

RESEARCH ARTICLE

Open Access

Genome-wide distribution comparative and composition analysis of the SSRs in Poaceae

Yi Wang^{1†}, Chao Yang^{1†}, Qiaojun Jin³, Dongjie Zhou², Shuangshuang Wang², Yuanjie Yu¹ and Long Yang^{1,2*}

Abstract

Background: The Poaceae family is of great importance to human beings since it comprises the cereal grasses which are the main sources for human food and animal feed. With the rapid growth of genomic data from Poaceae members, comparative genomics becomes a convenient method to study genetics of different species. The SSRs (Simple Sequence Repeats) are widely used markers in the studies of Poaceae for their high abundance and stability.

Results: In this study, using the genomic sequences of 9 Poaceae species, we detected 11,993,943 SSR loci and developed 6,799,910 SSR primer pairs. The results show that SSRs are distributed on all the genomic elements in grass. Hexamer is the most frequent motif and AT/TA is the most frequent motif in dimer. The abundance of the SSRs has a positive linear relationship with the recombination rate. SSR sequences in the coding regions involve a higher GC content in the Poaceae than that in the other species. SSRs of 70-80 bp in length showed the highest AT/GC base ratio among all of these loci. The result shows the highest polymorphism rate belongs to the SSRs ranged from 30 bp to 40 bp. Using all the SSR primers of Japonica, nineteen universal primers were selected and located on the genome of the grass family. The information of SSR loci, the SSR primers and the tools of mining and analyzing SSR are provided in the PSSRD (Poaceae SSR Database, <http://biodb.sdau.edu.cn/pssrd/>).

Conclusions: Our study and the PSSRD database provide a foundation for the comparative study in the Poaceae and it will accelerate the study on markers application, gene mapping and molecular breeding.

Background

The Poaceae (Grass family) is one of the largest families of flowering plants, comprised of approximately 600 genera and 10,000 species [1]. The Poaceae plants are the world's most important food source for human beings. SSRs (Simple Sequence Repeats) are short tandem repetitive (at least 2-7) sequences of a basic unit with less than seven base pairs [2]. SSRs are widely distributed in the genomes and are widely used in biological applications such as breeding, gene location, evolution, etc. [2-4].

SSR markers are co-dominant, abundant, multi-allelic, uniformly distributed, and can be detected by simple reproducible assays [5,6]. SSR markers are more

polymorphic than RFLPs (restriction fragment length polymorphism) and RAPDs (random amplified polymorphic DNAs) [3]. The high polymorphism, abundance, and convenience of SSRs make them a useful tool for genetic diversity and evolution studies [7-9].

SSR markers are ideal molecular tools for developing high density genetic maps [10]. In the past few years, an increasing number of genetic maps based on SSR markers have been developed, e.g., on rice [11], bread wheat [12] and triticale [13]. Currently, SSR markers are being used to integrate genetic maps and physical maps in plants and represent an efficient tool for breeders and geneticists to link phenotypic variations with genotypic variations [4]. Moreover, as tools for studying molecular evolution, SSR markers are used in investigating the origin, genetic diversity and dynamics of population evolution [8,14]. Furthermore, SSR markers are the commonly used markers for molecular marker assisted selection (MAS), which in turn directs molecular breeding [15,16]. The rapidly developing related studies

* Correspondence: lyang@sdau.edu.cn

[†]Equal contributors

¹Key Laboratory of Crop Biology of China, Shandong Agricultural University, Taian 271018, China

²Agricultural Big-Data Research Center, Shandong Agricultural University, Taian 271018, China

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

also need more information about the SSRs. In summary, it is clear that SSR markers have significant advantages in extensive applications in plants.

The development of SSR markers requires the sequencing of the target species. Initially, SSRs were developed from expressed sequence tags (ESTs) and bacterial artificial chromosome (BAC) sequences in most plants [2]. Currently, however, more and more markers are developed based on whole genome sequences. With the development of sequencing technology, a large number of EST sequences have been determined. Besides the *Oryza sativa L. ssp. Indica* with small genome (2002) [17], several plants belonging to the grass family with huge genomes also have been sequenced, such as the *moso bamboo* (2013) [18], *Triticum aestivum* (hexaploid bread wheat) [19] and its relatives of *Triticum urartu* (wheat A-genome, 2013) [20], *Aegilops tauschii* (wheat D-genome, 2013) [21]. With the newly sequenced genomes, studies on the identification and distribution of SSRs have been performed on numbers of species [22–26]. However, genome-wide comparison of the characteristics of SSRs among different grass species have not been reported, hampering the utilization of known sequences in marker development in related species and the research on the evolution of SSRs among species.

