Genetics and Molecular Research

Identification of novel and useful EST-SSR markers from *de novo* transcriptome sequence of wheat (*Triticum aestivum* L.)

Z.J. Yang, Z.S. Peng and H. Yang

Key Laboratory of Southwest China Wildlife Resources Conservation,
Ministry of Education, China West Normal University, Nanchong, Sichuan, China

Corresponding author: Z.J. Yang
E-mail: yangzaijun1@126.com

Genet. Mol. Res. 15 (1): gmr.15017509
Received August 24, 2015
Accepted October 30, 2015
Published February 19, 2016
DOI http://dx.doi.org/10.4238/gmr.15017509

ABSTRACT. Simple sequence repeats (SSRs) are highly informative, polymorphic, and co-dominant Mendelian markers that provide an important genomic resource for genetic research. Recently, the use of large-scale transcriptome sequence has become a reliable and efficient approach for the identification and development of new EST-SSR markers. In this study, 8389 potential SSRs with a minimum of five repetitions for all motifs were identified from 121,210 unigenes. Gene ontology analysis indicated that the unigenes containing SSR loci participate in various biological processes of regulation, growth, development, metabolism, and apoptosis in wheat. As in many other plants, trinucleotide repeats were found to be the most abundant repeat units with a frequency of 62.33%. A subset of 300 EST-SSRs was randomly selected for the applicability of EST-SSRs to be evaluated. Of the 300 primer pairs tested, 177 (59%) yielded unambiguous PCR products among five wheat cultivars. Using the Chinese Spring nulli-tetrasomic line, 131 of the 177 EST-SSR primer pairs yielded products and 178 loci were found to be located on all the 21 wheat chromosomes. These findings suggest that the novel EST-SSR markers, as a basis for future

genetic linkage and gene tagging analysis, are a valuable tool for genetic mapping, marker assisted selection, and comparative genome analysis.

Key words: *Triticum aestivum* L.; Transcriptome sequence; Simple sequence repeats; Chromosome localization

INTRODUCTION

Microsatellites or simple sequence repeats (SSRs) are tandem repeat sequences of short units of 2-6 nucleotides that occur frequently in all prokaryotic and eukaryotic genomes studied to date (Koelling et al., 2012). Distributed randomly across the genomes of plants and animals, SSRs are highly informative, polymorphic, and co-dominant Mendelian markers (Powell et al., 1996). These characteristics make SSR markers useful in linkage mapping, genetic diversity analysis, parental identification, DNA fingerprinting, and functional gene tagging. These SSR markers can be isolated from both conserved coding regions and non-coding nucleotide sequences of all higher organisms (Sraphet et al., 2011). The development of genomic SSRs, however, is relatively time-consuming and expensive; but SSRs can alternatively be found in public sequence databases of expressed sequence tags (ESTs) or cDNA (Zeng et al., 2010; Huang et al., 2011; Sraphet et al., 2011; Koelling et al., 2012). Such SSRs are referred to as EST-SSRs. In recent years, EST-SSR mining from coding sequences determined by RNA-seq technology has become increasingly popular in plant research (Kaur et al., 2012; Li et al., 2012).

Common wheat, *Triticum aestivum* L. (2n = 6x = 42), is a major staple food crop in several parts of the world in terms of its cultivation area and prevalent use as a food source. Although a number of EST-SSRs have been developed from the published EST database (Yu et al., 2004; Chen et al., 2005; Li et al., 2008), the development and application of EST-SSR markers from transcriptome sequence in wheat are still largely limited compared, with those of EST-SSR markers from other crop species.

In previous studies, we generated 40.88 Gb clean sequence data using Illumina sequencing from wheat pistillody stamen (PS), pistil (P), and stamen (S), which corresponded to 121,210 putative unigenes (Yang et al., 2015).

We previously reported on the development of a comprehensive set of EST-SSRs based on *de novo* transcriptome sequence from the wheat PS, P, and S. A total of 8389 EST-SSRs from 121,210 unigenes were generated for the *de novo* transcriptome. The gene ontology (GO) classification and characteristics of those EST-SSRs were assessed in this study. To evaluate the applicability of the EST-SSRs developed, 300 EST-SSR primer pairs were randomly selected for amplification using five wheat cultivars and chromosome localization using a nulli-tetrasomic line. These newly developed EST-SSR markers will rapidly enrich the number of functional molecular markers directly related to expressed regions of the genes in wheat and will therefore be valuable in facilitating genetic mapping and comparative genome analysis in wheat.

MATERIAL AND METHODS

Plant materials

The pistillody wheat mutant HTS-1 and its sib-line CSTP were used for cDNA library preparation and Illumina sequencing. The PS and P of the HTS-1 mutant line as well as in the S of

the CSTP sib-line were selected at the heading stage for RNA extraction. Five wheat cultivars were used to test the validity of the new developed EST-SSRs: *Triticum aestivum* L. var. Chinese spring (CS), *Triticum aestivum* L. var. Three pistil (TP), *Triticum aestivum* L. var. Mianmai 45 (MM45), *Triticum aestivum* L. var. Neimai 9 (NM9), and *Triticum aestivum* L. var. Chuanmai 28 (CM28). The assignment of EST-SSR markers to the specific wheat chromosomes was carried out using a set of CS nulli-tetrasomic lines.

RNA isolation and cDNA library construction

For Illumina sequencing, total RNA was extracted from PS, P, and S using a modified cetyltrimethylammonium bromide-based method (Chang et al., 1993) with a high salt concentration. The RNA was further purified with the RNeasy Plant Mini Kit (Qiagen, Shanghai, China). The cDNA library construction and sequencing with an Illumina HiSeq 2000 were performed at the Novogene Bioinformatics Institute (Beijing, China). The assembled unigenes were annotated using BLASTx against the nonredundant protein database [National Center for Biotechnology Information (NCBI); http://www.ncbi.nlm.nih.gov/blast/Blast.cgi] and the unigenes were allocated to the corresponding functional categories based on GO terms by Blast2GO (Conesa et al., 2005) with GO weight 2 (Huang et al., 2011).

