

RESEARCH ARTICLE

Genome-Wide Analysis of Microsatellite Markers Based on Sequenced Database in Chinese Spring Wheat (*Triticum aestivum* L.)

Bin Han¹✉, Changbiao Wang^{*2}✉, Zhaohui Tang^{2*}, Yongkang Ren³, Yali Li², Dayong Zhang^{4*}, Yanhui Dong², Xinghua Zhao²

1 College of Bio-engineering, Shanxi University, Taiyuan, China, **2** Biotechnology Research Center, Shanxi Academy of Agricultural Sciences, Taiyuan, China, **3** Institute of Crop Science, Shanxi Academy of Agricultural Sciences, Taiyuan, China, **4** Provincial Key Laboratory of Agrobiolgy, Institute of Biotechnology, Jiangsu Academy of Agricultural Sciences, Nanjing, China

✉ These authors contributed equally to this work.

* tzhui126@126.com (ZHT); cotton.z@126.com (DYZ); wcbksl@126.com (CBW)



CrossMark
click for updates

OPEN ACCESS

Citation: Han B, Wang C, Tang Z, Ren Y, Li Y, Zhang D, et al. (2015) Genome-Wide Analysis of Microsatellite Markers Based on Sequenced Database in Chinese Spring Wheat (*Triticum aestivum* L.). PLoS ONE 10(11): e0141540. doi:10.1371/journal.pone.0141540

Editor: Hikmet Budak, Sabanci University, TURKEY

Received: June 12, 2015

Accepted: October 10, 2015

Published: November 4, 2015

Copyright: © 2015 Han et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: SSR data are available at (<http://dx.doi.org/10.6084/m9.figshare.1575662>). Primer data are available at (<http://dx.doi.org/10.6084/m9.figshare.1575663>).

Funding: This work was financially supported in part by grants from the Twelfth Five-year-plan in National Science and Technology for the Rural Development in China (2011AA100501), Shanxi Provincial Scientific and Technological Research Projects (20130311001-8), Shanxi Academy of Agricultural Sciences Breeding Project (11yzgc028), Shanxi Academy of Agricultural Sciences Breeding Base Project (yzjc1201), Shanxi Academy of Agricultural

Abstract

Microsatellites or simple sequence repeats (SSRs) are distributed across both prokaryotic and eukaryotic genomes and have been widely used for genetic studies and molecular marker-assisted breeding in crops. Though an ordered draft sequence of hexaploid bread wheat have been announced, the researches about systemic analysis of SSRs for wheat still have not been reported so far. In the present study, we identified 364,347 SSRs from among 10,603,760 sequences of the Chinese spring wheat (CSW) genome, which were present at a density of 36.68 SSR/Mb. In total, we detected 488 types of motifs ranging from di- to hexanucleotides, among which dinucleotide repeats dominated, accounting for approximately 42.52% of the genome. The density of tri- to hexanucleotide repeats was 24.97%, 4.62%, 3.25% and 24.65%, respectively. AG/CT, AAG/CTT, AGAT/ATCT, AAAAG/CTTTT and AAAATT/AATTTT were the most frequent repeats among di- to hexanucleotide repeats. Among the 21 chromosomes of CSW, the density of repeats was highest on chromosome 2D and lowest on chromosome 3A. The proportions of di-, tri-, tetra-, penta- and hexanucleotide repeats on each chromosome, and even on the whole genome, were almost identical. In addition, 295,267 SSR markers were successfully developed from the 21 chromosomes of CSW, which cover the entire genome at a density of 29.73 per Mb. All of the SSR markers were validated by reverse electronic-Polymerase Chain Reaction (re-PCR); 70,564 (23.9%) were found to be monomorphic and 224,703 (76.1%) were found to be polymorphic. A total of 45 monomorphic markers were selected randomly for validation purposes; 24 (53.3%) amplified one locus, 8 (17.8%) amplified multiple identical loci, and 13 (28.9%) did not amplify any fragments from the genomic DNA of CSW. Then a dendrogram was generated based on the 24 monomorphic SSR markers among 20 wheat cultivars and three species of its diploid ancestors showing that monomorphic SSR markers represented a promising source to increase the number of genetic markers available for the wheat genome. The results of this study will be useful for investigating the genetic diversity

Science Scientific and Technological Research Projects (2013GG51), the Natural Science Foundation of Shanxi (2009021032-1).

Competing Interests: The authors have declared that no competing interests exist.

and evolution among wheat and related species. At the same time, the results will facilitate comparative genomic studies and marker-assisted breeding (MAS) in plants.

Introduction

Wheat (*Triticum aestivum*) is one of the most important cereals worldwide. The consumption of wheat is greater than that of rice, especially in China and India [1]. Moreover, wheat has long served as a major renewable resource, providing both feed and industrial raw materials [2]. However, after experiencing explosive growth over the past 40 years, the annual increase in wheat yield has begun to slow or even stagnate in most countries of the world [3]. The deterioration of environmental conditions, such as drought, heat and flooding, and the increasing world population have increased the demand for wheat [3]. To face the increasing demand for wheat, it will be important to breed new varieties of wheat that can withstand biotic and abiotic stresses (heat, cold, drought, flooding and so on) while maintaining yields and quality under conservation agriculture management practices [2]. Unlike the traditional process of phenotypic selection, which is too expensive and labor-intensive, new genetic and genomic approaches have been adopted to improve germplasm characterization at the molecular level [4]. During the past decades, great efforts have been made to develop molecular markers in wheat to improve breeding strategies [5].

Markers can reveal new alleles, as well as original alleles that were reduced in the wheat gene pool during the process of evolution, thereby offering a deeper understanding of wheat during domestication and selection. Due to the very large size and polyploid complexity of wheat genome [6], progress in wheat research has been slow. However, numerous molecular markers including restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs), sequence-tagged sites (STS), DNA amplification fingerprinting (DAF), amplified fragment length polymorphisms (AFLPs), simple sequence repeats (SSRs)/microsatellites, expressed sequence tags (ESTs) and single nucleotide polymorphisms (SNPs) have been used for molecular development, marker-assisted selection and marker validation in various wheat breeding studies [7].

SSRs have become the best choice among markers used in plant breeding programs, as they are practical, convenient, easy to use and inexpensive. Among all available molecular markers, SSRs are easy to score and have wide genomic distribution, codominant inheritance and a multiallelic nature. In addition, SSRs are superior to SNP markers because SSR markers can reveal more information per locus than biallelic SNP markers [8], which explains why SSR markers remain popular. To date, more than 4,000 SSR markers have been developed and used in genetic mapping studies of wheat. These markers enabled the construction of consensus maps or comparative maps by facilitating increasing marker density in specific regions [9]. An amount of SSR markers have been identified from chromosomes in wheat, like 1AL and 5DS [10,11]. Lucas et al have identified 362 SSR markers and 6948 ISBP molecular markers from the long arm of *T.aestivum* chromosome 1A. Then 44 putative markers (eight SSRs, 26 ISBPs and ten ISBPs incorporating SSRs) were tested for polymorphism. 23 (52.3%) were found to be useful. These work will benefit to map chromosomes and further research in wheat marker assisted breeding.

SSRs, which are unevenly distributed in the genomes of prokaryotes and eukaryotes, are tandemly repeated sequences comprising 1–6 base pair (bp) [12]. SSRs derived from expressed sequence tags and genomic libraries are referred to as EST-SSRs and g-SSRs, respectively. To

date, numerous SSRs have served as powerful tools to assess genetic diversity, establish core collections, select hybrid parents, study population structures and map or tag functional genes [13]. The polymorphism rate in EST-SSRs is lower than that in g-SSRs [14]. And g-SSRs can serve as valuable complements to EST-SSRs. Numerous EST-SSRs have been generated for wheat, which have revealed high universality between wheat and other cereals, such as barley, maize, rice and sorghum [15,16]. However, few studies have focused on identifying and analyzing g-SSRs in wheat. The availability of the whole draft genome sequence of CSW [6] provides an opportunity to accelerate the process of germplasm evaluation and breeding line identification in wheat breeding programs.

In this study, we identified g-SSRs from the recently sequenced genomic sequence of wheat cv. Chinese spring. The objectives of this study were a) to characterize the density, type and distribution of g-SSR motifs in CSW; b) to develop and analyze Chinese spring genomic SSR markers from a collection of genomic sequences and c) to evaluate the efficiency of these markers in polymorphism identification for application in comparative genomic studies and breeding.