In this study, we present the genome-wide detection of SSR sequences from nine completely sequenced grass species (*Oryza sativa L. ssp. japonica*, *Oryza sativa L. ssp. Indica*, *Zea mays*, *Sorghum bicolor*, *Brachypodium distachyon*, *Foxtail millet*, *Moso bamboo*, *Triticum urartu* and *Aegilops tauschii*). Specifically, the characteristics of the SSRs are described based on the abundance, density, and base ratio of different motifs and the genomic elements (EXON, INTRON, UTR and whole GENOME). Finally, we develop a freely available Poaceae SSR database which is a comprehensive platform for genetic studies and MAS.

Results

Statistics of the SSR loci and primers

In total, 11,993,943 SSR loci with a minimum length of 12 bp were detected from the nine grass species, and were categorized into seven types according to the unit length: monomer, dimer, trimer, tetramer, pentamer, hexamer, and heptamer (Table 1). Among these, hexamers are the most abundant (58.8%) type. There are no significant differences in the SSR abundance composition within the nine species. The SSR density of the grass family varies greatly between species, with the maximum value in rice (1611 per Mb) and the minimum value in wheat-A (771 per Mb).

A total of 6,799,910 (57%) SSR primers were designed from the identified loci (Table 2). The rest loci are not suitable for primer design either because the flanking sequences are too short or because containing high GC content. The density of SSR primers ranged from 411 per Mb (Wheat-D) to 929 per Mb (Japonica) in different species. The SSR marker density in the centromere region (469 per Mb, maize) was lower than in the chromosome arms (516 per Mb).

Characteristic of SSR loci

To analyze the frequency of different SSR motifs, SSRs are standardized first. For example, SSRs with motifs of ATG, TGA, GAT, TAC, ACT and CAT are analyzed as ATG. Among all the ten species (nine Poaceae and Arabidopsis), the most frequent motifs are in high similarity (Table 3), and most of them are abundant with AT bases except the trimer which has a higher percentage of GC bases.

The motifs of dimer were classified into four groups - I: AT/TA, II: AC/CA/TG/GT, III: AG/GA/TC/CT, and IV: CG/GC. The percentage of group I in Poaceae species (39.8%) is much lower than in Arabidopsis (57.1%) (Figure 1). However, the percentages of the other three

Table 1 Occurrence of different SSR motifs in Arabidopsis and 9 Poacead species

Species Motif	Arabidopsis	Indica	Bamboo	Bread wheat	Maize	Millet	Japonica	Sorghum	Wheat D	Wheat A	Percentage
Monomer	13,678	18,689	33,840	7,987	30,918	11,288	16,456	14,292	49,289	44,524	1.94%
Dimer	9,389	37,542	112,470	9,430	65,250	12,420	37,447	38,202	128,784	154,577	4.89%
Trimer	17,509	83,231	130,017	36,997	186,749	44,075	82,572	80,378	263,191	272,427	9.66%
Tetramer	11,196	51,793	150,923	32,191	122,672	33,871	49,725	78,057	198,272	216,215	7.63%
Pentamer	3,825	16,953	45,839	7,970	38,654	9,245	17,120	16,652	43,472	45,621	1.98%
Hexamer	87,737	321,203	102,4250	182,645	138,5114	251,640	302,713	492,967	154,1607	1,698,582	58.82%
Heptamer	27,342	91,677	292,745	47,776	305,914	69,959	87,499	142,410	392,833	411,405	15.09%
All	170,676	621,088	1,790,084	324,996	2,135,271	432,498	593,532	862,958	2,617,448	2,843,351	100.00%
Size(MB)	117.01	490.02	1,996.8	262.57	2,007.04	387.96	368.27	714.44	3,256.32	3,686.4	
No. of SSRs/Mb	1,458.6	1,267.5	896.5	1,237.7	1,063.9	1,114.8	1,611.7	1,207.9	803.8	771.3	

Table 2 Statistics of the SSR primers in Poaceae

Species	Size (MB)	Primers	Primer/MB
Indica	513.82	295,100	574.3
Bamboo	2,086.05	1,238,576	593.7
Bread wheat	275.32	244,157	886.8
Maize	2,100.87	1,066,470	507.6
Millet	406.80	308,876	759.3
Japonica	386.16	358,640	928.7
Sorghum	749.09	390,055	520.7
Wheat A	3,800.82	1,514,778	398.5
Wheat D	3,361.20	1,383,258	411.5
Average	1,520.02	755,545.56	620.1

groups, which contain C/G bases, are much higher in Poaceae family, especially for the group IV. The percentage of group IV is 90 times higher in the Poaceae than in Arabidopsis (2.6% vs 0.03%) (Figure 1). Our results clearly showed that the C/G content in the dimer type SSRs is much higher in the Poaceae (31.4%) than in Arabidopsis (21.5%).

Except in the UTRs, no significant difference of SSR coverage rate was observed in the genomic elements including exons, introns and intergenic regions (Table 4). The UTRs have a higher coverage rate than other

elements (such as gene, EST, intron and CDS). For the 10 species examined, the coverage rate of the genome ranged from 1.0% (*Triticum urartu*) to 2.2% (*Japonica*). The SSRs appear on every element, and the distribution of SSRs across all the elements is relatively uniform.

