African Journal of Biotechnology Vol. 11(69), pp. 13269-13371, 28 August, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.655
ISSN 1684–5315 © 2012 Academic Journals

Full Length Research Paper

In silico comparative analysis of EST-SSRs in three cotton genomes

Daojun Yuan[#], Shaoguang Liang[#], Zhongxu Lin* and Xianlong Zhang

National Key Laboratory of Crop Genetic Improvement and National Centre of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, Hubei, China.

Accepted 7 September, 2011

In this study, expressed sequence tags- simple sequence repeat (EST-SSRs) were surveyed in three cotton genomes (*Gossypium arboreum*, Ga; *Gossypium raimondii*, Gr and *Gossypium hirsutum*, Gh). The frequency of EST-SSRs was highest in Gr, and motif type for hexanucleotide was obviously abundant in Gr. Trinucleotide repeats were the most abundant motif; AT and AG, AAG and ATC were the most frequent motifs for dinucleotide and trinucleotide, respectively. The repeat number was greatly diverse between the three genomes with the highest variation in Gh. AG and AAG had a high frequency both in homologue groups (HGs) with and without repeat number change between genomes. The range of repeat number change in each HG was wider in Gr-Gh. The annotation of the SSR-ESTs showed that more Gene Ontology (GO) items targeted by SSR-ESTs of Ga and Gr than those of Gh. This study gave us new insights into the difference between the three cotton genomes, which will be more helpful to understand the differentiation and evolution of the three genomes.

Key words: Cotton, simple sequence repeat, expressed sequence tags, motif, gene ontology.

INTRODUCTION

Microsatellites or simple sequence repeats (SSRs) are DNA sequences consisting of tandemly repeated arrays of short (1-6 nucleotides) motifs, which are widespread in both eukaryotic and prokaryotic genomes (Field and Wills, 1996; Tóth et al., 2000). SSRs are abundant across genomes, showing high levels of polymorphism, and are usually considered as evolutionarily neutral DNA sequences. This polymorphism is probably due to slipped-strand mispairing, and simple tandem repeats may be predisposed to further length changes by unequal crossing over and (or) subsequent replication, repair, or recombination errors (Levinson and Gutman, 1987). As a result, microsatellite loci often mutate by insertions or deletions of one or more repeat elements, and the mutation rates generally increase with an increase in the length of the repeats (Wierdl et al., 1997; Xu et al., 2000).

Microsatellite markers were developed for a wide range of plant species, including many crops. The advantage of microsatellites as molecular markers is that they detect polymorphisms at highly variable loci (Powell et al., 1996; Hedrick, 2001). Furthermore, they are polymerase chain reaction (PCR)-based marker, co-dominant, and distribute throughout the genome. Together, these characteristics make microsatellite loci one of the best genetic markers for mapping purposes (Oliveira et al., 2006).

Microsatellites were developed from both coding and noncoding regions of plant genomes (Temnykh et al., 2001; Morgante et al., 2002). Several resources can be used to search for SSRs, including a variety of DNA libraries such as genomic-enriched for SSRs, bacterial artificial chromosome (BAC) and cDNA libraries, as well as public database, including expressed sequence tags (ESTs) database (Broughton et al., 2003; Li et al., 2004). In recent years, a number of random and targeted gene discovery programs have been initiated for a range of species, and this has led to a dramatic increase in the number of ESTs in public and specific species databases. ESTs database is a fast and efficient way to analyze the transcribed portion of the genome, and now becomes a potentially valuable source for genetic marker development.

---

*Corresponding author. E-mail: linzhongxu@mail.hzau.edu.cn. Tel: +86-27-87280510. Fax: +86-27-87280196.

# Both authors contributed equally to the work

There are over 375,700 cotton ESTs deposited in GenBank(http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html, 1st December, 2010). Among these ESTs, *Gossypium hirsutum* (upland cotton) had the maximum number (268,797) because it is widely planted in the world, followed by *Gossypium raimondii* (63,577), *Gossypium arboreum* (41,768), *Gossypium barbadense* (1,356) and *Gossypium herbaceum* var. africanum (247). Except for transcriptome analysis in cotton, these ESTs provide a robust way to develop SSR markers for genetic mapping the cotton genome. Qureshi et al. (2004) developed 84 primers for 133 SSR-containing ESTs from 9,948 sequences of known genes belonging to *G. hirsutum* L. in GenBank. Han et al. (2004) developed 544 EST-SSR primer pairs from 931 ESTs derived from *G. arboreum* fibers at 7 to 10 days post anthesis. After that 489 EST-SSR primer pairs were isolated from 13,505 ESTs which were developed from two cotton fiber/ovule cDNA libraries constructed for upland cotton (Han et al., 2006). Park et al. (2005) developed 1,232 microsatellite markers using 1,557 ESTs-containing SSRs (10 bp) and 5,794 ESTs-containing Complex Sequence Repeats (CSRs) (12 bp) obtained from ~14,000 consensus sequences derived from fiber ESTs generated from the cultivated diploid species *G. arboreum* L. cv AKA8401. Talierco et al. (2006) identified 192 EST-SSR primers from multiple *G. hirsutum* tissues. Wang et al. (2006) developed 1,554 EST-SSRs from 58,906 nonredundant *G. raimondii* EST sequences. Guo et al. (2007) additionally developed 664 *G. hirsutum*-derived EST-SSRs from 12,463 *G. hirsutum* acc.TM-1 sequences and 11,692 *G. hirsutum* cv. Xuzhou142 sequences. Zhang et al. (2007) developed 119 EST-SSRs based on 98 unique ESTs from a cDNA library constructed using developing fibers from *G. barbadense* acc. 3-79. In total, 5,318 EST-SSRs have been developed, accounting for 44.56% of total cotton SSRs (11,938) in the database of cottonmarker (http://www.cottonmarker.org/Primer.shtml, 31 December, 2010).

Cotton (*Gossypium* spp.) belongs to the genus *Gossypium* of the family *Malvaceae*, and it is an important cash crop and the second largest source of textile fiber and edible oil throughout the world. The genus *Gossypium* L. comprises about 45 diploid and 5 tetraploid species, and these are assigned to nine genome types on the basis of chromosome pairing affinities: eight diploid genomes (A, B, C, D, E, F, G and K; 2n=2x=26) and one tetraploid genome (AD; 2n=4x=52) (Endrizzi et al., 1985; Percival et al., 1999). At the tetraploid level, there are five species, designated $(AD)_1$ through $(AD)_5$ for their genome constitutions (Cronn and Wendel, 1998). The allotetraploid species are thought to have formed about 1-2 million years ago (MYA) after a polyploidization event that brought together the genomes of diploids closely related to *G. herbaceum* L. (2n=2x=26, A_1 genome) or *G. arboreum* L. (2n=2x=26, A_2 genome), and *G. raimondii* L. (2n=2x=26, D_5 genome), and they have been domesticated through extensive human selection (Wendel and Cronn, 2003).

Although the relationship of A, D and AD genome have been clearly defined, the abundantly existing EST sequences of the three genomes may provide an alternative way to study the three genomes. In this study, microsatellites were globally analyzed in ESTs of the three species to reveal their commonness and difference between A, D and AD genome and to discuss its evolution.

MATERIALS AND METHODS

Sources of EST sequences

The assembly EST sequences of *G. arboreum* (Ga), *G. raimondii* (Gr) and *G. hirsutum* (Gh) were downloaded from TIGR (The Institute for Genomic Research) PTAs (Plant Transcript Assemblies) database (http://plantta.tigr.org; 15 February, 2010) (Childs et al., 2006). The EST sequences of *G. arboreum* (Release 2) including 4,591 assemblies and 22,292 singletons amounted to 39,157 ESTs in GenBank with 19,283 kb; the EST sequences of *G. raimondii* (Release 2) including 8,665 assemblies and 19,977 singletons amounted to 63,265 ESTs in GenBank with 22,879 kb. The EST sequences of *G. hirsutum* (Release 2) including 24,797 assemblies and 45,870 singletons amounted to 176,521 ESTs in GenBank with 48,609 kb.

Identification of SSRs in ESTs

A windows-based robust SSR pipeline named Serafer developed in our laboratory was used to search SSRs (ftp://ensembl.genomics.org.cn/other/Serafer_1.9.5.zip or http://www.tgir.org/download/software/Serafer_1.9.5.zip). Serafer takes a FASTA formatted sequence file as an input, and it can produce an Excel file with number of SSRs, sequence name, SSR type, SSR motif, SSR position, repeat length, repeat number, repeat score and the length of the sequence.

In the present study, perfect EST-SSRs were considered to contain motifs of di- to hexanucleotides in size. The minimum repeat unit was defined as 9 for dinucleotides, 6 for trinucleotides, 5 for tetra- and pentanucleotides and 4 for hexanucleotides. These limits were chosen based on earlier studies that investigated plant EST-SSRs (Bérubé et al., 2007).

Compositional analysis of EST-SSRs

Microsatellites were classified by type (example, di- to hexanucleotide) and motif (example, AC, AG, AT, and CG) in Microsoft Excel®. What's more, the motifs were classified considering circular permutations and complementarities, example, AG, GA, TC and CT were considered as a single category. The analysis of the occurrence and frequency of EST-SSRs among Ga, Gr and Gh was carried out in Microsoft Excel®. Results on repeat type, motif type, repeat number, repeat length and frequency were collected for the three genomes. Frequency of EST-SSR refers to kilo-base pairs of EST sequences containing one SSR.

Identification of homologous SSR-ESTs through blast analysis

Sequences of SSR-ESTs (ESTs containing SSR) were respectively extracted from each database and were assembled into three new

Table 1. Number and frequencies of EST-SSR repeats in three cotton species.

Repeat type	Ga	Gr	Gh
Dinucleotide	183(9.49)	299(13.07)	274(5.64)
Trinucleotide	289(14.99)	499(21.81)	1036(21.31)
Tetranucleotide	32(1.66)	59(2.58)	190(3.91)
Pentanucleotide	10(0.52)	25(1.09)	38(0.78)
Hexanucleotide	90(4.67)	108(4.72)	251(5.16)
Total SSRs	604	990	1789
Total length (Kb)	19,283	22,878.9	48,608.7
Average distance (Kb)	31.93	23.11	27.17

Numbers in parentheses show frequency of repeat unit per 100kb. Ga, *Gossypium arboreum*; Gr, *Gossypium raimondii*; Gh, *Gossypium hirsutum*.

individual databases for Blast analysis. The Blast program MegaBLAST (Altschul et al., 1990) was downloaded from NCBI to identify homologous SSR-ESTs with the criterion for a sequence match based on the E-value threshold of $1e^{-15}$.

Annotation and functional classification of SSR-ESTs

Annotation of gene ontology (GO) terms were performed using Blast2GO (Conesa et al., 2005) with default parameters. Homolog analysis against the NCBI nr protein database was performed using a BLASTX (Altschul et al., 1990) program (E-value$\leq 1e^{-3}$) and top 20 best BLAST hits were kept. Then, InterProScan (Mulder and Apweiler, 2007) was performed and InterProScan GOs was merged to annotation. The functional categories were simplified by using the Plant GOslim (http://www.geneontology.org/GO.slims.shtml). Finally, group 2nd level and multilevel GO terms of the three main categories (biological process, molecular function and cellular component) was obtained and classified.

RESULTS

Comparison of EST-SSRs between the three genomes on the whole

Frequency and distribution of repeat type

There were 604, 990 and 1,789 EST-SSRs present in the EST sequences of Ga, Gr and Gh, respectively (Table S1). The frequency of dinucleotide was higher in Gr than that in Gh, but the number was similar between them; the frequency of trinucleotide was similar between Gr and Gh, but the number was very different (Table 1). The frequency of EST-SSRs was one in 31.93 kb in Ga, 23.11 kb in Gr and 27.17 kb in Gh, respectively. Among all the repeat types, trinucleotide repeats were the most abundant SSR types with up to 21.81 trinucleotide repeat units in 100 kb ESTs (Table 1).

Frequency and distribution of repeat motif

All the di- and trinucleotide motifs (except CG in Gr and Gh) showed different abundance in the three cotton genomes (Table 2). The most frequent dinucleotide motif was AT in Ga and Gr, whereas AG in Gh. The frequency of AC was low in all three genomes; however, AT was very frequent in Gr, which may be an important characteristics of this genome. For trinucleotide, AAG and ATC were the most frequent motifs in all the three genomes; AAT was also frequent in Gr and Gh.

A total of 9 tetranucleotide motifs, 4 pentanucleotide motifs and 31 hexanucleotide motifs were found in Ga; 21, 20 and 106 in Gr, respectively; 13, 13 and 52 in Gh, respectively. It was obvious that motif type was very abundant in Gr, especially for hexanucleotide. Although the tetra- to hexanucleotide motifs were less in Ga, the AGCCTG was very abundant with 35 ones found. Some tetranucleotide motifs were also very abundant in Gr, such as AAAT, AAAC, AAAG and ATAC. However, no abundant motifs were found to exist in Gh (Table S2).

Repeat number distribution

With an increase in their repeat number, all the perfect repeats detected in the three cotton genomes showed a decrease in their frequency as shown in Figure 1. The three cotton genomes showed the same tendency with a peak at the repeat number of 6, but the frequency was very low in Ga. However, the repeat number was greatly different between the three genomes with more variance in Gh and less in Gr (Figure 1).

To reveal which repeat unit(s) contributes to the variety of repeat number between the three cotton genomes, the major repeat units (di-, tri- and hexanucleotide) were separately counted. Figure 2 shows that di- and hexanucleotide contributed little to the repeat number frequency (Figure 2a and 2c), but dinucleotide contributed a lot to the repeat number variation (Figure 2a). Trinucleotide contributed a lot to the frequencies of repeat number but little to the repeat number variation, and the high frequency of repeat number of 6 was mainly con-tributed by trinucleotide (Figure 2b). A further exploring of the repeat number of different motifs revealed that the repeat number variation of dinucleotide

Table 2. Frequency of di- and trinucleotide repeats in three cotton species[a].

Repeat type	Motif	Ga	Gr	Gh
Dinucleotide	AT	4.77	8.65	1.05
	AC	0.52	0.87	0.76
	AG	4.10	3.54	3.83
	CG	0.10	0	0
Trinucleotide	AAT	1.76	3.63	2.37
	AAC	0.57	1.49	1.52
	AAG	4.51	6.51	6.03
	ATC	2.85	4.28	4.01
	ACT	0.21	0.48	0.25
	ACC	1.71	1.79	2.47
	ACG	0.41	0.44	0.64
	AGC	1.61	2.01	2.08
	AGG	0.73	0.61	0.88
	CCG	0.62	0.57	1.07

[a] Frequency of motif per 100 kb.

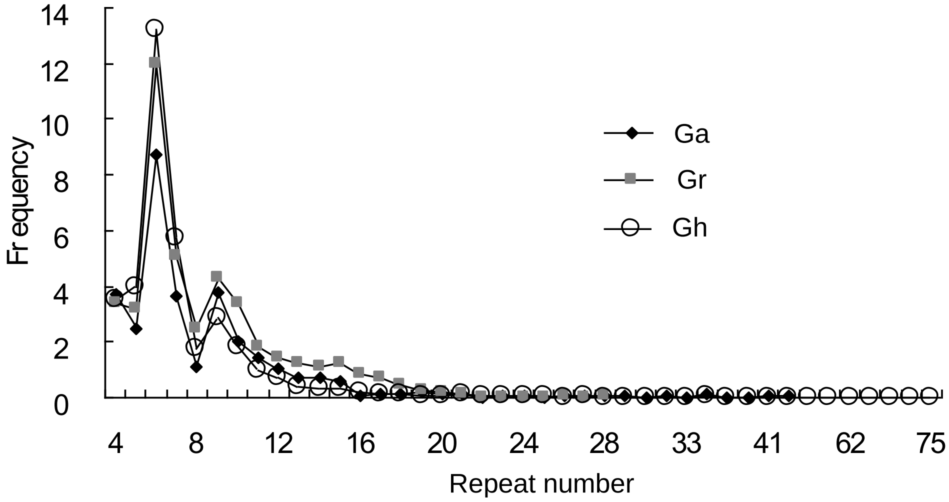


Figure 1. Distribution of the repeat number showed by number/kb in the three cotton species.

was mainly contributed by AG motif in Ga and Gh, and AT in Gr; the frequency of repeat number of 6 was mainly contributed by AAG and ATC motif.

Characteristics of EST-SSRs between A, D and AD genomes by homologous sequences

Homologous SSR-ESTs between A and AD genomes

Blast (E-value≤1e^{-15}) analysis was performed to search homologous SSR-ESTs between Ga and Gh, resulting in 196 homologous groups (HG) including 36 dinucleotides, 104 trinucleotides, 12 tetranucleotides and 44 hexanucleotides. In the 195 HGs, one EST of Ga had 1-9 ESTs with an average of 3 of Gh; in the exceptional HG, two Ga sequences with 9 and 10 AT repeats were corresponded to two Gh sequences with 10 and 11 AT repeats. The 196 HGs were divided into three groups: no repeat number change between Ga and Gh with 74 HGs (Group I) (37.8%), more repeat number in Ga than Gh with 31 HGs (Group II) (15.8%) and less repeat number in Ga than Gh with 91 HGs (Group III) (46.4%) (Table S3).

In Group I, all the 5 dinucleotides were AG motif, and ATC and AAG were the most frequent motifs in the trinucleotides. In Group II, the dinucleotides included 1 AC, 3 ATs and 10 AGs; the frequencies of different motifs were not obviously different in the trinucleotides. In Group

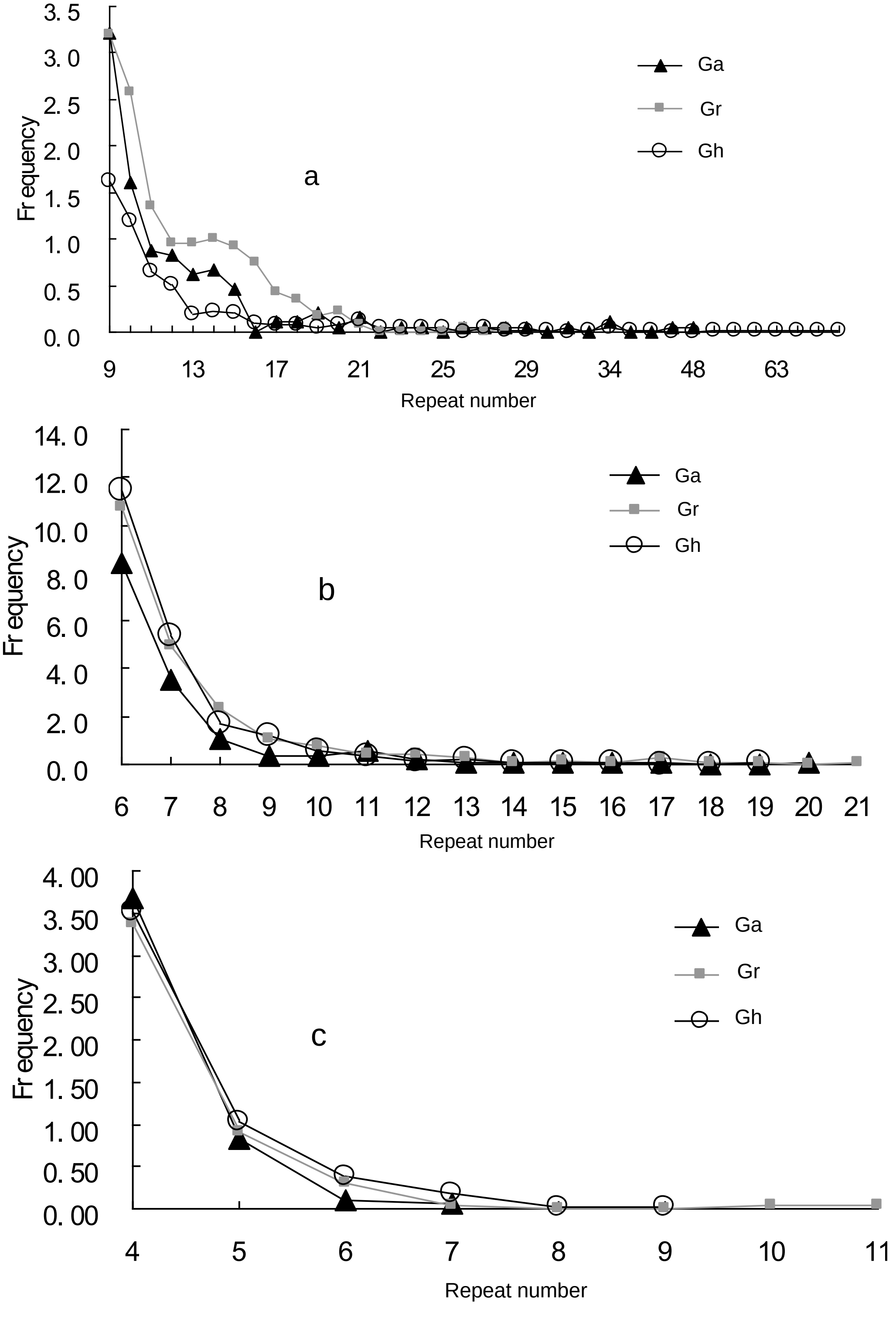


Figure 2. The repeat number of di-, tri- and hexanucleotide between the three cotton species. (a) the repeat number of dinucleotide; (b) the repeat number of trinucleotide; (c) the repeat number of hexanucleotide.

III, the dinucleotides repeats included 4 AGs and 13 ATs, and the most frequent motifs in the trinucleotides repeats were AGC and AAG; among the 35 hexanucleotides, 30 of them were TCAGGC motif (Table S3 and Figure 3a). Because multiple sequences were usually included in one HG and the repeat number varied between

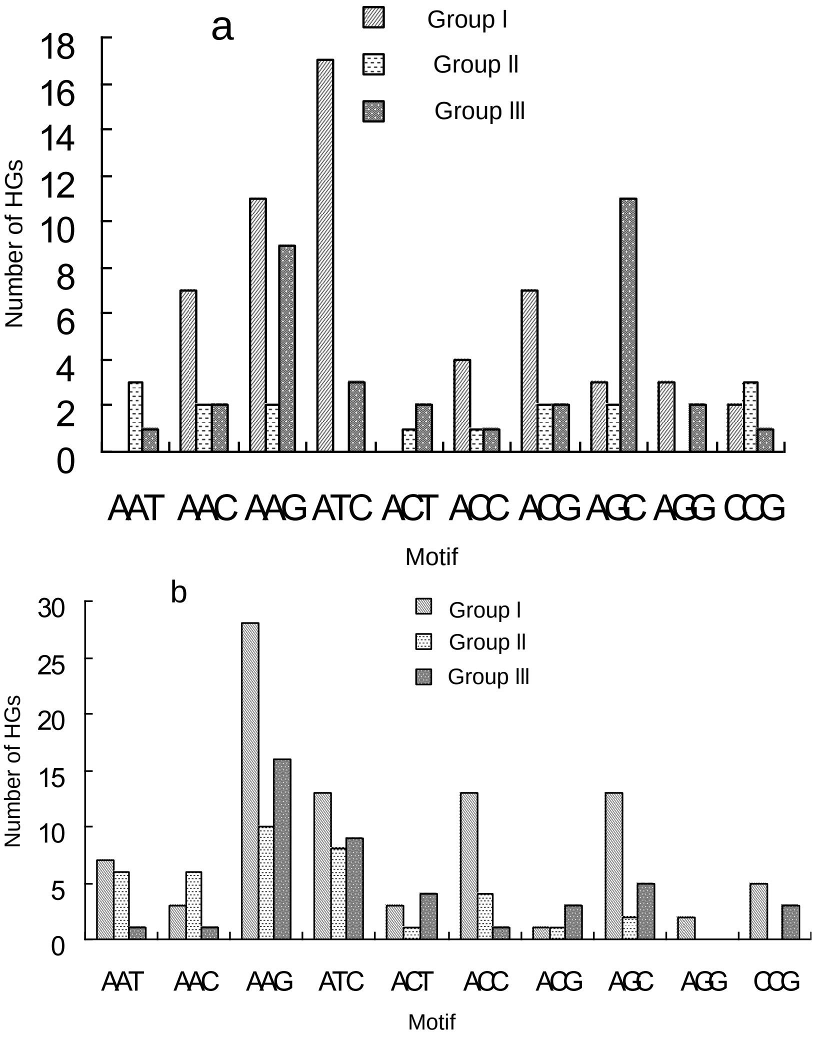


Figure 3. Number of HGs of different trinucleotides motifs in (a) Ga-Gh and (b) Gr-Gh.

sequences, when comparing difference between different genomes, the repeat numbers were averaged. Table 3 shows that the differences of repeat number between Ga and Gh were less than 2 in most HGs.

Homologous SSR-ESTs between D and AD genomes

A total of 249 HGs were identified between Gr and Gh, including 22 dinucleotides, 170 trinucleotides, 15 tetranucleotides, 7 pentanucleotides and 35 hexanucleotides. All the HGs except two HGs had one EST of Gr versus 1-7 ESTs with an average of 1.5 of Gh. In Group I, 123 HGs (49.4%) were included; the 5 dinucleotides included 3 AGs, 1 AT and 1 AC. AAG, ATC, ACC and AGC were the most frequent motifs in the trinucleotides. Among the 12 dinucleotides in Group II, ten of them were AGs and the other two were ACs; the most frequent motif in the trinucleotides was AAG. In Group III, the dinucleotides included 2 ATs, 2 AGs and 1 AC, and the most frequent motif in the trinucleotides was still AAG (Table S4 and Figure 3b). Compared to the differences of repeated number between Ga and Gh, the differences of repeated number between Gr and Gh were also less than 2 in most HGs, but the range was wider (Table 3).

Homologous SSR-ESTs between A and D genomes

A total of 43 HGs were identified between Ga and Gr including 3 dinucleotides, 33 trinucleotides, 2 tetranucleotides, 3 pentanucleotides and 2 hexanucleotides. One EST of Ga had 1-4 ESTs with an average of 1.4 of Gr in these HGs. The most frequent trinucleotides were ACC and AAG in Group I, ATC in Group II, and only AAG in Group III, respectively. The difference of repeat number

Table 3. HGs with difference in repeat number in Ga-Gh and Gr-Gh.

Repeat number	Number of HGs in Ga-Gh		Number of HGs in Gr-Gh	
	Group II	Group III	Group II	Group III
0 ~ 1	20	64	47	41
1 ~ 2	3	24	10	13
2 ~ 3	2	2	4	1
3 ~ 4	6	0	3	2
5	0	0	1	0
6	0	1	1	0
9	0	0	0	3

between Ga and Gr were 1-2 (Table S5).

Homologous SSR-ESTs between A, D and AD genomes

A total of 31 HGs were identified between Ga, Gr and Gh, including 6 dinucleotides, 21 trinucleotides, 2 tetranucleotides and 2 hexanucleotides. The repeat numbers were identical in 6 HGs among the three genomes. Among the HGs with variance in repeat number, all the motifs of dinucleotides were AGs; AAG was the most frequent motif in the trinucleotides (Table S6).