Materials and Methods

Plant material

The 23 samples used to validate the polymorphic nature of genic-SSR candidate markers included 20 wheat cultivars and three species of its diploid ancestors (wheat A,B,D -genome progenitor *Triticum urartu*, *Aegilops speltoides* Tausch and *Aegilops tauschii*). The 20 wheat cultivars, *Triticum urartu* and *Aegilops speltoides* Tausch were provided by the Institute of Crop Science, Shanxi Academy of Agricultural Sciences, Taiyuan, China. *Aegilops tauschii* was provided by the Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing, China.

Source of genomic sequences

The genome sequence of model wheat (*Triticum aestivum* cv. Chinese spring) was obtained in FASTA format from URGI (<https://urgi.versailles.inra.fr/download/iwgs/>). A total of 10,603,760 sequences were downloaded and studied.

SSR mining and primer design

The identification and localization of g-SSRs were carried out using MicroSatellite (MISA, <http://www.pgrc.ipk-gatersleben.de/misa>) and Primer 3.0 for large-scale primer design. The criteria used to search SSRs with the MISA script were as follows: motifs between two and six nucleotides long, with a minimum of ten repeats for dinucleotides, seven repeats for trinucleotides, five repeats for tetranucleotides and four repeats for penta- and hexanucleotides. The major parameters for primer design were as follows: primer length, 18–22 bp, with 20 bp being optimal; PCR product size, 100–800 bp; an annealing temperature of 50–65°C, with 57°C being optimal; and an optimal GC content of 50%, with 45% being the minimum.

Analysis of SSR polymorphism

Analysis of the uniqueness and specificity of the designed SSR markers in the Chinese spring genome was performed using the re-PCR strategy (<http://www.ncbi.nlm.nih.gov/tools/epcr/>). Re-PCR can be used to map STSs (sequencing tagged site) or short primers in sequence database. It is a version of e-PCR searching for STSs within DNA sequences. Those parameters were: re-PCR—S hash-file—n1 -g1 -r +. Subsequently, the corresponding amplicons were

analyzed and the previously obtained SSR markers were classified as definitely polymorphic or monomorphic. Polymorphic SSR markers amplified multiple identical loci in the Chinese spring genome, while monomorphic markers tended to amplify one locus. These data were analyzed with Excel microsoftware and plotted.

The validation of monomorphic SSR markers in the CSW genome

A total of 45 pairs of g-SSR primers were selected randomly for the validation of the designed monomorphic SSR markers in the CSW genome. Genomic DNA was extracted from fresh, young leaves of CSW using an improved cetyltrimethyl ammonium bromide (CTAB) method [17]. After extraction, the DNA quality and concentration were further assessed using an eppendorf biophotometer. Polymerase chain reaction (PCR) was performed in a total volume of 20.0 µl containing 1 µl of 50 ng/µl template DNA, 2 µl of 10× PCR buffer containing 20 mM MgCl₂, 0.4 mM of dNTPs, 0.3 U of Taq polymerase and sterile distilled water and 0.8 µl of 10 µmol/L each of forward and reverse primers. The reactions were performed using the following conditions: 94°C for 2 min; 35 cycles of 94°C for 40 s, 55°C for 45 s, and 72°C for 60 s; and a final step at 72°C for 7 min. Then, 2 µl of the PCR product and a 600bp molecular size marker were loaded onto an 8% denaturing polyacrylamide (PAGE) gel in 1×TBE buffer, run at 100 V, and visualized using silver staining. SSR analysis was performed at least twice to confirm primer amplification.

Phylogenetic relationship [18] among 20 wheat cultivars and three species of its diploid ancestors was constructed in a dendrogram coefficients using the program of NTSYS-pc Version 2.10 to estimate the SSRs monomorphic SSR markers.

Results

Identification and characterization of SSRs in the whole genome of CSW

A total of 364,347 SSRs were identified from among 10,603,760 sequences covering 9,932,960,273 bp of the Chinese spring genome, which were present at a density of 36.68 SSR/Mb. We detected 488 types of SSR motifs in total. Among these types, there were four, 10, 32, 102 and 340 types of di- to hexanucleotide repeats, respectively. The number of each type of motif increased with increasing length, as shown in Table 1. The percentage of dinucleotides was the lowest, while the percentage of hexanucleotides was the highest, accounting for 0.82% and 69.67% of the total, respectively. We further characterized the identified SSRs (Fig 1).

Table 1. Characteristics of SSRs in CSW.

SSR mining		Total	
Total length of analyzed equences (bp)		9932960273	
Number of identified SSRs		364347	
SSR/Mb		36.68	
Type	Kinds	number	percentage(%)
Di	4	154913	42.52
Tri	10	90960	24.97
Tetra	32	16823	4.62
Penta	102	11828	3.25
Hexa	340	89823	24.65
Total	488	364347	100.00

doi:10.1371/journal.pone.0141540.t001

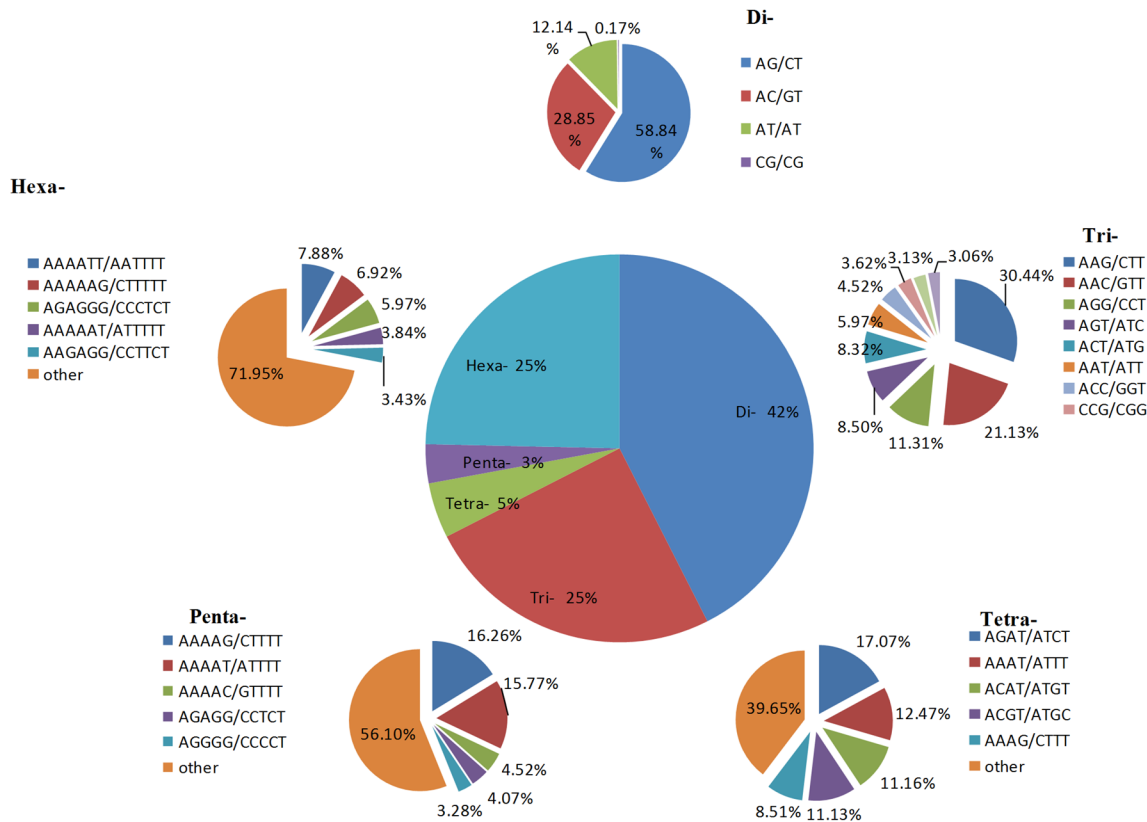


Fig 1. Distribution of repeat types in the Chinese spring wheat.