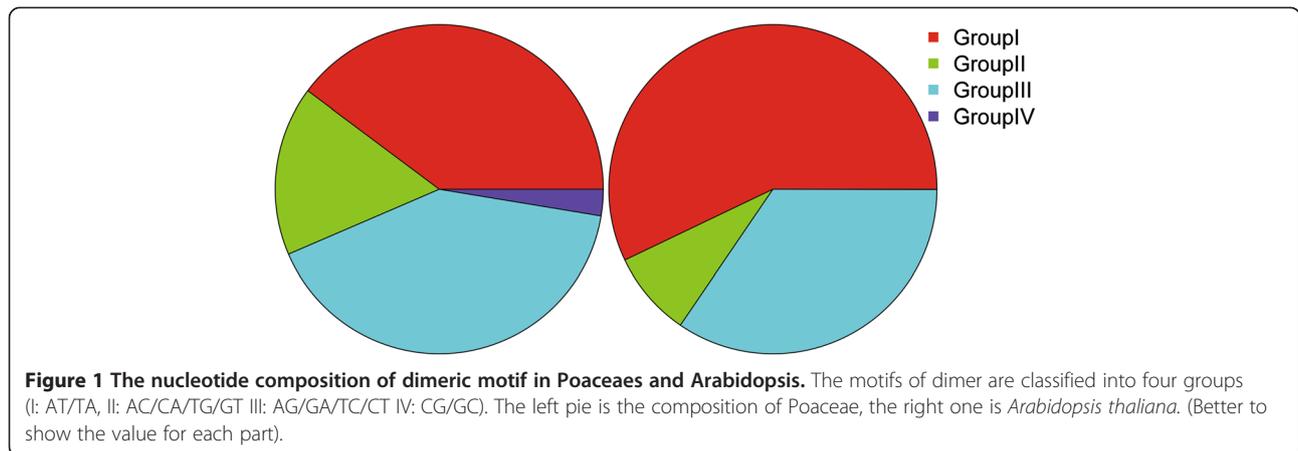
To elucidate the relationship between SSR abundance and the recombination rate, we conducted a correlation analysis between them of the maize chromosome 1 and 3. The results showed significant positive correlation (p -value <0.001) between SSR abundance and the recombination rate.

In this study, we defined the base ratio as the number of AT/the number of CG. The base ratios of CDS (coding sequences), introns, genes and the genome in Arabidopsis are higher than those in the Poaceae (Figure 2). For all the species examined, except Indica, the base ratio of CDS is the lowest, while that of the intron elements is the highest. Interestingly, the base ratios of all the elements are the same in Indica. In summary, the base ratio of the coding sequence is lower than that of the non-coding sequence.

We further divided the SSR sequences into 10 groups according to their length, and analyzed the relationship between SSR length and base ratio. For the SSRs shorter than 80 bp, the length of the SSR sequences is positively correlated with the base ratio, but for the SSRs longer

Table 3 The most frequent motif based on length

Motif length	Dimer	Trimer	Tetramer	Pentamer	Hexamer	Heptamer
Species						
<i>Oryza sativa L. ssp. indica</i>	motif ta ga ggc gag gatc tatt aaaag taaaa tctttt aaaata ttttatt aagaaaa	number 207,422 134,238 166,702 48,150 18,631 17,477 6,159 5,168 26,483 22,309 6,381 5,432				
<i>Oryza sativa L. ssp.japonica</i>	motif ta ga ggc gag gatc tatt aaaag taaaa tctttt aaaata ttttatt aagaaaa	number 216,603 132,553 169,394 49,836 17,713 16,079 6,275 5,189 24,246 21,107 6,381 4,891				
<i>Phyllostachys heterocycla</i>	motif ta ga ggc ctt tatt atac taaaa aaaag aaaata tctttt ttttatt gaggagg	number 521,497 469,501 118,594 108,529 70,239 49,365 29,112 13,984 72,939 52,814 12,708 8,818				
<i>Brachypodium distachyon</i>	motif ga ta ggc ctt tgca tatt taaaa aaaag tctttt aaaata aagaaaa ttttatt	number 38,348 21,589 52,062 26,219 13,804 10,600 2,838 2,686 12,970 7,543 2,621 1,640				
<i>Zea mays</i>	motif ta ga ctt ggc tatt tgca aaaag taaaa tctttt gagccc aggaagc ctgcttc	number 285,957 244,863 190,564 103,855 77,682 33,766 16,133 13,116 113,450 51,720 11,119 11,111				
<i>Foxtail millet</i>	motif ga ta ggc gag agaa gagg aaaag taaaa tctttt aaaata ttttatt aagaaaa	number 90,655 45,569 63,656 35,468 11,457 10,635 3,308 1,537 15,511 10,527 3,349 3,064				
<i>Sorghum bicolor</i>	motif ta ga ggc att tatt tgca taaaa aaaag aaaata tctttt aggggcc ggaccgg	number 275,228 102,404 66,278 65,832 60,105 18,540 4,857 4,831 54,185 35,085 12,765 12,699				
<i>Aegilops tauschii</i>	motif ga gt ctt gag tatt tgca taaaa aaaag tctttt aaaata ttttatt aagaaaa	number 556,203 301,146 297,281 208,816 80,376 61,754 30,819 19,290 109,440 83,782 21,296 20,412				
<i>Triticum urartu</i>	motif ga ta ctt gag tatt tgca taaaa aaaag tctttt aaaata ggtggtg ttttatt	number 629,789 308,771 285,749 191,802 109,305 54,928 26,034 20,862 109,022 107,694 32,306 22,936				
<i>Arabidopsis thaliana</i>	motif ta ga ctt aac tatt agaa taaaa gttttt aaaata tctttt taaacc ttttatt	number 49,195 29,738 35,167 9,971 10,479 6,064 3,514 1,623 11,160 9,407 2,845 2,498				



