.19 ($P < 0.001$), suggesting that G1 is largely divergent from G2. The levels of differentiation between subgroups were varied, with F_{st} ranging from 0.19 (P1 and P2, $P < 0.001$) to 0.41 (P2 and P5, $P < 0.001$) (Table 3). A similar pattern of differentiation among subgroups was also observed using Nei's minimum distance, which

ranged from 0.12 to 0.47 with the correlation coefficient to F_{st} being 0.704 ($P < 0.05$) (Table 3).

The genetic diversity in inferred subgroups was also assessed and compared using MAF, gene diversity, heterozygosity and PIC (Table 2). Compared to the entire panel, P2 had significantly lower gene diversity, allele number per locus, heterozygosity and PIC ($P < 0.05$ or $P < 0.01$). P5 had the highest level of MAF among all subgroups, followed by P4, P3, P1 and P2. P3 exhibited a similar level of MAF, gene diversity and PIC to P1 and P4, but higher level of heterozygosity ($P < 0.01$).

Allele frequencies and alleles distribution in different sesame cultivars in China

To more deeply dissect the genetic differentiation among different set of sesame cultivars in China, comparative analysis of allele frequencies was performed (Additional



file 4: Table S4). Of the 325 alleles, allele frequencies difference larger than 10% ($P > 0.01$) were observed for 117 (36.0%) alleles in the WIC(L) versus WLR subgroup (Figure 5A), and 133 (40.9%) alleles in the BIC versus BLR subgroup (Figure 5B). In comparison with the WLR subgroup, there were 22 missing alleles and 7 unique alleles identified in WIC(L). And 21 missing alleles and 6 unique alleles were identified in BIC subgroup compared to BLR. (Additional file 4: Table S4).

We also compared the allele frequencies of sesame cultivars that were released in different timelines to reveal their genetic difference. In the Y1980s versus Y1970s and Y1990s versus Y1980s comparisons, respectively, 125 (38.5%) and 134 (41.2%) alleles showed an allele frequency difference larger than 10% ($P < 0.01$) (Figure 5C, D). Only 88 (27.1%) and 68 (20.9%) alleles had an allele frequency larger than 10% in the comparisons of Y2000s versus Y1990s and Y2010s versus Y2000s, respectively (Figure 5E, F). Compared to the Y2000s subset, only 1 unique allele but 25 missing alleles were identified in the Y2010s subset (Additional file 4: Table S4). These results indicate distinct genetic differences among the four pairwise comparisons, with the strongest differentiation between Y1980s and Y1970s lines or Y1990s and Y1980s, the second between Y2000s and Y1990s, and the least between Y2010s and Y2000s (Figure 5C to F).

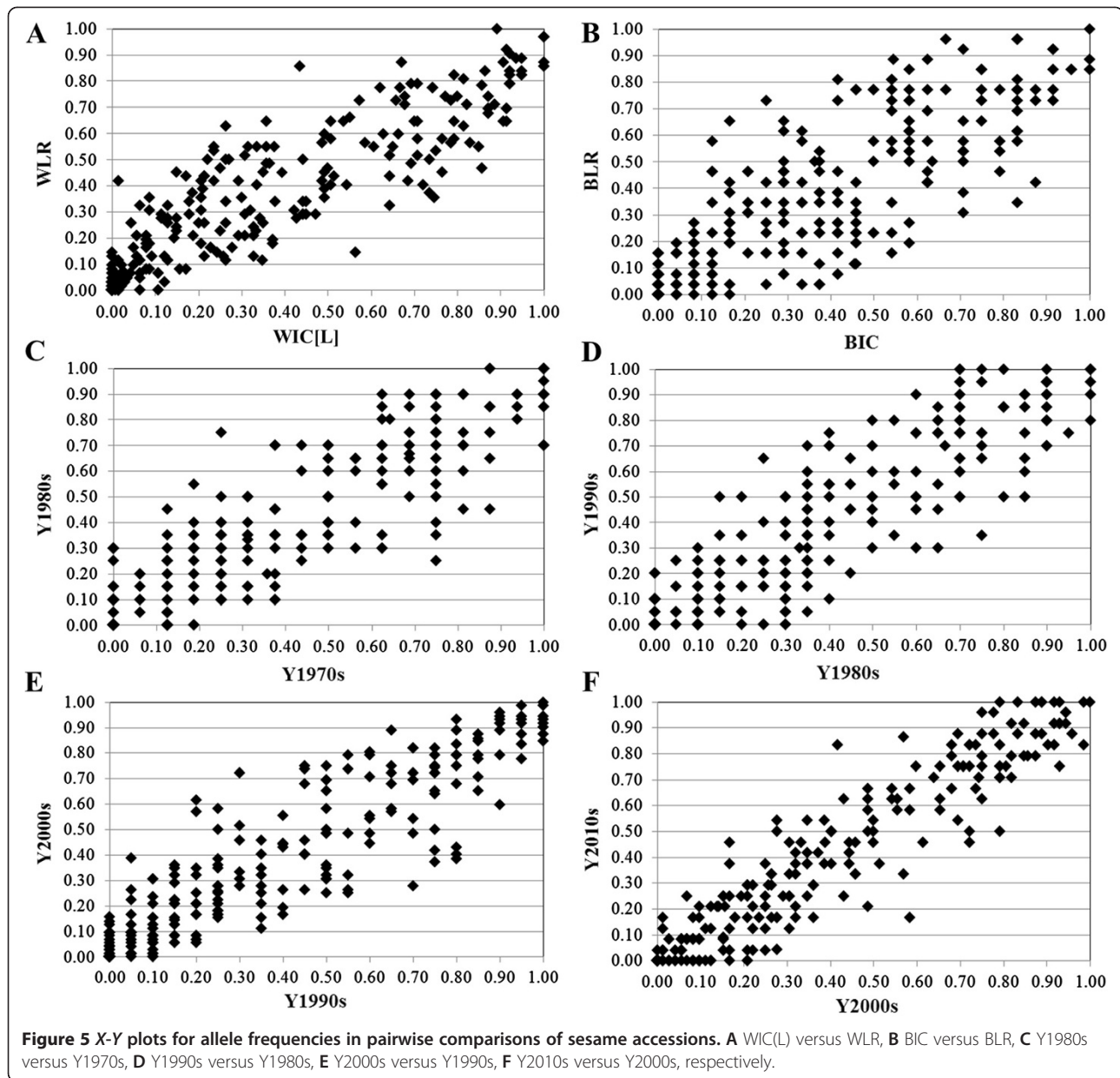
Moreover, we also compared the distribution of 325 alleles in four important Chinese sesame cultivar families with four different parental/progenitor lines (Table 4). In family I with the common parental/progenitor of