doi:10.1371/journal.pone.0141540.g001

Among the different unit sizes, dinucleotide repeats dominated over other types of repeats, accounting for approximately 42.52% of the genome. The density of SSRs decreased stepwise with increasing motif length (di- to pentanucleotide), except for hexanucleotide repeats. The density of tri- to hexanucleotide repeats was 24.97%, 4.62%, 3.25% and 24.65% among all SSRs, respectively. Surprisingly, we found that the ratio of trinucleotides was nearly equal to that of hexanucleotides. Among dinucleotide repeats, AG/CT had the highest occurrence (58.84%), followed by AC/GT (28.85%) and AT/AT (12.14%). CG/CG (0.17%) were so rare that they could almost be discounted. Among trinucleotide repeats, AAG/CTT made up the highest proportion (30.44%), followed by nine types of repeats, including AAC/GTT (21.13%), AGG/CCT (11.31%), AGT/ATC (8.50%), ACT/ATG (8.32%), AAT/ATT (5.97%), ACC/GGT (4.52%), CCG/CGG (3.62%), AGC/CGT (3.13%) and ACG/CTG (3.06%). The most common tetranucleotide repeats were AGAT/ATCT (17.07%), AAAT/ATTT (12.47%) and ACAT/ATGT (11.16%). AT-rich repeat patterns were the most abundant among penta- and hexanucleotides, such as AAAAG/CTTTT, AAAAT/ATTTT and AAAAC/GTTTT for pentanucleotides and AAAATT/AATTTT, AAAAAG/CTTTTT and AGAGGG/CCCTCT for hexanucleotides ([S1 Table](#)).

Characterization of SSRs on each chromosome of CSW

We then analyzed the distribution of SSRs on each chromosome of the Chinese spring genome ([Table 2](#)). The number of SSRs on each chromosome was as follows: 1A to 7A: 14,703, 20,822,

Table 2. Chromosome-wide distribution of SSRs in the CSW genome.

Genome	Chromosome	bp	SSR	SSR/Mb
A	1A	424409652	14703	34.64
	2A	576846926	20822	36.10
	3A	445759266	14289	32.06
	4A	600765939	20728	34.50
	5A	511490084	17477	34.17
	6A	428557670	16211	37.83
	7A	446879367	17699	39.61
B	1B	507010493	18876	37.23
	2B	686985855	25495	37.11
	3B	632916551	24717	39.05
	4B	548472594	20049	36.55
	5B	582228367	21523	36.97
	6B	464691465	18005	38.75
	7B	461943202	18814	40.73
D	1D	379056246	12264	32.35
	2D	421689632	17406	41.28
	3D	327157283	11670	35.67
	4D	325658742	11052	33.94
	5D	380292678	14072	37.00
	6D	351893008	12302	34.96
	7D	428255253	16173	37.76

doi:10.1371/journal.pone.0141540.t002

14,289, 20,728, 17,477, 16,211 and 17,699, respectively; 1B to 7B: 18,876, 25,495, 24,717, 20,049, 21,523, 18,005 and 18,814, respectively; 1D to 7D: 12,264, 17,406, 11,670, 11,052, 14,072, 12,302 and 16,173, respectively. Therefore, chromosome 2B and 3B had the largest number of SSRs among Chinese spring chromosomes, followed by chromosome 5B, 2A and 4A, while chromosome 4D contained the fewest SSRs. The top three longest chromosomes are 2B, 3B and 4A and the shortest is 4D, which is in accordance with the number of SSRs per chromosome. Overall, the differences in the densities of SSRs on different chromosomes were not significant, ranging from 41.28 SSR/Mb to 32.06 SSR/Mb. The density on chromosome 2D was the highest (41.28 SSR/Mb), while that on chromosome 3A was the lowest (32.06 SSR/Mb). As mentioned above, the average density of SSRs on the entire genome was 36.68 SSR/Mb, which is close to the density on chromosomes 4B and 5B. We then analyzed the repeats on each chromosome (Table 3). The percentage of di-, tri-, tetra-, penta- and hexanucleotide

Table 3. Number of different repeat classes in Chinese spring.

SSR repeat	Genome							Total
	1	2	3	4	5	6	7	
Di-	18923	26587	22370	21373	21541	20468	23651	154913
Tri-	11822	16168	12405	13278	13542	11404	12341	90960
Tetra-	2121	2948	2236	2427	2505	1964	2622	16823
Penta-	1464	2187	1621	1708	1837	1353	1658	11828
Hexa-	11513	15833	12044	13043	13647	11329	12414	89823
Total-	45843	63723	50676	51829	53072	46518	52686	364347

doi:10.1371/journal.pone.0141540.t003

repeats on every chromosome (and even on the whole genome) was nearly identical. The percentage of dinucleotide repeats was the highest, followed by tri-, hexa- and tetranucleotide repeats, while the percentage of pentanucleotide repeats was so low that they could almost be discounted. The ratios of tri- and hexanucleotides were nearly equivalent (Fig 2). In addition, AG/CT and AC/GT were the most abundant dinucleotide repeats on each chromosome, while AAG/CTT, AAC/GTT, AGG/CCT, AGT/ATC and ACT/ATG were the most abundant trinucleotide repeats on each chromosome. The largest proportion of tetranucleotide repeats included AGAT/ATCT, AAAT/ATTT, ACAT/ATGT and ACGT/ATGC, and the most abundant penta- and hexanucleotide repeats were AAAAN and AAAAAAN, respectively (S2 Table).

Genome-wide SSR markers development and polymorphism analysis

All SSRs were selected for SSR marker development, and a total of 295,267 SSR markers were successfully designed from the 21 chromosomes of CSW, covering the whole genome at a density of 29.73 per Mb. The densities of SSR markers on each chromosome were similar, ranging from 25.58 to 31.34 per Mb (Table 4). Among the chromosomes that contained SSR markers, the highest density of SSR markers was found on chromosome 5D, followed by 2D and 7D. Furthermore, all SSR markers were validated and subjected to polymorphism analysis via re-PCR. Markers that amplified prominent PCR products were classified as either polymorphic or monomorphic based on the number of corresponding loci. Of the markers amplified, 70,564 were monomorphic and 224,703 were polymorphic. The monomorphic markers included 2,387 (3.38%) present in compound formation and 8,177 (99.96%) present in perfect