than 80 bp, the length is negatively correlated with the base ratio (Figure 3).

Using Soybean, Watermelon, Sweet Orange, Apple and Arabidopsis as an out-group, the SSR loci types were categorized into two types: the monomers and the non-monomers. The base ratio has a very wide range for the monomers (Figure 4I). In the Poaceae species, it ranges from 0.11 (Wheat A) to 2.19 (*Indica*), and the base ratio of the others are similar among the nine Poaceae species and the control group (Figure 4II). For both the monomers and the others, the base ratio is much higher in the control group than in the Poaceae species.

Polymorphism of SSR primer

The polymorphism rate of primers varies with the length of the SSR; when the length of the SSRs is longer than 20 bp, they tend to have a higher polymorphism rate. The rate peaks at the length range from 30 to 40 bp and then oscillates but remains at a relatively stable level.

Universal primers

Seventy-three universal primers were developed, and 19,607 e-PCR products were found on the nine genomes. Nineteen universal primers belonging to Japonica were

selected randomly and their productions were anchored to the genomes of six species by MapDraw. The positions of the productions indicated that the markers have several duplications in the genomes of the grass. Additionally, some of the products were located in the mitochondrion and chloroplast. NCBI Blast also found these segments in the organelles and some unknown mRNAs, which further confirmed our findings. The sequences of some of the *in silico* amplification products are highly conserved; there are only a few SNPs (single nucleotide polymorphisms). Thus, the universal primers can be used in all the examined Poaceae species and can be used to localize genes or motifs that are linked with these sequences.

Table 4 Cover rate of SSRs as per genomic element

Species	Genome	Gene	CDS	Intron	UTR
<i>Arabidopsis thaliana</i>	1.90%	1.63%	1.44%	1.68%	2.24%
<i>Oryza sativa L. ssp. indica</i>	1.67%	1.87%	1.83%	1.88%	
<i>Phyllostachys heterocycla</i>	1.26%	1.68%	1.51%	1.60%	
<i>Brachypodium distachyon</i>	1.55%	1.50%	1.49%	1.04%	1.62%
<i>Zea mays</i>	1.33%	1.61%	1.38%	0.97%	
<i>Setaria italica</i>	1.47%	1.55%	1.68%	1.44%	
<i>Oryza sativa L. ssp. japonica</i>	2.16%	1.96%	1.56%	1.37%	2.48%
<i>Sorghum bicolor</i>	1.59%	2.10%	2.02%	1.69%	3.24%
<i>Aegilops_tauschii</i>	1.06%				
<i>Triticum_urartu</i>	1.02%				