Comparison of functional SSR-ESTs between A, D and AD genomes

Overview of the annotated SSR-ESTs

A total of 585, 937 and 1680 SSR-ESTs of Ga, Gr and Gh were annotated, respectively (level 2). All the SSR-ESTs of the three genomes targeted more ‘cellular process’, ‘metabolic process’, ‘cell’, ‘organelle’, ‘binding’ and ‘catalytic activity’ than other GO items. SSR-ESTs of Ga targeted more ‘cellular component organization’, ‘extracellular region’ and ‘structural molecule activity’ than the other two genomes. SSR-ESTs of Gr targeted more ‘biological regulation’, ‘cellular process’, ‘metabolic process’, ‘signaling’ ‘cell’, ‘organelle’ and ‘binding’ than the other two genomes. SSR-ESTs of Gh targeted less ‘cell’, ‘membrane-enclosed lumen’, ‘organelle’ and ‘catalytic activity’ than the other two genomes (Table S7 and Figure 4a).

Comparison of the annotated SSR-ESTs of di- and trinucleotides

SSR-ESTs of dinucleotides showed similar GO item distribution compared to all the SSR-ESTs with some exceptions: SSR-ESTs of Ga targeted more ‘cellular component organization’, ‘developmental process’, ‘multicellular organismal process’, ‘reproduction’ and less ‘metabolic process’, ‘organelle’, ‘binding’, ‘catalytic activity’, ‘structural molecule activity’, ‘transporter activity’; SSR-ESTs of Gr targeted less ‘growth’, ‘enzyme regular activity’; SSR-ESTs of Gh targeted less ‘response to stimulus’. However, SSR-ESTs of trinucleotides showed same GO item distribution as that of all the SSR-ESTs (Figure 4b-c).

Comparison of the annotated EST-SSRs with single motif

Dinucleotide motifs: for ESTs of AC motif, 5 GO items were absent in all three genomes; 5 GO items including 4 of ‘biological process’ and 1 of ‘molecular function’ were only present in Gh genome; ‘cellular component organization’ was only present and ‘signaling’, ‘molecular transducer activity’ absent in Ga genome.

For ESTs of AG motif, only ‘death’ was absent in all three genomes; ‘multi-organism process’ and ‘molecular transducer activity’ were also absent in Ga-ESTs, ‘structural mole-cule activity’ was absent in Gr-ESTs; ‘enzyme regulator activity’ was only present in Gh-ESTs. For ESTs of AT motif, ‘multi-organism process’ and ‘structural molecule activity’ were only present in Gr-ESTs; ‘death’ and ‘macro-molecular complex’ were only absent in Gh-ESTs; ‘trans-porter activity’ was also absent in Ga-ESTs (Figure S1).

Trinucleotide motifs: dramatic differences existed in the three genomes, especially for ACG, ACT, AGG and CCG motifs which were rare SSR-ESTs among the trinucleotide motifs. For example, GO items obviously abundant in Gr-ESTs and scarce in Ga-ESTs for ACG motif (Figure S2e). Details of these differences could be seen from Table S7 and Figure S1-2.

DISCUSSION

The frequency, distribution and abundance of SSRs can

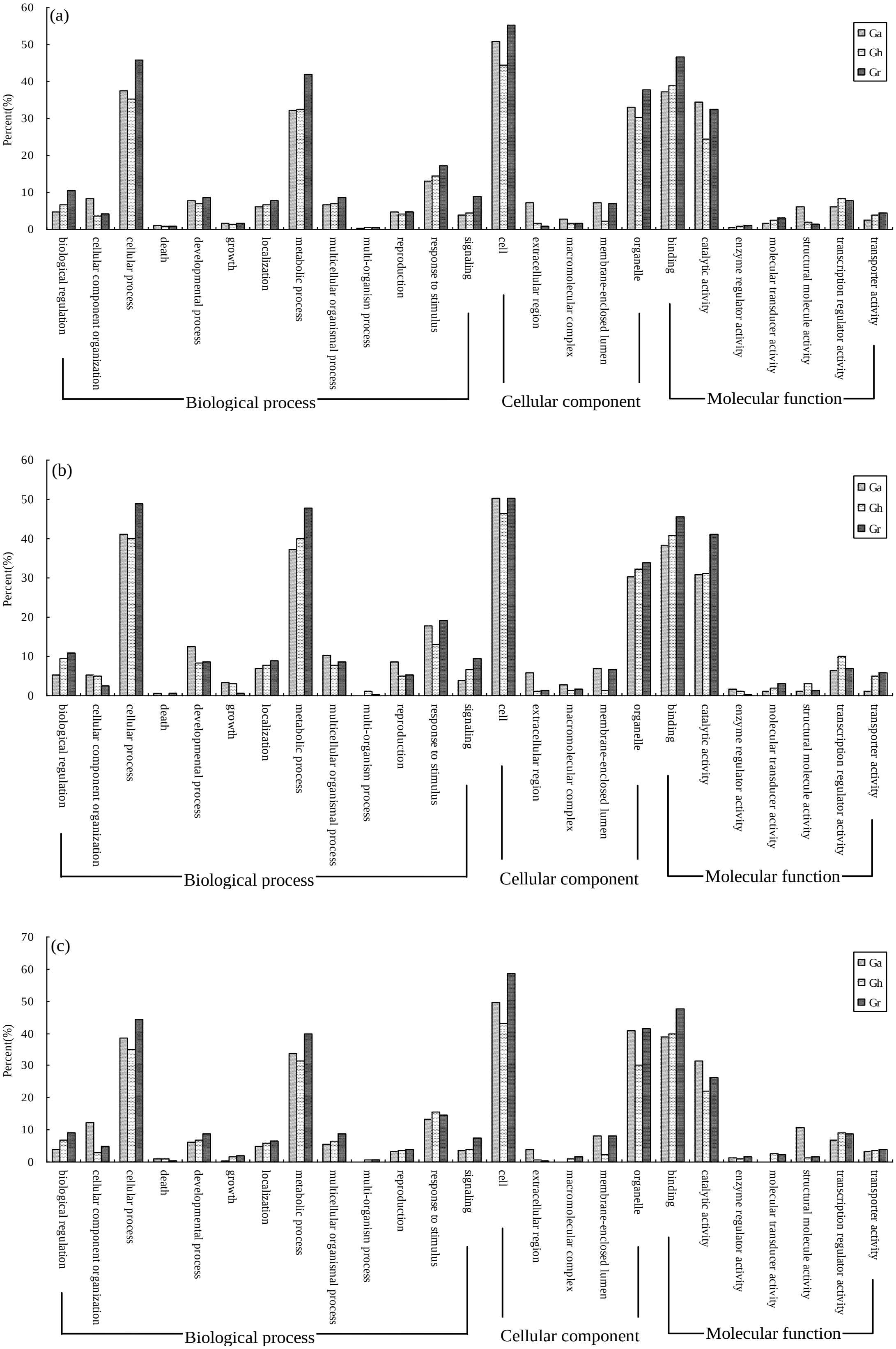


Figure 4. Function classifications of SSR-ESTs. ((a).all SSR-ESTs; (b).SSR-ESTs of Dinucleotides; (c).SSR-ESTs of Trinucleotides).

be highly variable depending on the SSR search criteria, the size of the dataset, and the database-mining tools (Varshney et al., 2005). In this study, the criteria described by Bérubé et al. (2007) were taken for identifying perfect SSR repeats, which resulted in a lower frequency compared to the study of Cardle et al. (2000).Although, more EST-SSRs were identified from Gh than Ga and Gr, the average distance (in kb) between SSRs was least in Gr but not in Gh, which indicated that less selection pressure was put on it. It is true for this species because *G. raimondii* (Gr) is a wild cotton species, while *G. arboreum* (Ga) is an Old World cultivated species and *G. hirsutum* (Gh) a New World cultivated species. The obviously high frequency of dinucleotides and the extremely high frequency of AT motif also supported this standpoint: the activity of dinucleotides could increase frame shift mutations; AT dimeric repeats were usually found in the untranslated regions of many species which are more polymorphic than other regions in the genome (Morgante et al., 2002; Jung et al., 2005).

Among all the repeat types, trinucleotide repeats were found to be the most abundant, which is in agreement with previous reports on several crop plants (Cardle et al., 2000; Varshney et al., 2002; Gao et al., 2003; Maia et al., 2009; Victoria et al., 2011) and some partly reports on cotton (Han et al., 2004; Han et al., 2006; Wang et al., 2006; Zhang et al., 2007). The predominance of trinucleotides in ESTs might reflect motifs of repetitive amino-acid sequences (Cardle et al., 2000), and could reduce frame shift mutations (Metzgar et al., 2000). Morgante et al. (2002) also suggested that mutation pressure and positive selection for specific single amino acid stretches are possible reasons for the high frequency of trinucleotide repeats.

In all the four dinucleotide repeats, AT and AG were most common, CG the lowest, and AC also had a low frequency. The rarity of GC/CG repeats seems to be common for all species; however, the most abundant motif in this study was not in accordance with some reports (Cardle et al., 2000; Kantety et al., 2002; Thiel et al., 2003; Jung et al., 2005; Chen et al., 2006). Interestingly, although the Gh derived from the combination of Ga and Gr, AT was very low in Gh among the three genomes which indicated that stronger selection was put on this species during its domestication. The more variance of repeat number in Gh also confirmed this point.

Among the trinucleotide repeats, AAG was the most frequent motifs in all three genomes, which was the same as some reports on some plants (Cardle et al., 2000; Varshney et al., 2002; Gao et al., 2003; Jung et al., 2005; Kumpatla and Mukhopadhyay, 2005; Aggarwal et al., 2007; Maia et al., 2009; Victoria et al., 2011). Conversely, CCG, the most common motif in monocotyledonous species (Chin et al., 1996; Temnykh et al., 2000; Cordeiro et al., 2001; Morgante et al., 2002; Varshney et al., 2002), was rare in cotton. It seemed that AAG might be the most abundant EST-SSR motif in dicots. Additionally, ATC was also abundant, which was similar to *Arabidopsis* (Cardle et al., 2000). The similar results between cotton and *Arabidopsis* were rooted from that the Malvaceae and Brassicaceae have diverged from a common ancestor less than 100 million years ago (Bowers et al., 2003), a high level of DNA sequence conservation is expected within protein coding regions (Fulton et al., 2002). However, AAT, ACC and AGC also had a high frequency, but least often in *Arabidopsis* (Cardle et al., 2000). The frequencies of different trinucleotide motifs were different in the three cotton genomes, and no motifs was most frequent in Ga, which is consistent with the fact that this species has been cultivated compared to Gr, but less domesticated compared to Gh.

In the three cotton genomes, the SSR frequency decreased with the increase of repeat number, which had been verified by some other reports (Kumpatla and Mukhopadhyay, 2005; Bérubé et al., 2007). The most frequent repeat number was also similar to the results of Kumpatla and Mukhopadhyay, (2005). However, the range of repeat number was wider than that in pine and spruce (Bérubé et al., 2007). Further survey confirmed that the variance was mainly caused by AG motif of dinucleotide, and the most frequent repeat number by AAG and ATC motif of trinucleotide.

Blast analysis revealed more homologue groups (HGs) in Gr-Gh than in Ga-Gh; however, more subject sequences were in Ga-Gh. Previous studies have proved that Ga is more closed to Gh than Gr (Wendel and Albert, 1992; Zhang et al., 2007; Zhu et al., 2009). Although, more HGs were identified in Gr-Gh, the more subject sequences in HGs of Ga-Gh and the more conserved repeat number change in HGs were the stronger evidence for their relationship. Among all the identified HGs, AG and AAG had a high frequency in all kinds of HGs. These results obtained from this study indicated that AG and AAG motif was an important characteristic of cotton.

The annotated SSR-ESTs showed some differences between the three genomes. The more GO items targeted by SSR-ESTs of Ga and Gr, and the less GO items targeted by SSR-ESTs of Gh indicated that some genes lost their function or were deleted during the polyploidization and domestication. However, the results of single motifs showed that some GO items were only found in SSR-ESTs of Gh, which indicated that non-functional genes in Ga or Gr gained new function in Gh. Because the mutation rate of SSRs is very high which results in the length changes of alleles (Wierdl et al., 1997; Xu et al., 2000), it is possible that some genes gain or loss functions because of the frame shift mutation. The differences of GO items between Ga and Gr could be a reflection of difference between the diploid genomes.

EST-derived SSR has become a novel way for developing SSRs in many plants, and they have been widely applied in germplasm evaluation, comparative mapping, tagging important traits of interest, and additional map-based cloning of important genes. Some of these EST-SSRs identified in this study have been mapped in our interspecific linkage map derived from *G. hirsutum* and *G. barbadense* (Zhang et al., 2008), and the others are under mapping, which would be helpful in fatherly exploring the mysteries of polyploidy cotton.

ACKNOWLEDGEMENT

This work was financially supported by National Basic Research Plan (2011CB109303) and the National Science Foundation of China (30871559).

REFERENCES

Aggarwal R, Hendre P, Varshney R, Bhat P, Krishnakumar V, Singh L (2007). Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. Theor. Appl. Genet. 114(2):359-372.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol. 215(3):403-410.

Bérubé Y, Zhuang J, Rungis D, Ralph S, Bohlmann J, Ritland K (2007). Characterization of EST-SSRs in loblolly pine and spruce. Tree Gen. Genomes 3(3):251-259.

Bowers JE, Chapman BA, Rong J, Paterson AH (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422(6930):433-438.

Broughton W, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003). Beans (*Phaseolus* spp.) Cmodel food legumes. Plant Soil. 252(1):55-128.

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000). Computational and Experimental Characterization of Physically Clustered Simple Sequence Repeats in Plants. Genetics 156(2):847-854.

Chen C, Zhou P, Choi Y, Huang S, Gmitter F (2006). Mining and characterizing microsatellites from citrus ESTs. Theor. Appl. Genet. 112(7):1248-1257.

Childs KL, Hamilton JP, Zhu W, Ly E, Cheung F, Wu H, Rabinowicz PD, Town CD, Buell CR, Chan AP (2006). The TIGR Plant Transcript Assemblies database. Nucleic Acids Res. 35(suppl 1):D846-D851.

Chin EC, Senior ML, Shu H, Smith JS (1996). Maize simple repetitive DNA sequences: abundance and allele variation. Genome 39(5):866-873.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674-3676.

Cordeiro GM, Casu R, McIntyre CL, Manners JM, Henry RJ (2001). Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to erianthus and sorghum. Plant Sci. 160(6):1115-1123.

Cronn RC, Wendel JF (1998). Simple methods for isolating homoeologous loci from allopolyploid genomes. Genome 41(6):756-762.

Endrizzi J, Turcotte E, Kohel R (1985). Genetics, cytology, and evolution of *Gossypium*. Adv. Genet. 23:271-375.

Field D, Wills C (1996). Long, Polymorphic Microsatellites in Simple Organisms. Proc. Royal Acad. London B. 263(1367):209-215.

Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD (2002). Identification, Analysis, and Utilization of Conserved Ortholog Set Markers for Comparative Genomics in Higher Plants. Plant Cell. 14(7): 1457-1467.

Gao L, Tang J, Li H, Jia J (2003). Analysis of microsatellites in major crops assessed by computational and experimental approaches. Mol. Breed 12(3):245-261.

Guo W, Cai C, Wang C, Han Z, Song X, Wang K, Niu X, Wang C, Lu K, Shi B, Zhang T (2007). A microsatellite-based, gene-rich linkage map reveals genome structure, function, and evolution in *Gossypium*. Genetics 176(1):527-541.

Han Z, Wang C, Song X, Guo W, Gou J, Li C, Chen X, Zhang T (2006). Characteristics, development and mapping of *Gossypium hirsutum* derived EST-SSRs in allotetraploid cotton. Theor. Appl. Genet. 112(3): 430-439.

Han ZG, Guo WZ, Song XL, Zhang TZ (2004). Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Mol. Genet. Genomics 272(3):308-327.

Hedrick PW (2001). Conservation genetics: where are we now? Trends Ecol. Evol. 16(11):629-636.

Jung S, Abbott A, Jesudurai C, Tomkins J, Main D (2005). Frequency, type, distribution and annotation of simple sequence repeats in Rosaceae ESTs. Funct. Integr. Genomics 5(3):136-143.

Kantety RV, La Rota M, Matthews DE, Sorrells ME (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol. Biol. 48(5):501-510.

Kumpatla SP, Mukhopadhyay S (2005). Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. Genome 48(6): 85-998.

Levinson G, Gutman GA (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4(3):203-221.

Li Y, Korol AB, Fahima T, Nevo E (2004). Microsatellites Within Genes: Structure, Function and Evolution. Mol. Biol. Evol. 21(6):991-1007.

Maia L, Souza V, Kopp M, Carvalho F, Oliveira A (2009). Tandem repeat distribution of gene transcripts in three plant families. Genet. Mol. Biol. 32(4):822-833.

Metzgar D, Bytof J, Wills C (2000). Selection Against Frameshift Mutations Limits Microsatellite Expansion in Coding DNA. Genome Res. 10(1):72-80.

Morgante M, Hanafey M, Powell W (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat. Genet. 30(2):194-200.

Mulder N, Apweiler R (2007). InterPro and InterProScan: tools for protein sequence classification and comparison. Methods Mol. Biol. 396:59-70.

Oliveira E, P¨¢dua J, Zucchi M, Vencovsky R, Vieira M (2006). Origin, evolution and genome distribution of microsatellites. Genet. Mol. Biol. 29:294-307.

Park Y, Alabady M, Ulloa M, Sickler B, Wilkins T, Yu J, Stelly D, Kohel R, El-Shihy O, Cantrell R (2005). Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol. Genet. Genomics 274(4):428-441.

Percival AE, Wendel JF, Stewart JM (1999). Cotton: origin, history, technology, and production. In: Smith CW and Cothren JT (ed) Taxonomy and gemplasm resources. John Wiley & Sons, Inc, New York, pp. 33-63.

Powell W, Machray GC, Provan J (1996). Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1(7):215-222.

Qureshi S, Saha S, Kantety R, Jenkins J (2004). EST-SSR: a new class of genetic markers in cotton. J. Cotton Sci. 8:112-123.

Tóth G, Gáspári Z, Jurka J (2000). Microsatellites in Different Eukaryotic Genomes: Survey and Analysis. Genome Res. 10(7):967-981.

Talierco E, Allen R, Essenberg M, Klueva N, Nguyen H, Patil M, Payton P, Millena A, Phillips A, Pierce M (2006). Analysis of ESTs from multiple *Gossypium hirsutum* tissues and identification of SSRs. Genome, 49(4):306-319.

Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR (2000). Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). Theor. Appl. Genet. 100(5):697-712.

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001). Computational and Experimental Analysis of Microsatellites in Rice (*Oryza sativa* L.): Frequency, Length Variation, Transposon Associations, and Genetic Marker Potential. Genome Res. 11(8):1441-1452.

Thiel T, Michalek W, Varshney RK, 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(3):411-422.

Varshney R, Thiel T, Stein N, Langridge P, Graner A (2002). *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cell Mol. Biol. Lett. 7(2A):537-546.

Varshney RK, Graner A, Sorrells ME (2005). Genic microsatellite markers in plants: features and applications. Trends Biotechnol. 23(1): 48-55.

Victoria F, Maia L, Costa de Oliveira A (2011). *In silico* comparative analysis of SSR markers in plants. BMC Plant Biol. 11(1):15.

Wang C, Guo W, Cai C, Zhang T (2006). Characterization, development and exploitation of EST-derived microsatellites in *Gossypium raimondii* Ulbrich. Chin Sci Bull. 51(5):557-561.

Wendel J, Albert V (1992). Phylogenetics of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Syst. Bot. 17(1):115-143.

Wendel JF, Cronn RC (2003). Polyploidy and the evolutionary history of cotton. Adv. Agron. 78:139-186.

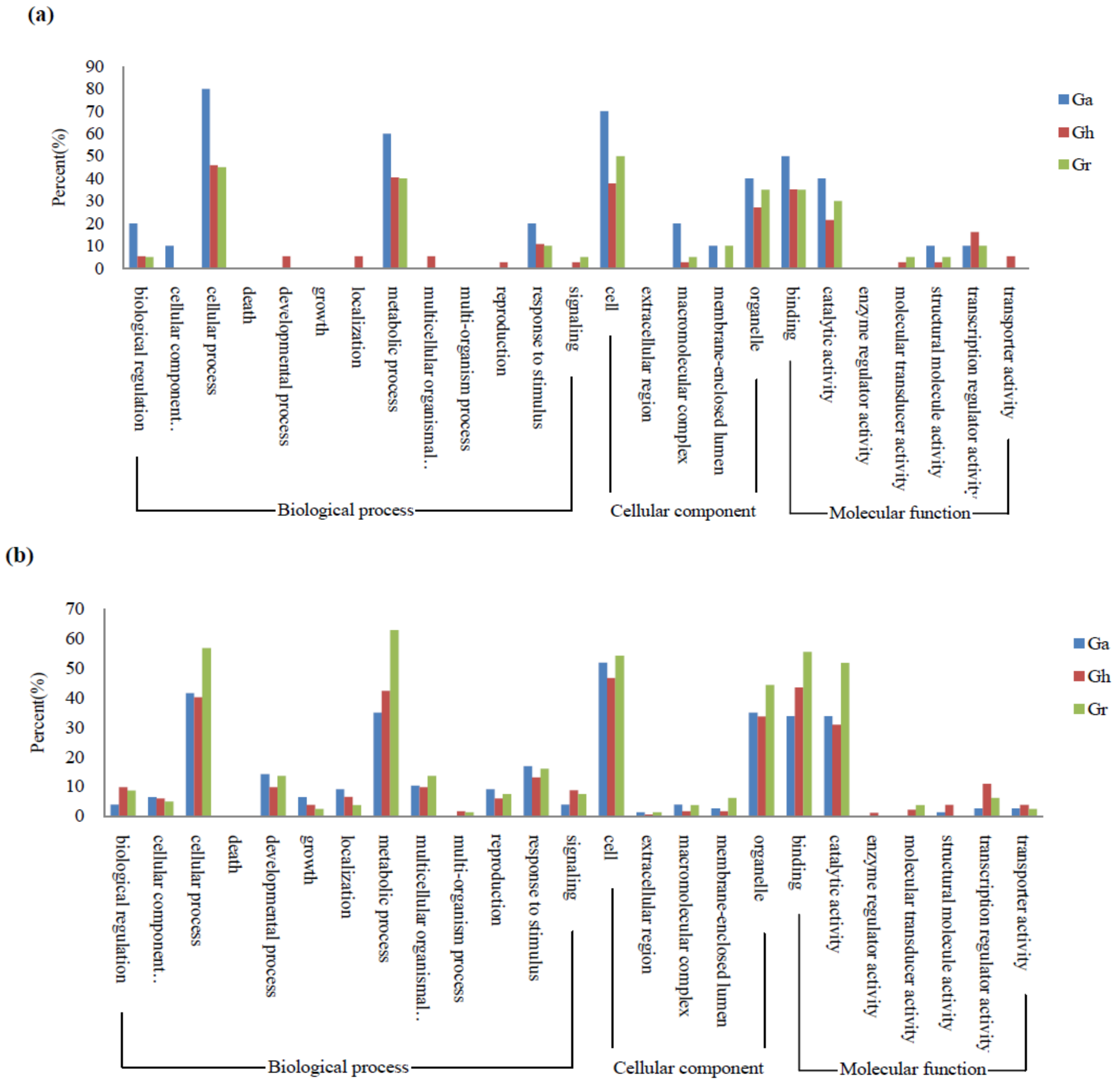
Wierdl M, Dominska M, Petes TD (1997). Microsatellite Instability in Yeast: Dependence on the Length of the Microsatellite. Genetics, 146(3):769-779.

Xu X, Peng M, Fang Z (2000). The direction of microsatellite mutations is dependent upon allele length. Nat. Genet. 24(4):396-399.

Zhang Y, Lin Z, Li W, Tu L, Nie Y, Zhang X (2007). Studies of new EST-SSRs derived from *Gossypium barbadense*. Chin. Sci. Bull. 52(18): 2522-2531.

Zhang Y, Lin Z, Xia Q, Zhang M, Zhang X (2008). Characteristics and analysis of simple sequence repeats in the cotton genome based on a linkage map constructed from a BC_1 population between *Gossypium hirsutum* and *G. barbadense*. Genome 51(7):534-546.

Zhu H, Zhang T, Yang L, Guo W (2009). EST-SSR sequences revealed the relationship of D-genome in diploid and tetraploid Species in *Gossypium*. Plant Sci. 176(3):397-405.