Yiyangbai, two cultivars were from subgroup P1, three from P2, and 5 from Pmix. They shared 27 common alleles, such as SBM073.5, HS050.2, ZM0740.1 and SBI009.3 (Table 4). Cultivars from the family II with Yuzhi No.4 as the common donor shared 22 alleles, most of which were from P2 subgroup, except for Wanzhi No.1, Zhuzhi No.11 and other four lines (Table 4). The family III of Zhongzhi No.1 included 4 cultivars from P1, 3 from Pmix and one from P1, with 21 shared alleles (Table 4). While the black seed-type family IV of Wuninghei had 4 cultivars with 19 shared alleles. On the whole, three EST-SSRs alleles and three genomic InDels alleles were shared in four families, including SBM073.8, SBM768.6, HS050.2, SBI014.1, SBI017.2 and SBI019.2. And six alleles including SBM750.3, SBM1111.1, HS137.4, Y1972.1, ZHY01.3 and SBI060.1

Table 3 Genetic distance, as measured by Nei's (1973) minimum distance (top diagonal) and pairwise F_{st} comparisons (bottom diagonal) among inferred sesame subgroups

Group	Subgroup	G1	G2			
		P1	P2	P3	P4	P5
G1	P1		0.12	0.19	0.28	0.43
	P2	0.19**		0.21	0.32	0.46
	P3	0.24**	0.29**		0.30	0.40
G2	P4	0.30**	0.37**	0.30**		0.47
	P5	0.33**	0.41**	0.28**	0.29**	

**Significant at $P < 0.01$ after 1,000 permutations.



were found to be specially shared in family I. Four alleles including HS225.1, ZM1179.2, SBI023.2 and SBI034.1 were specially shared in family II. Other four alleles of GSSR007.2, SBI036.2, SBI050.1 and SBI071.2 were specially shared in family III. Eight alleles specially shared in family IV were also identified, including SBM768.5, ZM1413.2, ZM1488.1, SBI005.1, SBI007.4, SBI025.1, SBI027.2 and SBI051.3 (Table 4). These alleles identified above with different allelic frequency, even miss, unique or family special, can be combined and used for characterization of sesame cultivars and for sesame molecular breeding.

Discussion

Development and utilization of sesame SSR and InDel markers for sesame genetic diversity analysis

In this study, we developed 315 EST-SSR markers from 1,688 unigenes from sequencing a cDNA library of Zhongzhi 14. Combined with 466 earlier EST-SSR and 134 earlier genomic-SSR markers in sesame, only 5.17% EST-SSRs and 12.15% genomic-SSRs (gSSRs) showed polymorphism between 'Zhongzhi 14' and 'Miaoqianzhima', which were two parents of an important RIL population for other works. Such polymorphism rate of EST-SSRs is lower than that in an intraspecific cross (7.5% or

Table 4 Comparison of cultivars from four different families using 89 molecular markers

Family ^a	Cultivar	Time ^b	Parentage	Group	Shared alleles ^d
I	Yiyangbai ^c	1970s	Selection from variety of "Zhongxiang Huangzhima"	P1	SBM073.5, SBM073.8, SBM750.3, SBM768.6, SBM1111.1, SBM1120.2, HS050.2, HS137.4, HS142.3, HS176.3, Y1972.1, ZHY01.3, ZHY023.4, ZM0740.1, ZM0961.4, GB182.3, SBI009.3, SBI012.2, SBI014.1, SBI017.2, SBI019.2, SBI030.1, SBI041.4, SBI043.2, SBI054.2, SBI060.1, SBI064.1
	Ningzhi No.1	1980s	Selection from variety of "Yiyangbai"	Pmix	
	Yuzhi No.4	1980s	Yiyangbai × Zhuzhi No. 1	P2	
	Yuzhi No.7	1990s	Zhongzhi No.7 × Yiyangbai	Pmix	
	Yuzhi 18	2000s	(Variety of Yiyangbai × Yuzhi No. 11)F ₃ × Zhenzhi 958	P2	
	Ezhi No. 6	2000s	Ezhi No. 1 × Yiyangbai	Pmix	
	Zhongzhi 12	2000s	CLSU-9 (Philippine) × Yiyangbai	P1	
	Zhongzhi 18	2010s	(Yiyangbai × Ezhi No. 1)F ₂ × Zhongzhi 11	Pmix	
	Zhongzhi 21	2010s	[(Yiyangbai × Zhushanbai) F ₄] × Fufengzhima	Pmix	
	Zhongzhiza No.2	2010s	95 ms-2 (male sterile) × Zhongzhi 12	P2	
II	Yuzhi No.4 ^c	1980s	Yiyangbai × Zhuzhi No. 1	P2	SBM064.3, SBM073.5, SBM073.8, SBM768.6, SBM1120.2, HS050.2, HS123.3, HS142.3, HS176.3, HS225.1, ZHY023.4, ZM030.2, ZM0740.1, ZM0961.4, ZM1179.2, SBI014.1, SBI017.2, SBI019.2, SBI023.2, SBI034.1, SBI043.2, SBI064.1
	Luozhi 12	2000s	Zhen 89H142 × Yuzhi No. 4	Pmix	
	Luozhi No. 15	2000s	Variety of Yuzhi No. 4	P2	
	Wanzhi No. 1	2000s	0176A (male sterile) × Yuzhi No. 4	P1	
	Wanzhi No. 2	2000s	Fuyang Xiaozibai × Yuzhi No. 4	P2	
	Zhongzhi 11	2000s	Space mutant of Yuzhi No. 4	P2	
	Zhongzhi 13	2000s	Space mutant of Yuzhi No. 4	P2	
	Zhongzhi 14	2000s	85-411 × Yuzhi No. 4	Pmix	
	Zhongzhi 15	2000s	Yuzhi No. 4 × Suxianzhima	Pmix	
	Zhuzhi No. 11	2000s	Zhu 81043 × Zhu 7801 (variety of Yuzhi No. 4)	P1	
	Zhuzhi No. 14	2000s	Zhu 86036 × Zhu 7801 (variety of Yuzhi No. 4)	P2	
	Zhuzhi No. 18	2000s	Zhu 893 × Zhu 7801 (variety of Yuzhi No. 4)	P2	
	Ezhi No. 7	2010s	Ezhi No. 3 × Yuzhi No. 4	Pmix	
	III	Zhongzhi No.1 ^c	1970s	Selection from "Enshi Baizhima"	
Zhongzhi No.7		1970s	Xiangyang Xiniujiang × Zhongzhi No.1	P1	
Zhongzhi No.8		1980s	Zhongzhi No. 7 × Jiangling Yongguangxingzhima	Pmix	
Zhongzhi No.9		1990s	Xinjiang Heizhima × Zhongzhi No. 7	Pmix	
Zhongzhi No.10		1990s	{Zhongzhi No. 5 × [(Zihuayeersan × Zhongzhi No. 1) × Suiping Xiaozihuang]} × (Zhongzhi No. 5 × Zhecheng Tiegucha)	P1	
Zhongzhi 17		2010s	Zhu86-207 × Zhongzhi No. 10	P2	
Zhongzhi 22		2010s	Zhongzhi No. 10 × Ezhi No. 1	P1	
Zhongzhi 23		2010s	(Zhongzhi No. 10 × Zhu 04) × Zhenzhi 98 N09	P1	
IV	Wuninghei ^c	1980s	An important landrace	P4	SBM064.3, SBM073.8, SBM768.5, SBM768.6, HS050.2, HS123.3, ZM1413.2, ZM1488.1, GSSR090.4, SBI005.1, SBI007.4, SBI012.2, SBI014.1, SBI017.2, SBI019.2, SBI025.1, SBI027.2, SBI030.1, SBI051.3
	Ganzhi No. 6	2000s	Yujiangheizhima × Wuninghei	Pmix	
	Ganzhi No. 9	2000s	Co ⁶⁰ radiation mutant of "Wuninghei"	P4	
	Jiheizhi No. 1	2000s	Jizhi No. 1 × Wuninghei	P3	
	Zhuzhi No. 10	2000s	7801H (variety of Yuzhi No. 4) × Wuninghei	Pmix	