Novel EST-SSR identification and primer design

All unigenes generated by deep transcriptome sequencing in wheat were screened for SSRs using the SSRIT (Simple Sequence Repeat Identification Tool; www.gramene.org/db/markers/ssrtool) software. In this study, SSRs containing a minimum five repetitions for all motifs were included in the study. Primer pairs flanking the SSRs were designed using Primer 3 (Rozen and Skaletsky, 2000) in accordance with core criteria for primers: predicted product size ranging from 100 to 500 bp, GC content of between 40 and 60%, optimum primer length of 22 bp, and melting temperature of between 50° and 60°C.

DNA extraction, PCR amplification and detection

Genomic DNA from five wheat cultivars and from Chinese Spring nulli-tetrasomic lines were extracted from young leaves using a modified CTAB protocol (Liu et al., 2003). Amplification by PCR was performed using a T-100 ThermalCycler (BIO-RAD) and reactions consisting of 2.0 mM $MgCl_2$, 0.2 mM dNTPs, 0.75 U Taq DNA polymerase (TaKaRa Bio Inc., Dalian, China), 0.1 µM of each primer, and 10 ng template DNA in a final reaction volume of 15 µL. The PCR conditions were as follows: initial denaturation at 94°C for 2 min; 35 cycles of 94°C for 40 s, 55°-60°C (depending on the T_m of the primer set used) for 45 s, and 72°C for 1 min; and final extension at 72°C for 10 min. The resulting PCR products were separated on 8% non-denaturing polyacrylamide gels at 200 V for 2-2.5 h and were then visualized using a rapid silver staining method (Liu et al., 2008).

RESULTS

Unigene sequences and GO analysis

The wheat cDNA library was sequenced using the Illumina Hiseq 2000 platform, yielding

a total of 43.46 Gb raw reads. After removing short and low-quality reads and trimming off adapter sequences, approximately 40.88 Gb high-confidence reads remained and were assembled into a total of 121,210 unigenes with an average length of 695 bp. These unigenes were subjected to further analyses.

GO was employed to identify the functional categories of the annotated unigenes and to classify and annotate the transcripts with known proteins. Of the 8389 unigenes containing SSR loci, 4818 unigenes were successfully associated with GO terms. The GO-annotated unigenes were found to belong to the biological process, cellular component, and molecular function groups and were classified into 47 categories at process level 2 (Figure 1). Among the biological process group, cellular process (60.4%), metabolic process (57.9%), and single-organism process (46.1%) were the most strongly represented categories. The cell (40.9%), cell part (40.9%), organelle (29.7%), and macromolecular complex (23.3%) categories within the cellular function group contributed the largest proportion of all annotations. In the molecular function group, binding (63.2%) and catalytic activity (46.1%) constituted the two major categories, followed by transporter activity (6.67%), nucleic acid binding transcription factor activity (5.21%), and structural molecule activity (3.03%). Other components were represented at less than 3%. These results suggest that the analyzed unigenes take part in various biological processes of regulation, growth, development, metabolism, and apoptosis in wheat.

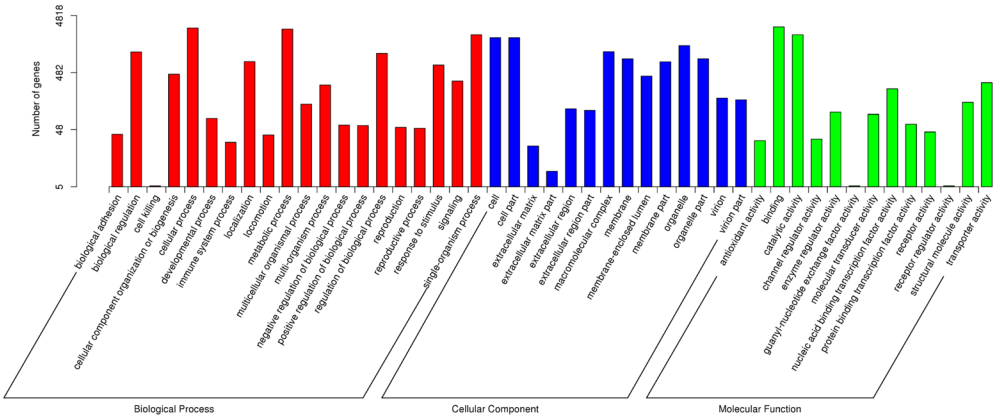

Figure 1. Gene ontology classification of unigene sequences containing SSR loci from wheat. The results are summarized in three main categories: biological process, cellular component, and molecular function.

Characteristics of EST-SSRs in the wheat transcriptome

SSRs were found to be highly abundant in the unigene dataset assembled in this study. In total, 8389 potential SSRs with a minimum of five repetitions for all motifs were identified from 121,210 unigenes. The frequency of occurrence of SSR loci was one in every 10.04 kb of unigene sequence. Among all repeat types, SSR length was distributed from 10 to 213 bp (average 16.76 bp). Incidences of different repeat types were assessed and 95.24% of the SSRs were found to exist as either dinucleotide repeats (DNRs) or trinucleotide repeats (TNRs). The most abundant repeat type among the SSRs was the TNR, comprising 62.3% of the total SSRs, followed by DNR

(32.9%), tetranucleotide repeat (4.4%), pentanucleotide repeat (0.3%), and hexanucleotide repeat (0.1%). Overall, the repeat unit number in the SSR loci was found to range from 5 to 20 (Figure 2A and B). Most (98.96%) of the DNRs and TNRs were found to have 4-10 repeat units, while motifs with more than 10 reiterations were rare (1.04%).