(a)
Percent(%)
90
80
70
60
50
40
30
20
10
0
Ga
Gh
Gr
biological regulation
cellular component..
cellular process
death
developmental process
growth
localization
metabolic process
multicellular organismal..
multi-organism process
reproduction
response to stimulus
signaling
cell
extracellular region
macromolecular complex
membrane-enclosed lumen
organelle
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
structural molecule activity
transcription regulator activity
transporter activity
Biological process
Cellular component
Molecular function
(b)
Percent(%)
70
60
50
40
30
20
10
0
Ga
Gh
Gr
biological regulation
cellular component..
cellular process
death
developmental process
growth
localization
metabolic process
multicellular organismal..
multi-organism process
reproduction
response to stimulus
signaling
cell
extracellular region
macromolecular complex
membrane-enclosed lumen
organelle
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
structural molecule activity
transcription regulator activity
transporter activity
Biological process
Cellular component
Molecular function

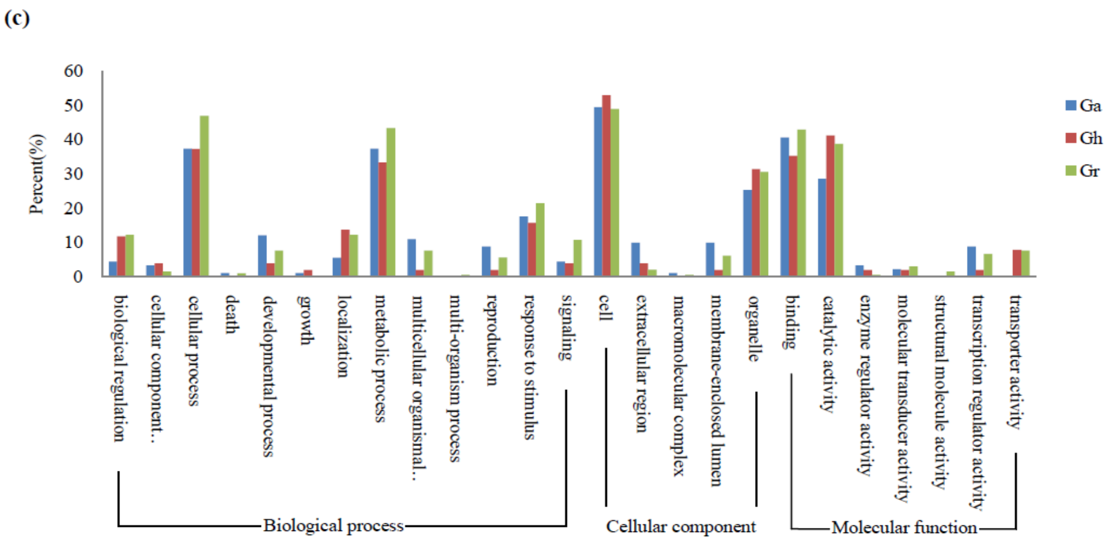
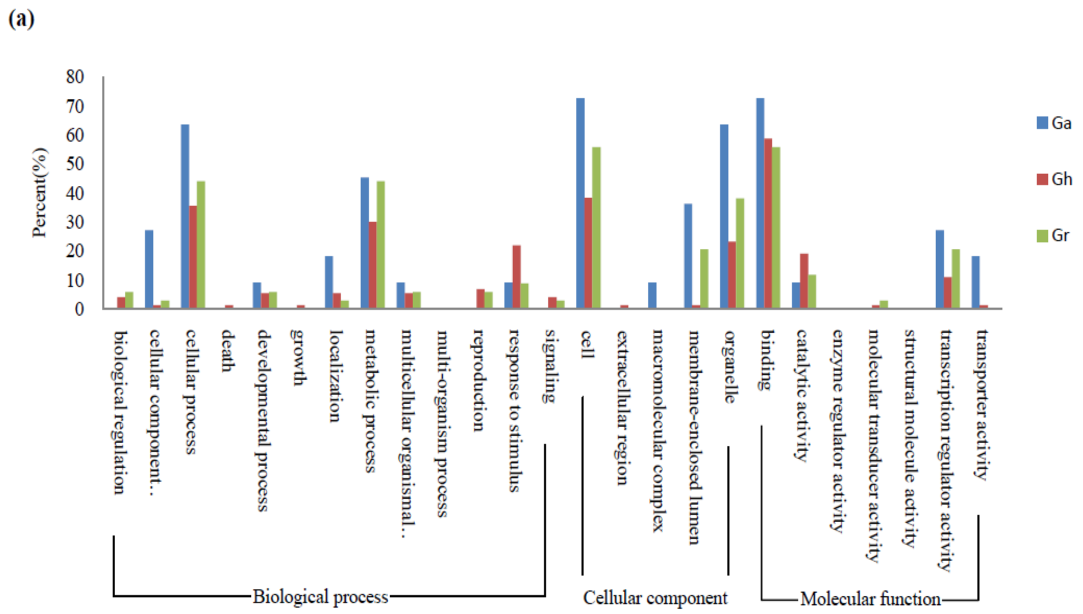
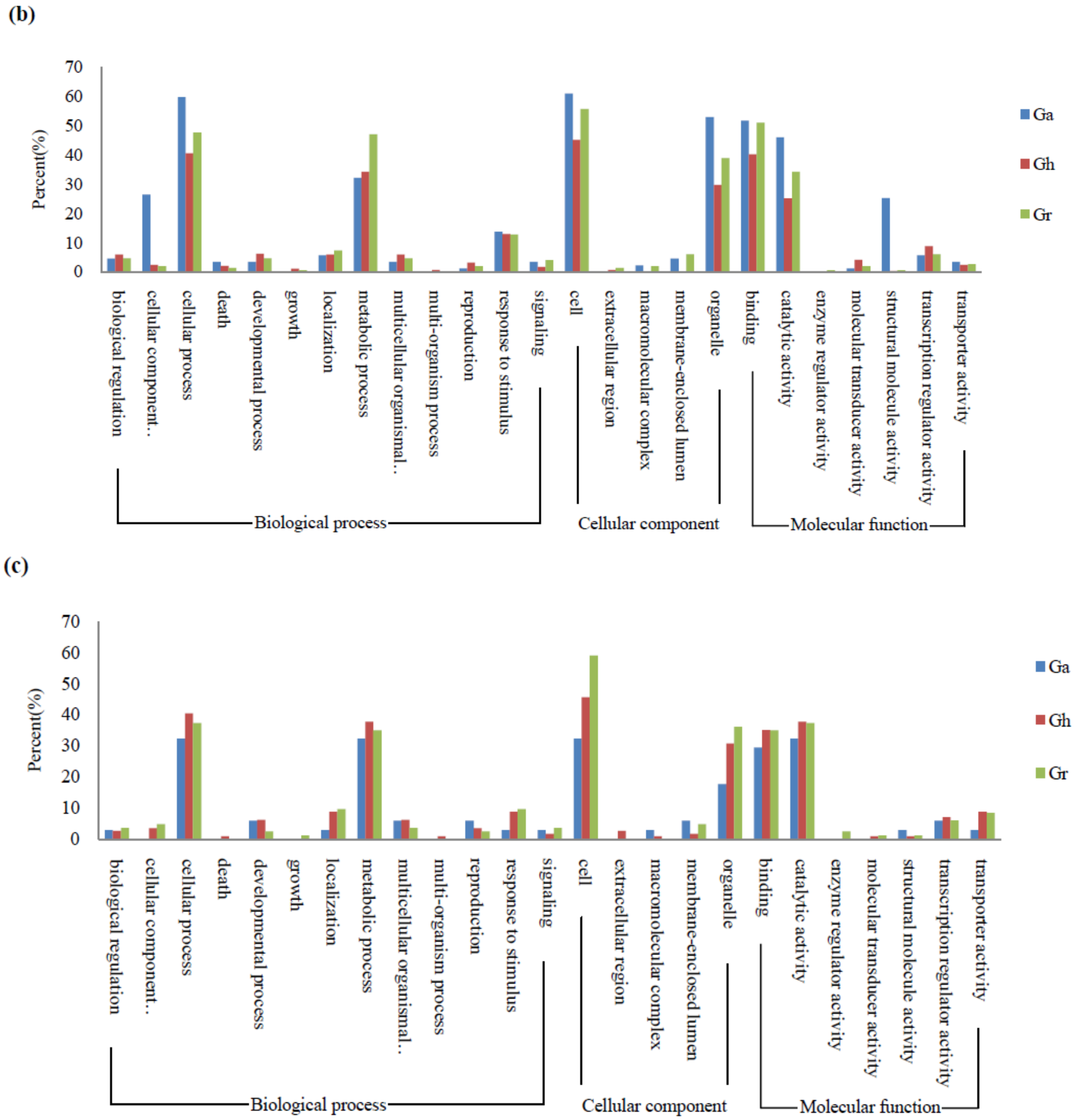
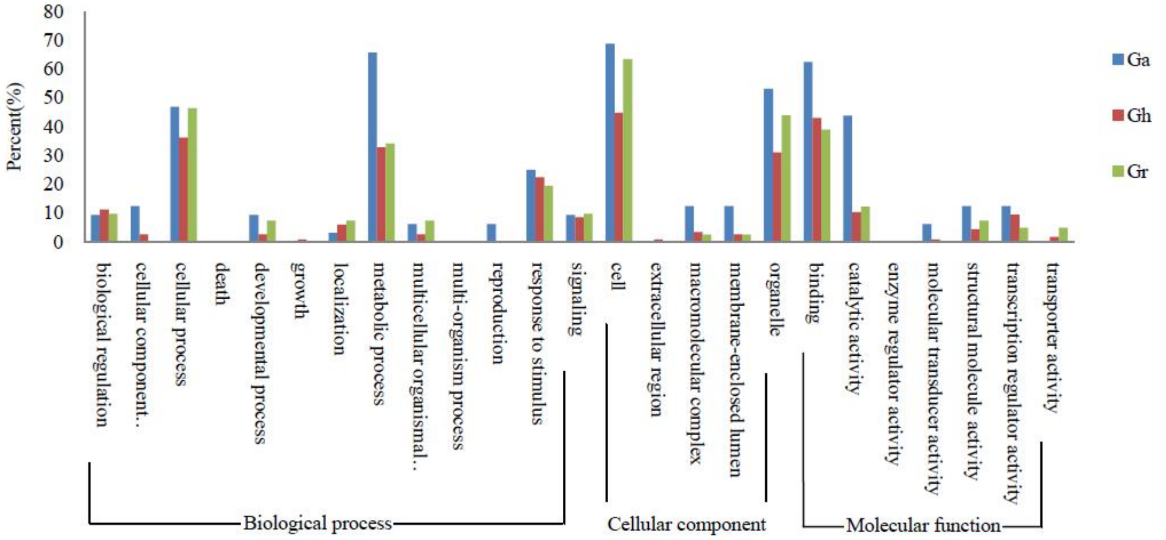


Figure S1. Function classifications of dinucleotides motif SSR-ESTs. (a) AC motif; (b) AG motif; (c) AT motif.

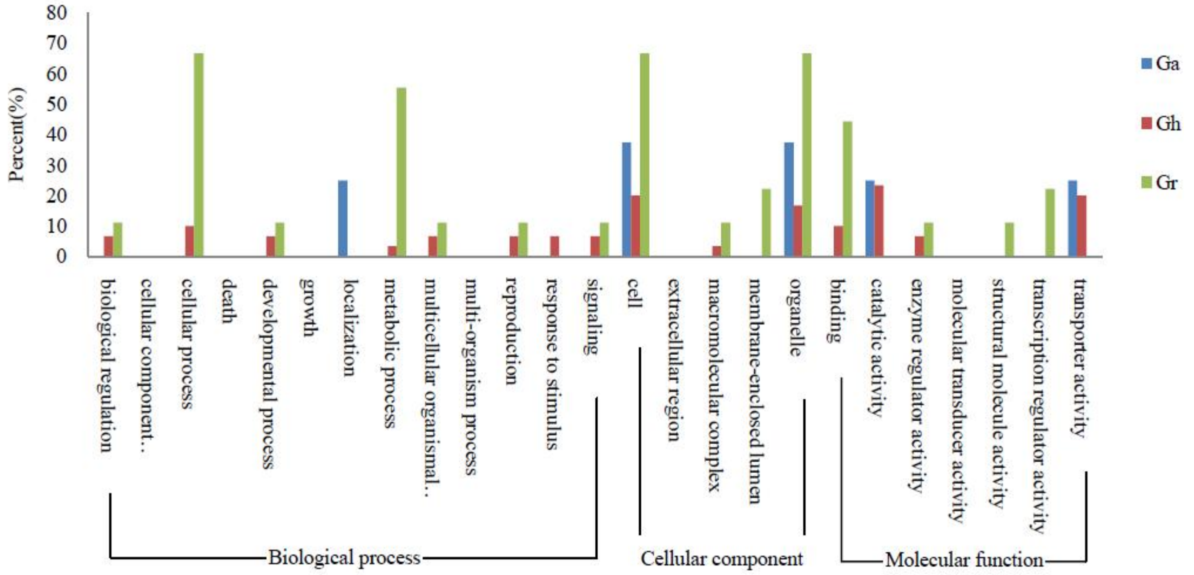



(b)
Percent(%)
70
60
50
40
30
20
10
0
Ga
Gh
Gr
biological regulation
cellular component..
cellular process
death
developmental process
growth
localization
metabolic process
multicellular organismal..
multi-organism process
reproduction
response to stimulus
signaling
cell
extracellular region
macromolecular complex
membrane-enclosed lumen
organelle
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
structural molecule activity
transcription regulator activity
transporter activity
Biological process
Cellular component
Molecular function
(c)
Percent(%)
70
60
50
40
30
20
10
0
Ga
Gh
Gr
biological regulation
cellular component..
cellular process
death
developmental process
growth
localization
metabolic process
multicellular organismal..
multi-organism process
reproduction
response to stimulus
signaling
cell
extracellular region
macromolecular complex
membrane-enclosed lumen
organelle
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
structural molecule activity
transcription regulator activity
transporter activity
Biological process
Cellular component
Molecular function

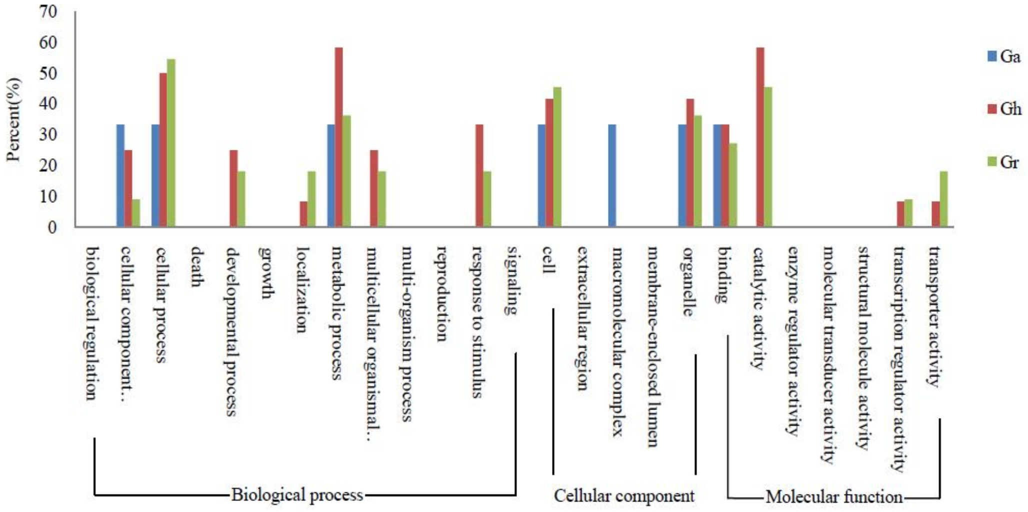
(d)


Percent(%)
80
70
60
50
40
30
20
10
0
Ga
Gh
Gr
biological regulation
cellular component..
cellular process
death
developmental process
growth
localization
metabolic process
multicellular organismal..
multi-organism process
reproduction
response to stimulus
signaling
cell
extracellular region
macromolecular complex
membrane-enclosed lumen
organelle
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
structural molecule activity
transcription regulator activity
transporter activity
Biological process
Cellular component
Molecular function

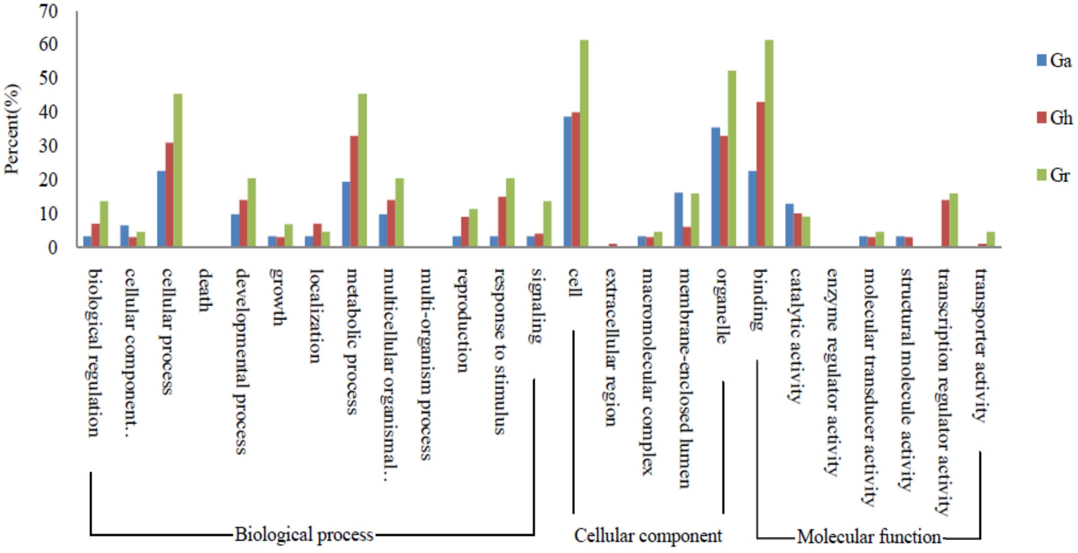
(e)


Percent(%)
80
70
60
50
40
30
20
10
0
Ga
Gh
Gr
biological regulation
cellular component..
cellular process
death
developmental process
growth
localization
metabolic process
multicellular organismal..
multi-organism process
reproduction
response to stimulus
signaling
cell
extracellular region
macromolecular complex
membrane-enclosed lumen
organelle
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
structural molecule activity
transcription regulator activity
transporter activity
Biological process
Cellular component
Molecular function

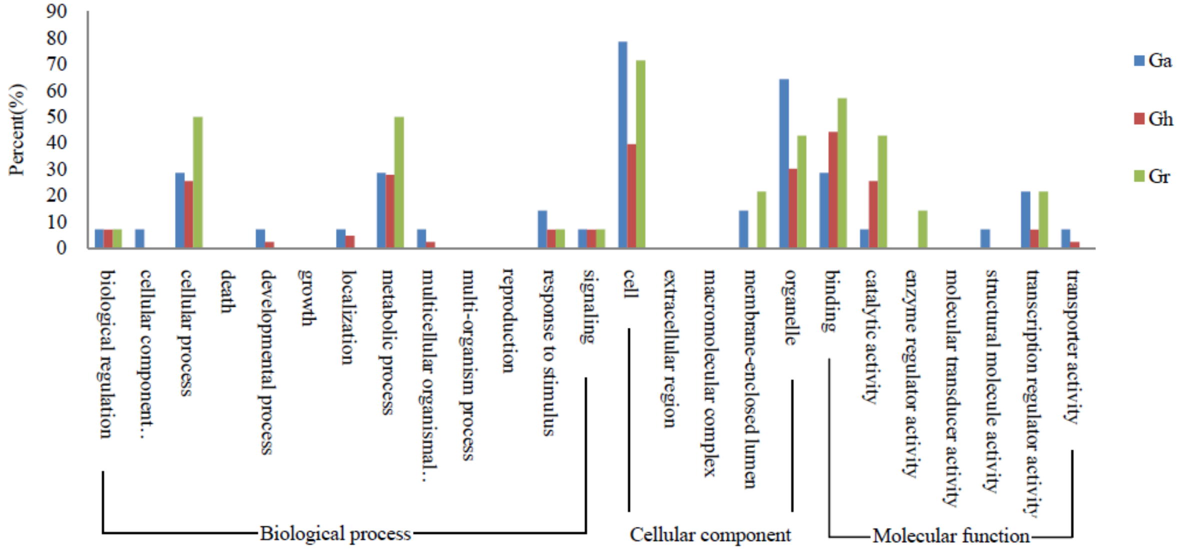
(f)


Percent(%)
70
60
50
40
30
20
10
0
Ga
Gh
Gr
biological regulation
cellular component..
cellular process
death
developmental process
growth
localization
metabolic process
multicellular organismal..
multi-organism process
reproduction
response to stimulus
signaling
cell
extracellular region
macromolecular complex
membrane-enclosed lumen
organelle
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
structural molecule activity
transcription regulator activity
transporter activity
Biological process
Cellular component
Molecular function

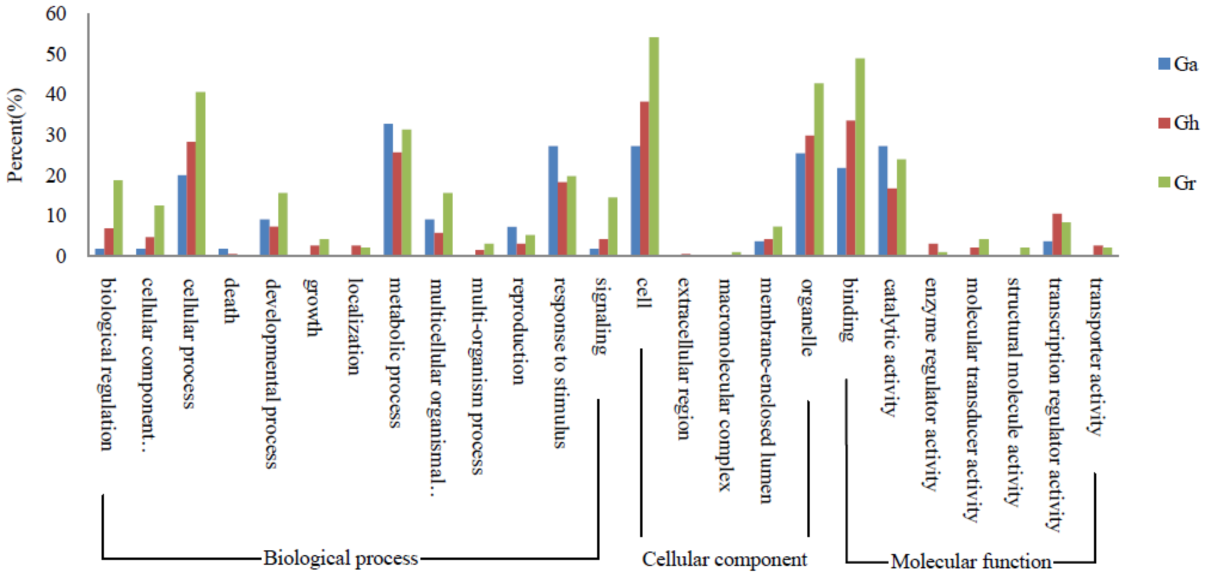
(g)


Percent(%)
70
60
50
40
30
20
10
0
Ga
Gh
Gr
biological regulation
cellular component..
cellular process
death
developmental process
growth
localization
metabolic process
multicellular organismal..
multi-organism process
reproduction
response to stimulus
signaling
cell
extracellular region
macromolecular complex
membrane-enclosed lumen
organelle
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
structural molecule activity
transcription regulator activity
transporter activity
Biological process
Cellular component
Molecular function

(h)


Percent(%)
90
80
70
60
50
40
30
20
10
0
Ga
Gh
Gr
biological regulation
cellular component..
cellular process
death
developmental process
growth
localization
metabolic process
multicellular organismal..
multi-organism process
reproduction
response to stimulus
signaling
cell
extracellular region
macromolecular complex
membrane-enclosed lumen
organelle
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
structural molecule activity
transcription regulator activity
transporter activity
Biological process
Cellular component
Molecular function

(i)


Percent(%)
60
50
40
30
20
10
0
Ga
Gh
Gr
biological regulation
cellular component..
cellular process
death
developmental process
growth
localization
metabolic process
multicellular organismal..
multi-organism process
reproduction
response to stimulus
signaling
cell
extracellular region
macromolecular complex
membrane-enclosed lumen
organelle
binding
catalytic activity
enzyme regulator activity
molecular transducer activity
structural molecule activity
transcription regulator activity
transporter activity
Biological process
Cellular component
Molecular function

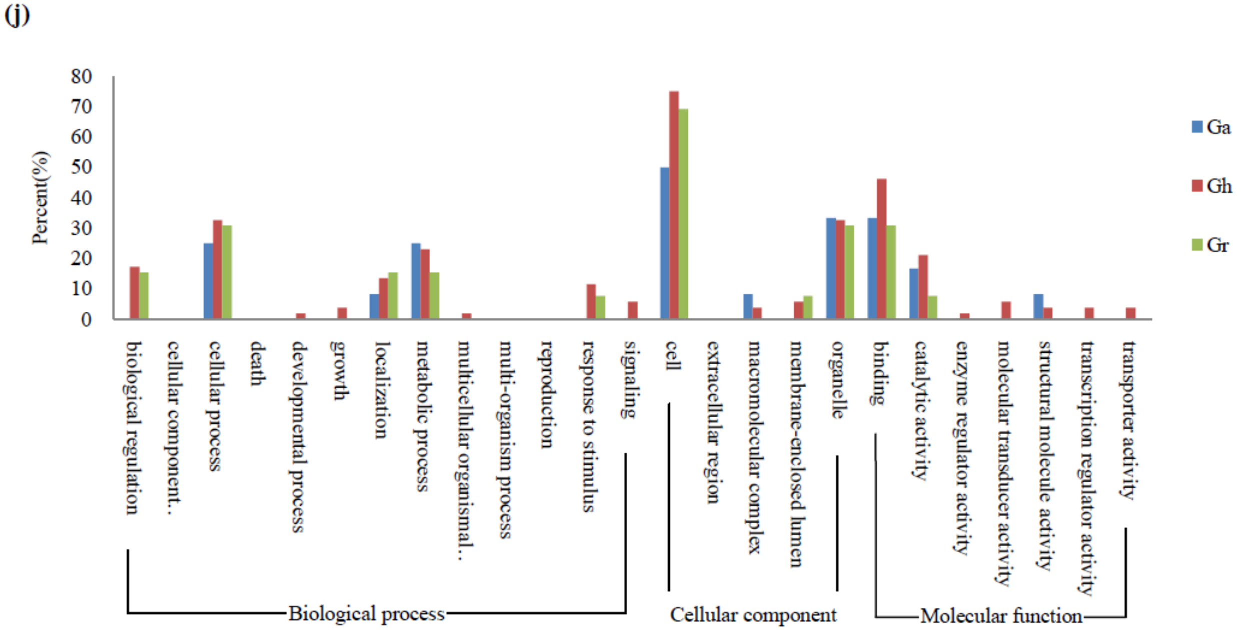


Figure S2. Function classifications of trinucleotides motif SSR-ESTs. (a) AAC motif; (b) AAG motif; (c) AAT motif; (d) ACC motif; (e) ACG motif; (f) ACT motif; (g) AGC motif; (h) AGG motif; (i) ATC motif; (j)CCG motif.

Supplementary Table 1.

Sequence ID	Period Size	Motif	Repeate Start	Repeate End	Repeate Length	Copy Number	Repeat
TA6694_29729	2	AC	144	172	28	14	ACACACACACACACACACACACACACACA
TA7789_29729	2	AC	122	139	18	9	ACACACACACACACACAC
TA8986_29729	2	AC	297	317	20	10	ACACACACACACACACACACA
BG441478	2	AC	53	73	20	10	ACACACACACACACACACACA
BG446513	2	AC	260	277	18	9	ACACACACACACACACAC
BQ404793	2	AC	11	31	20	10	ACACACACACACACACACACA
TA5554_29729	2	AG	1	19	18	9	AGAGAGAGAGAGAGAGAGA
TA6910_29729	2	AG	473	491	18	9	AGAGAGAGAGAGAGAGAGA
TA7767_29729	2	AG	621	638	18	9	AGAGAGAGAGAGAGAGAG
BE055792	2	AG	1	38	38	19	AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
BF269115	2	AG	28	95	68	34	AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
BG440058	2	AG	1	43	42	21	AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
BG442218	2	AG	2	30	28	14	AGAGAGAGAGAGAGAGAGAGAGAGAGAGA
BG444906	2	AG	2	54	52	26	AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
BG445882	2	AG	1	27	26	13	AGAGAGAGAGAGAGAGAGAGAGAGAGA
BM360136	2	AG	6	73	68	34	AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
BQ403493	2	AG	47	65	18	9	AGAGAGAGAGAGAGAGAGA
BQ404091	2	AG	49	67	18	9	AGAGAGAGAGAGAGAGAGA
BQ407993	2	AG	302	320	18	9	AGAGAGAGAGAGAGAGAGA
BQ409027	2	AG	10	55	46	23	AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
BQ410390	2	AG	17	35	18	9	AGAGAGAGAGAGAGAGAGA
TA4981_29729	2	AT	36	57	22	11	ATATATATATATATATATATAT
TA4988_29729	2	AT	1159	1187	28	14	ATATATATATATATATATATATATATATA
TA5179_29729	2	AT	570	591	22	11	ATATATATATATATATATATAT
TA6091_29729	2	AT	40	67	28	14	ATATATATATATATATATATATATATAT
TA6183_29729	2	AT	565	591	26	13	ATATATATATATATATATATATATATA
TA6310_29729	2	AT	744	781	38	19	ATATATATATATATATATATATATATATATATATATAT
TA7299_29729	2	AT	34	51	18	9	ATATATATATATATATAT
TA7596_29729	2	AT	1116	1135	20	10	ATATATATATATATATATAT
TA7860_29729	2	AT	377	399	22	11	ATATATATATATATATATATATA
TA8909_29729	2	AT	296	322	26	13	ATATATATATATATATATATATATATA
TA9389_29729	2	AT	682	711	30	15	ATATATATATATATATATATATATATATAT
AW730564	2	AT	877	896	20	10	ATATATATATATATATATAT
BE052757	2	AT	444	463	20	10	ATATATATATATATATATAT
BE055780	2	AT	467	495	28	14	ATATATATATATATATATATATATATATA
BF270284	2	AT	68	85	18	9	ATATATATATATATATAT
BF272647	2	AT	190	212	22	11	ATATATATATATATATATATATA
BF273977	2	AT	301	326	26	13	ATATATATATATATATATATATATAT
BF275697	2	AT	169	192	24	12	ATATATATATATATATATATATAT
BF276023	2	AT	681	701	20	10	ATATATATATATATATATATA
BG441793	2	AT	493	517	24	12	ATATATATATATATATATATATATA

Supplementary Table 1. Contd.