^aFour different families with one common parent or progenitor. ^brelease or application time of these cultivars. ^cthe corresponding common parents or progenitors of each family. ^dthe shared genetic alleles in one family.

6.52%) [25,33], but higher than that of 36 sesame accessions (4.01%) [26]. Polymorphism rate of gSSRs in this study is lower than reported in two earlier studies [22,45], which were 20% and 26.3% respectively. The relative low level of SSR polymorphism between 'Zhongzhi 14' and 'Miaoqianzhima' is obviously inconsistent with their obviously morphological variations, which might be interpreted by InDel, SNP (single nucleotide polymorphism), methylation or other genomic variation. And more polymorphic SSR markers might be identified by using more genomic sequence and more DNA template of sesame accessions.

A total of 75 genomic InDel markers were also developed, making use of RAD sequencing of 'Zhongzhi 14' and 'Miaoqianzhima'. The InDel markers showed much higher ability to discern genetic diversity, as the rate of polymorphism is as high as 48.0%. In the collection of cultivars, landraces even wild germplasm with different chromosome numbers, most InDel markers yielded single PCR fragments and showed polymorphisms. Such high efficiency of InDel markers was also reported in *Brassica rapa*, *Arabidopsis*, *Helianthus annuus* and *Citrus* [35,36,38,39,41]. Furthermore, the average allele number per locus, H_e , gene diversity and PIC of SSR markers were significant higher than those of InDel markers in the whole panel, as opposed to MAF and F_{st} values, which were similar between InDel and SSR markers. The distribution of H_e , MAF and F_{st} further confirmed that InDel markers showed similar differentiation between sesame accessions with more polymorphic than SSR markers. Similar pattern was also reported in cultivated citrus [41]. Therefore, this set of novel PCR-based SSR and InDel markers will be valuable for genetic studies and breeding in sesame. In addition, most of these polymorphic SSR and InDel markers showed normal segregation in a RIL population (data not shown), based on which a project toward high density genetic mapping employing these SSRs, InDels plus some SNP markers is now underway in our lab.

Genetic diversity and population structure in sesame panel

A thorough understanding of genetic diversity, population structure and familiar relatedness in a given panel is very important for successful association studies. For this purpose, a large number of DNA markers that are genome-wide distributed, reproducible, cost-effective, selectively neutral and highly polymorphic are necessary. SSRs and InDels are two nice choices of this kind. In this study, 88 polymorphic markers including EST-SSRs, genomic-SSRs and InDels randomly distributed in *Sesamum indicum* L. genome were selected to evaluate 130 sesame cultivars, landraces or wild germplasm. A total of 325 alleles, with an average of 3.69 alleles per locus, were detected in this

sesame panel. The number of polymorphic markers used in this study is higher than in most earlier reports, but the number of allele per locus is lower than that detected in 150 [24], 453 [7], 545 [46], 216 [47] sesame accession and 67 sesame cultivars in China [6]. The difference of allelic richness between our panel and other germplasm collections may be caused by the differences of materials analyzed, but the use of only site-specific SSR and InDel markers may also account for this.

More importantly, a larger number of loci (in particular, the use of dinucleotide repeat SSRs than tri- or higher) will lead to a higher number of alleles and thus a higher apparent level of genetic diversity [48]. The average PIC value and gene diversity across all lines in this panel were 0.365 and 0.432, respectively. They were much higher than some reported values [14,16,47,49,50], but lower than those of Yue et al. (2012) and Cho et al. (2011) [24,46], even excluding four wild germplasm. We also found that the diversity level in this panel was much lower than that of rice [51,52] and wheat [32,53,54], which are also self-pollinating crops. That might be ascribed to the lower frequency of gene flow by introduction and utilization of external genetic accessions in Chinese sesame breeding programs [47]. Furthermore, 130 sesame lines could be classified into five types, including WIC(L), WLR, BIC, BLR and wild germplasm according to their sources. All subsets showed similar MAF, gene diversity, heterozygosity and PIC except for four wild germplasm collections. WIC(L) showed the lowest but quite wide variation of gene diversity and PIC than other subsets, which indicated a relatively narrow genetic basis in Chinese white seeded improved cultivars or inbred lines.