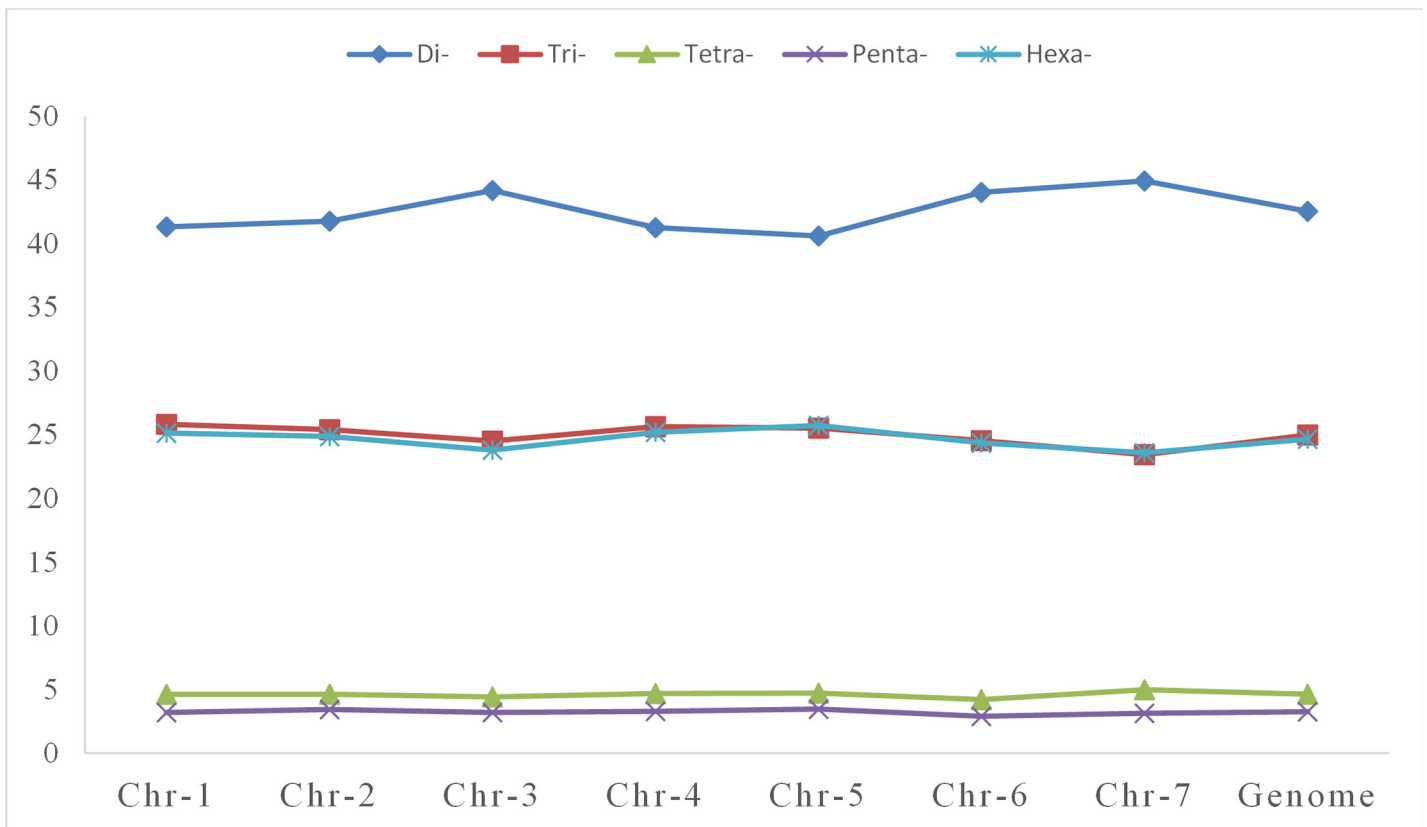


Fig 2. Frequency (%) of di- to hexanucleotide motifs in the Chinese spring wheat.

doi:10.1371/journal.pone.0141540.g002

Table 4. Summary of SSR markers in the Chinese spring genome.

Genome	Chromosome	bp	SSR marker	SSR marker/Mb
A	1A	424409652	12364	29.13
	2A	576846926	17163	29.75
	3A	445759266	11898	26.69
	4A	600765939	17113	28.49
	5A	511490084	14274	27.91
	6A	428557670	13308	31.05
	7A	446879367	13922	31.15
B	1B	507010493	15589	30.75
	2B	686985855	21001	30.57
	3B	632916551	19583	30.94
	4B	548472594	16250	29.63
	5B	582228367	17570	30.18
	6B	464691465	14126	30.40
	7B	461943202	14343	31.05
D	1D	379056246	10218	26.96
	2D	421689632	13193	31.29
	3D	327157283	8370	25.58
	4D	325658742	9093	27.92
	5D	380292678	11919	31.34
	6D	351893008	10611	30.15
	7D	428255253	13359	31.19
Whole genome		9932960273	295267	29.73

doi:10.1371/journal.pone.0141540.t004

formation, whose dinucleotide motifs (34.46%) were the most common, followed by hexanucleotide (28.29%), trinucleotide (23.72%), tetranucleotide (7.88%) and pentanucleotide motifs (5.65%), respectively (Fig 3). Moreover, we also examined the distribution of monomorphic markers on the Chinese spring chromosomes. Chromosome 3B had 5,387 monomorphic markers, which was considerably higher than that of the other chromosomes, followed by chromosome 2B and 5B, containing 5,112 and 4,346 monomorphic markers, respectively. Chromosome 1D contained the fewest monomorphic markers (2,082), while 3B contained the largest number of di-, tetra-, penta- and hexanucleotide motifs and 2B contained the largest number of trinucleotide motifs. Chromosome 1D had the fewest di- and hexanucleotide motifs, 3D had the fewest tri- and pentanucleotide motifs and 4D had the fewest tetranucleotide motifs (Table 5).

The validation of monomorphic SSR markers in the CSW genome

A sub set of 45 monomorphic markers were selected randomly for validation in CSW genome (Table 6). Of the markers, 24 (53.3%) amplified one locus, which could be used for marker-assisted breeding in wheat. 8 (17.8% monomorphic SSR markers amplified multiple identical loci and 13 (28.9%) monomorphic SSR markers amplified no fragment from genome of CSW (Fig 4). These data will provide a solid base for our follow-up study. The 24 monomorphic SSR markers amplified one locus were used to analysis genetic relationship among 20 wheat cultivars and three species of its diploid ancestors. The phylogenetic relationship constructed in a dendrogram coefficients using Numerical Taxonomy System of Multivariate Programs (NTSYS) cluster analysis (Fig 5). At a similarity coefficient ≥ 0.6 , the largest group consisted of hexaploid wheat cultivars and the diploid ancestor of the B genome. The diploid ancestor of the

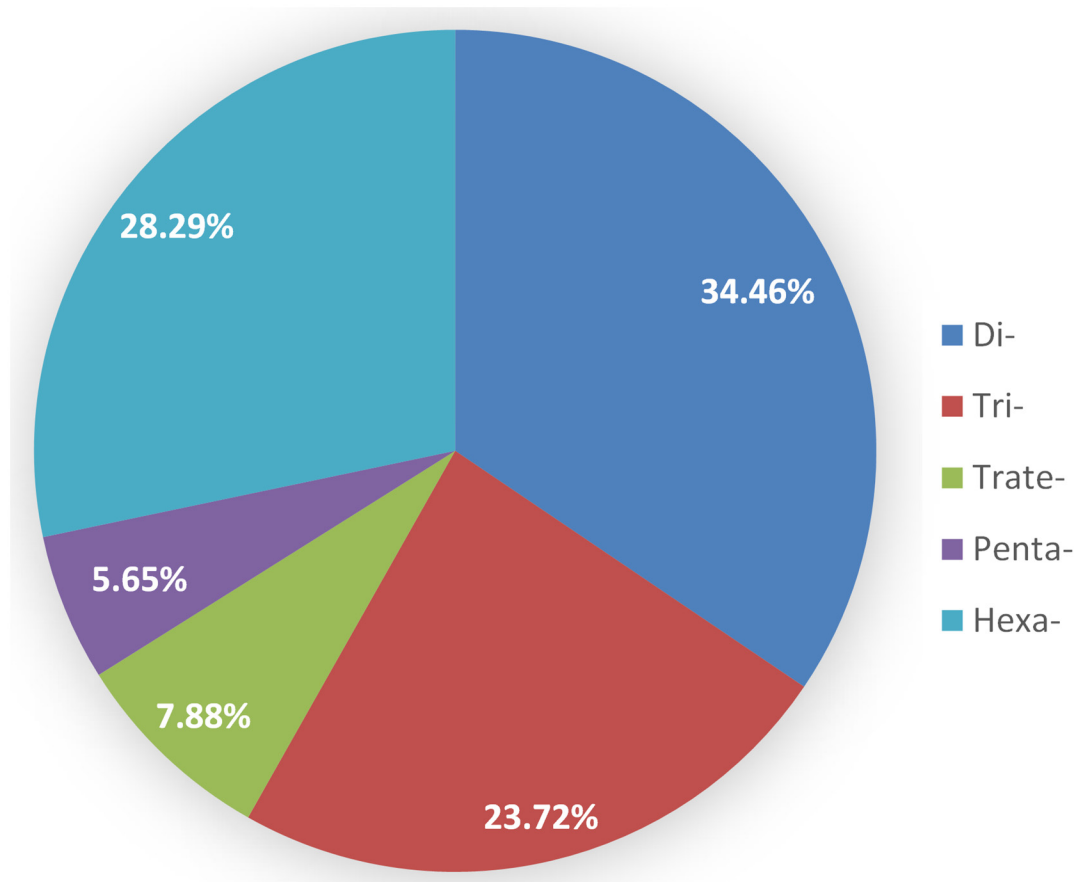


Fig 3. Distribution of monomorphic markers in the Chinese spring wheat genome.

doi:10.1371/journal.pone.0141540.g003

A and D genomes were clustered into subgroups at similarity values of 0.67. Hexaploid wheat cultivars were clustered into two subgroups as well at the same values. Our results indicate that the monomorphic SSR markers had the ability to assess molecular diversity and potential for use in fingerprinting analysis.