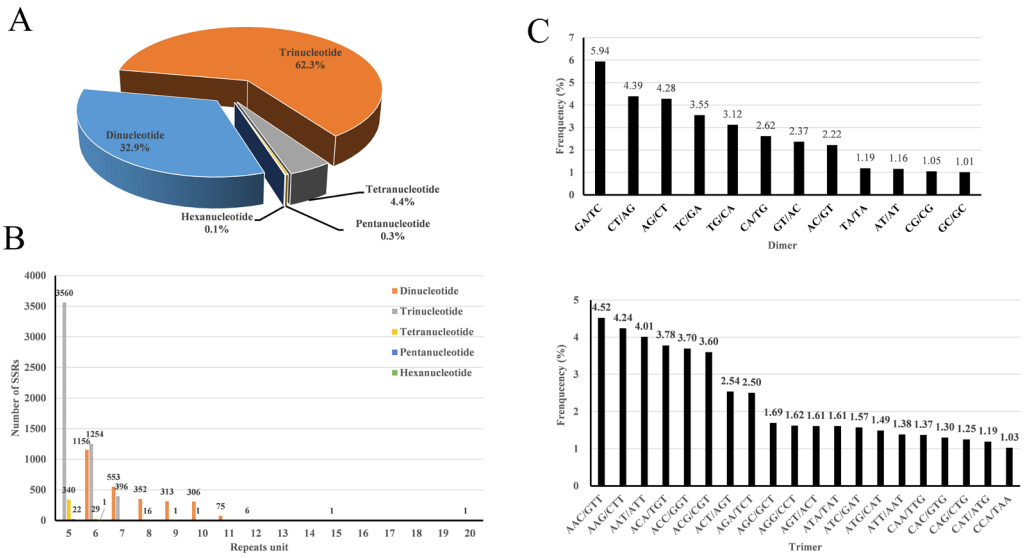

Figure 2. Characterization of SSRs in the wheat transcriptome. **A.** Distribution of different SSR repeat motif types; **B.** Number of different repeat motifs; **C.** Frequency distribution of major SSRs based on main motif sequence type.

The 20 types of major motifs and the frequencies of individual SSR units are shown in Figure 2C. In this study, the most common type of all the motifs detected in wheat unigenes was GA/TC (9.50%), followed by AG/CT (8.6%), TG/AC (5.75%), GT/AG (4.59%), AAC/GTT (4.52%), AAG/CTT (4.24%), and AAT/ATT (4.01%). Other motifs (Figure 2C) were represented by less than 4.0% of the total SSRs.

SSR marker development and genetic diversity analysis

A total of 300 EST-SSR primer pairs located on 300 unigenes were randomly selected and amplified using DNA templates extracted from five wheat cultivars (CS, TP, MM45, NM9, and CM28). Of these EST-SSR primer pairs, 177 (59%) exhibited stable and repeatable amplification (Figure 3). Despite multiple attempts at optimization of PCR conditions, 123 primer pairs did not yield any product. This observation was attributed to sequence assembly errors and to primers extending across splice sites with large introns (Dutta et al., 2011). This result highlights the complexity of the common hexaploid wheat genome.

Of the 177 primer pairs, 60 EST-SSR markers generated a unified and poor polymorphic band. A total of 117 of the primer pairs analyzed showed allelic polymorphisms and 401 alleles were detected in total. The number of alleles per locus ranged from 1 to 6 (average 3.4; Figure 3).

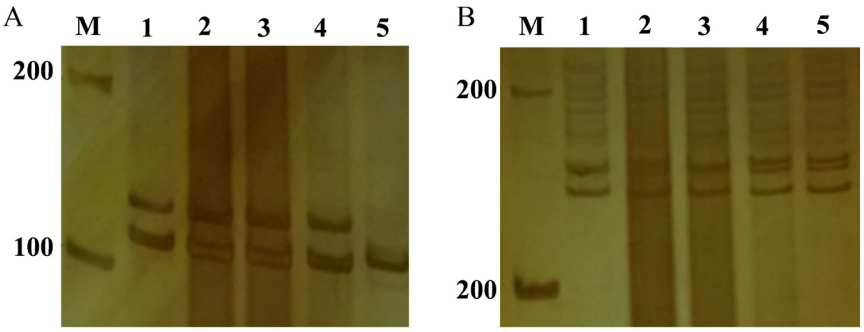

Figure 3. PCR products from wheat obtained using EST-SSR primer pairs detected by 8% non-denaturing polyacrylamide gels. **A.** Primer pair comp94; **B.** primer pair comp167. *Lane 1* = CS; *lane 2* = TP; *lane 3* = MM45; *lane 4* = NM9; *lane 5* = CM28.

Chromosome localization of EST-SSR markers using a nulli-tetrasomic line

Using the Chinese Spring nulli-tetrasomic wheat line to further assess the abovementioned 177 primers, 131 EST-SSR primer pairs and 178 loci were located on all the 21 wheat chromosomes (Tables 1 and 2; Figure 4), where 100 primer pairs yielded products with one chromosome, 18 with two chromosomes, 11 with three chromosomes, and two with five chromosomes (Table 1). Of all the 178 loci, only one locus was located on chromosome 4B, only two on chromosome 6D, while 13 loci were located on chromosome 1D. The chromosomes 2B, 2D, 3B, 4D, 5A, 5B, 5D, 6A, 7A, and 7D also had more loci than others. Sixty loci were located on the A genome, 54 on the B genome, and 64 on the D genome. The most loci (35) were found to be located on group 5, while the least (18 each) were located on groups 4 and 6 (Table 2).

DISCUSSION

Owing to the steady decrease in cost per sequenced nucleotide and increase in throughput data, NGS technologies have become a powerful approach for the high-throughput discovery of genes; they generate a large amount of sequence data for molecular marker identification (Iorizzo et al., 2011; Silva et al., 2013). Recently, *de novo* transcriptome assembly using Illumina sequences has been successfully developed and widely applied in various important plant species including rice, wheat, and maize (Lu and Lu. 2010; Li et al., 2010; Yang et al., 2015). Large-scale stamen- and pistil-specific transcriptome analysis may provide useful reference data for systemic gene expression profiling and for elucidation of the genetic mechanisms underlying wheat stamen and pistil formation (Yang et al., 2015). In this study, the EST-SSR markers were developed based on wheat pistillody stamen-, stamen-, and pistil-specific transcriptome sequence.

In this study, 8389 EST-SSRs were identified with motifs of 2-6 bp from 121,210 unigenes. The abundance of SSRs was 10.4 kb/SSR on average, which is lower than the abundance (14.0 kb/SSR) in poplar and *Arabidopsis* (Cardle et al., 2000) as well as the abundances (20.0 and 23.8 kb/SSR, respectively) in cotton (Jena et al., 2011) and in soybean (Gao et al., 2003). The findings reported here, however, differ from the abundance of 5.4 kb/SSR reported for wheat by Peng and Lapitan (2005). This discrepancy may be due to the different sources of the ESTs used: Peng and Lapitan used ESTs from an EST database, while the ESTs used in this study were from pistillody stamen-, stamen-, and pistil-specific transcriptome sequence.

Table 1. Sequence and chromosome locations of 131 EST-SSR primer pairs.