To get detailed knowledge of genetic relatedness among individuals (especially cultivars) in this panel, model-based STRUCTURE analyses were conducted and revealed the existence of two main groups in this sesame panel. The division of these two groups (G1 and G2) generally corresponds to their seed colors (white VS black) (Additional file 3: Table S3). Significant divergence between the two main groups was reflected by F_{st} . Five subpopulations were identified within the 130 sesame accessions, which was cross-validated by STRUCTURE, PCA, NJ phylogenetic tree based analysis and AMOVA. Furthermore, most previous related studies in sesame revealed certain relationship between population structure and geographical distribution [24,46,47,49]. Our study of population structure revealed limited correlation with geographical distribution in P1, P2 and P4. Some earlier studies also indicated limited association between ecological or geographical origin and population differentiation in sesame [14,46]. Furthermore, 48 lines (36.9%) in this sesame panel were assigned into a

mixed subgroup (Pmix) for low membership probability (< 0.60). Cho et al. [24] also categorized 27.3% of 150 sesame accessions as admixed forms with varying levels of membership shared among three genetic groups. 20.5% of 527 maize collection (a global germplasm) [55] and 35.5% of 155 maize inbred lines (mainly temperate germplasm) [48] were classified into a mixed group. This varied percentage of mixed lines may indicate various degree of gene flow by hybridization and introgression events.

Impacts of selection and breeding on genetic diversity of Chinese sesame cultivars

Genetic diversity in sesame as in other crops has been reduced during domestication and breeding [56-58]. Nyongesa et al. (2013) also reported the genetic divergence between sesame and related wild species ($2n = 32$) in East Africa using ISSR markers. In this study, four wild germplasm accessions showed highest MAF, gene diversity, heterozygosity and PIC. Population structure and differentiation analysis indicated that they (P5) were genetically far away from other sesame accessions in our panel. These wild germplasm accessions would therefore be useful in broadening genetic basis of traditional landraces and cultivars in China.

Furthermore, the genetic diversity and PIC of improved sesame cultivars was found to be lower than those of landraces, especially the white seeded cultivars. Greater differentiation of allele frequency was observed between BIC and BLR than WIC(L) and WLR lines. Compared to WLR or BLR lines, much more missing alleles than unique alleles were identified in WIC(L) or BIC lines, which indicates that the genetic basis was narrowed down during domestication and selection of mordent cultivars from landraces. Molecular genetic indices, such as MAF, gene diversity, PIC and allele frequency, all support that a declining genetic diversity occurred during the past five decades (from 1970s to 2010s) in China. Especially, compared to the Y2000s data, 7.7% missing alleles but only 1 unique allele were identified in the Y2010s.

In Chinese sesame breeding history, several important sesame cultivars had been developed and widely grown. The relationship among four families sharing common parents or progenitors, as well as among five subgroups suggested by STRUCTURE seems ambiguous, which might be caused by intercross of accessions belonging to different subgroups. In this study, several common alleles were identified in the four families, which can be used as important identification indexes of parentage or DNA fingerprinting for Chinese sesame cultivars. In addition, the common or unique alleles identified in different type, subgroups and families will be an important resource for marker-assisted breeding,

in particular marker-assisted backcross or pyramiding breeding if more functional information are added by linkage or association mapping of important QTLs/genes. The five subgroups suggested by STRUCTURE in this study may provide breeders with more advices for broadening genetic basis of sesame cultivars toward better adaptability.

Conclusions

This report presents the by far most comprehensive characterization of the molecular and genetic diversity of available sesame cultivars in China. We developed 349 SSRs and 79 InDel markers by a cDNA library and reduced-representation sequencing. Comparison of genetic diversity assessed by SSR and InDel markers confirmed that InDels are more polymorphic than SSRs but both showed comparable abilities for deciphering genetic diversity. Comparison of molecular marker information indicates that the genetic basis was narrowed down and the genetic diversity was declining during domestication and selection of mordent cultivars from landraces. Comparative analysis of allele distribution revealed genetic divergence between improved cultivars and landraces, even between cultivars released in different timelines. These results will be useful for assessing cultivars and for marker-assisted breeding in sesame.

Methods

Plant materials

Eighty two important Chinese improved sesame cultivars or lines (*Sesamum indicum* L., $2n = 26$), including 70 white seeded improved cultivars or inbred lines [WIC(L)] and 12 black seeded improved cultivars (BIC) from major production areas, 44 landraces (*S. indicum* L., $2n = 26$) representing geographically and phenotypically different sesame accession, and 4 wild germplasm accessions (putatively identified as *S. schinzianum*, *S. radiatum*, *S. malabaricum*, and *S. prostratum*) were used in this study (Additional file 3: Table S3). All of these lines had been self-pollinated for over five generations in Wuhan and Sanya to decrease the residual heterozygosity. Among these accessions, Zhongzhi 14 and Miaoqianzhima were chosen as a couple of templates for development of polymorphic markers and references for genotype determination, which show obviously different morphology in plant height, plant type, capsule shape, leaf shape and color, mature period, resistance and so on. All of these accessions were collected from the breeding units or the Sesame Middle Term Gene Bank at the Oil Crop Research Institute, Chinese Academy of Agricultural Sciences.

Microsatellite marker development

In previous cooperative study, 1,949 non-redundant SSRs were identified from 1,688 unigenes in a cDNA library of

Zhongzhi 14 [44]. Only SSR loci of perfect di-, tri-, tetra-, penta-, and hexanucleotide motifs with a minimum of 6, 4, 4, 4, and 4 repeats respectively were evaluated. Flanking oligonucleotide primers were designed using Primer 3 (<http://bioinfo.ut.ee/primer3/>), based on the following major parameters: PCR product size of 100-400 bp (optimal 200 bp), GC content of 40-70% (optimal 50%), annealing temperatures of 50-60°C (optimal 55°C), and primer length of 18-23 bases (optimal 20 bases).