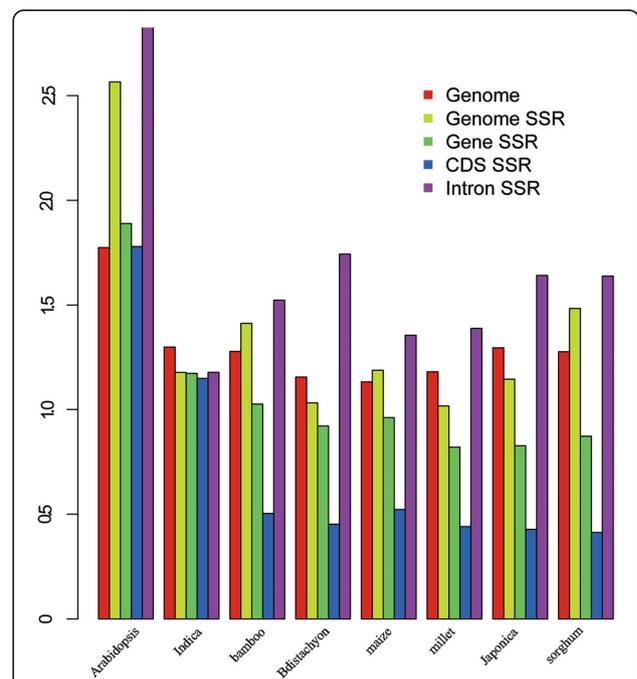
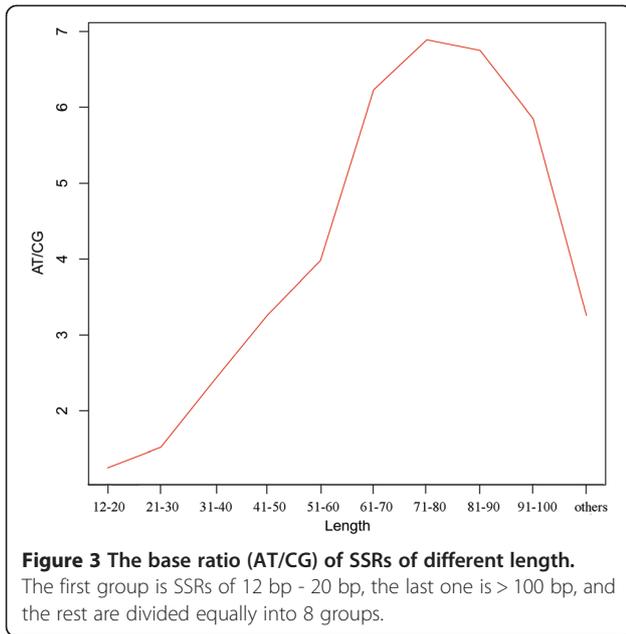


Figure 2 The base ratio (AT/CG) of the genomic elements. The genome elements include CDS SSRs, intron SSRs, gene SSRs, genome SSRs and genome.



Database content and web interface

All the information on SSR loci and SSR markers can be found in the PSSRD (Poaceae SSR Database, <http://biodb.sdau.edu.cn/pssrd/index.html>). PSSRD has several main pages with different functions. In the search page, SSRs can be searched by species, chromosome, motif, units, and even the start and end loci. The primers and universal primers can also be searched. An online tool for developing SSR primers is available in the tool page. The SSR loci and primers information can be downloaded in the download page.

Discussion

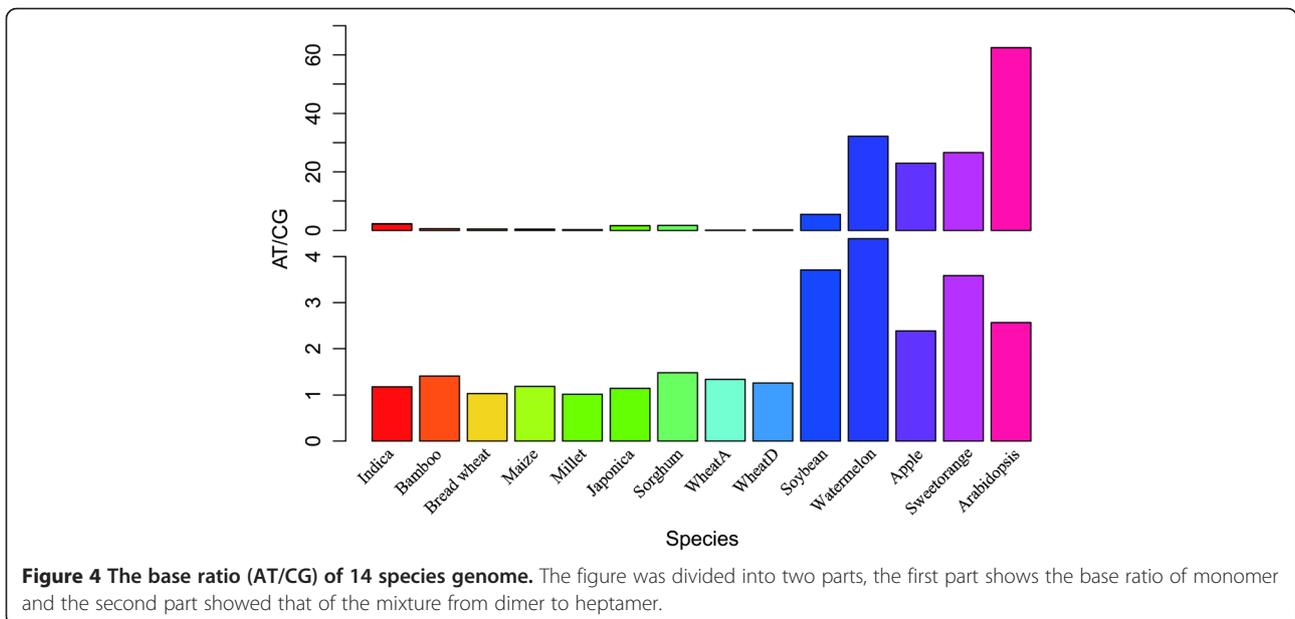
Distribution of SSRs

In the grass family, SSRs appear in all elements of the genome and are distributed throughout the chromosome. The SSRs appeared in each of the elements of the genome, indicating a wide usage of the genome. Our results are similar to those of Lawson MJ [27]. Combined with other markers, SSRs are ideal tools for studying the genome.