Discussion

Comparative characterization of SSRs between CSW and other crops

We compared the density and distribution of SSRs in the Chinese spring genome with those in other plants. A total of 364,347 SSRs were identified from among 10,603,760 unique sequences in Chinese spring wheat, at a density of 36.68 SSR/Mb. This density is extremely low compared with the reported number of SSRs in the monocot species *Brachypodium* (191.3 SSR/Mb), sorghum (175.4 SSR/Mb) and rice (363.3 SSR/Mb), as well as *Arabidopsis* (418.6 SSR/Mb), *Medicago* (495.8 SSR/Mb), the chromosome 5D of *Aegilops tauschii* [19] and *Populus* (667.9 SSR/Mb). [20]. These data may reflect the differences in DNA levels between genomes [8]. The distribution and density of SSRs are highly variable, perhaps due to differences in search criteria and database mining tools [21]. Though differently-sized genomes may also contribute to affecting repetitiveness of microsatellites, the density of SSRs were not significantly related to genome size[22,23] Moreover, the density of SSRs is considerably higher in dicot species than in monocots. We found that dinucleotide repeats were the most frequent motifs on each

Table 5. Characteristics of monomorphic SSR markers in the CSW genome.

Genome	Chromosome	compound	Di-	Tri-	Trate-	Penta-	Hexa-	Total
A	1A	87	883	676	232	155	856	2889
	2A	129	1306	964	340	213	1171	4123
	3A	93	977	626	225	146	787	2854
	4A	122	1319	890	319	225	1113	3988
	5A	78	932	619	258	180	776	2843
	6A	102	1153	791	236	161	960	3403
	7A	140	1257	696	322	170	916	3501
B	1B	131	1162	810	232	165	864	3364
	2B	182	1629	1308	359	278	1356	5112
	3B	266	1812	1242	377	305	1385	5387
	4B	115	1112	816	233	189	961	3426
	5B	133	1352	1113	329	259	1160	4346
	6B	141	1234	674	199	155	813	3216
	7B	147	1278	727	261	179	888	3480
D	1D	51	629	480	172	152	598	2082
	2D	110	1230	784	240	206	991	3561
	3D	71	749	478	192	97	613	2200
	4D	62	704	514	156	117	620	2173
	5D	80	948	711	244	180	853	3016
	6D	41	703	532	166	135	681	2258
	7D	106	1128	719	283	183	923	3342
Total		2387	23497	16170	5375	3850	19285	70564

doi:10.1371/journal.pone.0141540.t005

chromosome in CSW on the whole genome, which was also reported for the sweet orange genome [24]. Biswas *et al.* calculated the g-SSR frequencies in 11 plant genomes and found that dinucleotide repeats were predominant in both monocot and dicot genomes [20]. This conclusion is consistent with the current results. Like the genomes of *Arabidopsis thaliana* and rice, CSW contained the most AG/TC dinucleotide repeats, followed by AC/GT and AT/AT repeats [25,26]. The basic SSR composition of Chinese spring wheat gives priority to A and T in all types of repeats; for example, CG/CG occurred at the lowest density among dinucleotide repeat motifs. This is also the case for human, *Drosophila melanogaster* and other eukaryotic genomes [27]. Hong *et al.* [28] reported that GAA/TCC and AGA/TCT are the most frequent trinucleotide patterns in the *Solanaceae*, and CCG/CGG are the most abundant trinucleotide patterns in the coccolithophore *Emiliania huxleyi* [29]. Although the densities of trinucleotide repeat pattern are different among various species, the most abundant patterns in CSW, *Arabidopsis* and *Brassicarapa* are identical; the greatest number of trinucleotide repeats comprise AAG/CTT [28]. Among tetra- to hexanucleotide repeats, AAAN, AAAAN and AAAAAAN are much more common than other repeat motifs [28], which is also true for other plant genomes.

Comparative characterization of SSRs between CSW and its related species

The occurrence of SSRs in genomes mainly result from mutations during evolution, such as replication slippage, addition or removal of one or several repeat motifs. Therefore, the particular number and lengths of SSRs can serve as an index of genetic variation during the process of evolution. Chinese spring wheat, an allohexaploid *Triticum aestivum* cultivar, contains three homologous genomes (A, B and D) [30]. The number of SSRs in the A, B and D genomes was

Table 6. Characteristics of 45 SSR markers used in the present study.

ID of markers	Repeat motifs	Forward(5'-3')	Reverse(5'-3')	Tm(°C)
1	(GAA)7	AAACCCCTTCTCCCATCTAGG	TAACCGCTAGGTTTAGTCCG	57.112
2	(TAG)22	GAGGTGACTGAGGCTCTTGT	CTACTACCACCACCACCACC	57.771
3	(TTC)6	ACGGTTGTCTCTTTAGGTGG	GACCGCAACAAAGTTACCTC	57.328
4	(CATCCT)4	CAATACTCAAGCTGGCAATG	GCATATGGTTTAGTCGGGTC	57.028
5	(CCA)6	ATCTCCCGCATGAGACCT	GAGCTACTGCTGCTCAAAGC	58.159
6	(CGGCGA)3	GAGCAAGTCATCAAGTCACG	GTGCAGTTTGGTCAAGTCAG	56.833
7	(CA)19	GTCAAAGTTGTGGCATGTG	ATCAGAGCACCTCCAGAGTC	56.908
8	(GTC)7	CTGCACAACACTTATCGCTC	GAAACTCTGAAGCTGGTGGT	56.934
9	(CCG)6	AGTGCAAGACTAGGGTGGAG	GAAATGCGTTTGGTGGTTAG	58.147
10	(AAAAAG)3	TGGACATGCTATAGTGTGGG	GATGGGGATATAGCGAAATG	57.013
11	(CCCTCT)3	ATTCAGACACCGGAAGTGTT	TAGTCCTACATGGAAGGGGA	57.052
12	(TATT)7	CTTGTGAATCCCCAGTTTCT	GATGAAAATGGACAGAGAGG	57.063
13	(CCA)7	GCCGCGTACAGATAGAAGAT	GGTGTGCTTACCTCTGCTT	57.078
14	(TTA)6	TAGAAAATGGGGTGTCCAG	TAAGATAAGGACCTGGACGG	56.761
15	(CCG)7	CACACAGAGATCGAGTCAGG	AGCTAGGGTTTCTCATGG	56.78
16	(GGA)7	CTGCAGATGATGAAGACGAG	TACCTTTTACCCCAAGCAG	56.938
17	(TTC)8	GGGGAGGATATCTTGTTCCT	CCCAGCCAATCCTCTACTAA	56.914
18	(TG)11	ATCACATGTTTGTACAACGC	TGGACCCCTTAGTTGTGAGT	57.069
19	(GGA)6	GACCTAGCCATTTCTGTGCT	ATATCTTCCCCCTTTGGAC	56.941
20	(AAAAAC)3	GGACAATGCTGGACGAGTA	AGGAACATAGGATGCGAACT	56.343
21	(AAAAG)4	GCTATTTTGTTTAGCTGCCC	GAGGGTGTGTAGGATTCGAG	57.162
22	(GA)9	GTTATGCGGTAGGAGCTTGT	CAGGTCCTATTTTGCTCGAT	56.909
23	(GA)9	CAGGAAGGTAGCTTCAAGGA	GAATGAATGAATGGGTGGAG	57.775
24	(TTG)8	ACCTCTAGCCTACCCCAACT	GACAGAGGCTCCTTCCACTA	57.032
25	(GTGCCC)3	CCGCAAGATAGTGTACATGC	TAGCAGCGATGTGATGTAGC	57.63
26	(TATAGA)3	AACAGAAAAACTGTAGCGCC	CTGGGCAAAAGAAGGTTAAG	56.672
27	(GTT)6	TGTGATGAAGCTCAAGGAGA	AGAGAGAGGAAGGATTCGGT	56.966
28	(CT)11	AGAGAAGGATGACTGGCAGA	TTAATCCCTTAGTCCCTGC	58.136
29	(ACGGTG)3	TACATAGAGCTGGGATAGCG	CTGTTACCGTTGATTCCCC	56.944
30	(TG)16	GAGCACATGTCGTACCTCAA	CGTCGGGTGAAGAATAACTT	56.822
31	(TG)16	ACTACCATTTGGGTGGTAGGA	ACCTTTAGTACCGGTTGGTG	56.613
32	(TTCA)5	ATTCAATACCACCAGCCTTC	TGTAGTTTTAGCCCCTCACA	56.881
33	(TC)9	CAACTGCCCTATACCCAATC	GGATTCATGGAGAAACCTTG	57.017
34	(CAA)10	GTGCTGCAGTAGCAGTGTTT	AAGAGGAAGGAGATGCCTTT	57.075
35	(GGA)6	CCCTTTCCTGTTCCAAGTT	GAATAGTTTGGGTGCTTTCG	57.217
36	(CTTCTC)3	CCTAACAAAGCCCAGTTCTC	GAAGACTACATCCCCTGACC	57.158
37	(CT)10	TCTTCTTCTCGAACAAAC	ACATAACTGAAGGACGGGAA	57.113
38	(TAC)6	AGCTCTAGGAAATCCCCTC	GAACAAACAGGCTCAGGTCT	56.934
39	(AAG)6	GACACTTGCCCTCTGTTTT	GGGAGAGGTTGCTTGGTAT	56.613
40	(GGCTCC)3	ATGCTGGTACCTTTGAGAGC	ACGCCTCACAGAGAAAAATC	56.982
41	(ATAC)5	TCCTAGCACACACACACACA	GTAGCCCCACTGTGTAATCC	56.982
42	(AAAATA)3	GCGGCAGGATCTACTGTTAT	GGTACGTAGCTGATGATCCC	57.074
43	(TCATCG)3	CATCGTCATCATCTTCTCA	ACGAAAGGATAGAACCACCA	57.113
44	(ATGTCG)4	TACCACTTGACGACTTGAG	AAAAGAGAGCTCCAGCAAGA	57.018
45	(TC)14	GCGATCACATCTAGAACCCT	ACAAAGCTCGTGGTACATCA	56.753