Other published 466 EST-SSR markers and 134 genomic-SSR markers were also used in study (Table 1). The former included 342, 25, and 99 EST-SSR markers from HS [25,46], ZHY [33], and ZM [26,59] series respectively. The latter included 23 and 111 genomic-SSR markers of 'GBssr' [22,24] and 'no.' series (we named 'GSSR') [45], respectively.

RAD sequencing and InDel marker development

We have combined the RAD approach with Illumina DNA sequencing for rapid and effective discovery of InDel markers for sesame. Genomic DNA of Zhongzhi 14 and Miaoqianzhima was extracted from leaves of three-week-old seedlings using the DNA extraction kit (TIANGEN Co. Ltd, Beijing), following the manufacturer's instructions. The RAD library was constructed according to the protocol described by Baird et al. [60], restriction enzymes used were *EcoR* I and *Pst* I. Sequencing was carried out using the Illumina NGS platform HiSeq2000 at Major Biological Medicine Technology Co., Ltd. (Shanghai, China).

Solexa sequences at minimum coverage of 6X (about 2.4Gb each) were segregated by the barcode assigned to each sample. Reads of low quality (including reads with < 93 bp after trimming) or with ambiguous barcodes were discarded. After trimming each raw sequence read to 93 nucleotides from the 3' end. For the RAD pair end based InDel calling, sequence reads from two materials were first grouped into clusters of identical sequences (RAD tags) and clusters using *Stacks* [61], with <7 or >200 sequences were discarded. Forward reads of two materials were grouped and the reads of other side (reverse reads) can also be grouped at this step. The reverse reads of each cluster of two materials were de novo assembled by *phrap* separately [62]. Then BLAST was used to compare the contigs generated by *phrap* from two materials. InDels (> 2 bp) were identified by gaps in alignment results, and regarded as true polymorphisms when each allele was observed at least three times.

Genomic DNA extraction and PCR

Genomic DNA of 130 sesame accessions was all extracted from young leaves using the DNA extraction kit (TIANGEN Co. Ltd, Beijing). Polymerase chain reactions (PCR) for SSRs and InDels were performed in a 10 μ l

reactions, containing 10 ng DNA, 2 pmol of each primers, 2 nmol dNTPs, 15 nmol MgCl₂, 0.2 U Taq DNA polymerase (Fermentas, Canada) and 1X PCR buffer supplied together with the enzyme. The PCR cycles were 94°C 3 min, 36 cycles of 94°C 20 s, 55°C ~ 60°C 30 s, 72°C 40 s, and a 5 min at 72°C for final extension. PCR products were separated in 8% non-denaturing polyacrylamide gels (Acr:Bis = 19:1 or 29:1) on a constant voltage of 180 V for 2 ~ 3 h, visualized by silver staining [63].

Genotypic data analysis

For each polymorphic marker, the alleles present in each genotype were scored visually. The number of alleles, minor allele frequency, gene diversity, observed heterozygosity (H_e), group-specific alleles, family shared alleles, polymorphic information content (PIC) and Nei's genetic distance [64] were calculated using Powermarker version 3.25 [65,66]. Heterozygosity is simply the proportion of heterozygous individuals in the population. At a single

locus it is estimated as $H = 1 - \sum_{i=0}^k x_i^2$. Gene diversity

often referred to as expected heterozygosity, is defined as the probability that two randomly chosen alleles from the population are different. An unbiased estimator of

gene diversity at the l th locus is $\hat{D}_l = \left(1 - \sum_{u=1}^k \tilde{p}_{lu}^2\right) / \left(1 - \frac{1+f}{n}\right)$. The polymorphism information content (PIC)

is estimated as $PIC = 1 - \sum_{i=1}^k p_i^2 - \sum_{i=1}^{k-1} \sum_{j=i+1}^k 2p_i^2 p_j^2$. The sig-

nificance of difference in gene diversity, PIC, allele frequency and other statistics was based on P value from Fisher's exact test [67]. An analysis of F -statistics (F_{st}) among populations was calculated using GENEPOP V4.2 [68]. The definition of F -statistics used here is $F_{ST} \equiv \frac{Q_2 - Q_3}{1 - Q_3}$, Where the Q are probabilities of identity in state, Q_2 among genes in different individuals within groups (populations), and Q_3 among groups (populations). The model-based program STRUCTURE 2.34 [69,70] was used to infer population structure with SSRs and InDels. Five independent runs were performed setting the number of subpopulations (k) from 1 to 10, with 500,000 MCMC (Markov chain Monte Carlo) replications and a model for admixture model and correlated allele frequencies. The k value was determined by the log likelihood of the data ($\text{LnP}(D)$) in the STRUCTURE output and an ad hoc statistic Δk based on the rate of change in $\text{LnP}(D)$ between successive k [71]. Results of replicate runs from STRUCTURE were integrated by using the CLUMPP software [72]. Sesame accessions with membership probabilities ≥ 0.60 were assigned to the corresponding subgroup and accessions with membership probabilities ≤ 0.60 were

assigned to a mixed subgroup [73]. In addition, principal component analysis (PCA) was conducted using the modules EIGEN implemented in NTSYS-pc 2.10 [74], and a neighbor-joining dendrogram was also constructed using the unweighted pair-group method (UPGMA) in NTSYS-pc 2.10. The hierarchical analysis of molecular variance (AMOVA) across all groups, subgroups and pairwise subgroups was performed using Arlequin V3.11 [75], with 1,000 permutations and sum of squared size differences as molecular distance.

Additional files

Additional file 1: Table S1. Primer sequences, repeat types, repeat units, annealing temperature and expected PCR product sizes of 341 EST-SSRs identified from a cDNA library of Zhongzhi 14.

Additional file 2: Table S2. Primer sequences, gap size, annealing temperature and expected PCR product sizes of 79 InDels identified from RAD sequencing of Zhongzhi 14 and Miaoqianzhima.

Additional file 3: Table S3. Cultivar or accession, origin, type, releasing period, parentage and assignment of the genotypes assayed in this study.

Additional file 4: Table S4. Allele frequencies of the whole sesame panel, different types, cultivars of different releasing period, and subgroups in this study.