Primers	SSR	Left primer (5'-3')	Right primer (5'-3')	Tm (°C)	Chromosome	Primers	SSR	Left primer (5'-3')	Right primer (5'-3')	Tm (°C)	Chromosome
comp03	$(TAA)_7$	CATAAGGCGGGC CGAGTAAT	TGCATGCTGTTT GGAATCCG	59	5A\4DL	comp61	$(GAG)_7$	AATCAGGAACG AGCAGAGGC	GCAGGAGCAGA GGATGTGG	60	1D\7A
comp07	$(CTTG)_5$	TGCTCTTGACCAA TTTGTGCA	GGTAGAGAACA GCACCACCA	59	3D	comp62	$(AC)_{11}$	GCTGGCAAGAT CCTCAGGAA	ACACATCATCGA ACGAGCGA	59	7D
comp08	$(GATC)_5$	GCACAAGCCAAG CCCTAAAC	ATCATCATCAGC AGAGCCGG	60	5A	comp63	$(GGA)_7$	GATCTCGAGGA GGAGGTCGA	CACCACTTCACT CCCCACTC	60	5B
comp10	$(TTCT)_5$	ACTGCATCATGGA TTGGATGGT	AGCAAAACAGC AGCTTGACC	60	1B	comp67	$(ACAT)_5$	GGTTCTCCGGT TCTGTTCCA	AGGTACATATG CCAGCCAGC	59.5	7A
comp15	$(AAC)_7$	GGCATCACATCA CACAGACC	AAGATGCAGCC AGCTCAGAG	58	2D\6A	comp68	$(CAC)_7$	ACTCGAGACTC TTCTCCGCT	GGGGAGAGGAC CATGGATCT	60	3B
comp17	$(CTATT)_5$	CGGTGTTGGTGA CCAGATCA	ACCTTGAGCAC GTCTACTGC	60	6A	comp71	$(GTGC)_5$	CGGGTGTGAAA TCGGACTCA	CACCAAGGGCA AGCCTATGA	60	5A
comp18	$(AGCA)_5$	ACCAGACTAGCC GACTCTGT	ATTGCTTCCGG ATCGCATCT	59	3B	comp72	$(GAAT)_5$	GATCTCGCTGG CTGGATCTC	TAGTCACGCGA AATCGGAGG	60	7A
comp19	$(GTCC)_5$	ATCCACACTCGCA GCCATAC	CCGAGCAAAGG AGGAGGAAG	60	7A\7B\7D	comp77	$(GCTC)_5$	CGTGAGGGAAA GGCTAAGCA	CTCCTTCCTCCC CCTTCCTT	60	5D
comp20	$(AGAT)_5$	TCCCTCCGCCCC TTTTAAAC	GCCTTCCTTGG TCTCTTCCC	60	1D	comp79	$(CTG)_7$	AGTGAACTGAG TGCGCCTAC	CGGAGCACTTT GGAGACCTT	60	3A\3B\3D
comp23	$(ATAA)_5$	TGCTTGAGGTTGT GGGAAGA	AACATGTGCGA CCATACCCC	60	6B	comp80	$(ACGC)_5$	CCACGCTCCGT ATGTTCCTT	CCCCGTGAACT GTTGCTTTG	60	2As\2B\2D
comp24	$(GATC)_5$	TCCTTTCAGGTCC TCGTTGC	CTCTGTGTCAGT CGCTTCGA	60	4DS	comp81	$(GGA)_7$	TAGTAGGAGGT GGTGGTCCG	AACCATGGCTC GTCGAACAT	60	1D
comp25	$(TA)_{10}$	CCAGAAACACAC ACAGCCAC	ACCACCGACAA ACCATGACA	60	1B	comp84	$(TAGC)_5$	CGGAATTAGAG CAGCGCAAG	GCACAATCCTTC CCTCCCTC	60	1D
comp27	$(CA)_{10}$	AACACGCATTTGC ACACCAA	CTCCTTCTGCAA CCGATGGT	60	4A\6A	comp85	$(AAAC)_5$	TTATACAGGCG ACGGGCATC	GCTGGGCCTTT TTCGTTTCT	60	1D\7D
comp29	$(AGGA)_5$	TGGATGCATGCAT GTAGGGT	GGTATGGTAGT GGTGTGGTGG	60	1A\5A	comp91	$(CGA)_7$	CAGCAGGAGCA GCATGATCT	AGCAACCAAAC CGAACCTCT	60	1A
comp31	$(CAGA)_5$	TCACTCTTCCAGG TCCCCTC	TGGCCGACATT GGGATTTCT	60	2B\7B	comp93	$(CTCC)_5$	CGGAATCAGAC CAAATGGCG	GAATGGACGGA GGAGCAGAG	59.5	3B
comp33	$(CAG)_7$	GAAATGAATTCCG ACGGGCG	GAAGCCGATGA GGAGGGAAG	60	1D	comp94	$(CAT)_7$	CGAGAACCCAT GAACACCCA	CAGAAGTAGCT AGGCGAGGG	60	1A\1B
comp36	$(GATC)_5$	CCCCACACAAATC GAGTCGA	TTGGATGTAGG AGGGAGGGG	60	5A	comp95	$(GAA)_7$	CCACAGCAAAC AAGACGACG	GAGGTGACCAG TGCATCCAC	60	4DL\6A\6B
comp39	$(CT)_{10}$	GCTGAGTAAATG GGCGAGGA	AAGCATTCCTTC CCTCTCGC	60	3A	comp97	$(TCC)_7$	GCTCCACGGTG TGAGTAGTC	GAAGGGGATCT GAGCGATGG	60	1D
comp41	$(CAC)_7$	TCTTCCTCCTCCT CCCGTC	TGCCTGTTGAG GTCGAAGAG	60	2D	comp98	$(AGCCA)_5$	TTCCTCTACGTG CTCATCGC	CTCCATGTGGC TAGGGTGTG	60	2B

Continued on next page

Table 1. Continued.