BG444976	2	AT	614	635	22	11	ATATATATATATATATATATAT
BG446948	2	AT	322	339	18	9	ATATATATATATATATAT
BM358285	2	AT	125	147	22	11	ATATATATATATATATATATATA
BM359332	2	AT	566	584	18	9	ATATATATATATATATATA
BM360635	2	AT	359	378	20	10	ATATATATATATATATATAT
BQ407035	2	AT	468	496	28	14	ATATATATATATATATATATATATATATA
BQ408652	2	AT	192	211	20	10	ATATATATATATATATATAT
BQ408794	2	AT	348	369	22	11	ATATATATATATATATATATAT
BQ408877	2	AT	297	320	24	12	ATATATATATATATATATATATAT
BQ409420	2	AT	199	233	34	17	ATATATATATATATATATATATATATATATATATA
BQ409544	2	AT	449	466	18	9	ATATATATATATATATAT
BQ409727	2	AT	382	404	22	11	ATATATATATATATATATATATA
BQ409934	2	AT	227	280	54	27	ATATATATATATATATATATATATATATATATATATATATATATATATATATAT
BQ410502	2	AT	23	45	22	11	ATATATATATATATATATATATA
BQ410638	2	AT	584	601	18	9	ATATATATATATATATAT
BQ411575	2	AT	20	38	18	9	ATATATATATATATATATA
BQ412548	2	AT	598	625	28	14	ATATATATATATATATATATATATATAT
BQ413176	2	AT	88	105	18	9	ATATATATATATATATAT
BQ414437	2	AT	7	27	20	10	ATATATATATATATATATATA
BQ414511	2	AT	201	230	30	15	ATATATATATATATATATATATATATATAT
BQ415304	2	AT	384	403	20	10	ATATATATATATATATATAT
BQ415702	2	AT	440	460	20	10	ATATATATATATATATATATA
BQ407641	2	CA	202	220	18	9	CACACACACACACACACAC
AW727558	2	CG	844	861	18	9	CGCGCGCGCGCGCGCGCG
TA5890_29729	2	CT	31	60	30	15	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
TA5977_29729	2	CT	143	163	20	10	CTCTCTCTCTCTCTCTCTCTC
TA5984_29729	2	CT	63	81	18	9	CTCTCTCTCTCTCTCTCTC
TA7789_29729	2	CT	95	121	26	13	CTCTCTCTCTCTCTCTCTCTCTCTCTC
BE052113	2	CT	1	27	26	13	CTCTCTCTCTCTCTCTCTCTCTCTCTC
BF272268	2	CT	21	50	30	15	CTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
BF277085	2	CT	1	18	18	9	CTCTCTCTCTCTCTCTCT
BF277357	2	CT	1	22	22	11	CTCTCTCTCTCTCTCTCTCTCT
BG440757	2	CT	35	60	26	13	CTCTCTCTCTCTCTCTCTCTCTCTCT
BG444045	2	CT	154	174	20	10	CTCTCTCTCTCTCTCTCTCTC
BG447298	2	CT	52	75	24	12	CTCTCTCTCTCTCTCTCTCTCTCT
BG447299	2	CT	51	74	24	12	CTCTCTCTCTCTCTCTCTCTCTCT
BQ404264	2	CT	139	156	18	9	CTCTCTCTCTCTCTCTCT

Supplementary Table 1. Contd.

TA4881_29729	2	GA	1198	1217	20	10	GAGAGAGAGAGAGAGAGAGA
BE052657	2	GA	1	21	20	10	GAGAGAGAGAGAGAGAGAGAG
BE053217	2	GA	9	64	56	28	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
BF272810	2	GA	209	251	42	21	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
BF274512	2	GA	1	59	58	29	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
BF276023	2	GA	702	719	18	9	GAGAGAGAGAGAGAGAGA
BF276490	2	GA	132	168	36	18	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
BG440327	2	GA	76	93	18	9	GAGAGAGAGAGAGAGAGA
BG444071	2	GA	1	83	82	41	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
BG444077	2	GA	1	96	96	48	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
BG445402	2	GA	1	65	64	32	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
BG446518	2	GA	295	312	18	9	GAGAGAGAGAGAGAGAGA
BM358395	2	GA	1	29	28	14	GAGAGAGAGAGAGAGAGAGAGAGAGAGAG
BM358617	2	GA	1	39	38	19	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
BM360126	2	GA	8	45	38	19	GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
BQ405858	2	GA	565	582	18	9	GAGAGAGAGAGAGAGAGA
BQ407002	2	GA	3	28	26	13	GAGAGAGAGAGAGAGAGAGAGAGAGA
AW727558	2	GC	1040	1058	18	9	GCGCGCGCGCGCGCGCGCG
TA6582_29729	2	GT	398	416	18	9	GTGTGTGTGTGTGTGTGTG
TA5092_29729	2	TA	340	358	18	9	TATATATATATATATATAT
TA5674_29729	2	TA	671	693	22	11	TATATATATATATATATATATAT
TA5883_29729	2	TA	628	648	20	10	TATATATATATATATATATAT
TA5896_29729	2	TA	628	645	18	9	TATATATATATATATATA
TA6055_29729	2	TA	551	568	18	9	TATATATATATATATATA
TA6309_29729	2	TA	300	328	28	14	TATATATATATATATATATATATATATAT
TA6558_29729	2	TA	1025	1046	22	11	TATATATATATATATATATATA
TA6559_29729	2	TA	429	450	22	11	TATATATATATATATATATATA
TA6826_29729	2	TA	541	559	18	9	TATATATATATATATATAT
TA7329_29729	2	TA	764	806	42	21	TATATATATATATATATATATATATATATATATATATATATAT
TA7400_29729	2	TA	1108	1126	18	9	TATATATATATATATATAT
TA7553_29729	2	TA	575	593	18	9	TATATATATATATATATAT
TA7596_29729	2	TA	1007	1024	18	9	TATATATATATATATATA
TA7731_29729	2	TA	575	594	20	10	TATATATATATATATATATA
TA7857_29729	2	TA	325	345	20	10	TATATATATATATATATATAT
TA7930_29729	2	TA	50	74	24	12	TATATATATATATATATATATATAT

Supplementary Table 1. Contd.

TA8217_29729	2	TA	153	173	20	10	TATATATATATATATATATAT
TA8382_29729	2	TA	667	684	18	9	TATATATATATATATATA
TA8392_29729	2	TA	268	286	18	9	TATATATATATATATATAT
TA8913_29729	2	TA	560	578	18	9	TATATATATATATATATAT
BE052625	2	TA	205	223	18	9	TATATATATATATATATAT
BE053700	2	TA	357	377	20	10	TATATATATATATATATATAT
BE054013	2	TA	381	411	30	15	TATATATATATATATATATATATATATATAT
BE054303	2	TA	142	160	18	9	TATATATATATATATATAT
BE054872	2	TA	266	284	18	9	TATATATATATATATATAT
BE055619	2	TA	345	363	18	9	TATATATATATATATATAT
BF268601	2	TA	246	270	24	12	TATATATATATATATATATATATAT
BF268682	2	TA	530	551	22	11	TATATATATATATATATATATA
BF270944	2	TA	60	79	20	10	TATATATATATATATATATA
BF276865	2	TA	343	364	22	11	TATATATATATATATATATATA
BF276973	2	TA	369	387	18	9	TATATATATATATATATAT
BG441636	2	TA	344	362	18	9	TATATATATATATATATAT
BG442302	2	TA	311	329	18	9	TATATATATATATATATAT
BG443157	2	TA	346	364	18	9	TATATATATATATATATAT
BG444416	2	TA	212	230	18	9	TATATATATATATATATAT
BG444964	2	TA	287	305	18	9	TATATATATATATATATAT
BG445759	2	TA	191	214	24	12	TATATATATATATATATATATATA
BQ402139	2	TA	456	475	20	10	TATATATATATATATATATA
BQ405728	2	TA	590	614	24	12	TATATATATATATATATATATATAT
BQ407767	2	TA	116	134	18	9	TATATATATATATATATAT
BQ408241	2	TA	597	620	24	12	TATATATATATATATATATATATA
BQ410148	2	TA	467	487	20	10	TATATATATATATATATATAT
BQ412586	2	TA	59	76	18	9	TATATATATATATATATA
BQ413236	2	TA	510	546	36	18	TATATATATATATATATATATATATATATATATATAT
BQ413280	2	TA	446	466	20	10	TATATATATATATATATATAT
BQ414241	2	TA	535	575	40	20	TATATATATATATATATATATATATATATATATATATATAT
BQ414392	2	TA	504	528	24	12	TATATATATATATATATATATATAT
BQ414409	2	TA	594	616	22	11	TATATATATATATATATATATAT
BQ415692	2	TA	8	36	28	14	TATATATATATATATATATATATATATAT
BQ416174	2	TA	533	557	24	12	TATATATATATATATATATATATAT
TA4987_29729	2	TC	358	385	28	14	TCTCTCTCTCTCTCTCTCTCTCTCTCTC
TA5977_29729	2	TC	41	58	18	9	TCTCTCTCTCTCTCTCTC
TA7156_29729	2	TC	3	31	28	14	TCTCTCTCTCTCTCTCTCTCTCTCTCTCT

Supplementary Table 1. Contd.

TA7299_29729	2	TC	15	33	18	9	TCTCTCTCTCTCTCTCTCT
TA8009_29729	2	TC	118	135	18	9	TCTCTCTCTCTCTCTCTC
TA8184_29729	2	TC	46	64	18	9	TCTCTCTCTCTCTCTCTCT
TA8562_29729	2	TC	1	21	20	10	TCTCTCTCTCTCTCTCTCTCT
AW731151	2	TC	252	272	20	10	TCTCTCTCTCTCTCTCTCTCT
BE053003	2	TC	1	24	24	12	TCTCTCTCTCTCTCTCTCTCTCTC
BE053023	2	TC	325	342	18	9	TCTCTCTCTCTCTCTCTC
BF268139	2	TC	72	101	30	15	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
BF268191	2	TC	1	30	30	15	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
BF268204	2	TC	10	38	28	14	TCTCTCTCTCTCTCTCTCTCTCTCTCTCT
BF270284	2	TC	49	67	18	9	TCTCTCTCTCTCTCTCTCT
BF271444	2	TC	16	40	24	12	TCTCTCTCTCTCTCTCTCTCTCTCT
BF271486	2	TC	16	40	24	12	TCTCTCTCTCTCTCTCTCTCTCTCT
BF271488	2	TC	16	40	24	12	TCTCTCTCTCTCTCTCTCTCTCTCT
BF272117	2	TC	520	546	26	13	TCTCTCTCTCTCTCTCTCTCTCTCTCT
BF272287	2	TC	31	57	26	13	TCTCTCTCTCTCTCTCTCTCTCTCTCT
BG440412	2	TC	16	36	20	10	TCTCTCTCTCTCTCTCTCTCT
BG440825	2	TC	1	34	34	17	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
BG442970	2	TC	98	146	48	24	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
BG443314	2	TC	47	73	26	13	TCTCTCTCTCTCTCTCTCTCTCTCTCT
BG444045	2	TC	52	69	18	9	TCTCTCTCTCTCTCTCTC
BG445733	2	TC	604	621	18	9	TCTCTCTCTCTCTCTCTC
BG446768	2	TC	76	95	20	10	TCTCTCTCTCTCTCTCTCTC
BM359473	2	TC	44	70	26	13	TCTCTCTCTCTCTCTCTCTCTCTCTCT
BQ402652	2	TC	370	389	20	10	TCTCTCTCTCTCTCTCTCTC
BQ402783	2	TC	11	41	30	15	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
BQ403786	2	TC	91	121	30	15	TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
BQ405745	2	TC	1	19	18	9	TCTCTCTCTCTCTCTCTCT
BQ405963	2	TC	5	24	20	10	TCTCTCTCTCTCTCTCTCTC
BQ406414	2	TC	1	18	18	9	TCTCTCTCTCTCTCTCTC
BQ409912	2	TC	200	221	22	11	TCTCTCTCTCTCTCTCTCTCTC
TA8998_29729	2	TG	948	966	18	9	TGTGTGTGTGTGTGTGTGT
BQ409179	2	TG	592	610	18	9	TGTGTGTGTGTGTGTGTGT
TA6479_29729	3	AAC	589	607	18	6	AACAACAACAACAACAACA
TA6818_29729	3	AAG	44	66	21	7	AAGAAGAAGAAGAAGAAGAAGAA
TA8763_29729	3	AAG	253	284	30	10	AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAA
BF270051	3	AAG	322	344	21	7	AAGAAGAAGAAGAAGAAGAAGAA
BF274071	3	AAG	323	346	24	8	AAGAAGAAGAAGAAGAAGAAGAAG

Supplementary Table 1. Contd.

BF276199	3	AAG	18	38	21	7	AAGAAGAAGAAGAAGAAGAAG
BF276717	3	AAG	18	35	18	6	AAGAAGAAGAAGAAGAAG
BM360670	3	AAG	35	52	18	6	AAGAAGAAGAAGAAGAAG
BQ402652	3	AAG	523	540	18	6	AAGAAGAAGAAGAAGAAG
TA5627_29729	3	AAT	386	432	45	15	AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA
BF269744	3	AAT	510	531	21	7	AATAATAATAATAATAATAATA
BF272815	3	AAT	248	268	21	7	AATAATAATAATAATAATAAT
BM359991	3	AAT	3	21	18	6	AATAATAATAATAATAATA
BQ409090	3	AAT	68	88	21	7	AATAATAATAATAATAATAAT
TA9317_29729	3	ACA	526	551	24	8	ACAACAACAACAACAACAACAACAAC
TA5263_29729	3	ACC	66	83	18	6	ACCACCACCACCACCACC
TA8615_29729	3	ACC	174	193	18	6	ACCACCACCACCACCACCAC
BE052321	3	ACC	21	38	18	6	ACCACCACCACCACCACC
BE054357	3	ACC	394	411	18	6	ACCACCACCACCACCACC
BE055367	3	ACC	1	18	18	6	ACCACCACCACCACCACC
BG440558	3	ACC	10	27	18	6	ACCACCACCACCACCACC
BG443450	3	ACC	30	47	18	6	ACCACCACCACCACCACC
BM360664	3	ACC	55	72	18	6	ACCACCACCACCACCACC
BQ403357	3	ACC	1	19	18	6	ACCACCACCACCACCACCA
BQ407239	3	ACT	254	275	21	7	ACTACTACTACTACTACTACTA
BG444624	3	AGA	464	481	18	6	AGAAGAAGAAGAAGAAGA
TA5416_29729	3	AGC	478	500	21	7	AGCAGCAGCAGCAGCAGCAGCAG
BE054201	3	AGC	478	500	21	7	AGCAGCAGCAGCAGCAGCAGCAG
BE055322	3	AGC	473	495	21	7	AGCAGCAGCAGCAGCAGCAGCAG
BF277776	3	AGC	314	333	18	6	AGCAGCAGCAGCAGCAGCAG
BG441547	3	AGC	333	355	21	7	AGCAGCAGCAGCAGCAGCAGCAG
BG441749	3	AGC	462	484	21	7	AGCAGCAGCAGCAGCAGCAGCAG
BG442685	3	AGC	443	465	21	7	AGCAGCAGCAGCAGCAGCAGCAG
BG443059	3	AGC	111	133	21	7	AGCAGCAGCAGCAGCAGCAGCAG
BG443578	3	AGC	401	423	21	7	AGCAGCAGCAGCAGCAGCAGCAG
BM359805	3	AGC	376	398	21	7	AGCAGCAGCAGCAGCAGCAGCAG
TA8740_29729	3	AGG	209	226	18	6	AGGAGGAGGAGGAGGAGG
BQ402549	3	AGG	244	264	21	7	AGGAGGAGGAGGAGGAGGAGG
TA8270_29729	3	AGT	182	199	18	6	AGTAGTAGTAGTAGTAGT
AW728773	3	AGT	107	135	27	9	AGTAGTAGTAGTAGTAGTAGTAGTAGTAG
AW728773	3	AGT	137	162	24	8	AGTAGTAGTAGTAGTAGTAGTAGTAG
TA7230_29729	3	ATA	482	500	18	6	ATAATAATAATAATAATAA

Supplementary Table 1. Contd.

TA8375_29729	3	ATA	167	207	39	13	ATAATAATAATAATAATAATAATAATAATAATAATAATAAT
BF271371	3	ATA	247	265	18	6	ATAATAATAATAATAATAA
BG440604	3	ATA	333	350	18	6	ATAATAATAATAATAATA
BM357858	3	ATA	4	37	33	11	ATAATAATAATAATAATAATAATAATAATAATAA
BQ402072	3	ATA	17	35	18	6	ATAATAATAATAATAATAA
BQ409709	3	ATA	4	26	21	7	ATAATAATAATAATAATAATAAT
TA6817_29729	3	ATC	493	510	18	6	ATCATCATCATCATCATC
BM360646	3	ATC	464	481	18	6	ATCATCATCATCATCATC
BQ402818	3	ATC	49	72	24	8	ATCATCATCATCATCATCATCATC
BQ404712	3	ATC	357	375	18	6	ATCATCATCATCATCATCA
BQ412889	3	ATC	399	416	18	6	ATCATCATCATCATCATC
BG443471	3	ATG	843	867	24	8	ATGATGATGATGATGATGATGATGA
BM359739	3	ATG	480	515	36	12	ATGATGATGATGATGATGATGATGATGATGATGATG
BQ403752	3	ATG	394	414	21	7	ATGATGATGATGATGATGATG
BQ404588	3	ATG	675	692	18	6	ATGATGATGATGATGATG
BQ414418	3	ATG	497	514	18	6	ATGATGATGATGATGATG
TA5694_29729	3	ATT	638	658	21	7	ATTATTATTATTATTATTATT
TA5927_29729	3	ATT	660	678	18	6	ATTATTATTATTATTATTA
TA6888_29729	3	ATT	563	580	18	6	ATTATTATTATTATTATT
TA8122_29729	3	ATT	606	623	18	6	ATTATTATTATTATTATT
BE053044	3	ATT	465	485	21	7	ATTATTATTATTATTATTATT
BQ408202	3	ATT	527	549	21	7	ATTATTATTATTATTATTATTAT
BQ409456	3	ATT	519	578	60	20	ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT
BQ411088	3	ATT	477	527	51	17	ATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT
BQ413100	3	ATT	23	50	27	9	ATTATTATTATTATTATTATTATTATTA
TA9170_29729	3	CAA	478	499	21	7	CAACAACAACAACAACAACAAC
TA9404_29729	3	CAA	76	96	21	7	CAACAACAACAACAACAACAA
BM359324	3	CAA	603	620	18	6	CAACAACAACAACAACAA
BQ402724	3	CAA	515	537	21	7	CAACAACAACAACAACAACAACA
TA5763_29729	3	CAC	203	235	33	11	CACCACCACCACCACCACCACCACCACCACCAC
BE054767	3	CAC	204	236	33	11	CACCACCACCACCACCACCACCACCACCACCAC
BF270534	3	CAC	172	191	18	6	CACCACCACCACCACCACCA
BF273461	3	CAC	145	177	33	11	CACCACCACCACCACCACCACCACCACCACCAC
BF273463	3	CAC	143	175	33	11	CACCACCACCACCACCACCACCACCACCACCAC
BG440954	3	CAC	448	467	18	6	CACCACCACCACCACCACCA
BQ403459	3	CAC	226	244	18	6	CACCACCACCACCACCACC
BQ404164	3	CAC	467	486	18	6	CACCACCACCACCACCACCA

Supplementary Table 1. Contd.

BQ406809	3	CAC	17	38	21	7	CACCACCACCACCACCACCACC
BQ407414	3	CAC	138	156	18	6	CACCACCACCACCACCACC
BQ407429	3	CAC	495	514	18	6	CACCACCACCACCACCACCA
BQ410306	3	CAC	343	362	18	6	CACCACCACCACCACCACCA
TA6708_29729	3	CAG	19	51	33	11	CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG
TA9237_29729	3	CAG	9	31	21	7	CAGCAGCAGCAGCAGCAGCAGCA
BF269259	3	CAG	173	195	21	7	CAGCAGCAGCAGCAGCAGCAGCA
BF276122	3	CAG	441	458	18	6	CAGCAGCAGCAGCAGCAG
BG443126	3	CAG	158	175	18	6	CAGCAGCAGCAGCAGCAG
BG446140	3	CAG	249	266	18	6	CAGCAGCAGCAGCAGCAG
BM359410	3	CAG	156	173	18	6	CAGCAGCAGCAGCAGCAG
BQ406334	3	CAG	476	498	21	7	CAGCAGCAGCAGCAGCAGCAGCA
BQ406746	3	CAG	257	279	21	7	CAGCAGCAGCAGCAGCAGCAGCA
BQ413346	3	CAG	569	586	18	6	CAGCAGCAGCAGCAGCAG
TA5303_29729	3	CAT	123	142	18	6	CATCATCATCATCATCATCA
TA7168_29729	3	CAT	496	514	18	6	CATCATCATCATCATCATC
AW726061	3	CAT	131	150	18	6	CATCATCATCATCATCATCA
BE052471	3	CAT	238	256	18	6	CATCATCATCATCATCATC
BE053330	3	CAT	107	126	18	6	CATCATCATCATCATCATCA
BF275536	3	CAT	195	216	21	7	CATCATCATCATCATCATCATC
BF275537	3	CAT	9	28	18	6	CATCATCATCATCATCATCA
BF275563	3	CAT	13	35	21	7	CATCATCATCATCATCATCATCA
BG441669	3	CAT	90	109	18	6	CATCATCATCATCATCATCA
BG442727	3	CAT	81	100	18	6	CATCATCATCATCATCATCA
BG442903	3	CAT	111	130	18	6	CATCATCATCATCATCATCA
BG445378	3	CAT	399	418	18	6	CATCATCATCATCATCATCA
BQ410319	3	CAT	178	196	18	6	CATCATCATCATCATCATC
BQ410999	3	CAT	461	479	18	6	CATCATCATCATCATCATC
BQ411009	3	CAT	412	431	18	6	CATCATCATCATCATCATCA
BQ411446	3	CAT	271	293	21	7	CATCATCATCATCATCATCATCA
BQ412920	3	CAT	61	80	18	6	CATCATCATCATCATCATCA
BQ413309	3	CAT	465	482	18	6	CATCATCATCATCATCAT
BQ413494	3	CAT	98	122	24	8	CATCATCATCATCATCATCATCATC
BQ415559	3	CAT	470	489	18	6	CATCATCATCATCATCATCA
BF276058	3	CCA	55	75	21	7	CCACCACCACCACCACCACCA
BG443488	3	CCA	56	76	21	7	CCACCACCACCACCACCACCA
BQ403357	3	CCA	371	388	18	6	CCACCACCACCACCACCA

Supplementary Table 1. Contd.

BQ403789	3	CCA	78	98	21	7	CCACCACCACCACCACCACCA
BQ406649	3	CCA	206	223	18	6	CCACCACCACCACCACCA
BE053484	3	CCG	438	464	27	9	CCGCCGCCGCCGCCGCCGCCGCCGCCG
BE053679	3	CCG	398	415	18	6	CCGCCGCCGCCGCCGCCG
BM360104	3	CCG	15	41	27	9	CCGCCGCCGCCGCCGCCGCCGCCGCCG
BF277601	3	CGC	315	334	18	6	CGCCGCCGCCGCCGCCGCCG
BM360088	3	CGG	96	126	30	10	CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC
BE055500	3	CTC	205	224	18	6	CTCCTCCTCCTCCTCCTCCT
BF276934	3	CTG	485	503	18	6	CTGCTGCTGCTGCTGCTGC
TA6151_29729	3	CTT	464	482	18	6	CTTCTTCTTCTTCTTCTTC
TA6231_29729	3	CTT	377	416	39	13	CTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC
TA7513_29729	3	CTT	118	140	21	7	CTTCTTCTTCTTCTTCTTCTTCT
TA7685_29729	3	CTT	97	119	21	7	CTTCTTCTTCTTCTTCTTCTTCT
TA8040_29729	3	CTT	44	62	18	6	CTTCTTCTTCTTCTTCTTC
BE052283	3	CTT	49	67	18	6	CTTCTTCTTCTTCTTCTTC
BE053407	3	CTT	164	186	21	7	CTTCTTCTTCTTCTTCTTCTTCT
BF268719	3	CTT	8	30	21	7	CTTCTTCTTCTTCTTCTTCTTCT
BF275766	3	CTT	28	52	24	8	CTTCTTCTTCTTCTTCTTCTTCTTC
BF276118	3	CTT	167	189	21	7	CTTCTTCTTCTTCTTCTTCTTCT
BG445306	3	CTT	99	128	30	10	CTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
BG446135	3	CTT	97	114	18	6	CTTCTTCTTCTTCTTCTT
BQ405354	3	CTT	318	339	21	7	CTTCTTCTTCTTCTTCTTCTTC
BQ407215	3	CTT	184	203	18	6	CTTCTTCTTCTTCTTCTTCT
BQ415123	3	CTT	160	177	18	6	CTTCTTCTTCTTCTTCTT
BE054572	3	GAA	401	423	21	7	GAAGAAGAAGAAGAAGAAGAAGA
BF270826	3	GAA	211	232	21	7	GAAGAAGAAGAAGAAGAAGAAG
BF272603	3	GAA	196	225	30	10	GAAGAAGAAGAAGAAGAAGAAGAAGAAGAA
BE053470	3	GAC	192	214	21	7	GACGACGACGACGACGACGACGA
BF270454	3	GAC	375	397	21	7	GACGACGACGACGACGACGACGA
BG442800	3	GAC	92	109	18	6	GACGACGACGACGACGAC
BG446506	3	GAC	80	99	18	6	GACGACGACGACGACGACGA
BG446521	3	GAC	80	99	18	6	GACGACGACGACGACGACGA
BQ401418	3	GAC	53	70	18	6	GACGACGACGACGACGAC
BQ406180	3	GAC	225	242	18	6	GACGACGACGACGACGAC
BQ411879	3	GAC	133	150	18	6	GACGACGACGACGACGAC
BM359245	3	GAG	130	147	18	6	GAGGAGGAGGAGGAGGAG
BM360137	3	GAG	334	352	18	6	GAGGAGGAGGAGGAGGAGG

Supplementary Table 1. Contd.