Abbreviations

AFLP: Amplified fragment length polymorphism; AMOVA: Analysis of molecular variance; BIC: Black seeded improved cultivars; BLR: Black seeded landraces; EST: Expressed sequence tag; F_{st} : F -statistics; H_e : Heterozygosity; InDel: Insertion-Deletion; ISSR: Inter-simple sequence repeat; MAF: Minor allele frequency; PCA: Principal component analysis; PIC: Polymorphic information content; RAD: Restriction-site associated DNA; RAPD: Random amplified polymorphic DNA; SRAP: Sequence-related amplified polymorphism; SSR: Simple sequence repeat; SNP: Single nucleotide polymorphism; UPGMA: Unweighted pair-group method; WIC(L): White seeded improved cultivars or inbred lines; WLR: White seeded landraces.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

K.W. and Y.Z.Z. designed research and wrote the manuscript; M.M.Y. performed SSR markers analysis; H.Y.L. constructed this sesame panel and collected information of sesame cultivars; Y.T. performed RAD sequencing and InDel markers development; J.M. performed InDel markers analysis; K.W. performed EST-SSR markers development, analyzed data and result. All authors read and approved the final manuscript.

Acknowledgments

The authors are grateful to Dr. Pei Xu from Zhejiang Academy of Agricultural Sciences for assistance with manuscript editing. We thank the anonymous referees and the editor for their comments and suggestions that helped improve the manuscript. This study was supported by the National Science Foundation of China (No. 31201243), National Program on Key Basic Research Project of China (No. 2011CB109304), National Science Foundation of Hubei Province (No. 2012FFB06703), Open Project of Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, P. R. China (No. 201210), the China Agriculture Research System (No. CARS-15) and Director Foundation of Oil Crops Research Institute of CAAS (No. 1610172011007).

Author details

¹Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Sesame Genetic Improvement Laboratory, Oil Crops Research

Institute of the Chinese Academy of Agricultural Sciences (OCRI-CAAS), Wuhan, Hubei 430062, China. ²Shanghai Major Biological Medicine Technology Co., Ltd, Shanghai 201203, China.

Received: 18 November 2013 Accepted: 10 March 2014

Published: 19 March 2014

References

1. Bedigian D: Evolution of sesame revisited: domestication, diversity and prospects. *Genet Resour Crop Ev* 2003, **50**(7):779–787.
2. Zhang H, Miao H, Wang L, Qu L, Liu H, Wang Q, Yue M: Genome sequencing of the important oilseed crop *Sesamum indicum* L. *Genome Biol* 2013, **14**(1):401.
3. Anilakumar KR, Pal A, Khanum F, Bawa AS: Nutritional, medicinal and industrial uses of sesame (*Sesamum indicum* L.) seeds-an overview. *Agric conspec sci* 2010, **75**(4):159–168.
4. Namiki M: The Chemistry and Physiological Functions of Sesame. *Food Rev Int* 1995, **11**(2):281–329.
5. Moazzami AA, Kamal-Eldin A: Sesame seed is a rich source of dietary lignans. *J Am Oil Chem Soc* 2006, **83**(8):719–723.
6. Zhang YX, Sun JA, Zhang XR, Wang LH, Che Z: Analysis on Genetic Diversity and Genetic Basis of the Main Sesame Cultivars Released in China. *Agr Sci China* 2011, **10**(4):509–518.
7. Zhang YX, Zhang XR, Che Z, Wang LH, Wei WL, Li DH: Genetic diversity assessment of sesame core collection in China by phenotype and molecular markers and extraction of a mini-core collection. *BMC Genet* 2012, **13**:102.
8. Reif JC, Zhang P, Dreisigacker S, Warburton ML, Van Ginkel M, Hoisington D, Bohn M, Melchinger AE: Wheat genetic diversity trends during domestication and breeding. *Theor Appl Genet* 2005, **110**(5):859–864.
9. Lu YL, Yan JB, Guimaraes CT, Taba S, Hao ZF, Gao SB, Chen SJ, Li JS, Zhang SH, Vivek BS, Magorokosho C, Mugo S, Makumbi D, Parentoni SN, Shah T, Rong TZ, Crouch JH, Xu YB: Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. *Theor Appl Genet* 2009, **120**(1):93–115.
10. Qian W, Ge S, Hong DY: Assessment of genetic variation of *Oryza granulata* detected by RAPDs and ISSRs. *Acta Bot Sin* 2000, **42**(7):741–750.
11. Plaschke J, Ganai MW, Roder MS: Detection of Genetic Diversity in Closely-Related Bread Wheat Using Microsatellite Markers. *Theor Appl Genet* 1995, **91**(6–7):1001–1007.
12. Manifesto MM, Schlatter AR, Hopp HE, Suarez EY, Dubcovsky J: Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. *Crop Sci* 2001, **41**(3):682–690.
13. Xu P, Wu XH, Wang BG, Liu YH, Qin DH, Ehlers JD, Close TJ, Hu TT, Lu ZF, Li GJ: Development and polymorphism of *Vigna unguiculata* ssp. *unguiculata* microsatellite markers used for phylogenetic analysis in asparagus bean (*Vigna unguiculata* ssp. *sesquipedalis* (L.) Verdc.). *Mol Breeding* 2010, **25**(4):675–684.
14. Laurentin HE, Karlovsky P: Genetic relationship and diversity in a sesame (*Sesamum indicum* L.) germplasm collection using amplified fragment length polymorphism (AFLP). *BMC Genet* 2006, **7**:10.
15. Laurentin H, Karlovsky P: AFLP fingerprinting of sesame (*Sesamum indicum* L.) cultivars: identification, genetic relationship and comparison of AFLP informativeness parameters. *Genet Resour Crop Ev* 2007, **54**(7):1437–1446.
16. Zhang YX, Zhang XR, Hua W, Wang LH, Che Z: Analysis of genetic diversity among indigenous landraces from sesame (*Sesamum indicum* L.) core collection in China as revealed by SRAP and SSR markers. *Genes Genom* 2010, **32**(3):207–215.
17. Bhat KV, Babrekar PP, Lakhnpaul S: Study of genetic diversity in Indian and exotic sesame (*Sesamum indicum* L.) germplasm using random amplified polymorphic DNA (RAPD) markers. *Euphytica* 1999, **110**(1):21–33.
18. Ercan AG, Taskin M, Turgut K: Analysis of genetic diversity in Turkish sesame (*Sesamum indicum* L.) populations using RAPD markers. *Genet Resour Crop Ev* 2004, **51**(6):599–607.
19. Salazar B, Laurentin H, Davila M, Castillo MA: Reliability of the RAPD technique for germplasm analysis of sesame (*Sesamum indicum* L.) from Venezuela. *Interciencia* 2006, **31**(6):456–460.
20. Kim DH, Zur G, Danin-Poleg Y, Lee SW, Shim KB, Kang CW, Kashi Y: Genetic relationships of sesame germplasm collection as revealed by inter-simple sequence repeats. *Plant Breed* 2002, **121**(3):259–262.