Primers	SSR	Left primer (5'-3')	Right primer (5'-3')	Tm (°C)	Chromosome	Primers	SSR	Left primer (5'-3')	Right primer (5'-3')	Tm (°C)	Chromosome
comp42	$(AGCG)_5$	GCGGCCAGAGTT AGGAAGAA	GACTCTACCGC CGACCAAG	60	5Bl5D	comp99	$(AAGA)_5$	GGGAGGAGAGG CTAGTTGGA	CCGAGGAGGTG ATGCAGATT	60	7A\7B\7D
comp43	$(GGGA)_5$	GCTGTTTCTCCT CAGCGGA	GTCGTTTCTCCT CTCACCCG	59.5	3A	comp102	$(AGCA)_5$	GAGCAAGAGAC CGCCTTCC	AACACAGCAAA ATGGTCCGC	60	5B
comp44	$(GAC)_7$	ATCGCAAATCCTC CTAGCCG	CCCCTCCTCGT CGTCTTTTC	60	6B	comp108	$(CGTC)_5$	GGAGAGGAGAC GACTGCAAC	CTTTTCCCCTCC CACCAAGG	60	1D
comp45	$(CATT)_5$	TAGCCAACTCGC ACACGTAA	GCCGGCTGATC GATCGATAT	60	2B	comp112	$(GAGG)_5$	AGATCCCTCATC GTGCGTTC	CTCTCTCATCCC CTCCCCTC	60	5D
comp47	$(TACA)_5$	GCCGGGTAGATT ATTGCCCA	GCCGTTGCGTA TCGACAATC	60	4A	comp113	$(GCA)_7$	GCAGCACCAGC AACAAATGA	ACGATCCGTCT CCTCTCACA	60	5A\5B\5D
comp48	$(GCGA)_5$	CCGAGGTCCACC AGATCTG	CACCTCACCGC CTCAGAAAG	60	1B	comp114	$(CAA)_7$	CGATGCCGATC TCCCACTC	CATGTCCTCGA AGCTCACGT	60	4A\5D
comp50	$(CTGT)_5$	AAGACGCGAGGA ATGGTGTT	CAACAAGATAA GCAGCGCCC	60	4A	comp115	$(CAGG)_5$	CTCAGAGGCAG AGATCCACG	ACAAGAGAGAA GCGAAGGCC	60	4DS
comp52	$(GCCT)_5$	CCCCTTTTTGGTG GTGGAGA	CAACACCCAAT CCCATCCCA	60	4Ds	comp117	$(GTAT)_5$	AGCACACCTAT GAACGCAAGA	AGATGTACGGC TCCTGTCCT	60	7A
comp53	$(CTGG)_5$	GGTAGATCTGCC GAGCTGAC	AAGCAGAGAGG AGGGAGGAG	60	3A\3B\3D	comp118	$(ATC)_7$	GCATCAAAGGA GCCCTCTGT	ACCCACCCAAA AGGCTTTGA	60	1D
comp54	$(CCAAC)_5$	GCCCTGTTTTCGT TGGTTCC	GACAGACAGAC AGACGGACG	60	5D	comp120	$(GGA)_7$	AGCAAGAACCAA CAGCCAAAG	TTCTTTCCCGTT ACGTCCCC	60	4DL
comp57	$(CTG)_7$	AGCATCCTGGAG TAGAGCCT	TGGGTGGTACG TACTCCAGT	60	1B\1D	comp122	$(GCA)_7$	CCCCAGCTCCA AAACCCTAG	GTTGCTGCGAG AACGAGTTC	60	1A
comp58	$(ATC)_7$	ACCGAACACACC AACAACCT	GCCCATGACAC CATCAGGAA	60	6A\6B\6D	comp123	$(ATTG)_5$	GACTGAGTAGT GCGCACGAT	TTAAACACACAG GGCAGCCA	60	2B
comp59	$(AAGC)_5$	CAACAGCTCACC AACACAGC	TTTTACGTGCCG TCCTCCTC	60	2As\2B\2D	comp124	$(TGGA)_5$	AGGTGTTCTTTT GTGCGTGC	TATTCCTCCCG GCAACAACC	60	3A\3D
comp125	$(GAAG)_5$	AAGAAACAGAGC ACAGCCGA	GAGGGAGGAGG AGGACATGT	60	2As\2B\2D\5A\5B	comp196	$(GCC)_7$	ACTTTGAAGGG GAGCTTGGG	GCATCGGCCGA ACCAAAATT	60	3D
comp126	$(GTCC)_5$	ATCGAGCTTGGC TTTGGGAA	GGTACTCCCAT AATGCCCGG	60	3B	comp197	$(TGGC)_5$	GACGGACTCCC ACCAGGA	CTCATCTCGTCT TCTCCGCC	60	2D\4A
comp132	$(CTTT)_6$	CTTCCGTTCCCTC CCTGTTC	GGCGGTATCTC CTCCCTGTA	60	1A	comp199	$(TGAA)_5$	ACCCGTACCGT ACTCTCAGT	ACAGGATGTCA GACAGAGGT	60	6D
comp134	$(GATT)_5$	AGGGTTCGTTCG CTCGTTC	AATCGTGGAAG GGAGGCATG	60	7B\7D	comp203	$(ATGC)_5$	AACGATCGACTT GGCTGGTT	GGTGTGGACAT GTGCAGAGA	60	3A\3B
comp135	$(TATC)_5$	GAAGCAACTCTA GGACCCCC	CCCTGCCATCT CCCTTTTGT	59.5	4DS	comp206	$(GGGA)_5$	CTGGTACATCC CGACCATCG	CACCCCTTGCA GACCCTAAC	60	3B
comp138	$(AGGA)_5$	GCCAGGTTAGGT CCCATACAG	CCACCTTTCTTC ACCAAAGCG	60	2D	comp207	$(GGT)_7$	CTTTGTTAATGC GGGCTCGG	AGGATGAGGAG GACAACTGGA	60	6A

Table 1. Continued.