In our study, the abundance of SSRs is highly correlated with the recombination rate and is consistent with the studies on human SSRs [28]. There might be a direct link if recombination is mutagenic to microsatellite sequences or if the simple sequence repeats participate in the recombination to some extent; for example, if SSRs act as recombination signals or if their special composition accelerates the recombination in some way. Alternatively, recombination could exert an indirect effect by the uncoupling of natural selection at linked loci, promoting polymorphism [28]. All these data are limited by the density of the molecular linkage maps. The availability of more high density molecular linkage maps and more suitable standards for the SSRs will clarify the relationship between SSR abundance and the recombination rate.

Implications of the base ratio

Compared with other families, the SSRs of the Poaceae species have a higher percentage of C/G bases, which means that SSRs in the grass family are more stable [29]. Stable genetic materials lead to a lower mutation rate and a longer period of evolution. It might also be the reason that the Poaceae plants contain longer



chromosomes than other species. All these results are consistent with the previous studies in the monocots [29].

The higher percentage of C/G in the coding regions means a lower probability of mutation for the SSRs in the coding sequence. On the other hand, the non-coding sequences show a higher base ratio (AT/GC) and therefore have a higher probability of being mutated.

It is possible that there are excessive amounts of redundant sequences in the grass family, causing the base ratio to more closely reflect a random composition. The base ratio of the monomer varies across species. It has a very wide range, which may be due to the species' own characteristics. And the result shows that the composition of the monomers is not random, perhaps because the number of the monomers for the coding sequences of the grass family is too small. The total number of monomers in the grass family is not less than that in the control group, but the number of monomers in the coding sequences is much less than in the other species.

The higher GC-content of the grass SSR, especially of the coding sequences, may lead to a lower expression of the grass genes compared with the controls. GC content has been proven to be correlated with recombination during meiosis [30,31]. The significant positive linear relationship of SSR abundance to recombination rate further proves our point. The high GC base ratio indicates a special evolutionary status of the Poaceae. The result also points to a new way of studying evolution.

Poaceae primers

The huge number of primers designed in this database will be useful for developing genetic maps, assembling

genomes, locating the genes, and breeding and will promote the study of the Poaceae.

High conservation implies a very important function and lesser evolutionary pressure to change [32]. The appearance of universal primer products in the genomes of all the examined species implies that they play essential roles in those species. Their appearance in the organelles shows the relation between the nuclear genome and the organelle genome. It also provides new evidence for the theory that the nuclear genome guides the synthesis of the organelles. The universal primer can also be used to detect the positions of some related genes, some of which might be housekeeping ones.

The polymorphism rate shows that longer SSRs may cause more polymorphisms because of the instability of long SSRs and there being many productions per primer pair in this range, which provides more opportunities for the SSRs to change. The SSRs longer than 40 bp have lower polymorphism rates, perhaps because they are too long to have enough loci per primer, and thus the chances of changes are low. This result suggests that a primer that contains a SSRs range from 30 to 40 bp is more efficacious.

Poaceae SSR database

To our knowledge, PSSRD is the first comprehensive database for SSR loci and primers in Poaceae. Data in this database is free for all academic and non-commercial users. This database will provide a huge number of markers and related information for the researchers involved in the study of Poaceae. Researchers are expected to give feedback after using the data in this

Table 5 Sources of genome sequences used in this study

Species	Version	Data source
<i>Arabidopsis thaliana</i>	TAIR10	http://www.arabidopsis.org/
<i>Oryza sativa</i> L. <i>ssp. japonica</i>	Release 7.0	http://rice.plantbiology.msu.edu/
<i>Zea mays</i>	Release 5a	http://www.maizegdb.org/
<i>Sorghum bicolor</i>	Sbi1_4	http://www.phytozome.net/sorghum.php
<i>Oryza sativa</i> L. <i>ssp. indica</i>	full Release	http://rice.genomics.org.cn/rice/link/download.jsp
<i>Brachypodium distachyon</i>	192	ftp://ftp.jgi-psf.org/pub/comp/gen/phytozome/v9.0/Bdistachyon/
Foxtail millet	v2.0	http://foxtailmillet.genomics.org.cn/page/species/download.jsp
<i>Phyllostachys heterocycla</i>	v1.0	http://202.127.18.221/bamboo/down.php
<i>Triticum urartu</i>	v1	http://www.ncbi.nlm.nih.gov/sra/SRX047327
<i>Aegilops tauschii</i>	v1	http://www.ncbi.nlm.nih.gov/sra/SRX047430
<i>Oryza sativa</i> (<i>indica</i> cultivar-group) PA64	2013.8.13	ftp://public.genomics.org.cn/BGI/rice_seq/PA64s/
<i>Cucumis sativus</i>	v2.3	http://cucumber.genomics.org.cn/page/cucumber/index.jsp
<i>Glycine max</i>	Phytozome v7.0	http://www.soybase.org/dlpages/
Sweetorange <i>Citrus sinensis</i>	2012.11.29	http://citrus.hzau.edu.cn/orange/download/data.php
<i>Malus apple</i>	v1	http://www.rosaceae.org/data/download

database to facilitate updates and further provide more useful information to the users.