21. Kumar H, Kaur G, Banga S: Molecular Characterization and Assessment of Genetic Diversity in Sesame (*Sesamum indicum* L.) Germplasm Collection Using ISSR Markers. *J Crop Improv* 2012, **26**(4):540–557.
22. Dixit A, Jin MH, Chung JW, Yu JW, Chung HK, Ma KH, Park YJ, Cho EG: Development of polymorphic microsatellite markers in sesame (*Sesamum indicum* L.). *Mol Ecol Notes* 2005, **5**(4):736–738.
23. Hongyan L, Kun W, Minmin Y, Yang Z, Yingzhong Z: DNA Fingerprinting of Sesame (*Sesamum indicum* L.) Varieties (Lines) from Recent National Regional Trials in China. *Acta Agron Sin* 2012, **38**(04):596–605.
24. Cho Yi, Park JH, Lee CW, Ra WH, Chung JW, Lee JR, Ma KH, Lee SY, Lee KS, Lee MC, Park YJ: Evaluation of the genetic diversity and population structure of sesame (*Sesamum indicum* L.) using microsatellite markers. *Genes Genom* 2011, **33**(2):187–195.
25. Zhang HY, Wei LB, Miao HM, Zhang TD, Wang CY: Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC Genomics* 2012, **13**:316.
26. Wang L, Zhang Y, Qi X, Gao Y, Zhang X: Development and characterization of 59 polymorphic cDNA-SSR markers for the edible oil crop *Sesamum indicum* (Pedaliaceae). *Am J Bot* 2012, **99**(10):e394–398.
27. Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A: The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol Breeding* 1996, **2**(3):225–238.
28. Zhang P, Li J, Li X, Liu X, Zhao X, Lu Y: Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. *PLoS One* 2011, **6**(12):e27565.
29. Patzak J, Papirstein F, Henychova A, Sedlak J: Comparison of genetic diversity structure analyses of SSR molecular marker data within apple (*Malus x domestica*) genetic resources. *Genome* 2012, **55**(9):647–665.
30. Emanuelli F, Lorenzi S, Grzeskowiak L, Catalano V, Stefanini M, Troggio M, Myles S, Martinez-Zapater JM, Zyprian E, Moreira FM, Grandi MS: Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biol* 2013, **13**:39.
31. Frascaroli E, Schrag LA, Melchinger AE: Genetic diversity analysis of elite European maize (*Zea mays* L.) inbred lines using AFLP, SSR, and SNP markers reveals ascertainment bias for a subset of SNPs. *Theor Appl Genet* 2013, **126**(1):133–141.
32. Wurschum T, Langer SM, Longin CF, Korzun V, Akhunov E, Ebmeyer E, Schachschneider R, Schacht J, Kazman E, Reif JC: Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. *Theor Appl Genet* 2013, **126**(6):1477–1486.
33. Wei LB, Zhang HY, Zheng YZ, Miao HM, Zhang TZ, Guo WZ: A Genetic Linkage Map Construction for Sesame (*Sesamum indicum* L.). *Genes Genom* 2009, **31**(2):199–208.
34. Britten RJ, Rowen L, Williams J, Cameron RA: Majority of divergence between closely related DNA samples is due to indels. *Proc Natl Acad Sci USA* 2003, **100**(8):4661–4665.
35. Liu B, Wang Y, Zhai W, Deng J, Wang H, Cui Y, Cheng F, Wang XW, Wu J: Development of InDel markers for *Brassica rapa* based on whole-genome re-sequencing. *Theor Appl Genet* 2013, **126**(1):231–239.
36. Pacurar DI, Pacurar ML, Street N, Bussell JD, Pop TI, Gutierrez L, Bellini C: A collection of INDEL markers for map-based cloning in seven *Arabidopsis* accessions. *J Exp Bot* 2012, **63**(7):2491–2501.
37. Ollitrault F, Terol J, Martin AA, Pina JA, Navarro L, Talon M, Ollitrault P: Development of Indel Markers from *Citrus Clementina* (Rutaceae) Bac-End Sequences and Interspecific Transferability in *Citrus*. *Am J Bot* 2012, **99**(7):E268–E273.
38. Hou XH, Li LC, Peng ZY, Wei BY, Tang SJ, Ding MY, Liu JJ, Zhang FX, Zhao YD, Gu HY, Qu LJ: A platform of high-density INDEL/CAPS markers for map-based cloning in *Arabidopsis*. *Plant J* 2010, **63**(5):880–888.
39. Heesacker A, Kishore VK, Gao WX, Tang SX, Kolkman JM, Gingle A, Matvienko M, Kozik A, Michelmore RM, Lai Z, Rieseberg LH, Knapp SJ: SSRs and INDELS mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. *Theor Appl Genet* 2008, **117**(7):1021–1029.
40. Hayashi K, Yoshida H, Ashikawa I: Development of PCR-based allele-specific and InDel marker sets for nine rice blast resistance genes. *Theor Appl Genet* 2006, **113**(2):251–260.
41. Garcia-Lor A, Luro F, Navarro L, Ollitrault P: Comparative use of InDel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity: a perspective for genetic association studies. *Mol Genet Genomics* 2012, **287**(1):77–94.
42. Raman H, Raman R, Wood R, Martin P: Repetitive indel markers within the ALMT1 gene conditioning aluminium tolerance in wheat (*Triticum aestivum* L.). *Mol Breeding* 2006, **18**(2):171–183.
43. Vasemagi A, Gross R, Palm D, Paaver T, Primmer CR: Discovery and application of insertion-deletion (INDEL) polymorphisms for QTL mapping of early life-history traits in *Atlantic salmon*. *BMC Genomics* 2010, **11**:156.
44. Ke T, Dong CH, Mao H, Zhao YZ, Chen H, Liu HY, Dong XY, Tong CB, Liu SY: Analysis of expression sequence tags from a full-length-enriched cDNA library of developing sesame seeds (*Sesamum indicum*). *BMC Plant Biol* 2011, **11**:180.
45. Spandana B, Reddy VP, Prasanna GJ, Anuradha G, Sivaramakrishnan S: Development and characterization of microsatellite markers (SSR) in Sesamum (*Sesamum indicum* L.) species. *Appl Biochem Biotechnol* 2012, **168**(6):1594–1607.
46. Yue WD, Wei LB, Zhang TD, Li C, Miao HM, Zhang HY: Analysis of genetic diversity and population structure of germplasm resources in sesame (*Sesamum indicum* L.) by SSR markers. *Acta Agronomica Sinica (Chinese)* 2012, **38**(12):2286–2296.
47. Wei W, Zhang Y, Lu H, Li D, Wang L, Zhang X: Association Analysis for Quality Traits in a Diverse Panel of Chinese Sesame (*Sesamum indicum* L.) Germplasm. *J Integr Plant Biol* 2013, **55**(8):745–758.
48. Yang X, Yan J, Shah T, Warburton ML, Li Q, Li L, Gao Y, Chai Y, Fu Z, Zhou Y, Xu S, Bai G, Meng Y, Zheng Y, Li J: Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. *Theor Appl Genet* 2010, **121**(3):417–431.
49. Wei W, Zhang Y, Lu H, Wang L, Li D, Zhang X: Population Structure and Association Analysis of Oil Content in a Diverse Set of Chinese Sesame (*Sesamum indicum* L.) Germplasm. *Agr Sci China* 2012, **45**(10):1895–1903.
50. Nyongesa B, Were B, Gudu S, Dangasuk O, Onkware A: Genetic diversity in cultivated sesame (*Sesamum indicum* L.) and related wild species in East Africa. *J Crop Sci Biot* 2013, **16**(1):9–15.
51. Jin L, Lu Y, Xiao P, Sun M, Corke H, Bao J: Genetic diversity and population structure of a diverse set of rice germplasm for association mapping. *Theor Appl Genet* 2010, **121**(3):475–487.
52. Das B, Sengupta S, Parida SK, Roy B, Ghosh M, Prasad M, Ghose TK: Genetic diversity and population structure of rice landraces from Eastern and North Eastern States of India. *BMC Genet* 2013, **14**:71.
53. Chen X, Min D, Yasir TA, Hu YG: Genetic diversity, population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. *PLoS One* 2012, **7**(9):e44510.
54. Zhang L, Liu D, Guo X, Yang W, Sun J, Wang D, Sourdille P, Zhang A: Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers. *BMC Genet* 2011, **12**:42.
55. Yang X, Gao S, Xu S, Zhang Z, Prasanna BM, Li L, Li J, Yan J: Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol Breeding* 2011, **28**(4):511–526.
56. Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JS, Jaqueth J, Smith OS, Doebley J: An analysis of genetic diversity across the maize genome using microsatellites. *Genetics* 2005, **169**(3):1617–1630.
57. Roy A, Bandyopadhyay A, Mahapatra AK, Ghosh SK, Singh NK, Bansal KC, Koundal KR, Mohapatra T: Evaluation of genetic diversity in jute (*Corchorus* species) using STMS, ISSR and RAPD markers. *Plant Breed* 2006, **125**(3):292–297.
58. Zhao WG, Mia XX, Jia SH, Pan YL, Huang Y: Isolation and characterization of microsatellite loci from the mulberry, *Morus* L. *Plant Sci* 2005, **168**(2):519–525.
59. Wei W, Qi X, Wang L, Zhang Y, Hua W, Li D, Lv H, Zhang X: Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. *BMC Genomics* 2011, **12**:451.
60. Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA: Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 2008, **3**(10):e3376.
61. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH: Stacks: building and genotyping loci de novo from short-read sequences. *G3 (Bethesda)* 2011, **1**(3):171–182.
62. Lee WH, Vega VB: Heterogeneity detector: finding heterogeneous positions in Phred/Phrap assemblies. *Bioinformatics* 2004, **20**(16):2863–2864.