Primers	SSR	Left primer (5'-3')	Right primer (5'-3')	Tm (°C)	Chromosome	Primers	SSR	Left primer (5'-3')	Right primer (5'-3')	Tm (°C)	Chromosome
comp140	$(CGAG)_5$	CGAGTCCGATCC ATCCGATC	ATTCAGATTCGG GCTTGGCT	60	1A	comp208	$(GCGG)_5$	TTAGGTCTCCCT CCTCTCGC	ACCACAAGGCC ATGAACCAA	60	3B
comp141	$(ATAGA)_5$	CCAGTTGCTGAC ACAAGCAC	ACCAACGACCT CTCATCACG	60	3D	comp211	$(GTGC)_5$	TCTTTGTCTCTC ACACGCCC	CCTGCTCTACTC TGCTCTGC	60	5B
comp144	$(GGTT)_5$	TGTGTGTGAATGT GAGGGCA	ACCAGGTCGCA CGATCGG	61	7A	comp212	$(GGAT)_5$	CCTGGAGACGA ACACCACTC	CTAGCCAAAAC CCCAACCCT	60	5A
comp148	$(AGCC)_5$	CATTTGTGGACTT GCGGTGG	TGCTGGTAATGT TCGTGCCT	60	1D	comp213	$(ATTT)_5$	AGCATCTGGTA AGCAGGGTT	TCCACAAACCAT CTCCCGTG	60	6A
comp152	$(GAAA)_5$	ATTGTACTGGGC CTTGGTCG	TGGTGCGGTGT TCTACTCAC	60	1B	comp214	$(TGGC)_5$	ATTCTTGGGCCT GGTCTGTG	ACAATGCGATTA ACCCCCGT	60	7D
comp155	$(GCG)_7$	TGGAACTGCCCA TGGATGTC	TCCGCACCAGG AATCAAGAG	60	4BS	comp215	$(TAG)_7$	CTGCTGGGACC AAAGCAAAC	AATGCCAGATC CAGGGTTCC	60	5A
comp156	$(TCAT)_5$	TTTCTGGCAGTTC CAAGGGG	AGTTGCTTGGAT GTCGGGAG	60	2B	comp217	$(TCCG)_5$	AGCAACCCGGC GTATACAAA	GCGCCACGAGG AAGTAGTAG	60	6A
comp159	$(GA)_{10}$	ACCCTAGCAAGG TTTGACCG	GATAGACATCC ACTGGCCCG	60	5D	comp219	$(GGAG)_5$	CGGCCGGGTTT CTTTTCTTG	CAGGAGGAGGA TTGATCGGC	60	5B\5D
comp160	$(TTCT)_5$	AGTCGACAACCAT GGCATGT	GGGAACTTTGG CTTGTGCAC	60	2D	comp222	$(CGCA)_5$	CTGGGTGACAT TTTGACGCC	CACTACTGTGC CCCTCATCG	60	1B
comp162	$(GTCG)_5$	CCAGTTGGAGAC GAGATCCG	ACCCCACTCCT CCTCCATC	60	7A\7D	comp225	$(CGA)_7$	ACAGAACAGCA AGCGATCGA	ATGAGGATTGC GCCTTGACA	60	7D
comp164	$(ATTG)_5$	GGGGAGGACGCC ATTGTATT	AGGTTTGCATTC CACACCCT	60	2D	comp226	$(CTTC)_5$	GTATCCTGGTA GTGCCGTGG	TTGACTGGAAC GAGCTGACG	60	4DS
comp165	$(GATG)_5$	TCTGGGGGCTGA CTATCTCC	CCTCCACTGTT GTAGGCAGG	60	6A	comp229	$(AATC)_6$	CCTGACCCTGA TCGATCGTG	CGCAAGACCAC ACCATACCT	60	2D\5A\5D
comp167	$(CAGAG)_5$	GACGAGAGCAAA GAGCTCTTC	GGTTCGACTGG CAATCGGC	60	5A\5B\5D	comp231	$(CGGCA)_5$	TAGCAGCATTG ACCAGTCCG	GCTAGGTGCGC AGTACTAGG	60	3A
comp169	$(CTGG)_5$	GCCAGGTGGTGA GGAACTC	CGGAGGAGAGT GAGGAGGAG	60	7D	comp233	$(TGAT)_5$	CCTCCCCCTCTC AGATGAGCT	GTCCTCCTCCT CGATCTCCA	60	7A\7D
comp171	$(AG)_{10}$	TGTCAGACTTGCT AGGCGAC	TCAAGACCCAC ATGACACCT	60	1D	comp234	$(TTGT)_5$	CGGTCACCCTG GAAGGTTTT	GACACACGCAC ACAGAGAGA	60	6A
comp172	$(TTA)_7$	ATGGATCAGACC CTGCCTCT	AAACAGCTTGTC GCGCGCC	60	7A	comp236	$(GCAT)_5$	GCTGTGGACGA GCTAGCTAG	TCCACGTGTTG ATGATGCCA	60	5D
comp173	$(GAA)_7$	GTGGAGATGATG GTGTGGCA	CGTCAAGCACA ACGTTTACGA	60	2B	comp241	$(CGAT)_5$	AATCACAAGCAT GGCAACGG	TACTCTCCACCT CCTGCTCC	60	3A
comp175	$(TAC)_7$	ATACAATGCTGCC ACCACCA	TCGAATCGACA AGGCTTGCT	60	3B	comp243	$(GGCA)_5$	CATCGTCGCTC AGATCCCC	CTGGCTCCTAC TACTTGGCG	60	5D
comp178	$(CT)_{10}$	TCCCCTGACCTCT CGATCTC	GAATTGCTGTTG CTACCGGC	60	1A	comp245	$(TATC)_5$	CCCGCTGACCC TTTTGACTT	ACATCACGGAA GCAACTTGC	59.5	2B

Table 1. Continued.

Primers	SSR	Left primer (5'-3')	Right primer (5'-3')	Tm (°C)	Chromosome	Primers	SSR	Left primer (5'-3')	Right primer (5'-3')	Tm (°C)	Chromosome
comp179	$(GGAG)_5$	AGTGCCCCCTCC TATCAGAA	ATATCTGAAACA CAGCGCGC	59.5	1A	comp252	$(CCGAC)_5$	CAACAAGAACA AGTGCGGCA	ATCGGATCGGA TCGGAGGAT	60	6B
comp181	$(GTG)_7$	ATTGGTAGGTCG GGACTCGA	GCTGCCTAATCT CCTCGGAC	60	4DS	comp257	$(AACC)_5$	ACCTTCTCCATC CTCTCCCC	GCTTTCTCCTCT CTCGGCTG	60	4A
comp183	$(CTA)_{115}$	ATTCTAGCAGTAC GGCTCGC	GAACAGAACAG AGCTCGGCT	60	5B	comp260	$(TGCC)_5$	CGGCAGTGCAA TCATCCAAC	AATAACGGCGA TGGACTGCA	60	4DS
comp185	$(GGAG)_5$	CTCTCCCCTGCTT CTCCTCT	ATCCCCCAAGT CCTCTCCTC	60	3A	comp261	$(CTT)_7$	ATTGGTGACTGT GTGGGCAA	CGAGCGGAGAG GTTTGTTCT	60	7B
comp190	$(TTCT)_5$	GAATGACACCTGT TGCGCTG	GCTGGATAACA CAGCCACCT	60	4A	comp262	$(AAAG)_5$	AGGGCGAATCG GAATGGATC	GCGAGGATTGT CCAGTCCAG	60	7A
comp191	$(GCAA)_5$	GGAAGTTGGCTG CAGATTGC	ATGGGCAGCCA TTTTGGAGT	60	2B	comp263	$(GAGG)_5$	GACGAGCCGTA CGACTTGAT	CCTAGCTACGT ACGTGCCTG	60	5B
comp193	$(TA)_{10}$	GAGATGGGCATG TGCTGTCT	CGCCACACCTT GATTTCCAC	60	5B	comp265	$(CGG)_7$	GGAGATGTACT CGTCGCAGG	CGCAAAGCCGG AGTAAAACC	60	7B
comp195	$(CA)_{10}$	GAGCTTGAGGAT CTCGAGGC	GCTTTTGGTCG CTGTCGTTT	60	1D	comp267	$(AGT)_7$	ACCGAACCAAC CAATGATCCT	AGTACAAGCAG TTCACGGGG	60	6B
comp269	$(GGAGAA)_6$	TTGTGCAGAAGG AGACTCGC	GGTTTCAGCCA TCAGGTCCA	60	2As\2B\2D\5B						