BM360263	3	GAG	1	18	18	6	GAGGAGGAGGAGGAGGAG
BQ402314	3	GAG	374	391	18	6	GAGGAGGAGGAGGAGGAG
CB066413	3	GAG	101	118	18	6	GAGGAGGAGGAGGAGGAG
CK640533	3	GAG	404	423	18	6	GAGGAGGAGGAGGAGGAGGA
TA7199_29729	3	GAT	431	448	18	6	GATGATGATGATGATGAT
TA7800_29729	3	GAT	774	798	24	8	GATGATGATGATGATGATGATGATG
BF270416	3	GAT	144	163	18	6	GATGATGATGATGATGATGA
BF270826	3	GAT	241	260	18	6	GATGATGATGATGATGATGA
BG445086	3	GAT	511	528	18	6	GATGATGATGATGATGAT
BQ407920	3	GAT	406	423	18	6	GATGATGATGATGATGAT
BQ410998	3	GAT	304	322	18	6	GATGATGATGATGATGATG
BQ415217	3	GAT	398	422	24	8	GATGATGATGATGATGATGATGATG
BF277798	3	GCA	128	152	24	8	GCAGCAGCAGCAGCAGCAGCAGCAG
BM359389	3	GCA	583	606	24	8	GCAGCAGCAGCAGCAGCAGCAGCA
BG440543	3	GCC	192	212	21	7	GCCGCCGCCGCCGCCGCCGCC
BM358391	3	GCC	192	212	21	7	GCCGCCGCCGCCGCCGCCGCC
BM360106	3	GCC	36	57	21	7	GCCGCCGCCGCCGCCGCCGCCG
TA8195_29729	3	GCG	122	141	18	6	GCGGCGGCGGCGGCGGCGGC
BQ404704	3	GCT	266	287	21	7	GCTGCTGCTGCTGCTGCTGCTG
TA9435_29729	3	GGA	795	813	18	6	GGAGGAGGAGGAGGAGGAG
BM359886	3	GGA	112	133	21	7	GGAGGAGGAGGAGGAGGAGGAG
BF276904	3	GGC	219	238	18	6	GGCGGCGGCGGCGGCGGCGG
BG443893	3	GGC	160	177	18	6	GGCGGCGGCGGCGGCGGC
BG445086	3	GGC	421	443	21	7	GGCGGCGGCGGCGGCGGCGGCGG
AW667879	3	GTG	773	791	18	6	GTGGTGGTGGTGGTGGTGG
BE055118	3	GTG	127	151	24	8	GTGGTGGTGGTGGTGGTGGTGGTGG
BF275852	3	GTG	120	140	21	7	GTGGTGGTGGTGGTGGTGGTG
BQ405592	3	GTG	134	154	21	7	GTGGTGGTGGTGGTGGTGGTG
BQ406729	3	GTG	225	242	18	6	GTGGTGGTGGTGGTGGTG
BG445787	3	GTT	186	204	18	6	GTTGTTGTTGTTGTTGTTG
BQ415371	3	GTT	259	283	24	8	GTTGTTGTTGTTGTTGTTGTTGTTG
BG439856	3	TAA	562	582	21	7	TAATAATAATAATAATAATAA
BG444929	3	TAT	301	319	18	6	TATTATTATTATTATTATT
BQ413388	3	TAT	486	545	60	20	TATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAT
TA7643_29729	3	TCA	209	228	18	6	TCATCATCATCATCATCATC
BF268883	3	TCA	98	131	33	11	TCATCATCATCATCATCATCATCATCATCATCAT
BF272166	3	TCA	237	256	18	6	TCATCATCATCATCATCATC
BF278307	3	TCA	339	361	21	7	TCATCATCATCATCATCATCATC

Supplementary Table 1. Contd.

BG443154	3	TCA	234	253	18	6	TCATCATCATCATCATCATC
BG443992	3	TCA	225	244	18	6	TCATCATCATCATCATCATC
BM357862	3	TCA	297	334	36	12	TCATCATCATCATCATCATCATCATCATCATCATCATC
BQ404964	3	TCA	622	646	24	8	TCATCATCATCATCATCATCATCAT
BQ405578	3	TCA	155	188	33	11	TCATCATCATCATCATCATCATCATCATCATCAT
BQ405692	3	TCA	323	342	18	6	TCATCATCATCATCATCATC
TA6203_29729	3	TCC	227	248	21	7	TCCTCCTCCTCCTCCTCCTCCT
BF269750	3	TCC	805	826	21	7	TCCTCCTCCTCCTCCTCCTCCT
TA6398_29729	3	TCT	101	120	18	6	TCTTCTTCTTCTTCTTCTTC
TA7378_29729	3	TCT	345	363	18	6	TCTTCTTCTTCTTCTTCTT
TA7904_29729	3	TCT	66	84	18	6	TCTTCTTCTTCTTCTTCTT
TA8283_29729	3	TCT	119	138	18	6	TCTTCTTCTTCTTCTTCTTC
TA9261_29729	3	TCT	12	40	27	9	TCTTCTTCTTCTTCTTCTTCTTCTTCTTC
BF271967	3	TCT	332	353	21	7	TCTTCTTCTTCTTCTTCTTCTT
BF274343	3	TCT	67	89	21	7	TCTTCTTCTTCTTCTTCTTCTTC
BF275626	3	TCT	445	467	21	7	TCTTCTTCTTCTTCTTCTTCTTC
BF277995	3	TCT	90	110	21	7	TCTTCTTCTTCTTCTTCTTCT
BF279348	3	TCT	12	40	27	9	TCTTCTTCTTCTTCTTCTTCTTCTTCTTC
BG444328	3	TCT	113	132	18	6	TCTTCTTCTTCTTCTTCTTC
BG444426	3	TCT	81	106	24	8	TCTTCTTCTTCTTCTTCTTCTTCTTC
BG445591	3	TCT	19	38	18	6	TCTTCTTCTTCTTCTTCTTC
BM359726	3	TCT	218	237	18	6	TCTTCTTCTTCTTCTTCTTC
BM359821	3	TCT	114	133	18	6	TCTTCTTCTTCTTCTTCTTC
BQ402342	3	TCT	377	399	21	7	TCTTCTTCTTCTTCTTCTTCTTC
BQ404554	3	TCT	162	181	18	6	TCTTCTTCTTCTTCTTCTTC
BQ406206	3	TCT	266	310	45	15	TCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
BQ408399	3	TCT	268	295	27	9	TCTTCTTCTTCTTCTTCTTCTTCTTCTT
TA5514_29729	3	TGA	1111	1128	18	6	TGATGATGATGATGATGA
BF279120	3	TGA	293	312	18	6	TGATGATGATGATGATGATG
BQ408632	3	TGA	418	435	18	6	TGATGATGATGATGATGA
BQ410226	3	TGA	487	510	24	8	TGATGATGATGATGATGATGATGA
BQ410493	3	TGA	441	458	18	6	TGATGATGATGATGATGA
BQ411010	3	TGA	505	524	18	6	TGATGATGATGATGATGATG
CK640498	3	TGA	363	386	24	8	TGATGATGATGATGATGATGATGA
TA5823_29729	3	TGC	448	465	18	6	TGCTGCTGCTGCTGCTGC
TA5953_29729	3	TGC	206	223	18	6	TGCTGCTGCTGCTGCTGC
BF269659	3	TGC	249	267	18	6	TGCTGCTGCTGCTGCTGCT
BF270691	3	TGC	468	485	18	6	TGCTGCTGCTGCTGCTGC

Supplementary Table 1. Contd.

BF276313	3	TGC	56	73	18	6	TGCTGCTGCTGCTGCTGC
BG442229	3	TGC	266	283	18	6	TGCTGCTGCTGCTGCTGC
BQ401914	3	TGC	117	134	18	6	TGCTGCTGCTGCTGCTGC
BF271971	3	TGG	233	256	24	8	TGGTGGTGGTGGTGGTGGTGGTGG
BE052575	3	TGT	556	573	18	6	TGTTGTTGTTGTTGTTGT
TA5476_29729	3	TTA	886	935	48	16	TTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT
TA5626_29729	3	TTA	458	476	18	6	TTATTATTATTATTATTAT
TA5985_29729	3	TTA	1483	1501	18	6	TTATTATTATTATTATTAT
TA7034_29729	3	TTA	541	558	18	6	TTATTATTATTATTATTA
BF276846	3	TTA	668	705	36	12	TTATTATTATTATTATTATTATTATTATTATTATTATT
BF276847	3	TTA	661	698	36	12	TTATTATTATTATTATTATTATTATTATTATTATTATT
BG444301	3	TTA	520	539	18	6	TTATTATTATTATTATTATT
BQ405592	3	TTA	20	53	33	11	TTATTATTATTATTATTATTATTATTATTATTAT
BQ413644	3	TTA	25	43	18	6	TTATTATTATTATTATTAT
BQ415860	3	TTA	484	518	33	11	TTATTATTATTATTATTATTATTATTATTATTATT
TA4858_29729	3	TTC	18	37	18	6	TTCTTCTTCTTCTTCTTCTT
TA4859_29729	3	TTC	30	49	18	6	TTCTTCTTCTTCTTCTTCTT
TA5246_29729	3	TTC	691	708	18	6	TTCTTCTTCTTCTTCTTC
TA6371_29729	3	TTC	32	61	30	10	TTCTTCTTCTTCTTCTTCTTCTTCTTCTTC
TA6739_29729	3	TTC	176	193	18	6	TTCTTCTTCTTCTTCTTC
TA6740_29729	3	TTC	219	236	18	6	TTCTTCTTCTTCTTCTTC
TA7048_29729	3	TTC	101	121	21	7	TTCTTCTTCTTCTTCTTCTTC
TA7486_29729	3	TTC	44	64	21	7	TTCTTCTTCTTCTTCTTCTTC
TA9370_29729	3	TTC	209	226	18	6	TTCTTCTTCTTCTTCTTC
BE051955	3	TTC	24	43	18	6	TTCTTCTTCTTCTTCTTCTT
BE052040	3	TTC	18	37	18	6	TTCTTCTTCTTCTTCTTCTT
BE052859	3	TTC	30	49	18	6	TTCTTCTTCTTCTTCTTCTT
BE052916	3	TTC	123	141	18	6	TTCTTCTTCTTCTTCTTCT
BE053214	3	TTC	86	105	18	6	TTCTTCTTCTTCTTCTTCTT
BE053742	3	TTC	15	34	18	6	TTCTTCTTCTTCTTCTTCTT
BE054092	3	TTC	15	34	18	6	TTCTTCTTCTTCTTCTTCTT
BE054356	3	TTC	35	54	18	6	TTCTTCTTCTTCTTCTTCTT
BE055134	3	TTC	15	35	21	7	TTCTTCTTCTTCTTCTTCTTC
BE055541	3	TTC	30	49	18	6	TTCTTCTTCTTCTTCTTCTT
BF277771	3	TTC	98	118	21	7	TTCTTCTTCTTCTTCTTCTTC
BG440000	3	TTC	32	51	18	6	TTCTTCTTCTTCTTCTTCTT
BG440223	3	TTC	15	34	18	6	TTCTTCTTCTTCTTCTTCTT
BG440502	3	TTC	3	22	18	6	TTCTTCTTCTTCTTCTTCTT

Supplementary Table 1. Contd**.**

BG440605	3	TTC	30	49	18	6	TTCTTCTTCTTCTTCTTCTT
BG441191	3	TTC	20	39	18	6	TTCTTCTTCTTCTTCTTCTT
BG441343	3	TTC	12	31	18	6	TTCTTCTTCTTCTTCTTCTT
BG441357	3	TTC	12	31	18	6	TTCTTCTTCTTCTTCTTCTT
BG441898	3	TTC	130	159	30	10	TTCTTCTTCTTCTTCTTCTTCTTCTTCTTC
BG442145	3	TTC	28	47	18	6	TTCTTCTTCTTCTTCTTCTT
BG444819	3	TTC	15	34	18	6	TTCTTCTTCTTCTTCTTCTT
BG444885	3	TTC	20	39	18	6	TTCTTCTTCTTCTTCTTCTT
BG445222	3	TTC	20	39	18	6	TTCTTCTTCTTCTTCTTCTT
BG445388	3	TTC	6	25	18	6	TTCTTCTTCTTCTTCTTCTT
BG445581	3	TTC	26	45	18	6	TTCTTCTTCTTCTTCTTCTT
BM360153	3	TTC	195	212	18	6	TTCTTCTTCTTCTTCTTC
BQ403629	3	TTC	215	239	24	8	TTCTTCTTCTTCTTCTTCTTCTTCT
BQ404189	3	TTC	167	210	42	14	TTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
BQ404857	3	TTC	92	112	21	7	TTCTTCTTCTTCTTCTTCTTC
BQ405117	3	TTC	218	250	33	11	TTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTC
BQ405692	3	TTC	343	368	24	8	TTCTTCTTCTTCTTCTTCTTCTTCTT
BQ406743	3	TTC	2	26	24	8	TTCTTCTTCTTCTTCTTCTTCTTCT
BF274447	3	TTG	228	245	18	6	TTGTTGTTGTTGTTGTTG
BF277638	3	TTG	63	82	18	6	TTGTTGTTGTTGTTGTTGTT
BQ413337	4	AAAC	120	142	20	5	AAACAAACAAACAAACAAACAAA
BQ405865	4	AAAT	358	378	20	5	AAATAAATAAATAAATAAATA
TA9129_29729	4	AATA	221	242	20	5	AATAAATAAATAAATAAATAAA
BQ410598	4	AATA	63	84	20	5	AATAAATAAATAAATAAATAAA
BQ405038	4	AATT	565	586	20	5	AATTAATTAATTAATTAATTAA
BQ407187	4	ACAT	457	492	36	9	ACATACATACATACATACATACATACATACATACAT
BQ414500	4	ACAT	478	513	36	9	ACATACATACATACATACATACATACATACATACAT
BQ415843	4	ACAT	457	492	36	9	ACATACATACATACATACATACATACATACATACAT
TA7252_29729	4	AGCC	4	24	20	5	AGCCAGCCAGCCAGCCAGCCA
BG444331	4	AGGG	656	675	20	5	AGGGAGGGAGGGAGGGAGGG
TA8391_29729	4	ATAA	252	272	20	5	ATAAATAAATAAATAAATAAA
TA9339_29729	4	ATAA	16	36	20	5	ATAAATAAATAAATAAATAAA
BF278325	4	ATTG	322	345	24	6	ATTGATTGATTGATTGATTGATTG
TA6604_29729	4	ATTT	438	457	20	5	ATTTATTTATTTATTTATTT
BF268634	4	ATTT	394	413	20	5	ATTTATTTATTTATTTATTT
TA7499_29729	4	CTCC	9	30	20	5	CTCCCTCCCTCCCTCCCTCCCT
BQ411337	4	GAGG	479	503	24	6	GAGGGAGGGAGGGAGGGAGGGAGGG
BE052138	4	GGAG	740	759	20	5	GGAGGGAGGGAGGGAGGGAG

Supplementary Table 1. Contd.

BG440440	4	TAAA	21	42	20	5	TAAATAAATAAATAAATAAATA
BG447194	4	TAAT	356	376	20	5	TAATTAATTAATTAATTAATT
TA8897_29729	4	TATG	421	456	36	9	TATGTATGTATGTATGTATGTATGTATGTATGTATG
BE053131	4	TATG	337	359	20	5	TATGTATGTATGTATGTATGTAT
BM357948	4	TATG	642	664	20	5	TATGTATGTATGTATGTATGTAT
BM358426	4	TATG	538	565	28	7	TATGTATGTATGTATGTATGTATGTATG
BQ414527	4	TATT	9	29	20	5	TATTTATTTATTTATTTATTT
BG439886	4	TGTA	616	635	20	5	TGTATGTATGTATGTATGTA
BF272588	4	TTCT	16	37	20	5	TTCTTTCTTTCTTTCTTTCTTT
TA5862_29729	4	TTTA	455	477	20	5	TTTATTTATTTATTTATTTATTT
BQ409148	4	TTTC	130	156	24	6	TTTCTTTCTTTCTTTCTTTCTTTCTTT
BQ402637	4	TTTG	484	506	20	5	TTTGTTTGTTTGTTTGTTTGTTT
BF271164	5	AGAAA	44	68	25	5	AGAAAAGAAAAGAAAAGAAAAGAAA
TA7673_29729	5	AGAAG	68	93	25	5	AGAAGAGAAGAGAAGAGAAGAGAAGA
BG444055	5	AGAGA	20	53	30	6	AGAGAAGAGAAGAGAAGAGAAGAGAAGAGAAGAG
BQ412345	5	CAAAC	72	96	25	5	CAAACCAAACCAAACCAAACCAAAC
BF276655	5	TCCCC	64	90	25	5	TCCCCTCCCCTCCCCTCCCCTCCCCTC
BF276803	5	TCTAA	24	51	25	5	TCTAATCTAATCTAATCTAATCTAATCT
BF274421	5	TCTTT	445	470	25	5	TCTTTTCTTTTCTTTTCTTTTCTTTT
TA7484_29729	5	TTCTT	34	61	25	5	TTCTTTTCTTTTCTTTTCTTTTCTTTTC
BG445274	5	TTTCT	72	97	25	5	TTTCTTTTCTTTTCTTTTCTTTTCTT
TA7930_29729	5	TTTTC	330	354	25	5	TTTTCTTTTCTTTTCTTTTCTTTTC
BE052504	6	AAAAAT	1	24	24	4	AAAAATAAAAATAAAAATAAAAAT
BF278842	6	AAAAAT	44	73	30	5	AAAAATAAAAATAAAAATAAAAATAAAAAT
BG447357	6	AAAAAT	453	477	24	4	AAAAATAAAAATAAAAATAAAAATA
BQ408211	6	AAGCTC	35	59	24	4	AAGCTCAAGCTCAAGCTCAAGCTCA
BG440787	6	AAGGCA	48	80	30	5	AAGGCAAAGGCAAAGGCAAAGGCAAAGGCAAAG
BF272140	6	CAACAG	291	317	24	4	CAACAGCAACAGCAACAGCAACAGCAA
BG440800	6	CACAGG	201	228	24	4	CACAGGCACAGGCACAGGCACAGGCACA
TA5668_29729	6	CACTCT	107	131	24	4	CACTCTCACTCTCACTCTCACTCTC
TA9237_29729	6	CAGCAA	336	360	24	4	CAGCAACAGCAACAGCAACAGCAAC
BG444074	6	CAGGCA	132	166	30	5	CAGGCACAGGCACAGGCACAGGCACAGGCACAGGC
BM358018	6	CCACAG	15	47	30	5	CCACAGCCACAGCCACAGCCACAGCCACAGCCA
TA6665_29729	6	CCCCAA	91	119	24	4	CCCCAACCCCAACCCCAACCCCAACCCCA
BE052870	6	CCCCAA	212	240	24	4	CCCCAACCCCAACCCCAACCCCAACCCCA
BE054108	6	CCCCAA	64	92	24	4	CCCCAACCCCAACCCCAACCCCAACCCCA
BF277774	6	CCCCTC	40	64	24	4	CCCCTCCCCCTCCCCCTCCCCCTCC

Supplementary Table 1. Contd.

BE053289	6	CCCTTC	234	263	30	5	CCCTTCCCCTTCCCCTTCCCCTTCCCCTTC
BE054387	6	CCCTTT	140	169	30	5	CCCTTTCCCTTTCCCTTTCCCTTTCCCTTT
BQ414792	6	CCGAGA	364	410	42	7	CCGAGACCGAGACCGAGACCGAGACCGAGACCGAGACCGAGACCGAG
TA8669_29729	6	CCTCCA	42	66	24	4	CCTCCACCTCCACCTCCACCTCCAC
BQ402121	6	CGGCTC	458	488	30	5	CGGCTCCGGCTCCGGCTCCGGCTCCGGCTCC
BQ414131	6	CGGCTC	453	483	30	5	CGGCTCCGGCTCCGGCTCCGGCTCCGGCTCC
TA5668_29729	6	CTCTCA	139	163	24	4	CTCTCACTCTCACTCTCACTCTCAC
BF270137	6	CTCTCC	97	126	30	5	CTCTCCCTCTCCCTCTCCCTCTCCCTCTCC
BQ404244	6	CTGCCT	84	108	24	4	CTGCCTCTGCCTCTGCCTCTGCCTC
BE054746	6	CTTCCC	209	233	24	4	CTTCCCCTTCCCCTTCCCCTTCCCC
BG445180	6	CTTCCC	212	236	24	4	CTTCCCCTTCCCCTTCCCCTTCCCC
BF275274	6	GAAGAT	478	504	24	4	GAAGATGAAGATGAAGATGAAGATGAA
BG446849	6	GAAGAT	460	487	24	4	GAAGATGAAGATGAAGATGAAGATGAAG
BQ409057	6	GAGGAA	522	545	24	4	GAGGAAGAGGAAGAGGAAGAGGAA
TA8901_29729	6	GAGTCC	23	49	24	4	GAGTCCGAGTCCGAGTCCGAGTCCGAG
BG445177	6	GATAGG	276	301	24	4	GATAGGGATAGGGATAGGGATAGGGA
BG445815	6	GATAGG	561	586	24	4	GATAGGGATAGGGATAGGGATAGGGA
BQ414273	6	GATGGG	381	404	24	4	GATGGGGATGGGGATGGGGATGGG
BQ402465	6	GCCACA	561	585	24	4	GCCACAGCCACAGCCACAGCCACAG
TA4949_29729	6	GCTCAT	329	354	24	4	GCTCATGCTCATGCTCATGCTCATGC
BF272793	6	GCTCCT	60	95	36	6	GCTCCTGCTCCTGCTCCTGCTCCTGCTCCTGCTCCT
BM359711	6	GGAAGA	260	286	24	4	GGAAGAGGAAGAGGAAGAGGAAGAGGA
BQ402120	6	GGAGCC	588	611	24	4	GGAGCCGGAGCCGGAGCCGGAGCC
TA9009_29729	6	GGCTTC	144	173	30	5	GGCTTCGGCTTCGGCTTCGGCTTCGGCTTC
TA6589_29729	6	TATTTT	395	419	24	4	TATTTTTATTTTTATTTTTATTTTT
BG440224	6	TCAAGC	259	288	30	5	TCAAGCTCAAGCTCAAGCTCAAGCTCAAGC
BG440896	6	TCAAGC	267	296	30	5	TCAAGCTCAAGCTCAAGCTCAAGCTCAAGC
TA4950_29729	6	TCAGGC	267	293	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE052527	6	TCAGGC	183	209	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE052669	6	TCAGGC	256	282	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE052760	6	TCAGGC	260	286	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE053088	6	TCAGGC	215	241	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE053587	6	TCAGGC	215	241	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE053730	6	TCAGGC	261	287	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE053966	6	TCAGGC	266	292	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE054144	6	TCAGGC	262	288	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE054522	6	TCAGGC	266	292	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA

Supplementary Table 1. Contd.

BE054976	6	TCAGGC	229	255	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE055230	6	TCAGGC	256	282	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE055232	6	TCAGGC	265	291	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE055623	6	TCAGGC	265	291	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE055624	6	TCAGGC	265	291	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BE055626	6	TCAGGC	261	287	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BF272413	6	TCAGGC	274	300	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG440521	6	TCAGGC	257	283	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG441211	6	TCAGGC	242	268	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG442797	6	TCAGGC	241	267	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG443565	6	TCAGGC	309	335	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG443843	6	TCAGGC	257	283	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG443923	6	TCAGGC	258	284	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG444074	6	TCAGGC	167	192	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTC
BG444093	6	TCAGGC	243	269	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG444468	6	TCAGGC	258	284	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG444500	6	TCAGGC	121	147	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG444556	6	TCAGGC	2	28	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG444921	6	TCAGGC	265	291	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG444937	6	TCAGGC	257	283	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG445036	6	TCAGGC	267	293	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG445128	6	TCAGGC	257	283	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG445281	6	TCAGGC	242	268	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
BG446067	6	TCAGGC	266	292	24	4	TCAGGCTCAGGCTCAGGCTCAGGCTCA
TA6451_29729	6	TCCTCT	236	268	30	5	TCCTCTTCCTCTTCCTCTTCCTCTTCCTCTTCC
BG442199	6	TCTTCG	184	224	36	6	TCTTCGTCTTCGTCTTCGTCTTCGTCTTCGTCTTCGTCTTC
BQ410449	6	TCTTTT	94	118	24	4	TCTTTTTCTTTTTCTTTTTCTTTTT
BF279254	6	TGCAAT	78	106	24	4	TGCAATTGCAATTGCAATTGCAATTGCAA
TA4888_29729	6	TGGCTG	694	726	30	5	TGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGG
BQ409453	6	TGGCTG	513	545	30	5	TGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGG
BQ415466	6	TGGCTG	548	580	30	5	TGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGG
TA5994_29729	6	TGTGAG	29	54	24	4	TGTGAGTGTGAGTGTGAGTGTGAGTG
BG441253	6	TGTGAG	25	50	24	4	TGTGAGTGTGAGTGTGAGTGTGAGTG
BM358740	6	TGTGAG	25	50	24	4	TGTGAGTGTGAGTGTGAGTGTGAGTG
BG446247	6	TTTCCT	48	73	24	4	TTTCCTTTTCCTTTTCCTTTTCCTTT
TA9299_29729	6	TTTTGG	450	476	24	4	TTTTGGTTTTGGTTTTGGTTTTGGTTT
BF270158	6	TTTTTC	162	185	24	4	TTTTTCTTTTTCTTTTTCTTTTTC

Supplementary Table 2.

Repeat unit	Motif	Ga	Gr	Gh
Tetranucleotide	AAAT	10	52	5
	AAAC	2	22	4
	AAAG	2	31	11
	AATT	2	7	6
	AATC	1	3	1
	AATG		3	2
	AACC	1	1	
	AAGC		1	
	AAGG		2	2
	ATAC	9	39	17
	ATAG		5	3
	ATCC		4	3
	ATGC		2	2
	ACAG		2	2
	ACTC		4	
	ACCT		1	
	AGCT		1	
	AGCC	1	2	1
	AGCG		2	
	AGGC		5	
	AGGG	4	1	
Pentanucleotide	AAAAT		4	2
	AAAAC		2	1
	AAAAG	5	3	1
	AAATC		4	
	AAACC	1	2	3
	AAAGC			1
	AATAT		1	1
	AATAC		2	
	AATCT	1		
	AACTC		1	3
	AACTG		1	
	AACCT		1	
	AACCC		3	
	AACCG		3	

Supplementary Table 2.

	AAGAG	2	4	8
	ATACC			1
	ATACT			1
	ATATC		1	
	ATAGC		1	1
	ATCTC		1	
	ATGCC			1
	ACAGT		1	
	ACCAG			1
	ACCTC		1	
	AGCCT		1	
	AGGGG	1	1	
Hexanucleotide	AAAAAT	4	3	6
	AAAAAC		3	
	AAAAAG	2	4	5
	AAAATT		4	
	AAAATC		1	1
	AAAATG		1	5
	AAAAGG	1		2
	AAAACC	1	6	1
	AAATAT		1	
	AAATAC		1	
	AAATTT		1	
	AAATTG		1	
	AAATCC		3	
	AAATGG		2	
	AAATGG			1
	AAAGGG	1		2
	AAAGGC	1		
	AAACAG		1	
	AAACCC		3	
	AAAGAC		1	
	AAAGAG		2	
	AAAGCC		1	
	AAAGGC		9	
	AAAGGG		2	

Supplementary Table 2.

AATTCC		1	
AATTGC	1	1	
AATCAT		1	
AATCAG		1	
AATCCC		2	1
AATGAG		3	1
AATGTG			3
AATGGC		2	6
AATGGG		3	1
AACAAG		1	1
AACATG		1	
AACATC			2
AACACT		1	
AACAGC	2	8	2
AACTCC		3	1
AACTCG			1
AACTGG		1	
AACCAC			1
AACCAG		1	
AACCTC		8	1
AACCCT		2	3
AACCCC	3	4	
AACGCC		1	1
AACGGC		1	
AAGATG	2	6	1
AAGACG	1	1	
AAGAGG	3	8	3
AAGCAC		1	
AAGCAG		1	1
AAGCTC	3	1	
AAGCTG		1	
AAGCCT		1	
AAGCCG	1		
AAGGAG		7	1
AAGGTG		5	2
AAGGCC			2
AAGGGC		1	

Supplementary Table 2. Contd.