63. Liang HW, Wang CZ, Li Z, Luo XZ, Zou GW: Improvement of the silver-stained technique of polyacrylamide gel electrophoresis. *Yi Chuan* 2008, **30**(10):1379–1382.
64. Nei M: Genetic distance between populations. *Am Nat* 1972, **283**:292.
65. Liu K, Muse SV: PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 2005, **21**(9):2128–2129.
66. Laurentin H: Data analysis for molecular characterization of plant genetic resources. *Genet Resour Crop Ev* 2009, **56**(2):277–292.
67. Fisher RA: On the interpretation of χ^2 from contingency tables, and the calculation of P. *J Roy Stat Soc* 1922, **85**(1):87–94.
68. Rousset F: genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour* 2008, **8**(1):103–106.
69. Pritchard JK, Stephens M, Donnelly P: Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 2000, **155**(2):945–959.
70. Falush D, Stephens M, Pritchard JK: Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003, **164**(4):1567–1587.
71. Evanno G, Regnaut S, Goudet J: Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 2005, **14**(8):2611–2620.
72. Jakobsson M, Rosenberg NA: CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 2007, **23**(14):1801–1806.
73. Xu P, Wu X, Wang B, Luo J, Liu Y, Ehlers J, Close T, Roberts P, Lu Z, Wang S: Genome wide linkage disequilibrium in Chinese asparagus bean (*Vigna unguiculata* ssp. *sesquipedialis*) germplasm: implications for domestication history and genome wide association studies. *Heredity* 2012, **109**(1):34–40.
74. Rohlf FJ: *NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.1*. 2000.
75. Excoffier L, Laval G, Schneider S: Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform* 2005, **1**:47–50.

doi:10.1186/1471-2156-15-35

Cite this article as: Wu et al.: Genetic analysis and molecular characterization of Chinese sesame (*Sesamum indicum* L.) cultivars using Insertion-Deletion (InDel) and Simple Sequence Repeat (SSR) markers. *BMC Genetics* 2014 **15**:35.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