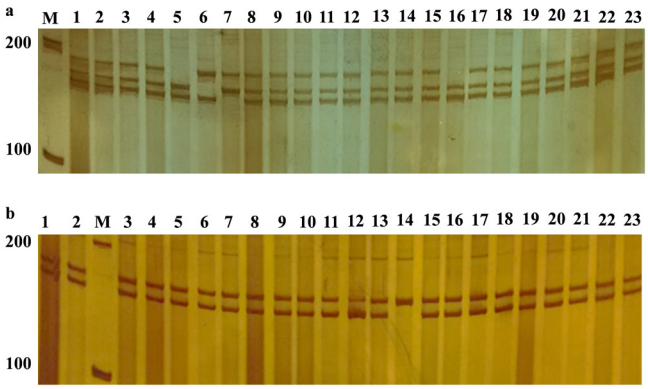

Figure 4. Chromosome locations of EST-SSR markers using the Chinese Spring nulli-tetrasomic wheat line. **A.** Primer comp 81; **B.** primer comp 115. *Lane 1* = CS; *lane 2* = N1AT1D; *lane 3* = N1BT1A; *lane 4* = N1DT1B; *lane 5* = Dt2AS; *lane 6* = N2BT2D; *lane 7* = N2DT2A; *lane 8* = N3AT3B; *lane 9* = N3BT3D; *lane 10* = N3DT3B; *lane 11* = N4AT4D; *lane 12* = Dt4BS; *lane 13* = Dt4DL; *lane 14* = Dt4DS; *lane 15* = N5AT5D; *lane 16* = N5BT5D; *lane 17* = N5DT5B; *lane 18* = N6AT6B; *lane 19* = N6BT6A; *lane 20* = N6DT6B; *lane 21* = N7AT7B; *lane 22* = N7BT7A; *lane 23* = N7DT7B.

Table 2. Distribution of EST-SSR markers on wheat chromosomes.

Homologous group	Genome			Total
	A	B	D	
1	8	7	13	28
2	4	12	11	27
3	9	10	6	25
4	7	1	10	18
5	11	12	12	35
6	10	6	2	18
7	11	6	10	27
Total	60	54	64	178

Because genes have characteristic temporal and spatial expression patters, there were less ESTs from the pistillody stamen-, stamen-, and pistil-specific transcriptome sequence than from the EST database; moreover, accordingly, the number of SSRs was also lower. The fact that the EST-SSRs in this study were tissue-specific means that the resulting data can be used for mapping of genes associated with flower development in wheat.

The TNR motif was previously shown to be the most abundant SSR motif in wheat (Gao et al., 2003; Thiel et al., 2003; Chen et al., 2005) and similar results were observed in the present study. Among TNRs, the AAC/GTT motif was the most frequent in wheat, which is in accordance with other reports on wheat (Gao et al., 2003; Chen et al., 2005). The DNR AG/CT and AC/GT motifs were reported to be the most frequent repeats in barley (Thiel et al., 2003), while the AG/CT repeat was frequently observed in wheat (Gao et al., 2003). In this study, the GA/TC motif was the most abundant, followed by AG/CT, TG/AC, and GT/AC. Although the functional significance of SSRs in plant transcript regions is not clear, the AG/CT motif, a homopurine-homopyrimidine stretch

present at a high frequency in the 5' untranslated region, reportedly plays a role in regulating gene expression and nucleic acid metabolism in plants (Martienssen and Colot, 2001; Scaglione et al., 2009; Wöhrmann and Weising, 2011).

A subset of 300 EST-SSRs was randomly selected for an evaluation of the applicability of the EST-SSRs. Of the 300 primer pairs, 177 (59%) yielded unambiguous PCR products across five wheat cultivars. The success rate of PCR amplification was lower than that reported for cucumber (88.6%) (Hua et al., 2010). This difference may be a result of the complexity of the genome of common hexaploid wheat. Of the 177 primer pairs, 117 showed allelic polymorphisms. Nulli-tetrasomic lines are widely employed to assign molecular markers and genes due to the precision associated with the line. Yu et al. (2004) located 80 EST-SSRs and 104 loci on wheat chromosomes using nulli-tetrasomic lines, and Chen et al. (2005) located 93 EST-SSRs (193 loci) on wheat chromosomes. Li et al. (2008) located 139 EST-SSRs (240 loci) on the 21 wheat chromosomes. In the present study, 131 EST-SSRs and 178 loci were located on the 21 wheat chromosomes using nulli-tetrasomic lines. The chromosomal locations of EST-SSR loci provide a basis for genetic mapping and gene identification.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (Grant #31301319 and #31540041), a key project of the Chinese Ministry of Education (Grant #211164).

REFERENCES

Cardle L, Ramsay L, Milbourne D, Macaulay M, et al. (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* 156: 847-854.