AAGGGG	3	2	
ATATAC		6	5
ATATAG		2	1
ATAGAC		1	
ATAGCC		1	
ATAGGG	2	1	1
ATCACC		4	3
ATCACG		1	
ATCAGC		3	3
ATCTCC		3	
ATCCTC		2	
ATCCCC	1	2	
ATCCGC		2	
ATCGCC		2	
ATCGGC		1	
ATGACC		4	
ATGAGC	1		1
ATGCCC		4	2
ACACTC	3		
ACACGC		1	
ACAGTG		2	
ACAGCC	5	2	2
ACAGGC	2	1	
ACAGGG		1	
ACTCTC	2	3	1
ACTCCT			1
ACTCGG	1	1	
ACTGCT		1	
ACTGGT		1	1
ACCAGC		5	
ACCAGG		1	
ACCTCC	1	3	2
ACCTCG		2	
ACCCTC		1	
ACCCCC		1	
ACCCCG			1
ACCGAG	1		1

Supplementary Table 2. Contd.

ACCGCC		4	2
ACGCTC		1	
ACGCCC		1	
ACGCCG		1	1
ACGGAG		2	
ACGGGC	1	1	
AGAGCG		1	
AGAGGC		1	
AGAGGG	1	3	3
AGCAGG	1	7	8
AGCTCC		2	1
AGCTGC		1	1
AGCCTC		3	
AGCCTG	35	14	
AGCCGG	3		
AGCCCC		1	2
AGCGGC		1	
AGCGGG		1	2
AGGCCG		2	
AGGCGC		1	
AGGCGG		5	4
AGGGGC		2	
AGGGGG	1		
CCCGGG			1

Supplementary Table 2.

Group I							
HG	**Sequence ID**	**Period Size**	**Period**	**Rep. Length**	**Copy Num.**	**Blast Strand**	**Species**
1	BQ403493 Putative ubiquitin-conjugating enzyme [Oryza sativa (japonica cultivar-group)]	2	AG	18	9	Plus	Gossypium_arboreum
	DR044194 Putative ubiquitin-conjugating enzyme [Populus canadensis (Carolina poplar)]	2	AG	18	9	Plus	Gossypium_hirsutum
	TA22106_3635 Putative ubiquitin-conjugating enzyme [Oryza sativa (japonica cultivar-group)]	2	AG	18	9	Plus	Gossypium_hirsutum
2	BQ404091 Putative ubiquitin-conjugating enzyme [Oryza sativa (japonica cultivar-group)]	2	AG	18	9	Plus	Gossypium_arboreum
	DR044194 Putative ubiquitin-conjugating enzyme [Populus canadensis (Carolina poplar)]	2	AG	18	9	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	TA22106_3635 Putative ubiquitin-conjugating enzyme [Oryza sativa (japonica cultivar-group)]	2	AG	18	9	Plus	Gossypium_hirsutum
3	BQ410390 Putative ubiquitin-conjugating enzyme [Oryza sativa (japonica cultivar-group)]	2	AG	18	9	Plus	Gossypium_arboreum
	DR044194 Putative ubiquitin-conjugating enzyme [Populus canadensis (Carolina poplar)]	2	AG	18	9	Plus	Gossypium_hirsutum
	TA22106_3635 Putative ubiquitin-conjugating enzyme [Oryza sativa (japonica cultivar-group)]	2	AG	18	9	Plus	Gossypium_hirsutum
4	TA5984_29729 Putative ubiquitin carrier protein UBC7 [Oryza sativa (japonica cultivar-group)]	2	CT	18	9	Plus	Gossypium_arboreum
	AI726604 Putative ubiquitin carrier protein UBC7 [Oryza sativa (japonica cultivar-group)]	2	CT	18	9	Plus	Gossypium_hirsutum
5	BG446518 Sucrose transporter [Malus domestica (Apple) (Malus sylvestris)]	2	GA	18	9	Plus	Gossypium_arboreum
	TA31646_3635 Putative sucrose carrier [Ricinus communis (Castor bean)]	2	GA	18	9	Plus	Gossypium_hirsutum
6	BF270051	3	AAG	21	7	Plus	Gossypium_arboreum
	TA26991_3635 RNA polymerase II transcriptional coactivator KELP [Arabidopsis thaliana (Mouse-ear cress)]	3	AAG	21	7	Plus	Gossypium_hirsutum
	DR455976	3	AAG	21	7	Plus	Gossypium_hirsutum
7	BQ402652 Eukaryotic translation initiation factor 2 subunit beta [Triticum aestivum (Wheat)]	3	AAG	18	6	Plus	Gossypium_arboreum
	TA28888_3635 Eukaryotic translation initiation factor 2 beta subunit-like [Solanum tuberosum (Potato)]	3	AAG	18	6	Plus	Gossypium_hirsutum
8	BM360670	3	AAG	18	6	Plus	Gossypium_arboreum
	TA41661_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	AAG	18	6	Plus	Gossypium_hirsutum
9	TA8615_29729	3	ACC	18	6	Plus	Gossypium_arboreum
	TA20920_3635 Heavy metal transport/detoxification protein; Harpin-induced 1 [Medicago truncatula (Barrel medic)]	3	ACC	18	6	Plus	Gossypium_hirsutum
10	BG444624	3	AGA	18	6	Plus	Gossypium_arboreum
	TA23448_3635 Adenylate translocator (Brittle-1)-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	18	6	Plus	Gossypium_hirsutum
11	BM359324 GRF1-interacting factor 1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CAA	18	6	Plus	Gossypium_arboreum
	DT555999 GRF1-interacting factor 1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CAA	18	6	Plus	Gossypium_hirsutum
12	BQ407429 Similarity to N-acetyltransferase [Arabidopsis thaliana (Mouse-ear cress)]	3	CAC	18	6	Plus	Gossypium_arboreum
	TA20324_3635 Similarity to N-acetyltransferase [Arabidopsis thaliana (Mouse-ear cress)]	3	CAC	18	6	Plus	Gossypium_hirsutum
13	BQ404164 Similarity to N-acetyltransferase [Arabidopsis thaliana (Mouse-ear cress)]	3	CAC	18	6	Plus	Gossypium_arboreum
	TA20324_3635 Similarity to N-acetyltransferase [Arabidopsis thaliana (Mouse-ear cress)]	3	CAC	18	6	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

14	BQ410306 Similarity to N-acetyltransferase [Arabidopsis thaliana (Mouse-ear cress)]	3	CAC	18	6	Plus	Gossypium_arboreum
	TA20324_3635 Similarity to N-acetyltransferase [Arabidopsis thaliana (Mouse-ear cress)]	3	CAC	18	6	Plus	Gossypium_hirsutum
15	BQ407414	3	CAC	18	6	Plus	Gossypium_arboreum
	DW242818	3	CAC	18	6	Plus	Gossypium_hirsutum
	DW242818	3	CAC	18	6	Plus	Gossypium_hirsutum
	DW242818	3	CAC	18	6	Plus	Gossypium_hirsutum
16	BF276122	3	CAG	18	6	Plus	Gossypium_arboreum
	DW244622	3	CAG	18	6	Plus	Gossypium_hirsutum
	TA21810_3635	3	CAG	18	6	Plus	Gossypium_hirsutum
	TA21808_3635	3	CAG	18	6	Plus	Gossypium_hirsutum
17	BM359410	3	CAG	18	6	Plus	Gossypium_arboreum
	DW244622	3	CAG	18	6	Plus	Gossypium_hirsutum
	TA21808_3635	3	CAG	18	6	Plus	Gossypium_hirsutum
	TA21810_3635	3	CAG	18	6	Plus	Gossypium_hirsutum
18	BG443126	3	CAG	18	6	Plus	Gossypium_arboreum
	DW244622	3	CAG	18	6	Plus	Gossypium_hirsutum
	TA21808_3635	3	CAG	18	6	Plus	Gossypium_hirsutum
	TA21810_3635	3	CAG	18	6	Plus	Gossypium_hirsutum
19	BQ414418	3	ATG	18	6	Plus	Gossypium_arboreum
	TA41967_3635	3	CAT	18	6	Minus	Gossypium_hirsutum
20	BG445378 Hypothetical protein [Cleome spinosa]	3	CAT	18	6	Plus	Gossypium_arboreum
	DR458947 Hypothetical protein [Cleome spinosa]	3	CAT	18	6	Plus	Gossypium_hirsutum
21	BE053330 Cysteine proteinase [Gossypium hirsutum (Upland cotton)]	3	CAT	18	6	Plus	Gossypium_arboreum
	TA21017_3635 Cysteine proteinase [Gossypium hirsutum (Upland cotton)]	3	CAT	18	6	Plus	Gossypium_hirsutum
22	BQ412920	3	CAT	18	6	Plus	Gossypium_arboreum
	DW514597	3	TGA	18	6	Minus	Gossypium_hirsutum
23	BG442727 Cysteine proteinase [Gossypium hirsutum (Upland cotton)]	3	CAT	18	6	Plus	Gossypium_arboreum
	TA21017_3635 Cysteine proteinase [Gossypium hirsutum (Upland cotton)]	3	CAT	18	6	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

24	BQ411009 Hypothetical protein [Cleome spinosa]	3	CAT	18	6	Plus	Gossypium_arboreum
	DR458947 Hypothetical protein [Cleome spinosa]	3	CAT	18	6	Plus	Gossypium_hirsutum
25	BG442903 Cysteine proteinase [Gossypium hirsutum (Upland cotton)]	3	CAT	18	6	Plus	Gossypium_arboreum
	TA21017_3635 Cysteine proteinase [Gossypium hirsutum (Upland cotton)]	3	CAT	18	6	Plus	Gossypium_hirsutum
26	BE052471	3	CAT	18	6	Plus	Gossypium_arboreum
	TA33369_3635	3	CAT	18	6	Plus	Gossypium_hirsutum
27	BQ410999	3	CAT	18	6	Plus	Gossypium_arboreum
	DR453293	3	GAT	18	6	Minus	Gossypium_hirsutum
	DT052017	3	CAT	18	6	Plus	Gossypium_hirsutum
28	BG441669 Cysteine proteinase [Gossypium hirsutum (Upland cotton)]	3	CAT	18	6	Plus	Gossypium_arboreum
	TA21017_3635 Cysteine proteinase [Gossypium hirsutum (Upland cotton)]	3	CAT	18	6	Plus	Gossypium_hirsutum
29	AW726061	3	CAT	18	6	Plus	Gossypium_arboreum
	TA21017_3635 Cysteine proteinase [Gossypium hirsutum (Upland cotton)]	3	CAT	18	6	Plus	Gossypium_hirsutum
30	BQ410319 Elicitor inducible protein [Nicotiana tabacum (Common tobacco)]	3	CAT	18	6	Plus	Gossypium_arboreum
	TA22715_3635 F9F8.20 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	CAT	18	6	Plus	Gossypium_hirsutum
	DW514790	3	CAT	18	6	Plus	Gossypium_hirsutum
31	TA7168_29729 Wound-induced protein 1 [Solanum tuberosum (Potato)]	3	CAT	18	6	Plus	Gossypium_arboreum
	DW514790	3	CAT	18	6	Plus	Gossypium_hirsutum
	TA22715_3635 F9F8.20 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	CAT	18	6	Plus	Gossypium_hirsutum
32	TA5303_29729 Cysteine proteinase [Gossypium hirsutum (Upland cotton)]	3	CAT	18	6	Plus	Gossypium_arboreum
	TA21017_3635 Cysteine proteinase [Gossypium hirsutum (Upland cotton)]	3	CAT	18	6	Plus	Gossypium_hirsutum
33	BQ403789 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_arboreum
	DN800615 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_hirsutum
	TA23709_3635 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_hirsutum
34	BG443488 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_arboreum
	DN800615 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	TA23709_3635 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_hirsutum
35	BF276058 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_arboreum
	DN800615 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_hirsutum
	TA23709_3635 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_hirsutum
36	BQ406649	3	CCA	18	6	Plus	Gossypium_arboreum
	TA20324_3635 Similarity to N-acetyltransferase [Arabidopsis thaliana (Mouse-ear cress)]	3	CAC	18	6	Plus	Gossypium_hirsutum
37	BE052283 Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K14A3 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	18	6	Plus	Gossypium_arboreum
	DW496704 Hypothetical protein [Brassica oleracea (Wild cabbage)]	3	CTT	18	6	Plus	Gossypium_hirsutum
38	TA7685_29729 Non-green plastid inner envelope membrane protein precursor [Brassica oleracea (Wild cabbage)]	3	CTT	21	7	Plus	Gossypium_arboreum
	TA24688_3635 Putative non-green plastid inner envelope membrane protein [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	21	7	Plus	Gossypium_hirsutum
39	BE054572 Calnexin [Glycine max (Soybean)]	3	GAA	21	7	Plus	Gossypium_arboreum
	TA22700_3635 Calnexin [Glycine max (Soybean)]	3	GAA	21	7	Plus	Gossypium_hirsutum
40	BG446521 Hypothetical protein T16K5.150 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_arboreum
	DT546230	3	GAC	18	6	Plus	Gossypium_hirsutum
	TA23939_3635 Predicted protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_hirsutum
41	BG446506 Predicted protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_arboreum
	DT546230	3	GAC	18	6	Plus	Gossypium_hirsutum
	TA23939_3635 Predicted protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_hirsutum
42	BG442800 AT5g64160/MHJ24_14 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_arboreum
	TA34306_3635 AT5g64160/MHJ24_14 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_hirsutum
43	BQ401418 Hypothetical protein OJ1288_G09.18 [Oryza sativa (japonica cultivar-group)]	3	GAC	18	6	Plus	Gossypium_arboreum
	TA34306_3635 AT5g64160/MHJ24_14 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_hirsutum
44	BQ406180 F2D10.15 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_arboreum
	DT562604 F2D10.15 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_hirsutum
	TA33154_3635 Hypothetical protein P0489D11.5 [Oryza sativa (japonica cultivar-group)]	3	GAC	18	6	Plus	Gossypium_hirsutum
45	BQ411879	3	GAC	18	6	Plus	Gossypium_arboreum

Supplementary Table 3. Contd.

	TA34306_3635 AT5g64160/MHJ24_14 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_hirsutum
46	BM359245 Putative tumor differentially expressed protein 1 [Oryza sativa (japonica cultivar-group)]	3	GAG	18	6	Plus	Gossypium_arboreum
	TA22207_3635 T24D18.26 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAG	18	6	Plus	Gossypium_hirsutum
47	BQ402314 T24D18.26 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAG	18	6	Plus	Gossypium_arboreum
	TA22207_3635 T24D18.26 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAG	18	6	Plus	Gossypium_hirsutum
48	CB066413 Putative tumor differentially expressed protein 1 [Oryza sativa (japonica cultivar-group)]	3	GAG	18	6	Plus	Gossypium_arboreum
	TA22207_3635 T24D18.26 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAG	18	6	Plus	Gossypium_hirsutum
49	BQ410998	3	GAT	18	6	Plus	Gossypium_arboreum
	DR453293	3	GAT	18	6	Plus	Gossypium_hirsutum
50	BG445086	3	GAT	18	6	Plus	Gossypium_arboreum
	DT555558 Gb|AAF36750.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_hirsutum
	TA31218_3635 Gb|AAF36750.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_hirsutum
51	BF270416 BZIP protein BZO2H2 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_arboreum
	TA26195_3635	3	GAT	18	6	Plus	Gossypium_hirsutum
52	BF276904 Putative 7-transmembrane G-protein-coupled receptor [Solanum chacoense (Chaco potato)]	3	GGC	18	6	Plus	Gossypium_arboreum
	TA33755_3635 Putative 7-transmembrane G-protein-coupled receptor [Solanum chacoense (Chaco potato)]	3	GGC	18	6	Plus	Gossypium_hirsutum
53	BG443893	3	GGC	18	6	Plus	Gossypium_arboreum
	TA28715_3635	3	GCC	18	6	Minus	Gossypium_hirsutum
54	BM359726	3	TCT	18	6	Plus	Gossypium_arboreum
	DT460417 Homeodomain protein Hfi22 [Nicotiana tabacum (Common tobacco)]	3	GAA	18	6	Minus	Gossypium_hirsutum
55	TA8283_29729 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TCT	18	6	Plus	Gossypium_arboreum
	TA32679_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TCT	18	6	Plus	Gossypium_hirsutum
56	BG442229 Von Willebrand factor, type A; Ubiquitin interacting motif [Medicago truncatula (Barrel medic)]	3	TGC	18	6	Plus	Gossypium_arboreum
	TA22965_3635 Von Willebrand factor, type A; Ubiquitin interacting motif [Medicago truncatula (Barrel medic)]	3	TGC	18	6	Plus	Gossypium_hirsutum
	DT571041 Von Willebrand factor, type A; Ubiquitin interacting motif [Medicago truncatula (Barrel medic)]	3	TGC	18	6	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

57	TA5246_29729 Fatty acid elongase [Persea americana (Avocado)]	3	TTC	18	6	Plus	Gossypium_arboreum
	DN799958 Fatty acid elongase [Persea americana (Avocado)]	3	TTC	18	6	Plus	Gossypium_hirsutum
	DT466005 Fatty acid elongase [Persea americana (Avocado)]	3	TTC	18	6	Plus	Gossypium_hirsutum
	TA20571_3635 Putative fatty acid elongase [Zea mays (Maize)]	3	TTC	18	6	Plus	Gossypium_hirsutum
	TA20576_3635 Putative beta-ketoacyl-CoA synthase [Oryza sativa (japonica cultivar-group)]	3	TTC	18	6	Plus	Gossypium_hirsutum
58	TA9370_29729 Fatty acid elongase [Persea americana (Avocado)]	3	TTC	18	6	Plus	Gossypium_arboreum
	DN799958 Fatty acid elongase [Persea americana (Avocado)]	3	TTC	18	6	Plus	Gossypium_hirsutum
	TA20571_3635 Putative fatty acid elongase [Zea mays (Maize)]	3	TTC	18	6	Plus	Gossypium_hirsutum
	DT466005 Fatty acid elongase [Persea americana (Avocado)]	3	TTC	18	6	Plus	Gossypium_hirsutum
	TA20576_3635 Putative beta-ketoacyl-CoA synthase [Oryza sativa (japonica cultivar-group)]	3	TTC	18	6	Plus	Gossypium_hirsutum
59	BF274447	3	TTG	18	6	Plus	Gossypium_arboreum
	TA28224_3635 Cyclin-like F-box [Medicago truncatula (Barrel medic)]	3	TTG	18	6	Plus	Gossypium_hirsutum
	DT573761 Cyclin-like F-box [Medicago truncatula (Barrel medic)]	3	TTG	18	6	Plus	Gossypium_hirsutum
60	BQ405038 Hypothetical protein T5J17.10 [Arabidopsis thaliana (Mouse-ear cress)]	4	AATT	20	5	Plus	Gossypium_arboreum
	DW235242	4	AATT	20	5	Plus	Gossypium_hirsutum
	DV850022 Hypothetical protein [Medicago truncatula (Barrel medic)]	4	AATT	20	5	Plus	Gossypium_hirsutum
61	BQ407187 Putative 1-aminocyclopropane-1-carboxylate deaminase [Oryza sativa (japonica cultivar-group)]	4	ACAT	36	9	Plus	Gossypium_arboreum
	TA28634_3635 Putative 1-aminocyclopropane-1-carboxylate deaminase [Oryza sativa (japonica cultivar-group)]	4	ACAT	36	9	Plus	Gossypium_hirsutum
62	BQ415843 Putative 1-aminocyclopropane-1-carboxylate deaminase [Oryza sativa (japonica cultivar-group)]	4	ACAT	36	9	Plus	Gossypium_arboreum
	TA28634_3635 Putative 1-aminocyclopropane-1-carboxylate deaminase [Oryza sativa (japonica cultivar-group)]	4	ACAT	36	9	Plus	Gossypium_hirsutum
63	BQ414500 F11A17.2 protein [Arabidopsis thaliana (Mouse-ear cress)]	4	ACAT	36	9	Plus	Gossypium_arboreum
	TA28634_3635 Putative 1-aminocyclopropane-1-carboxylate deaminase [Oryza sativa (japonica cultivar-group)]	4	ACAT	36	9	Plus	Gossypium_hirsutum
64	BF278325 Copper chaperone [Populus alba x Populus tremula var. glandulosa]	4	ATTG	24	6	Plus	Gossypium_arboreum
	TA28375_3635 Copper chaperone [Populus alba x Populus tremula var. glandulosa]	4	ATTG	24	6	Plus	Gossypium_hirsutum
	DW499145 Copper chaperone [Populus alba x Populus tremula var. glandulosa]	4	ATTG	24	6	Plus	Gossypium_hirsutum
65	BM357948 Mitogen-activated protein kinase [Euphorbia esula (Leafy spurge)]	4	TATG	20	5	Plus	Gossypium_arboreum
	TA23495_3635 Mitogen-activated protein kinase 1 [Lycopersicon esculentum (Tomato)]	4	TATG	20	5	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

66	TA8669_29729 Putative CDP-diacylglycerol--glycerol-3-phosphate 3- phosphatidyltransferase [Arabidopsis thaliana (Mouse-ear cress)]	6	CCTCCA	24	4	Plus	Gossypium_arboreum
	TA31034_3635 Putative CDP-diacylglycerol--glycerol-3-phosphate 3- phosphatidyltransferase [Arabidopsis thaliana (Mouse-ear cress)]	6	CCTCCA	24	4	Plus	Gossypium_hirsutum
67	BF275274 GT-like trihelix DNA-binding protein, putative [Arabidopsis thaliana (Mouse-ear cress)]	6	GAAGAT	24	4	Plus	Gossypium_arboreum
	DR459252	6	GAAGAT	24	4	Plus	Gossypium_hirsutum
68	BG446849 GT-like trihelix DNA-binding protein, putative [Arabidopsis thaliana (Mouse-ear cress)]	6	GAAGAT	24	4	Plus	Gossypium_arboreum
	DR459252	6	GAAGAT	24	4	Plus	Gossypium_hirsutum
69	BG445177	6	GATAGG	24	4	Plus	Gossypium_arboreum
	TA37112_3635 Zinc finger, RING-type; RINGv [Medicago truncatula (Barrel medic)]	6	GATAGG	24	4	Plus	Gossypium_hirsutum
70	BG445815 T5A14.7 protein [Arabidopsis thaliana (Mouse-ear cress)]	6	GATAGG	24	4	Plus	Gossypium_arboreum
	TA37112_3635 Zinc finger, RING-type; RINGv [Medicago truncatula (Barrel medic)]	6	GATAGG	24	4	Plus	Gossypium_hirsutum
71	BM359711 Hypothetical protein T8M16_190 [Arabidopsis thaliana (Mouse-ear cress)]	6	GGAAGA	24	4	Plus	Gossypium_arboreum
	TA34184_3635 RNA-binding protein AKIP1-like [Solanum tuberosum (Potato)]	6	GGAAGA	24	4	Plus	Gossypium_hirsutum
72	BG444556 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044185 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
73	BF279254	6	TGCAAT	24	4	Plus	Gossypium_arboreum
	TA44674_3635 Protein At3g56830 [Arabidopsis thaliana (Mouse-ear cress)]	6	CAATTG	24	4	Plus	Gossypium_hirsutum
74	TA9299_29729 Predicted protein [Populus alba x Populus tremula]	6	TTTTGG	24	4	Plus	Gossypium_arboreum
	TA40993_3635 Predicted protein [Populus alba x Populus tremula]	6	TGGTTT	24	4	Plus	Gossypium_hirsutum

Group II

HG	Sequence ID	Period Size	Period	Rep. Length	Copy Num.	Blast Strand	Species
1	BG441478 Thioredoxin domain-containing protein 9 homolog [Arabidopsis thaliana (Mouse-ear cress)]	2	AC	20	10	Plus	Gossypium_arboreum

Supplementary Table 3. Contd.

	TA29754_3635 Putative ATP binding protein associated with cell differentiation; Protein 1-4 [Oryza sativa (japonica cultivar-group)]	2	AC	18	9	Plus	Gossypium_hirsutum
2	TA6183_29729 Eukaryotic translation initiation factor 3 subunit 12 [Oryza sativa (Rice)]	2	AT	26	13	Plus	Gossypium_arboreum
	TA23639_3635 Eukaryotic translation initiation factor 3 subunit 12 [Oryza sativa (Rice)]	2	AT	24	12	Plus	Gossypium_hirsutum
3	BF272268 OSJNBb0039L24.7 protein [Oryza sativa (japonica cultivar-group)]	2	CT	30	15	Plus	Gossypium_arboreum
	DR458537 Putative LIM domain protein PLIM-2 [Oryza sativa (japonica cultivar-group)]	2	CT	22	11	Plus	Gossypium_hirsutum
	AI054962 Putative LIM domain protein PLIM-2 [Oryza sativa (japonica cultivar-group)]	2	CT	22	11	Plus	Gossypium_hirsutum
4	TA5890_29729 LIM-domain SF3 protein [Nicotiana tabacum (Common tobacco)]	2	CT	30	15	Plus	Gossypium_arboreum
	DR458537 Putative LIM domain protein PLIM-2 [Oryza sativa (japonica cultivar-group)]	2	CT	22	11	Plus	Gossypium_hirsutum
	AI054962 Putative LIM domain protein PLIM-2 [Oryza sativa (japonica cultivar-group)]	2	CT	22	11	Plus	Gossypium_hirsutum
5	BG445759	2	TA	24	12	Plus	Gossypium_arboreum
	DT548108	2	TA	18	9	Plus	Gossypium_hirsutum
6	TA6309_29729 E-class P450, group I [Medicago truncatula (Barrel medic)]	2	TA	28	14	Plus	Gossypium_arboreum
	TA23514_3635 E-class P450, group I [Medicago truncatula (Barrel medic)]	2	TA	26	13	Plus	Gossypium_hirsutum
7	BM359473 F20N2.15 [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	26	13	Plus	Gossypium_arboreum
	DR456327 BTB/POZ; TRAF-like [Medicago truncatula (Barrel medic)]	2	TC	18	9	Plus	Gossypium_hirsutum
8	BG443314 F20N2.15 [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	26	13	Plus	Gossypium_arboreum
	DR456327 BTB/POZ; TRAF-like [Medicago truncatula (Barrel medic)]	2	TC	18	9	Plus	Gossypium_hirsutum
9	BF272117 Gb|AAF02129.1 [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	26	13	Plus	Gossypium_arboreum
	DT051614 Gb|AAF02129.1 [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	24	12	Plus	Gossypium_hirsutum
10	TA7156_29729 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	TC	28	14	Plus	Gossypium_arboreum
	DN802296	2	CT	20	10	Plus	Gossypium_hirsutum
	TA35543_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	TC	22	11	Plus	Gossypium_hirsutum
11	BG446768 T16O11.17 protein [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	20	10	Plus	Gossypium_arboreum
	TA24594_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	TC	18	9	Plus	Gossypium_hirsutum
12	BF268204	2	TC	28	14	Plus	Gossypium_arboreum

Supplementary Table 3. Contd.