Conclusions

Our studies of the SSRs from nine Poaceae plants show that the GC base ratio in the grass family is higher than other families. The SSRs have a high positive correlation with the recombination rate in maize. PSSRD database provides an useful resource to the comparative SSRs study in the Poaceae and will accelerate the study on markers application, the gene mapping and the molecular breeding.

Methods

Genome sequences source

The whole genome sequences of nine grass species were downloaded from public databases (Table 5). The genomes of *Arabidopsis thaliana*, soybean, watermelon, sweet orange, apple and PA64 (a variety of *Oryza sativa* L. ssp. *Indica*) were used as control groups.

Genome-wide SSRs analysis

Using a Perl script, SSR loci were detected and classified into seven types according to their copy number, and all of them tandemly arranged: monomer (one copies, at least twelve repeats), dimer (two copies, at least seven repeats), trimer (three copies, at least six repeats), tetramer (four copies, at least five repeats), pentamer (five copies, at least four repeats), hexamer (six copies, at least three repeats), and heptamer (seven copies, at least two repeats).

The distribution, abundance, and base ratio (AT/CG) of the SSRs were analyzed by several Perl scripts. Based on the genetic map and the physical map [33], chromosomes 1 and 3 of maize were chosen to test the correlation between the recombination rate and the SSR abundance.

Primers development

The SSR primers flanking 60 bp of the SSR loci were designed using eprimer3 [34]. Using the genome sequences of nine grass species, all the primers were tested and filtered by e-PCR [35]. The universal primers were selected using all the SSR primers amplified in the nine genomes, and then a polymorphism analysis was performed in silico on the genomes of Japonica, Indica and PA64. Some of the universal primers were also located on the genome using Mapdraw [36].

Database architecture

Based on all these data, a database named PSSRD has been established. The PSSRD database (Poaceae SSR Database, <http://biodb.sdau.edu.cn/pssrd/index.html>) consists of some interrelated relational databases implemented in

MySQL. The data handling and analysis parts of the database used pipelines written in Perl Scripts. The web interface, running on an Apache web server, was implemented in HTML. The PSSRD database was set up on a World Wide Web server allowing internet access with a web client.

Abbreviations

SSR: Simple sequence repeats; EST: Expressed sequence tags; UTR: Untranslated region; CDS: Coding sequence.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LY and YC conceived and designed the experiments. YW and YY performed the experiments. YW, LY, QJ, DZ and SW wrote and modified the paper. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by National Natural Science Foundation of China (31201231), National Basic Research Program of China (2013CB127702), and Foundation of Excellent Young Scholar of Shandong Province of China (BS2011SW027).

Author details

¹Key Laboratory of Crop Biology of China, Shandong Agricultural University, Taian 271018, China. ²Agricultural Big-Data Research Center, Shandong Agricultural University, Taian 271018, China. ³College of Plant Protection, Northwest Agriculture and Forestry University, Yangling 712100, China.

Received: 8 September 2014 Accepted: 5 February 2015

Published online: 15 February 2015

References

- Clayton WD, Renvoize SA. Genera graminum. Grasses of the World. Kew Bulletin Additional Series. 1986;13:389.
- 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(6):e21298.
- Liu ZJ, Cordes JF. DNA marker technologies and their applications in aquaculture genetics. Aquaculture. 2004;238:1–37.
- Gupta PK, Varshney RK. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica. 2000;113:163–85.
- Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet. 2002;30(2):194–200.
- Wayne Powell GCM, Jim P. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1996;1:215–22.
- Zhang Y, Zalapa JE, Jakubowski AR, Price DL, Acharya A, Wei Y, et al. Post-glacial evolution of *Panicum virgatum*: centers of diversity and gene pools revealed by SSR markers and cpDNA sequences. Genetica. 2011;139(7):933–48.
- Zheng H, Wang H, Yang H, Wu J, Shi B, Cai R, et al. Genetic diversity and molecular evolution of chinese waxy maize germplasm. PLoS One. 2013;8(6):e66606.
- Lin HS, Chiang CY, Chang SB, Kuoh CS. Development of Simple Sequence Repeats (SSR) Markers in *Setaria italica* (Poaceae) and Cross-Amplification in Related Species. Int J Mol Sci. 2011;12(11):7835–45.
- Zhang Q, Ma B, Li H, Chang Y, Han Y, Li J, et al. Identification, characterization, and utilization of genome-wide simple sequence repeats to identify a QTL for acidity in apple. BMC Genomics. 2012;13:537.
- Orjuela J, Garavito A, Bouniol M, Arbelaez JD, Moreno L, Kimball J, et al. A universal core genetic map for rice. TAG. 2010;120(3):563–72.
- Garvin DF, McKenzie N, Vogel JP, Mockler TC, Blankenheim ZJ, Wright J, et al. An SSR-based genetic linkage map of the model grass *Brachypodium distachyon*. Genome. 2009;53(1):1–13.