doi:10.1371/journal.pone.0141540.t006

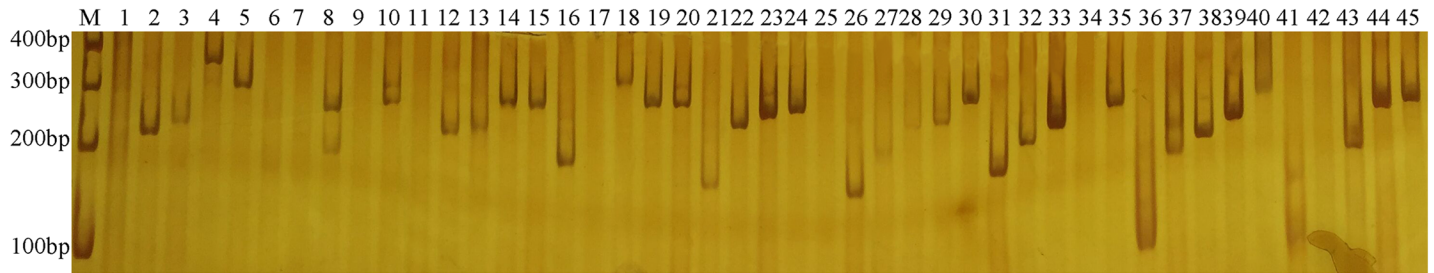


Fig 4. The validation of monomorphic SSR markers in the CSW genome.

doi:10.1371/journal.pone.0141540.g004

121,929, 147,479 and 94,939 respectively, with the B genome containing the largest number. These results suggest that there is the most variation in the B genome of CSW (Fig 6). This finding may at least partially explain why the draft genomes of the wheat A-genome progenitor *Triticum urartu* and D-genome progenitor *Aegilops tauschii* were sequenced, while there is currently no draft sequence for the B-genome progenitor [30,31]. We compared the g-SSR distribution in A-genome progenitor *Triticum urartu*, D-genome progenitor *Aegilops tauschii* and Chinese spring (Fig 7). In *Triticum urartu* and CSW, the dinucleotide repeat motifs were the predominant types (57.8% and 42.5%, respectively). By contrast, trinucleotide repeats were the most abundant motifs in *Aegilops tauschii* (39.13%). Trinucleotide repeats play an important role in specific selection against frameshift mutations in genetic regions. Trinucleotide repeats could refrain from selective pressures in coding regions owing to they had not generated frameshifts through expansion of triplet SSRs. But, non-triplet SSRs tended to face greater purifying selection with the frameshifts mutations. Therefore, the most percentage of trinucleotide repeats may related to high genic density in *Aegilops tauschii*[23,32,33]. Additionally, these variations of repeats are related to the different parameters used when mining SSRs of different species.

Genome-wide SSR markers development and polymorphism analysis

As Cavagnaro *et al.* [34] noted, mononucleotide repeats are not suitable for marker development. Therefore, we only developed primers based on di- to hexanucleotide repeats. A total of

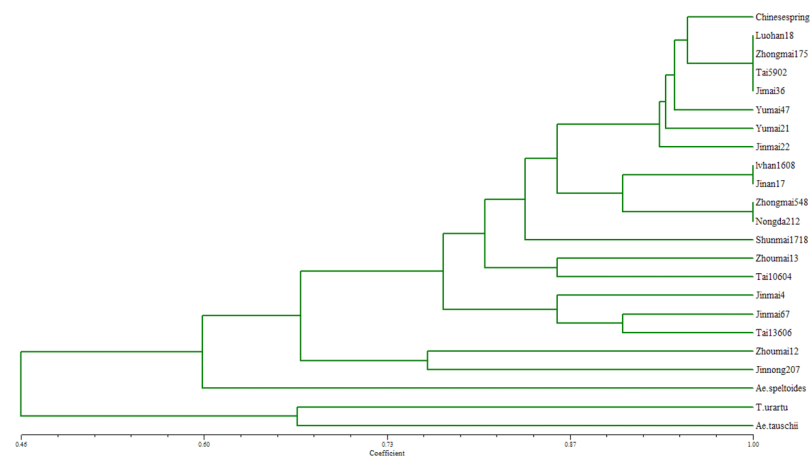


Fig 5. Phylogenetic relationship among 23 wheat cultivars was constructed using NTSYS cluster analysis.

doi:10.1371/journal.pone.0141540.g005

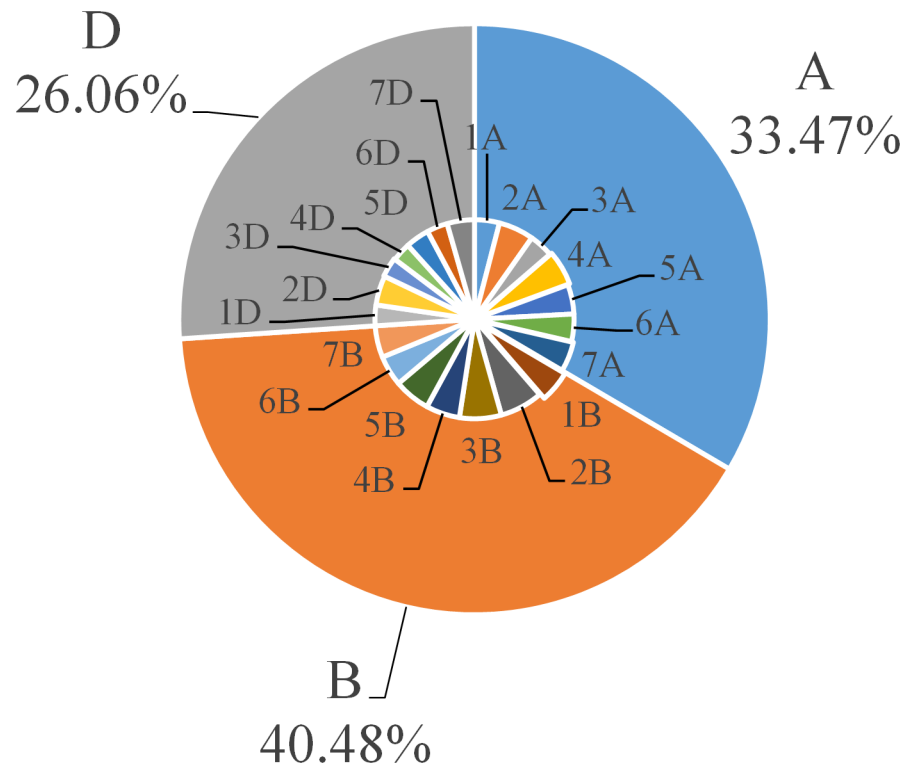


Fig 6. Distribution of g-SSRs in the Chinese spring homoeologous genomes.

doi:10.1371/journal.pone.0141540.g006

295,267 SSR markers were successfully designed, all of which were validated by re-PCR. Among the SSR markers, 70,564 (23.9%) were found to be monomorphic and 224,703 (76.1%) were found to be polymorphic. These monomorphic markers may serve as powerful tools for detecting sequence variation within a population of wheat and related species, examining the level of changes in genetic diversity and phylogenetic analyses [35]. Of the monomorphic markers, dinucleotide motifs (34.46%) were the most common, with chromosome 3B containing the most monomorphic markers, suggesting that the dinucleotide motifs on chromosome 3B may serve as better selectable markers for MAS in wheat. In addition, these markers represent an important genomic resource for use in many cereal crops and will benefit numerous genetic and genomic studies involving genetic diversity evaluation, population genetics, cloning functional genes related to agronomic and quality traits and comparative genomics in plants.