	DN802296	2	CT	20	10	Plus	Gossypium_hirsutum
	TA35543_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	TC	22	11	Plus	Gossypium_hirsutum
13	BE052113	2	CT	26	13	Plus	Gossypium_arboreum
	TA30762_3635 Targeting protein for Xklp2 containing protein, expressed [Oryza sativa (japonica cultivar-group)]	2	TC	22	11	Plus	Gossypium_hirsutum
14	BF272287 Hypothetical protein At3g56480/T5P19_130 [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	26	13	Plus	Gossypium_arboreum
	TA33492_3635 Hypothetical protein At3g56480/T5P19_130 [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	20	10	Plus	Gossypium_hirsutum
	DR044137	2	TC	22	11	Plus	Gossypium_hirsutum
15	BF274071 RNA polymerase II transcriptional coactivator KELP [Arabidopsis thaliana (Mouse-ear cress)]	3	AAG	24	8	Plus	Gossypium_arboreum
	DR455976	3	AAG	21	7	Plus	Gossypium_hirsutum
	TA26990_3635 RNA polymerase II transcriptional coactivator KELP [Arabidopsis thaliana (Mouse-ear cress)]	3	AAG	21	7	Plus	Gossypium_hirsutum
	TA26991_3635 RNA polymerase II transcriptional coactivator KELP [Arabidopsis thaliana (Mouse-ear cress)]	3	AAG	21	7	Plus	Gossypium_hirsutum
16	BF269744 Putative epoxide hydrolase [Arabidopsis thaliana (Mouse-ear cress)]	3	AAT	21	7	Plus	Gossypium_arboreum
	TA41918_3635 Soluble epoxide hydrolase [Citrus jambhiri (rough lemon)]	3	AAT	18	6	Plus	Gossypium_hirsutum
17	TA8375_29729 Sucrose synthase [Gossypium hirsutum (Upland cotton)]	3	ATA	39	13	Plus	Gossypium_arboreum
	TA23143_3635 Sucrose synthase [Gossypium hirsutum (Upland cotton)]	3	ATA	30	10	Plus	Gossypium_hirsutum
18	BQ408202 Hypothetical protein At3g26700/MLJ15_10 [Arabidopsis thaliana (Mouse-ear cress)]	3	ATT	21	7	Plus	Gossypium_arboreum
	DR457741 Protein kinase-like protein [Cicer arietinum (Chickpea) (Garbanzo)]	3	ATT	18	6	Plus	Gossypium_hirsutum
	DT462889 Hypothetical protein At3g26700/MLJ15_10 [Arabidopsis thaliana (Mouse-ear cress)]	3	ATT	21	7	Plus	Gossypium_hirsutum
19	BF269259	3	CAG	21	7	Plus	Gossypium_arboreum
	TA37795_3635	3	TGC	18	6	Minus	Gossypium_hirsutum
20	BF270454 Nascent polypeptide-associated complex NAC; UBA-like [Medicago truncatula (Barrel medic)]	3	GAC	21	7	Plus	Gossypium_arboreum
	AI731240 Nascent polypeptide-associated complex NAC; UBA-like [Medicago truncatula (Barrel medic)]	3	GAC	18	6	Plus	Gossypium_hirsutum
	DR457661 Nascent polypeptide-associated complex subunit alpha-like protein 5 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_hirsutum
	DN779978	3	GAC	18	6	Plus	Gossypium_hirsutum
	CD485985 Nascent polypeptide-associated complex subunit alpha-like protein 5 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_hirsutum
	AI730170 Nascent polypeptide-associated complex subunit alpha-like protein 5 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	21	7	Plus	Gossypium_hirsutum
	DR044193 Nascent polypeptide-associated complex subunit alpha-like protein 5 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	21	7	Plus	Gossypium_hirsutum
	TA22487_3635 Nascent polypeptide-associated complex NAC; UBA-like [Medicago truncatula (Barrel medic)]	3	GAC	21	7	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

21	BE053470 Nascent polypeptide associated complex alpha chain [Oryza sativa (japonica cultivar-group)]	3	GAC	21	7	Plus	Gossypium_arboreum
	AI731240 Nascent polypeptide-associated complex NAC; UBA-like [Medicago truncatula (Barrel medic)]	3	GAC	18	6	Plus	Gossypium_hirsutum
	DR457661 Nascent polypeptide-associated complex subunit alpha-like protein 5 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_hirsutum
	DN779978	3	GAC	18	6	Plus	Gossypium_hirsutum
	CD485985 Nascent polypeptide-associated complex subunit alpha-like protein 5 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	18	6	Plus	Gossypium_hirsutum
	AI730170 Nascent polypeptide-associated complex subunit alpha-like protein 5 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	21	7	Plus	Gossypium_hirsutum
	DR044193 Nascent polypeptide-associated complex subunit alpha-like protein 5 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAC	21	7	Plus	Gossypium_hirsutum
	TA22487_3635 Nascent polypeptide-associated complex NAC; UBA-like [Medicago truncatula (Barrel medic)]	3	GAC	21	7	Plus	Gossypium_hirsutum
22	BG440543	3	GCC	21	7	Plus	Gossypium_arboreum
	TA27320_3635 F10K1.20 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GCC	18	6	Plus	Gossypium_hirsutum
23	BM358391 Hypothetical protein P0015C02.10 [Oryza sativa (japonica cultivar-group)]	3	GCC	21	7	Plus	Gossypium_arboreum
	TA27320_3635 F10K1.20 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GCC	18	6	Plus	Gossypium_hirsutum
24	BG445086	3	GGC	21	7	Plus	Gossypium_arboreum
	DT555558 Gb|AAF36750.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	GGC	18	6	Plus	Gossypium_hirsutum
	TA31218_3635 Gb|AAF36750.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	GCG	24	8	Plus	Gossypium_hirsutum
25	BQ405592 unknown protein [Arabidopsis thaliana]	3	GTG	21	7	Plus	Gossypium_arboreum
	TA38018_3635 unknown protein [Arabidopsis thaliana]	3	GTG	18	6	Plus	Gossypium_hirsutum
26	BQ415371	3	GTT	24	8	Plus	Gossypium_arboreum
	TA22290_3635 Myb, DNA-binding [Medicago truncatula (Barrel medic)]	3	CAA	21	7	Minus	Gossypium_hirsutum
27	CK640498 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	24	8	Plus	Gossypium_arboreum
	AJ513319 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DT047070	3	TCA	21	7	Minus	Gossypium_hirsutum
	TA20630_3635 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DW225975 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	GAT	21	7	Plus	Gossypium_hirsutum
	DW499141 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	TA42103_3635	3	TGA	24	8	Plus	Gossypium_hirsutum
	CO493420 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	24	8	Plus	Gossypium_hirsutum
28	TA6371_29729 T7I23.17 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	30	10	Plus	Gossypium_arboreum
	TA25576_3635 T7I23.17 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	27	9	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

29	BQ402724	3	CAA	21	7	Plus	Gossypium_arboreum
	AI055346 Helicase, C-terminal [Medicago truncatula (Barrel medic)]	3	CAA	18	6	Plus	Gossypium_hirsutum
	TA27181_3635 DEAD-box ATP-dependent RNA helicase 8 [Arabidopsis thaliana (Mouse-ear cress)]	3	CAA	18	6	Plus	Gossypium_hirsutum
30	TA6708_29729	3	CAG	33	11	Plus	Gossypium_arboreum
	TA22510_3635	3	CAG	27	9	Plus	Gossypium_hirsutum
31	BQ411825 Phospholipase D delta isoform 1a [Gossypium hirsutum (Upland cotton)]	4	GTAT	40	10	Plus	Gossypium_arboreum
	TA20281_3635 Phospholipase D delta isoform 1a [Gossypium hirsutum (Upland cotton)]	4	TGTA	28	7	Plus	Gossypium_hirsutum
	TA20264_3635 Phospholipase D delta isoform 1a [Gossypium hirsutum (Upland cotton)]	4	TGTA	40	10	Plus	Gossypium_hirsutum
	TA20247_3635 Phospholipase D delta isoform 1a [Gossypium hirsutum (Upland cotton)]	4	TGTA	40	10	Plus	Gossypium_hirsutum

Group III

HG	**Sequence ID**	**Period Size**	**Period**	**Rep. Length**	**Copy Num.**	**Blast Strand**	**Species**
1	TA6910_29729	2	AG	18	9	Plus	Gossypium_arboreum
	DR462523	2	TC	18	9	Minus	Gossypium_hirsutum
	DT571123	2	TC	24	12	Minus	Gossypium_hirsutum
2	BQ407993	2	AG	18	9	Plus	Gossypium_arboreum
	DR462523	2	TC	18	9	Minus	Gossypium_hirsutum
	DT571123	2	TC	24	12	Minus	Gossypium_hirsutum
3	TA7299_29729 F6I1.14 protein [Arabidopsis thaliana (Mouse-ear cress)]	2	AT	18	9	Plus	Gossypium_arboreum
	TA44152_3635 F6I1.14 protein [Arabidopsis thaliana (Mouse-ear cress)]	2	AT	22	11	Plus	Gossypium_hirsutum
4	BF270284 Pentaxin [Medicago truncatula (Barrel medic)]	2	AT	18	9	Plus	Gossypium_arboreum
	TA44152_3635 F6I1.14 protein [Arabidopsis thaliana (Mouse-ear cress)]	2	AT	22	11	Plus	Gossypium_hirsutum
5	BG440327	2	GA	18	9	Plus	Gossypium_arboreum
	TA34307_3635 Gb|AAC24081.1 [Arabidopsis thaliana (Mouse-ear cress)]	2	AG	30	15	Plus	Gossypium_hirsutum
	DT574307 Gb|AAC24081.1 [Arabidopsis thaliana (Mouse-ear cress)]	2	AG	30	15	Plus	Gossypium_hirsutum
6	BG444416	2	TA	18	9	Plus	Gossypium_arboreum
	TA21656_3635 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049695 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	DT049714 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT053439 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053492 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053135 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053298 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
7	BE054872 Snakin-1 [Solanum tuberosum (Potato)]	2	TA	18	9	Plus	Gossypium_arboreum
	TA21656_3635 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049695 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049714 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT053439 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053492 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053298 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053135 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
8	BG444964 Snakin-1 [Solanum tuberosum (Potato)]	2	TA	18	9	Plus	Gossypium_arboreum
	TA21656_3635 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049695 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049714 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT053439 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053135 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053298 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053492 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
9	BE055619	2	TA	18	9	Plus	Gossypium_arboreum
	TA21656_3635 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049714 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049695 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT053298 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053492 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053439 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053135 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
10	BG443157 Snakin-1 [Solanum tuberosum (Potato)]	2	TA	18	9	Plus	Gossypium_arboreum
	TA21656_3635 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049695 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049714 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT053439 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053298 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	DT053492 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053135 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
11	BG441636 Snakin-1 [Solanum tuberosum (Potato)]	2	TA	18	9	Plus	Gossypium_arboreum
	TA21656_3635 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049714 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049695 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT053298 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053492 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053439 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053135 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
12	TA5092_29729 Snakin-1 [Solanum tuberosum (Potato)]	2	TA	18	9	Plus	Gossypium_arboreum
	TA21656_3635 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049695 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT049714 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT053439 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053492 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053298 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
	DT053135 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
13	BQ407767	2	TA	18	9	Plus	Gossypium_arboreum
	DT049714 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	TA21656_3635 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	18	9	Plus	Gossypium_hirsutum
	DT053439 Snakin-1 [Solanum tuberosum (Potato)]	2	AT	20	10	Plus	Gossypium_hirsutum
14	TA7596_29729 Hypothetical protein F12A12.60 [Arabidopsis thaliana (Mouse-ear cress)]	2	TA	18	9	Plus	Gossypium_arboreum
	TA7596_29729 Hypothetical protein F12A12.60 [Arabidopsis thaliana (Mouse-ear cress)]	2	AT	20	10	Plus	Gossypium_arboreum
	DR463781	2	TA	20	10	Plus	Gossypium_hirsutum
	DT456861	2	TA	22	11	Minus	Gossypium_hirsutum
15	TA6055_29729 CAMP response element binding (CREB) protein [Medicago truncatula (Barrel medic)]	2	TA	18	9	Plus	Gossypium_arboreum
	DR463833 CAMP response element binding (CREB) protein [Medicago truncatula (Barrel medic)]	2	TA	22	11	Plus	Gossypium_hirsutum
	DT049066 BZIP transcription factor ATB2 [Glycine max (Soybean)]	2	TA	22	11	Plus	Gossypium_hirsutum
16	BQ412586	2	TA	18	9	Plus	Gossypium_arboreum
	DR463781	2	TA	20	10	Minus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	DT456861	2	TA	22	11	Plus	Gossypium_hirsutum
17	BF270284 Pentaxin [Medicago truncatula (Barrel medic)]	2	TC	18	9	Plus	Gossypium_arboreum
	TA44152_3635 F6I1.14 protein [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	22	11	Plus	Gossypium_hirsutum
18	TA6479_29729 MYB transcription factor [Cucumis sativus (Cucumber)]	3	AAC	18	6	Plus	Gossypium_arboreum
	TA42751_3635	3	AAC	18	6	Plus	Gossypium_hirsutum
	DW008269 MYB transcription factor [Cucumis sativus (Cucumber)]	3	AAC	21	7	Plus	Gossypium_hirsutum
19	TA6818_29729 T1N6.1 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AAG	21	7	Plus	Gossypium_arboreum
	AI731713 T1N6.1 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AAG	24	8	Plus	Gossypium_hirsutum
20	BF276199 Putative early nodule-specific-like protein ENOD8 [Oryza sativa (japonica cultivar-group)]	3	AAG	21	7	Plus	Gossypium_arboreum
	TA35757_3635 Putative early nodule-specific-like protein ENOD8 [Oryza sativa (japonica cultivar-group)]	3	AAG	24	8	Plus	Gossypium_hirsutum
21	TA5263_29729 Cytochrome c1, heme protein, mitochondrial precursor [Solanum tuberosum (Potato)]	3	ACC	18	6	Plus	Gossypium_arboreum
	TA24730_3635 Putative ubiquinol-cytochrome-c reductase cytochrome c1 [Oryza sativa (japonica cultivar-group)]	3	ACC	24	8	Plus	Gossypium_hirsutum
22	BE055322	3	AGC	21	7	Plus	Gossypium_arboreum
	TA21426_3635	3	AGC	27	9	Plus	Gossypium_hirsutum
23	BE054201	3	AGC	21	7	Plus	Gossypium_arboreum
	TA21426_3635	3	AGC	27	9	Plus	Gossypium_hirsutum
24	BG441749	3	AGC	21	7	Plus	Gossypium_arboreum
	TA21426_3635	3	AGC	27	9	Plus	Gossypium_hirsutum
25	BM359805	3	AGC	21	7	Plus	Gossypium_arboreum
	TA21426_3635	3	AGC	27	9	Plus	Gossypium_hirsutum
26	BG443578	3	AGC	21	7	Plus	Gossypium_arboreum
	TA21426_3635	3	AGC	27	9	Plus	Gossypium_hirsutum
27	BG443059	3	AGC	21	7	Plus	Gossypium_arboreum
	TA21426_3635	3	AGC	27	9	Plus	Gossypium_hirsutum
28	BG441547	3	AGC	21	7	Plus	Gossypium_arboreum

Supplementary Table 3. Contd.

	TA21426_3635	3	AGC	27	9	Plus	Gossypium_hirsutum
29	BG442685	3	AGC	21	7	Plus	Gossypium_arboreum
	TA21426_3635	3	AGC	27	9	Plus	Gossypium_hirsutum
30	TA5416_29729	3	AGC	21	7	Plus	Gossypium_arboreum
	TA21426_3635	3	AGC	27	9	Plus	Gossypium_hirsutum
31	BF270534	3	CAC	18	6	Plus	Gossypium_arboreum
	DT050816	3	TGG	21	7	Minus	Gossypium_hirsutum
32	BG446140	3	CAG	18	6	Plus	Gossypium_arboreum
	TA21426_3635	3	AGC	27	9	Plus	Gossypium_hirsutum
33	BE055500 MYC2 [Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle)]	3	CTC	18	6	Plus	Gossypium_arboreum
	TA21503_3635	3	CTC	21	7	Plus	Gossypium_hirsutum
34	BF276934 Hypothetical protein At2g44300 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTG	18	6	Plus	Gossypium_arboreum
	DR462320 Hypothetical protein At2g44300 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTG	21	7	Plus	Gossypium_hirsutum
35	BF275766 T7I23.17 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	24	8	Plus	Gossypium_arboreum
	TA25576_3635 T7I23.17 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	27	9	Plus	Gossypium_hirsutum
36	BQ405354 Ultraviolet-B-repressible protein [Gossypium hirsutum (Upland cotton)]	3	CTT	21	7	Plus	Gossypium_arboreum
	TA39593_3635 Ultraviolet-B-repressible protein [Gossypium hirsutum (Upland cotton)]	3	CTT	24	8	Plus	Gossypium_hirsutum
37	BF270826 Hypothetical protein T20N10_200 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAA	21	7	Plus	Gossypium_arboreum
	DW476508	3	TCT	21	7	Minus	Gossypium_hirsutum
	DW476507	3	GAA	21	7	Plus	Gossypium_hirsutum
	TA28486_3635 Hypothetical protein T20N10_200 [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	39	13	Plus	Gossypium_hirsutum
38	BQ407920 Auxin and ethylene responsive GH3-like protein [Capsicum chinense (Scotch bonnet) (Bonnet pepper)]	3	GAT	18	6	Plus	Gossypium_arboreum
	DW505072 Auxin and ethylene responsive GH3-like protein [Capsicum chinense (Scotch bonnet) (Bonnet pepper)]	3	GAT	18	6	Plus	Gossypium_hirsutum
	DT054325 Probable indole-3-acetic acid-amido synthetase GH3.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	21	7	Plus	Gossypium_hirsutum
	TA21526_3635 Nt-gh3 deduced protein [Nicotiana tabacum (Common tobacco)]	3	GAT	21	7	Plus	Gossypium_hirsutum
39	TA7199_29729 Auxin and ethylene responsive GH3-like protein [Capsicum chinense (Scotch bonnet) (Bonnet pepper)]	3	GAT	18	6	Plus	Gossypium_arboreum

Supplementary Table 3. Contd.

	DW505072 Auxin and ethylene responsive GH3-like protein [Capsicum chinense (Scotch bonnet) (Bonnet pepper)]	3	GAT	18	6	Plus	Gossypium_hirsutum
	DT054325 Probable indole-3-acetic acid-amido synthetase GH3.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	21	7	Plus	Gossypium_hirsutum
	TA21526_3635 Nt-gh3 deduced protein [Nicotiana tabacum (Common tobacco)]	3	GAT	21	7	Plus	Gossypium_hirsutum
40	TA8195_29729 Phospho-2-dehydro-3-deoxyheptonate aldolase 2, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	3	GCG	18	6	Plus	Gossypium_arboreum
	TA32414_3635 Phospho-2-dehydro-3-deoxyheptonate aldolase 2, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	3	GCG	24	8	Plus	Gossypium_hirsutum
41	BM359886 Cleft lip and palate associated transmembrane protein-like [Arabidopsis thaliana (Mouse-ear cress)]	3	GGA	21	7	Plus	Gossypium_arboreum
	TA28051_3635 Cleft lip and palate associated transmembrane protein-like [Arabidopsis thaliana (Mouse-ear cress)]	3	GGA	24	8	Plus	Gossypium_hirsutum
	DT559190 Cleft lip and palate associated transmembrane protein-like [Arabidopsis thaliana (Mouse-ear cress)]	3	GGA	27	9	Plus	Gossypium_hirsutum
42	BF278307 Putative RING zinc finger protein-like protein [Thellungiella halophila (Salt cress)]	3	TCA	21	7	Plus	Gossypium_arboreum
	TA23903_3635 RING-H2 finger protein [Poncirus trifoliata (Hardy orange)]	3	TCA	24	8	Plus	Gossypium_hirsutum
43	BQ404554 Copper/zinc superoxide dismutase, putative [Medicago truncatula (Barrel medic)]	3	TCT	18	6	Plus	Gossypium_arboreum
	DT567072 Cu/Zn-superoxide dismutase copper chaperone precursor [Glycine max (Soybean)]	3	TCT	18	6	Plus	Gossypium_hirsutum
	TA23380_3635 Cu/Zn-superoxide dismutase copper chaperone precursor [Glycine max (Soybean)]	3	TCT	21	7	Plus	Gossypium_hirsutum
44	BG444426	3	TCT	24	8	Plus	Gossypium_arboreum
	TA21780_3635 At5g06390 [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	27	9	Plus	Gossypium_hirsutum
	DR044171 Gb|AAF02137.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	27	9	Plus	Gossypium_hirsutum
	TA21781_3635 Gb|AAF02137.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	27	9	Plus	Gossypium_hirsutum
45	TA5514_29729 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	18	6	Plus	Gossypium_arboreum
	AJ513319 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DW225975 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	GAT	21	7	Plus	Gossypium_hirsutum
	TA20630_3635 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DW499141 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DT047070	3	TCA	21	7	Minus	Gossypium_hirsutum
	CO493420 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	24	8	Plus	Gossypium_hirsutum
46	BQ410493 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	18	6	Plus	Gossypium_arboreum
	TA42103_3635	3	TGA	24	8	Plus	Gossypium_hirsutum
47	TA5823_29729 Histone deacetylase 2a [Solanum chacoense (Chaco potato)]	3	TGC	18	6	Plus	Gossypium_arboreum
	TA22749_3635	3	TGC	18	6	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	DW514526 Histone deacetylase 2a [Solanum chacoense (Chaco potato)]	3	TGC	18	6	Plus	Gossypium_hirsutum
	TA22748_3635	3	TGC	18	6	Plus	Gossypium_hirsutum
	TA22750_3635 Histone deacetylase 2a [Solanum chacoense (Chaco potato)]	3	GCT	21	7	Plus	Gossypium_hirsutum
48	BF270691 Histone deacetylase 2a [Solanum chacoense (Chaco potato)]	3	TGC	18	6	Plus	Gossypium_arboreum
	TA22749_3635	3	TGC	18	6	Plus	Gossypium_hirsutum
	DW514526 Histone deacetylase 2a [Solanum chacoense (Chaco potato)]	3	TGC	18	6	Plus	Gossypium_hirsutum
	TA22748_3635	3	TGC	18	6	Plus	Gossypium_hirsutum
	TA22750_3635 Histone deacetylase 2a [Solanum chacoense (Chaco potato)]	3	GCT	21	7	Plus	Gossypium_hirsutum
49	TA5985_29729 Protein At3g63460 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	18	6	Plus	Gossypium_arboreum
	TA26634_3635	3	TAA	18	6	Minus	Gossypium_hirsutum
	TA26636_3635	3	ATA	21	7	Minus	Gossypium_hirsutum
	TA26635_3635 Protein At3g63460 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	21	7	Plus	Gossypium_hirsutum
50	TA7048_29729 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	21	7	Plus	Gossypium_arboreum
	DR459816 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	27	9	Plus	Gossypium_hirsutum
	TA26423_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	27	9	Plus	Gossypium_hirsutum
51	BQ404857 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	21	7	Plus	Gossypium_arboreum
	DR459816 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	27	9	Plus	Gossypium_hirsutum
	TA26423_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	27	9	Plus	Gossypium_hirsutum
52	TA6604_29729	4	ATTT	20	5	Plus	Gossypium_arboreum
	TA33749_3635 Rapid alkalinization factor 2 [Solanum chacoense (Chaco potato)]	4	ATTT	20	5	Plus	Gossypium_hirsutum
	DT559940 Rapid alkalinization factor 2 [Solanum chacoense (Chaco potato)]	4	ATTT	24	6	Plus	Gossypium_hirsutum
53	BF268634 Rapid alkalinization factor 2 [Solanum chacoense (Chaco potato)]	4	ATTT	20	5	Plus	Gossypium_arboreum
	TA33749_3635 Rapid alkalinization factor 2 [Solanum chacoense (Chaco potato)]	4	ATTT	20	5	Plus	Gossypium_hirsutum
	DT559940 Rapid alkalinization factor 2 [Solanum chacoense (Chaco potato)]	4	ATTT	24	6	Plus	Gossypium_hirsutum
54	BG439886 Hypothetical protein [Medicago truncatula (Barrel medic)]	4	TGTA	20	5	Plus	Gossypium_arboreum
	TA31088_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	4	TGTA	24	6	Plus	Gossypium_hirsutum
55	BF272588 1-aminocyclopropane-1-carboxylate synthase [Momordica charantia (Bitter gourd) (Balsam pear)]	4	TTCT	20	5	Plus	Gossypium_arboreum
	TA23441_3635 1-aminocyclopropane-1-carboxylate synthase ACS6 [Gossypium hirsutum (Upland cotton)]	4	TTCT	24	6	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

56	BQ402637 Hypothetical protein F7A10.9 [Arabidopsis thaliana (Mouse-ear cress)]	4	TTTG	20	5	Plus	Gossypium_arboreum
	DW496475 Hypothetical protein F7A10.9 [Arabidopsis thaliana (Mouse-ear cress)]	4	TTTG	24	6	Plus	Gossypium_hirsutum
	TA29538_3635 Hypothetical protein F7A10.9 [Arabidopsis thaliana (Mouse-ear cress)]	4	TTTG	24	6	Plus	Gossypium_hirsutum
57	BQ408211 Protein At5g14170 [Arabidopsis thaliana (Mouse-ear cress)]	6	AAGCTC	24	4	Plus	Gossypium_arboreum
	DR453420	6	AAGCTC	30	5	Plus	Gossypium_hirsutum
58	BG440787 YT521-B-like protein [Medicago truncatula (Barrel medic)]	6	AAGGCA	30	5	Plus	Gossypium_arboreum
	DT554707 YT521-B-like protein [Medicago truncatula (Barrel medic)]	6	AAGGCA	36	6	Plus	Gossypium_hirsutum
59	BE054108	6	CCCCAA	24	4	Plus	Gossypium_arboreum
	TA25871_3635	6	CCCCAA	36	6	Plus	Gossypium_hirsutum
60	TA6665_29729	6	CCCCAA	24	4	Plus	Gossypium_arboreum
	TA25871_3635	6	CCCCAA	36	6	Plus	Gossypium_hirsutum
	TA25870_3635	6	ACCCCA	48	8	Plus	Gossypium_hirsutum
61	BG444093 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044185 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
62	BG444468 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
63	BG444074 At5g57100 [Arabidopsis thaliana (Mouse-ear cress)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DT571165 At5g57100 [Arabidopsis thaliana (Mouse-ear cress)]	6	CAGGCT	30	5	Plus	Gossypium_hirsutum
64	BG443843 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
65	BG443923 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044185 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
66	BG444500 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

67	BG445128 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
68	BG445281 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044185 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
69	BG446067 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
70	BG443565 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
71	BG445036 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
72	BG444921 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
73	BG444937 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
74	BG440521 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
75	BE054522 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
76	BE054651 E6-1 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
77	BE054976 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum

Supplementary Table 3. Contd.

	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
78	BE054144 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
79	BE053966 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
80	BE052527 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
81	BE053587 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044185 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
82	BE053730 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
83	BG442797 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044185 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
84	BE055230 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
85	TA4950_29729 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044185 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
6	BG441211 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044185 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
87	BE055232 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
88	BF272413 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
89	BE055623 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
90	BE055626 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum

Supplementary Table 3. Contd.