13. Tyrka M, Bednarek PT, Kilian A, Wedzony M, Hura T, Bauer E. Genetic map of triticale compiling DArT, SSR, and AFLP markers. *Genome*. 2011;54(5):391–401.
14. Teixeira H, Rodriguez-Echeverria S, Nabais C. Genetic diversity and differentiation of *Juniperus thurifera* in Spain and Morocco as determined by SSR. *PLoS One*. 2014;9(2):e88996.
15. Cao H, Wang Y, Xie Z, Huang L, Xu H, Zhang L, et al. TGB: the tobacco genetics and breeding database. *Mol Breeding*. 2013;31(3):655–63.
16. Qiu LJ, Xing LL, Guo Y, Wang J, Jackson SA, Chang RZ. A platform for soybean molecular breeding: the utilization of core collections for food security. *Plant Mol Biol*. 2013;83(1–2):41–50.
17. Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*. 2002;296(5565):79–92.
18. Peng Z, Lu Y, Li L, Zhao Q, Feng Q, Gao Z, et al. The draft genome of the fast-growing non-timber forest species moso bamboo (*Phyllostachys heterocycla*). *Nat Genet*. 2013;45(4):456–61. 461e451–452.
19. Consortium TIWGS. A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science*. 2014;345:6194.
20. Ling HQ, 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(7443):87–90.
21. 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(7443):91–5.
22. Singh R, Narzary D, Bhardwaj J, Singh AK, Kumar S, Kumar A. Molecular diversity and SSR transferability studies in Vetiver grass (*Vetiveria zizanioides* L. Nash). *Ind Crop Prod*. 2014;53:187–98.
23. Ho CW, Wu TH, Hsu TW, Huang JC, Huang CC, Chiang TY. Development of 12 genic microsatellite loci for a biofuel grass, *Miscanthus sinensis* (Poaceae). *Am J Bot*. 2011;98(8):e201–3.
24. Pandey G, Misra G, Kumari K, Gupta S, Parida SK, Chattopadhyay D, et al. Genome-wide development and use of microsatellite markers for large-scale genotyping applications in foxtail millet [*Setaria italica* (L.)]. *DNA Res*. 2013;20(2):197–207.
25. Novy A, Flory SL, Honig JA, Bonos S, Hartman JM. Characterization of polymorphic microsatellites for the invasive grass *Microstegium vimineum* (Poaceae). *Am J Bot*. 2012;99(2):e56–8.
26. Kantety RV, La Rota M, Matthews DE, Sorrells ME. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol*. 2002;48(5–6):501–10.
27. Lawson MJ, Zhang L. Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biol*. 2006;7(2):R14.
28. Brandstrom M, Bagshaw AT, Gemmill NJ, Ellegren H. The relationship between microsatellite polymorphism and recombination hot spots in the human genome. *Mol Biol Evol*. 2008;25(12):2579–87.
29. Šmarda P, Bureš P, Horová L, Leitch IJ, Mucina L, Pacini E, et al. Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc Natl Acad Sci*. 2014;111(39):E4096–102.
30. Tortereau F, Servin B, Frantz L, Megens HJ, Milan D, Rohrer G, et al. A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *BMC Genomics*. 2012;13:586.
31. Clément Y, Arndt PF. Meiotic recombination strongly influences GC-content evolution in short regions in the mouse genome. *Mol Biol Evol*. 2013;30(12):2612–8.
32. Assis R, Kondrashov AS. Conserved proteins are fragile. *Mol Biol Evol*. 2014;31(2):419–24.
33. Schaeffer ML, Harper LC, Gardiner JM, Andorf CM, Campbell DA, Cannon EK, et al. MaizeGDB: curation and outreach go hand-in-hand. *Database*. 2011;2011:bar022.
34. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res*. 2012;40(15):e115.
35. Lanar DE, Kain KC. Expression-PCR (E-PCR): overview and applications. *PCR Methods Appl*. 1994;4(2):S92–6.
36. Liu RH, Meng JL. MapDraw: a microsoft excel macro for drawing genetic linkage maps based on given genetic linkage data. *Yi Chuan*. 2003;25(3):317–21.

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