The validation of monomorphic SSR markers in the CSW genome was performed using 45 randomly selected monomorphic SSR markers. Eight amplified multiple loci. This may be a consequence of the fact that the full CSW genome sequence is not known and that there are large repetitive homologous sequences in non-homologous [6]. This can increase the frequency at which monomorphic makers amplify multiple loci. Monomorphic makers which amplify one locus are more useful for phylogenetic analysis of wheat cultivars [36]. Analyses of phylogenetic relationship by using SSR markers could provide a better understanding of genetic background of wheat cultivars and become a foundation of the genetic improvement of wheat. The results of the present study indicated the monomorphic SSR markers used in this study might provide useful information for genetic improvement and germplasm conservation, evaluation and utilization in wheat. Monomorphic makers would play an important role in many aspects

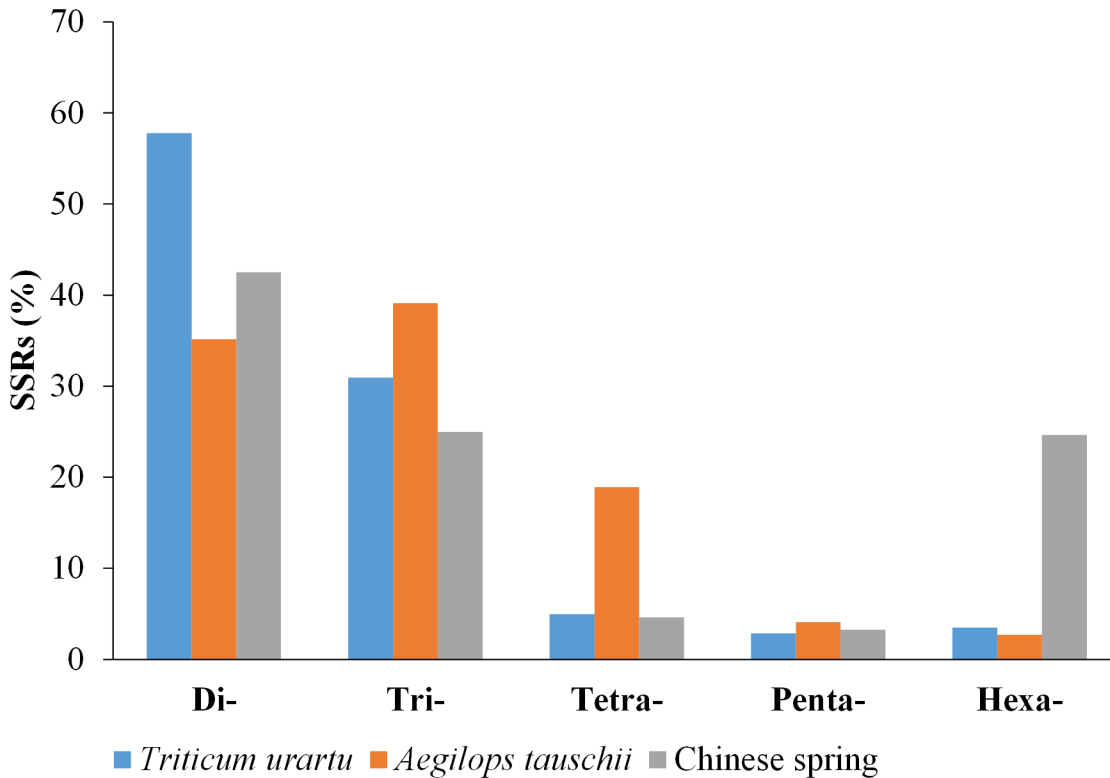


Fig 7. Comparison of SSRs among *Triticum urartu*, *Aegilops tauschii* and Chinese spring wheat.

doi:10.1371/journal.pone.0141540.g007

of wheat breeding, including in the identification of the genes responsible for desirable traits, and in the analysis of genetic relationships between and the diversity of wheat germplasm collections. Monomorphic markers should assist with improvements in wheat breeding[37].

Supporting Information

S1 Table. Summary of the frequency of different SSR repeat motif types in Chinese spring wheat.

(DOCX)

S2 Table. Distribution of different types of g-SSRs on Chinese spring chromosomes.

(DOC)

Acknowledgments

This work was financially supported in part by grants from the Twelfth Five-year-plan in National Science and Technology for the Rural Development in China (2011AA100501), Shanxi Provincial Scientific and Technological Research Projects (20130311001–8), Shanxi Academy of Agricultural Sciences Breeding Project (11yzgc028), Shanxi Academy of Agricultural Sciences Breeding Base Project (yzjc1201), Shanxi Academy of Agricultural Science Scientific and Technological Research Projects (2013GG51), the Natural Science Foundation of Shanxi (2009021032–1).

Author Contributions

Conceived and designed the experiments: ZHT DYZ. Performed the experiments: CBW. Analyzed the data: YKR YLL BH. Contributed reagents/materials/analysis tools: YHD XHZ CBW. Wrote the paper: BH.

References

1. Makharia GK, Mulder CJ, Goh KL, Ahuja V, Bai JC, Catassi C, et al. Issues associated with the emergence of coeliac disease in the Asia—Pacific region: a working party report of the World Gastroenterology Organization and the Asian Pacific Association of Gastroenterology. *Journal of gastroenterology and hepatology*. 2014; 29: 666–677. PMID: [24783246](#)
2. Paux E, Sourdille P, Mackay I, Feuillet C. Sequence-based marker development in wheat: advances and applications to breeding. *Biotechnology advances*. 2012; 30: 1071–1088. doi: [10.1016/j.biotechadv.2011.09.015](#) PMID: [21989506](#)
3. Baenziger P, Russell W, Graef G, Campbell B. Improving lives. *Crop science*. 2006; 46: 2230–2244.
4. Landjeva S, Korzun V, Börner A. Molecular markers: actual and potential contributions to wheat genome characterization and breeding. *Euphytica*. 2007; 156: 271–296.
5. Marone D, Laido G, Gadaleta A, Colasuonno P, Ficco DB, Giancaspro A, et al. A high-density consensus map of A and B wheat genomes. *Theoretical and Applied Genetics*. 2012; 125: 1619–1638. doi: [10.1007/s00122-012-1939-y](#) PMID: [22872151](#)
6. Mayer KF, Rogers J, Doležel J, Pozniak C, Eversole K, Feuillet C, et al. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*. 2014; 345: 1251788. doi: [10.1126/science.1251788](#) PMID: [25035500](#)
7. Gupta P, Varshney RK, Sharma P, Ramesh B. Molecular markers and their applications in wheat breeding. *Plant breeding*. 1999; 118: 369–390.
8. Xu J, Liu L, Xu Y, Chen C, Rong T, Ali F, et al. Development and characterization of simple sequence repeat markers providing genome-wide coverage and high resolution in maize. *DNA research*. 2013; 20: 497–509. doi: [10.1093/dnares/dst026](#) PMID: [23804557](#)
9. Ren J, Sun D, Chen L, You FM, Wang J, Peng Y, et al. Genetic diversity revealed by single nucleotide polymorphism markers in a worldwide germplasm collection of durum wheat. *International journal of molecular sciences*. 2013; 14: 7061–7088. doi: [10.3390/ijms14047061](#) PMID: [23538839](#)
10. Lucas SJ, Šimková H, Šafař J, Jurman I, Cattonaro F, Vautrin S, et al. Functional features of a single chromosome arm in wheat (1AL) determined from its structure. *Functional & integrative genomics*. 2012; 12: 173–182.
11. Wang Y, Hao C, Zheng J, Ge H, Zhou Y, Ma Z, et al. A haplotype block associated with thousand-kernel weight on chromosome 5DS in common wheat (*Triticum aestivum* L.). *Journal of integrative plant biology*. 2014.
12. Asp T, Frei UK, Didion T, Nielsen KK, Lübberstedt T. Frequency, type, and distribution of EST-SSRs from three genotypes of *Lolium perenne*, and their conservation across orthologous sequences of *Festuca arundinacea*, *Brachypodium distachyon*, and *Oryza sativa*. *BMC plant biology*. 2007; 7: 36. PMID: [17626623](#)
13. Yadav HK, Ranjan A, Asif MH, Mantri S, Sawant SV, Tuli R. EST-derived SSR markers in *Jatropha curcas* L.: development, characterization, polymorphism, and transferability across the species/genera. *Tree genetics & genomes*. 2011; 7: 207–219.
14. Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, et al. A worldwide bread wheat core collection arrayed in a 384-well plate. *Theoretical and Applied Genetics*. 2007; 114: 1265–1275. PMID: [17318494](#)
15. Tang J, Gao L, Cao Y, Jia J. Homologous analysis of SSR-ESTs and transferability of wheat SSR-EST markers across barley, rice and maize. *Euphytica*. 2006; 151: 87–93.
16. Li L, Wang J, Guo Y, Jiang F, Xu Y, Wang Y, et al. Development of SSR markers from ESTs of gramineous species and their chromosome location on wheat. *Progress in Natural Science*. 2008; 18: 1485–1490.
17. Del Sal G, Manfioletti G, Schneider C. The CTAB-DNA precipitation method: a common mini-scale preparation of template DNA from phagemids, phages or plasmids suitable for sequencing. *Biotechniques*. 1989; 7: 514–520. PMID: [2699240](#)
18. Mahmood A, Baenziger P, Budak H, Gill K, Dweikat I. The use of microsatellite markers for the detection of genetic similarity among winter bread wheat lines for chromosome 3A. *Theoretical and applied genetics*. 2004; 109: 1494–1503. PMID: [15290051](#)