Chang SJ, Puryear J and Cairney J (1993). A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* 11: 113-116. http://dx.doi.org/10.1007/BF02670468

Chen HM, Li LZ, Wei XY, Sishen L, et al. (2005). Development, chromosome location and genetic mapping of EST-SSR markers in wheat. *Chin. Sci. Bull.* 50: 2328-2336. http://dx.doi.org/10.1007/BF03183744

Conesa A, Götz S, García-Gómez JM, Terol J, et al. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674-3676. http://dx.doi.org/10.1093/bioinformatics/bti610

Dutta S, Kumawat G, Singh BP, Gupta DK, et al. (2011). Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol.* 11: 17-29. http://dx.doi.org/10.1186/1471-2229-11-17

Gao LF, Tang JF and Li HW (2003). Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol. Breed.* 12: 245-261. http://dx.doi.org/10.1023/A:1026346121217

Hua JB, Zhou XY and Li JW (2010). Development of novel EST-SSR markers for cucumber (*Cucumis sativus*) and their transferability to related species. *Sci. Hortic. (Amsterdam)* 125: 534-538. http://dx.doi.org/10.1016/j.scienta.2010.03.021

Huang H, Lu J, Ren Z, Hunter W, et al. (2011). Mining and validating grape (*Vitis* L.) ESTs to develop EST-SSR markers for genotyping and mapping. *Mol. Breed.* 28: 241-254. http://dx.doi.org/10.1007/s11032-010-9477-2

Iorizzo M, Senalik DA, Grzebelus D, Bowman M, et al. (2011). *De novo* assembly and characterization of the carrot transcriptome reveals novel genes, new markers, and genetic diversity. *BMC Genomics* 12: 389-402. http://dx.doi.org/10.1186/1471-2164-12-389

Jena SN, Srivastava A, Rai KM, Ranjan A, et al. (2012). Development and characterization of genomic and expressed SSRs for levant cotton (*Gossypium herbaceum* L.). *Theor. Appl. Genet.* 124: 565-576. http://dx.doi.org/10.1007/s00122-011-1729-y

Kaur S, Pembleton LW, Cogan NO, Savin KW, et al. (2012). Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics* 13: 104-115. http://dx.doi.org/10.1186/1471-2164-13-104

Koelling J, Coles MC, Matthews PD and Schwekendiek A (2012). Development of new microsatellite markers (SSRs) for *Humulus lupulus*. *Mol. Breed.* 30: 479-484. http://dx.doi.org/10.1007/s11032-011-9637-z

Li D, Deng Z, Qin B, Liu X, et al. (2012). *De novo* assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). *BMC Genomics* 13: 192-203. http://dx.doi.org/10.1186/1471-2164-13-192

Li LZ, Wang JJ, Guo Y, Jiang F, et al. (2008). Development of SSR markers from ESTs of gramineous species and their chromosome location on wheat. *Prog. Nat. Sci.* 18: 1485-1490. http://dx.doi.org/10.1016/j.pnsc.2008.05.012

Li P, Ponnala L, Gandotra N, Wang L, et al. (2010). The developmental dynamics of the maize leaf transcriptome. *Nat. Genet.* 42: 1060-1067. http://dx.doi.org/10.1038/ng.703

Liu L, Guo W, Zhu X and Zhang T (2003). Inheritance and fine mapping of fertility restoration for cytoplasmic male sterility in *Gossypium hirsutum* L. *Theor. Appl. Genet.* 106: 461-469.

Liu LW, Zhao LP, Gong YQ, Wang MX, et al. (2008). DNA fingerprinting and genetic diversity analysis of late-bolting radish cultivars with RAPD, ISSR and SRAP markers. *Sci. Hortic. (Amsterdam)* 116: 240-247. http://dx.doi.org/10.1016/j.scienta.2007.12.011

Lu T, Lu G, Fan D, Zhu C, et al. (2010). Function annotation of the rice transcriptome at single-nucleotide resolution by RNA-seq. *Genome Res.* 20: 1238-1249. http://dx.doi.org/10.1101/gr.106120.110

Martienssen RA and Colot V (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* 293: 1070-1074. http://dx.doi.org/10.1126/science.293.5532.1070

Peng JH and Lapitan NL (2005). Characterization of EST-derived microsatellites in the wheat genome and development of eSSR markers. *Funct. Integr. Genomics* 5: 80-96. http://dx.doi.org/10.1007/s10142-004-0128-8

Powell W, Machray GC and Provan J (1996). Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.* 1: 215-222. http://dx.doi.org/10.1016/S1360-1385(96)86898-0

<edb>Rozen S and Skaletsky HJ (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press.365-386.

Scaglione D, Acquadro A, Portis E, Taylor CA, et al. (2009). Ontology and diversity of transcript-associated microsatellites mined from a globe artichoke EST database. *BMC Genomics* 10: 454-470. http://dx.doi.org/10.1186/1471-2164-10-454

Silva PI, Martins AM, Gouvea EG, Pessoa-Filho M, et al. (2013). Development and validation of microsatellite markers for *Brachiaria ruziziensis* obtained by partial genome assembly of Illumina single-end reads. *BMC Genomics* 14: 17-25. http://dx.doi.org/10.1186/1471-2164-14-17

Sraphet S, Boonchanawiwat A, Thanyasiriwat T, Boonseng O, et al. (2011). SSR and EST-SSR-based genetic linkage map of cassava (*Manihot esculenta* Crantz). *Theor. Appl. Genet.* 122: 1161-1170. http://dx.doi.org/10.1007/s00122-010-1520-5

Thiel T, Michalek W, Varshney RK and Graner A (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106: 411-422.

Wöhrmann T and Weising K (2011). *In silico* mining for simple sequence repeat loci in a pineapple expressed sequence tag database and cross-species amplification of EST-SSR markers across Bromeliaceae. *Theor. Appl. Genet.* 123: 635-647. http://dx.doi.org/10.1007/s00122-011-1613-9

Yang Z, Peng Z, Wei S, Liao M, et al. (2015). Pistillody mutant reveals key insights into stamen and pistil development in wheat (*Triticum aestivum* L.). *BMC Genomics* 16: 211-220. http://dx.doi.org/10.1186/s12864-015-1453-0

Yu JK, Dake TM, Singh S, Benscher D, et al. (2004). Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. *Genome* 47: 805-818. http://dx.doi.org/10.1139/g04-057

Zeng S, Xiao G, Guo J, Fei Z, et al. (2010). Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. *BMC Genomics* 11: 94-104. http://dx.doi.org/10.1186/1471-2164-11-94