19. Akpinar BA, Lucas SJ, Vrana J, Dolezel J, Budak H. Sequencing chromosome 5D of *Aegilops tauschii* and comparison with its allopolyploid descendant bread wheat (*Triticum aestivum*). *Plant Biotechnol J*. 2015; 13: 740–752. doi: [10.1111/pbi.12302](https://doi.org/10.1111/pbi.12302) PMID: [25516153](https://pubmed.ncbi.nlm.nih.gov/25516153/)
20. Sonah H, Deshmukh RK, Sharma A, Singh VP, Gupta DK, Gacche RN, et al. Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*. *PLoS One*. 2011; 6: e21298. doi: [10.1371/journal.pone.0021298](https://doi.org/10.1371/journal.pone.0021298) PMID: [21713003](https://pubmed.ncbi.nlm.nih.gov/21713003/)
21. Varshney RK, Graner A, Sorrells ME. Genic microsatellite markers in plants: features and applications. *TRENDS in Biotechnology*. 2005; 23: 48–55. PMID: [15629858](https://pubmed.ncbi.nlm.nih.gov/15629858/)
22. Zhao X, Tian Y, Yang R, Feng H, Ouyang Q, Tian Y, et al. Coevolution between simple sequence repeats (SSRs) and virus genome size. *BMC genomics*. 2012; 13: 435. doi: [10.1186/1471-2164-13-435](https://doi.org/10.1186/1471-2164-13-435) PMID: [22931422](https://pubmed.ncbi.nlm.nih.gov/22931422/)
23. Behura SK, Severson DW. Motif mismatches in microsatellites: insights from genome-wide investigation among 20 insect species. *DNA Research*. 2014: dsu036.
24. Biswas MK, Xu Q, Mayer C, Deng X. Genome Wide Characterization of Short Tandem Repeat Markers in Sweet Orange (*Citrus sinensis*). *PLoS one*. 2014; 9: e104182. doi: [10.1371/journal.pone.0104182](https://doi.org/10.1371/journal.pone.0104182) PMID: [25148383](https://pubmed.ncbi.nlm.nih.gov/25148383/)
25. Lawson MJ, Zhang L. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome biology*. 2006; 7: R14. PMID: [16507170](https://pubmed.ncbi.nlm.nih.gov/16507170/)
26. Zhang L, Yuan D, Yu S, Li Z, Cao Y, Miao Z, et al. Preference of simple sequence repeats in coding and non-coding regions of *Arabidopsis thaliana*. *Bioinformatics*. 2004; 20: 1081–1086. PMID: [14764542](https://pubmed.ncbi.nlm.nih.gov/14764542/)
27. Katti MV, Ranjekar PK, Gupta VS. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Molecular Biology and Evolution*. 2001; 18: 1161–1167. PMID: [11420357](https://pubmed.ncbi.nlm.nih.gov/11420357/)
28. Hong CP, Piao ZY, Kang TW, Batley J, Yang T, Hur Y, et al. Genomic distribution of simple sequence repeats in *Brassica rapa*. *Molecules and cells*. 2007; 23: 349. PMID: [17646709](https://pubmed.ncbi.nlm.nih.gov/17646709/)
29. Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, Maumus F, et al. Pan genome of the phytoplankton *Emiliania huxleyi* underpins its global distribution. *Nature*. 2013; 499: 209–213. doi: [10.1038/nature12221](https://doi.org/10.1038/nature12221)
30. Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, et al. *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature*. 2013; 496: 91–95. doi: [10.1038/nature12028](https://doi.org/10.1038/nature12028) PMID: [23535592](https://pubmed.ncbi.nlm.nih.gov/23535592/)
31. Ling H-Q, Zhao S, Liu D, Wang J, Sun H, Zhang C, et al. Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature*. 2013; 496: 87–90. doi: [10.1038/nature11997](https://doi.org/10.1038/nature11997) PMID: [23535596](https://pubmed.ncbi.nlm.nih.gov/23535596/)
32. Metzgar D, Bytof J, Wills C. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome research*. 2000; 10: 72–80. PMID: [10645952](https://pubmed.ncbi.nlm.nih.gov/10645952/)
33. Qu J, Liu J. A genome-wide analysis of simple sequence repeats in maize and the development of polymorphism markers from next-generation sequence data. *BMC research notes*. 2013; 6: 403. doi: [10.1186/1756-0500-6-403](https://doi.org/10.1186/1756-0500-6-403) PMID: [24099602](https://pubmed.ncbi.nlm.nih.gov/24099602/)
34. Cavagnaro PF, Senalik DA, Yang L, Simon PW, Harkins TT, Kodira CD, et al. Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). *BMC genomics*. 2010; 11: 569. doi: [10.1186/1471-2164-11-569](https://doi.org/10.1186/1471-2164-11-569) PMID: [20950470](https://pubmed.ncbi.nlm.nih.gov/20950470/)
35. Nazareno AG, dos Reis MS. The same but different: monomorphic microsatellite markers as a new tool for genetic analysis. *American journal of botany*. 2011; 98: e265–e267. doi: [10.3732/ajb.1100163](https://doi.org/10.3732/ajb.1100163) PMID: [21926307](https://pubmed.ncbi.nlm.nih.gov/21926307/)
36. Mian MAR, Saha MC, Hopkins AA, Wang Z-Y. Use of tall fescue EST-SSR markers in phylogenetic analysis of cool-season forage grasses. *Genome*. 2005; 48: 637–647. PMID: [16094432](https://pubmed.ncbi.nlm.nih.gov/16094432/)
37. Olango TM, Tesfaye B, Pagnotta MA, Pè ME, Catellani M. Development of SSR markers and genetic diversity analysis in enset (*Ensete ventricosum* (Welw.) Cheesman), an orphan food security crop from Southern Ethiopia. *BMC genetics*. 2015; 16: 98. doi: [10.1186/s12863-015-0250-8](https://doi.org/10.1186/s12863-015-0250-8) PMID: [26243662](https://pubmed.ncbi.nlm.nih.gov/26243662/)