	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
91	BE055624 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_arboreum
	DR044210 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DR044175 Fb35 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20221_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20231_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	TA20267_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	24	4	Plus	Gossypium_hirsutum
	DT569706 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	CO499609 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	DR044205 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum
	TA20273_3635 Protein E6 [Gossypium hirsutum (Upland cotton)]	6	TCAGGC	30	5	Plus	Gossypium_hirsutum

Supplementary Table 4.

Group I

HG	Sequence ID	Period Size	Period	Rep. Length	Copy Num.	Blast Strand	Species
1	TA16319_29730 Emb|CAB87904.1 [Arabidopsis thaliana (Mouse-ear cress)]	2	AG	24	12	Plus	Gossypium_raimondii
	AI731677 Emb|CAB87904.1 [Arabidopsis thaliana (Mouse-ear cress)]	2	AG	24	12	Plus	Gossypium_hirsutum
	TA22041_3635 Hypothetical protein At3g26935 [Arabidopsis thaliana (Mouse-ear cress)]	2	AG	24	12	Plus	Gossypium_hirsutum
2	CO118048 Hypothetical protein At3g22280 [Arabidopsis thaliana (Mouse-ear cress)]	2	AT	20	10	Plus	Gossypium_raimondii
	DT047324	2	TA	20	10	Minus	Gossypium_hirsutum
3	TA9835_29730 F-box protein-like [Oryza sativa (japonica cultivar-group)]	2	CT	20	10	Plus	Gossypium_raimondii
	TA25552_3635 F-box protein-like [Oryza sativa (japonica cultivar-group)]	2	CT	20	10	Plus	Gossypium_hirsutum
	TA25551_3635 T3H13.1 protein [Arabidopsis thaliana (Mouse-ear cress)]	2	CT	20	10	Plus	Gossypium_hirsutum
4	TA30762_3635 Targeting protein for Xklp2 containing protein, expressed [Oryza sativa (japonica cultivar-group)]	2	TC	22	11	Plus	Gossypium_hirsutum
	TA12822_29730	2	CT	22	11	Plus	Gossypium_raimondii
5	CO079094 Hypothetical protein [Solanum tuberosum (Potato)]	2	TG	20	10	Plus	Gossypium_raimondii
	TA25279_3635 Hypothetical protein [Solanum tuberosum (Potato)]	2	TG	20	10	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

6	CO088220 Eukaryotic translation initiation factor 2 subunit beta [Triticum aestivum (Wheat)]	3	AAG	18	6	Plus	Gossypium_raimondii
	TA28888_3635 Eukaryotic translation initiation factor 2 beta subunit-like [Solanum tuberosum (Potato)]	3	AAG	18	6	Plus	Gossypium_hirsutum
7	CO084206	3	AAG	18	6	Plus	Gossypium_raimondii
	TA39376_3635 Hypothetical protein At4g35530 [Arabidopsis thaliana (Mouse-ear cress)]	3	AAG	18	6	Plus	Gossypium_hirsutum
8	TA14327_29730 Eukaryotic translation initiation factor 2 beta subunit-like [Solanum tuberosum (Potato)]	3	AAG	18	6	Plus	Gossypium_raimondii
	TA28888_3635 Eukaryotic translation initiation factor 2 beta subunit-like [Solanum tuberosum (Potato)]	3	AAG	18	6	Plus	Gossypium_hirsutum
9	TA10994_29730 DNA repair helicase XPB2 [Arabidopsis thaliana (Mouse-ear cress)]	3	AAG	18	6	Plus	Gossypium_raimondii
	CO495668 DNA repair helicase XPB1 [Arabidopsis thaliana (Mouse-ear cress)]	3	AAG	18	6	Plus	Gossypium_hirsutum
10	TA12940_29730 Esterase/lipase/thioesterase; Peptidase S9B, dipeptidylpeptidase IV N- terminal [Medicago truncatula (Barrel medic)]	3	AAT	18	6	Plus	Gossypium_raimondii
	DW506950 Putative dipeptidyl peptidase IV [Oryza sativa (japonica cultivar-group)]	3	AAT	18	6	Plus	Gossypium_hirsutum
11	TA10917_29730 T7N9.15 [Arabidopsis thaliana (Mouse-ear cress)]	3	ACC	27	9	Plus	Gossypium_raimondii
	TA23937_3635 T7N9.15 [Arabidopsis thaliana (Mouse-ear cress)]	3	ACC	27	9	Plus	Gossypium_hirsutum
12	TA16260_29730 Putative membrane transporter [Arabidopsis thaliana (Mouse-ear cress)]	3	ACC	18	6	Plus	Gossypium_raimondii
	TA21124_3635 Putative membrane transporter [Arabidopsis thaliana (Mouse-ear cress)]	3	ACC	18	6	Plus	Gossypium_hirsutum
13	TA13654_29730 SNF1 related protein kinase [Arabidopsis thaliana (Mouse-ear cress)]	3	ACC	18	6	Plus	Gossypium_raimondii
	TA32113_3635 SNF1 related protein kinase [Arabidopsis thaliana (Mouse-ear cress)]	3	ACC	18	6	Plus	Gossypium_hirsutum
14	CO096690 Hypothetical protein [Arabidopsis thaliana (Mouse-ear cress)]	3	ACC	18	6	Plus	Gossypium_raimondii
	TA31238_3635 Hypothetical protein [Arabidopsis thaliana (Mouse-ear cress)]	3	ACC	18	6	Plus	Gossypium_hirsutum
15	TA13896_29730 Contains similarity to ethylene responsive element binding factor [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	18	6	Plus	Gossypium_raimondii
	TA25013_3635 Ethylene-responsive transcription factor 5 [Nicotiana tabacum (Common tobacco)]	3	AGA	18	6	Plus	Gossypium_hirsutum
16	CO074836 TCP transcription factor [Medicago truncatula (Barrel medic)]	3	AGA	18	6	Plus	Gossypium_raimondii
	TA25064_3635 TCP transcription factor [Medicago truncatula (Barrel medic)]	3	AGA	18	6	Plus	Gossypium_hirsutum
17	TA12812_29730	3	AGA	18	6	Plus	Gossypium_raimondii
	DW235030	3	CTT	18	6	Minus	Gossypium_hirsutum
	TA22163_3635	3	CTT	18	6	Minus	Gossypium_hirsutum

Supplementary Table 4. Contd.

18	TA10286_29730 Gamma hydroxybutyrate dehydrogenase-like protein [Oryza sativa (japonica cultivar-group)]	3	AGA	18	6	Plus	Gossypium_raimondii
	DN804214 Gamma hydroxybutyrate dehydrogenase [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	18	6	Plus	Gossypium_hirsutum
19	CO117125 Telomere binding protein TBP1 [Nicotiana glutinosa (Tobacco)]	3	AGC	18	6	Plus	Gossypium_raimondii
	TA34162_3635 Telomere repeat-binding protein homolog [Arabidopsis thaliana (Mouse-ear cress)]	3	AGC	18	6	Plus	Gossypium_hirsutum
20	TA12177_29730 Telomere binding protein TBP1 [Nicotiana glutinosa (Tobacco)]	3	AGC	18	6	Plus	Gossypium_raimondii
	TA34162_3635 Telomere repeat-binding protein homolog [Arabidopsis thaliana (Mouse-ear cress)]	3	AGC	18	6	Plus	Gossypium_hirsutum
21	CO122920 Hypothetical protein P0456E05.23 [Oryza sativa (japonica cultivar-group)]	3	ATC	18	6	Plus	Gossypium_raimondii
	DR459558	3	TGA	18	6	Minus	Gossypium_hirsutum
22	CO128050 Hypothetical protein P0456E05.23 [Oryza sativa (japonica cultivar-group)]	3	ATC	18	6	Plus	Gossypium_raimondii
	DR459558	3	TGA	18	6	Minus	Gossypium_hirsutum
23	CO094728 Hypothetical protein T8G24.9 [Arabidopsis thaliana (Mouse-ear cress)]	3	ATG	18	6	Plus	Gossypium_raimondii
	DT553753 Hypothetical protein T8G24.9 [Arabidopsis thaliana (Mouse-ear cress)]	3	ATG	18	6	Plus	Gossypium_hirsutum
24	DR455003	3	ATT	18	6	Plus	Gossypium_hirsutum
	CO107954	3	ATT	18	6	Plus	Gossypium_raimondii
25	DR455003	3	ATT	18	6	Plus	Gossypium_hirsutum
	TA10229_29730	3	ATT	18	6	Plus	Gossypium_raimondii
26	CO100998	3	ATT	18	6	Plus	Gossypium_raimondii
	DR455003	3	ATT	18	6	Plus	Gossypium_hirsutum
27	TA13997_29730 Hypothetical protein F23E12.210 [Arabidopsis thaliana (Mouse-ear cress)]	3	CAA	18	6	Plus	Gossypium_raimondii
	DW504741 Hypothetical protein F23E12.210 [Arabidopsis thaliana (Mouse-ear cress)]	3	CAA	18	6	Plus	Gossypium_hirsutum
28	CO075719 F28L1.7 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	CAA	18	6	Plus	Gossypium_raimondii
	TA21684_3635	3	CAA	18	6	Plus	Gossypium_hirsutum
	DN802943 F28L1.7 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	CAA	18	6	Plus	Gossypium_hirsutum
	TA21681_3635	3	CAA	18	6	Plus	Gossypium_hirsutum
29	TA11100_29730	3	CAA	18	6	Plus	Gossypium_raimondii
	DN780342	3	TTG	18	6	Minus	Gossypium_hirsutum

Supplementary Table 4. Contd.

	TA25295_3635 Putative zinc finger protein ID1 [Oryza sativa (japonica cultivar-group)]	3	CAA	18	6	Plus	Gossypium_hirsutum
30	CO117165 Mutant cincinnata [Antirrhinum majus (Garden snapdragon)]	3	CAC	18	6	Plus	Gossypium_raimondii
	TA35391_3635 Mutant cincinnata [Antirrhinum majus (Garden snapdragon)]	3	CAC	18	6	Plus	Gossypium_hirsutum
31	TA12798_29730 Transcriptional factor B3; Auxin response factor; Aux/IAA_ARF_dimerisation [Medicago truncatula (Barrel medic)]	3	CAG	18	6	Plus	Gossypium_raimondii
	TA43550_3635	3	CAG	18	6	Plus	Gossypium_hirsutum
32	CO127350 Hypothetical protein Z97340.18 [Arabidopsis thaliana (Mouse-ear cress)]	3	CAG	18	6	Plus	Gossypium_raimondii
	TA30296_3635 Hypothetical protein Z97340.18 [Arabidopsis thaliana (Mouse-ear cress)]	3	CAG	18	6	Plus	Gossypium_hirsutum
33	CO099987 Hypothetical protein Z97340.18 [Arabidopsis thaliana (Mouse-ear cress)]	3	CAG	18	6	Plus	Gossypium_raimondii
	TA30296_3635 Hypothetical protein Z97340.18 [Arabidopsis thaliana (Mouse-ear cress)]	3	CAG	18	6	Plus	Gossypium_hirsutum
34	TA12116_29730 At3g04070 [Arabidopsis thaliana (Mouse-ear cress)]	3	CAG	18	6	Plus	Gossypium_raimondii
	CD485949	3	CAG	18	6	Plus	Gossypium_hirsutum
35	TA14299_29730 Eukaryotic transcription factor, DNA-binding [Medicago truncatula (Barrel medic)]	3	CAT	18	6	Plus	Gossypium_raimondii
	TA41967_3635	3	CAT	18	6	Plus	Gossypium_hirsutum
36	TA9645_29730 Auxin-regulated protein [Populus tremula x Populus tremuloides]	3	CAT	21	7	Plus	Gossypium_raimondii
	DW511571	3	CAT	21	7	Plus	Gossypium_hirsutum
37	CO121268 Nt-iaa28 deduced protein [Nicotiana tabacum (Common tobacco)]	3	CAT	21	7	Plus	Gossypium_raimondii
	DW511571	3	CAT	21	7	Plus	Gossypium_hirsutum
38	TA11701_29730 HyPRP1 [Gossypium hirsutum (Upland cotton)]	3	CCA	21	7	Plus	Gossypium_raimondii
	TA21454_3635 HyPRP1 [Gossypium hirsutum (Upland cotton)]	3	CCA	21	7	Plus	Gossypium_hirsutum
39	CO092245 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_raimondii
	DN800615 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_hirsutum
	TA23709_3635 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_hirsutum
40	CO088421 HyPRP1 [Gossypium hirsutum (Upland cotton)]	3	CCA	21	7	Plus	Gossypium_raimondii
	TA21454_3635 HyPRP1 [Gossypium hirsutum (Upland cotton)]	3	CCA	21	7	Plus	Gossypium_hirsutum
41	TA10530_29730 ATFP3 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCG	21	7	Plus	Gossypium_raimondii

Supplementary Table 4. Contd.

	DT570547 Hypothetical protein At5g63530 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCG	21	7	Plus	Gossypium_hirsutum
	DT560725 OSJNBa0053B21.14 protein [Oryza sativa (japonica cultivar-group)]	3	CCG	21	7	Plus	Gossypium_hirsutum
	DT548104 ATFP3 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCG	21	7	Plus	Gossypium_hirsutum
	TA21278_3635 Hypothetical protein At5g50740 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCG	21	7	Plus	Gossypium_hirsutum
42	CO074297 OSJNBa0053B21.14 protein [Oryza sativa (japonica cultivar-group)]	3	CCG	21	7	Plus	Gossypium_raimondii
	DT548104 ATFP3 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCG	21	7	Plus	Gossypium_hirsutum
	DT570547 Hypothetical protein At5g63530 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCG	21	7	Plus	Gossypium_hirsutum
	TA21278_3635 Hypothetical protein At5g50740 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCG	21	7	Plus	Gossypium_hirsutum
	DT560725 OSJNBa0053B21.14 protein [Oryza sativa (japonica cultivar-group)]	3	CCG	21	7	Plus	Gossypium_hirsutum
43	CO096636	3	CCG	18	6	Plus	Gossypium_raimondii
	TA21684_3635	3	CGG	18	6	Minus	Gossypium_hirsutum
44	CO074836 TCP transcription factor [Medicago truncatula (Barrel medic)]	3	CGA	18	6	Plus	Gossypium_raimondii
	DR454446	3	CGA	18	6	Plus	Gossypium_hirsutum
	TA25064_3635 TCP transcription factor [Medicago truncatula (Barrel medic)]	3	CGA	18	6	Plus	Gossypium_hirsutum
45	CO075719 F28L1.7 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	CGG	18	6	Plus	Gossypium_raimondii
	DN758979	3	CGG	18	6	Plus	Gossypium_hirsutum
	TA21683_3635 Protein At5g19090 [Arabidopsis thaliana (Mouse-ear cress)]	3	CGG	18	6	Plus	Gossypium_hirsutum
	TA21684_3635	3	CGG	18	6	Plus	Gossypium_hirsutum
	TA21681_3635	3	CGG	18	6	Plus	Gossypium_hirsutum
46	TA17138_29730 MYC2 [Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle)]	3	CTC	21	7	Plus	Gossypium_raimondii
	TA21503_3635	3	CTC	21	7	Plus	Gossypium_hirsutum
47	CO125521 Hypothetical protein F18O22_180 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	24	8	Plus	Gossypium_raimondii
	AI731904 Hypothetical protein F18O22_180 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	24	8	Plus	Gossypium_hirsutum
48	TA12795_29730 Putative exoglucanase [Oryza sativa (Rice)]	3	CTT	21	7	Plus	Gossypium_raimondii
	TA36696_3635 T15F16.6 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAA	21	7	Minus	Gossypium_hirsutum
49	TA10320_29730 Glutamate-1-semialdehyde 2,1-aminomutase 2, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	18	6	Plus	Gossypium_raimondii
	TA30107_3635 Glutamate-1-semialdehyde 2,1-aminomutase 2, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	18	6	Plus	Gossypium_hirsutum
50	CO111000	3	CTT	18	6	Plus	Gossypium_raimondii

Supplementary Table 4. Contd.

	DW235030	3	CTT	18	6	Plus	Gossypium_hirsutum
	TA22163_3635	3	CTT	18	6	Plus	Gossypium_hirsutum
51	CO080640 Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K14A3 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	18	6	Plus	Gossypium_raimondii
	DW496704 Hypothetical protein [Brassica oleracea (Wild cabbage)]	3	CTT	18	6	Plus	Gossypium_hirsutum
52	TA10526_29730 Hypothetical protein F10M23.360 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	18	6	Plus	Gossypium_raimondii
	DT048210 AT4g27020/F10M23_360 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	18	6	Plus	Gossypium_hirsutum
53	CO121921	3	CTT	18	6	Plus	Gossypium_raimondii
	DT047343	3	CTT	18	6	Plus	Gossypium_hirsutum
54	CO098330 GATA transcription factor 14 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAA	18	6	Plus	Gossypium_raimondii
	TA23453_3635 GATA transcription factor 14 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAA	18	6	Plus	Gossypium_hirsutum
	DT572945	3	GAA	18	6	Plus	Gossypium_hirsutum
	TA23455_3635 Hypothetical protein [Thellungiella halophila (Salt cress)]	3	GAA	18	6	Plus	Gossypium_hirsutum
55	TA11396_29730 GATA transcription factor 14 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAA	18	6	Plus	Gossypium_raimondii
	DT572945	3	GAA	18	6	Plus	Gossypium_hirsutum
	TA23455_3635 Hypothetical protein [Thellungiella halophila (Salt cress)]	3	GAA	18	6	Plus	Gossypium_hirsutum
	TA23453_3635 GATA transcription factor 14 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAA	18	6	Plus	Gossypium_hirsutum
	TA23454_3635	3	TTC	18	6	Minus	Gossypium_hirsutum
56	CO113948 Hypothetical protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAA	18	6	Plus	Gossypium_raimondii
	DR456101	3	GAA	18	6	Plus	Gossypium_hirsutum
	TA43955_3635 Hypothetical protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAA	18	6	Plus	Gossypium_hirsutum
	TA36873_3635 Hypothetical protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAA	18	6	Plus	Gossypium_hirsutum
57	TA9687_29730 Protein At5g20190 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAA	24	8	Plus	Gossypium_raimondii
	DN758687 Protein At5g20190 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAA	24	8	Plus	Gossypium_hirsutum
58	CO126475 Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K14A3 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAG	21	7	Plus	Gossypium_raimondii
	TA25142_3635 Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K14A3 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAG	21	7	Plus	Gossypium_hirsutum
	DT569099 Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K14A3 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAG	21	7	Plus	Gossypium_hirsutum
59	TA15485_29730 Hypothetical protein At3g20260 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_raimondii
	DT555558 Gb|AAF36750.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

	TA31218_3635 Gb|AAF36750.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_hirsutum
60	TA16069_29730 GTP1/OBG [Medicago truncatula (Barrel medic)]	3	GAT	18	6	Plus	Gossypium_raimondii
	DW230589 GTP1/OBG [Medicago truncatula (Barrel medic)]	3	GAT	18	6	Plus	Gossypium_hirsutum
61	TA15021_29730 Emb|CAA16573.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_raimondii
	TA44028_3635 Hypothetical protein F10N7.30 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_hirsutum
62	TA10905_29730 Rieske [2Fe-2S] domain, putative [Oryza sativa (japonica cultivar-group)]	3	GCA	18	6	Plus	Gossypium_raimondii
	DW499682 Rieske [2Fe-2S] domain, putative [Oryza sativa (japonica cultivar-group)]	3	GCA	18	6	Plus	Gossypium_hirsutum
63	CO074381 Amino acid/polyamine transporter II [Medicago truncatula (Barrel medic)]	3	GCA	18	6	Plus	Gossypium_raimondii
	TA44554_3635	3	GCA	18	6	Plus	Gossypium_hirsutum
64	CO074960 Rieske [2Fe-2S] domain, putative [Oryza sativa (japonica cultivar-group)]	3	GCA	18	6	Plus	Gossypium_raimondii
	DW499682 Rieske [2Fe-2S] domain, putative [Oryza sativa (japonica cultivar-group)]	3	GCA	18	6	Plus	Gossypium_hirsutum
65	CO105377 Predicted protein [Populus alba x Populus tremula]	3	GCT	18	6	Plus	Gossypium_raimondii
	TA33856_3635 Predicted protein [Populus alba x Populus tremula]	3	GCT	18	6	Plus	Gossypium_hirsutum
66	CO085114 Cleft lip and palate associated transmembrane protein-like [Arabidopsis thaliana (Mouse-ear cress)]	3	GGA	24	8	Plus	Gossypium_raimondii
	TA28051_3635 Cleft lip and palate associated transmembrane protein-like [Arabidopsis thaliana (Mouse-ear cress)]	3	GGA	24	8	Plus	Gossypium_hirsutum
67	TA12379_29730 Pre-mRNA splicing factor [Oryza sativa (Rice)]	3	GGC	21	7	Plus	Gossypium_raimondii
	TA35576_3635 Pre-mRNA splicing factor [Oryza sativa (Rice)]	3	GGC	21	7	Plus	Gossypium_hirsutum
68	TA11455_29730 50S ribosomal protein L5, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	3	GGT	18	6	Plus	Gossypium_raimondii
	TA25745_3635 50S ribosomal protein L5, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	3	GGT	18	6	Plus	Gossypium_hirsutum
69	TA10731_29730 unknown protein [Arabidopsis thaliana]	3	GTG	18	6	Plus	Gossypium_raimondii
	TA38018_3635 unknown protein [Arabidopsis thaliana]	3	GTG	18	6	Plus	Gossypium_hirsutum
70	TA10656_29730 Endo-1,4-beta-glucanase precursor [Lycopersicon esculentum (Tomato)]	3	TAC	18	6	Plus	Gossypium_raimondii
	AI055507 Endo-1,4-beta-glucanase [Glycine max (Soybean)]	3	TAC	18	6	Plus	Gossypium_hirsutum
	TA23759_3635 Endo-1,4-beta-glucanase precursor [Lycopersicon esculentum (Tomato)]	3	TAC	18	6	Plus	Gossypium_hirsutum
71	TA12398_29730 Patellin-3 [Arabidopsis thaliana (Mouse-ear cress)]	3	TAC	18	6	Plus	Gossypium_raimondii

Supplementary Table 4. Contd.

	TA24577_3635 Putative cytosolic factor [Trifolium pratense (Red clover)]	3	TAC	18	6	Plus	Gossypium_hirsutum
72	TA10655_29730 Endoglucanase 17 precursor [Arabidopsis thaliana (Mouse-ear cress)]	3	TAC	18	6	Plus	Gossypium_raimondii
	AI055507 Endo-1,4-beta-glucanase [Glycine max (Soybean)]	3	TAC	18	6	Plus	Gossypium_hirsutum
	TA23759_3635 Endo-1,4-beta-glucanase precursor [Lycopersicon esculentum (Tomato)]	3	TAC	18	6	Plus	Gossypium_hirsutum
73	TA15027_29730 Hypothetical protein [Cicer arietinum (Chickpea) (Garbanzo)]	3	TAT	18	6	Plus	Gossypium_raimondii
	TA26727_3635 Hypothetical protein [Cicer arietinum (Chickpea) (Garbanzo)]	3	TAT	18	6	Plus	Gossypium_hirsutum
74	TA13070_29730 unknown protein [Arabidopsis thaliana]	3	TAT	18	6	Plus	Gossypium_raimondii
	TA28431_3635 unknown protein [Arabidopsis thaliana]	3	TAT	18	6	Plus	Gossypium_hirsutum
75	CO096513 F10G19.3 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TCC	18	6	Plus	Gossypium_raimondii
	TA42124_3635 F10G19.3 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TCC	18	6	Plus	Gossypium_hirsutum
76	TA42124_3635 F10G19.3 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TCC	18	6	Plus	Gossypium_hirsutum
	CO116088 F10G19.3 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TCC	18	6	Plus	Gossypium_raimondii
77	CO079404	3	TCT	18	6	Plus	Gossypium_raimondii
	DT047343	3	CTT	18	6	Plus	Gossypium_hirsutum
78	TA11358_29730 Hypothetical protein MED24.18 [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	21	7	Plus	Gossypium_raimondii
	TA34299_3635 Hypothetical protein MED24.18 [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	21	7	Plus	Gossypium_hirsutum
79	TA13968_29730 At2g20670/F23N11.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	TGA	27	9	Plus	Gossypium_raimondii
	DT572217 Hypothetical protein F8B4.180 [Arabidopsis thaliana (Mouse-ear cress)]	3	TGA	27	9	Plus	Gossypium_hirsutum
80	TA15848_29730 Hypothetical protein F28P22.14 [Arabidopsis thaliana (Mouse-ear cress)]	3	TGA	18	6	Plus	Gossypium_raimondii
	DT549381	3	TGA	18	6	Plus	Gossypium_hirsutum
81	TA11085_29730 6b-interacting protein 1-like [Oryza sativa (japonica cultivar-group)]	3	TGA	18	6	Plus	Gossypium_raimondii
	DR458359	3	TGA	18	6	Plus	Gossypium_hirsutum
82	CO131026 Protein At2g27110 [Arabidopsis thaliana (Mouse-ear cress)]	3	TGA	18	6	Plus	Gossypium_raimondii
	DW519866 Protein At2g27110 [Arabidopsis thaliana (Mouse-ear cress)]	3	TGA	18	6	Plus	Gossypium_hirsutum
	TA42826_3635 Protein At2g27110 [Arabidopsis thaliana (Mouse-ear cress)]	3	TGA	18	6	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

83	CO105343 Von Willebrand factor, type A; Ubiquitin interacting motif [Medicago truncatula (Barrel medic)]	3	TGC	18	6	Plus	Gossypium_raimondii
	DT571041 Von Willebrand factor, type A; Ubiquitin interacting motif [Medicago truncatula (Barrel medic)]	3	TGC	18	6	Plus	Gossypium_hirsutum
	TA22965_3635 Von Willebrand factor, type A; Ubiquitin interacting motif [Medicago truncatula (Barrel medic)]	3	TGC	18	6	Plus	Gossypium_hirsutum
84	TA12115_29730	3	TGC	18	6	Plus	Gossypium_raimondii
	CD485949	3	CAG	18	6	Minus	Gossypium_hirsutum
85	CO088783 Von Willebrand factor, type A; Ubiquitin interacting motif [Medicago truncatula (Barrel medic)]	3	TGC	18	6	Plus	Gossypium_raimondii
	DT571041 Von Willebrand factor, type A; Ubiquitin interacting motif [Medicago truncatula (Barrel medic)]	3	TGC	18	6	Plus	Gossypium_hirsutum
	TA22965_3635 Von Willebrand factor, type A; Ubiquitin interacting motif [Medicago truncatula (Barrel medic)]	3	TGC	18	6	Plus	Gossypium_hirsutum
86	TA14161_29730 Probable WRKY transcription factor 48 [Arabidopsis thaliana (Mouse-ear cress)]	3	TGG	18	6	Plus	Gossypium_raimondii
	TA40466_3635	3	TGG	18	6	Plus	Gossypium_hirsutum
87	TA9645_29730 Auxin-regulated protein [Populus tremula x Populus tremuloides]	3	TGG	21	7	Plus	Gossypium_raimondii
	CD486500 Auxin-regulated protein [Populus tremula x Populus tremuloides]	3	TGG	21	7	Plus	Gossypium_hirsutum
	TA24844_3635 Auxin-regulated protein [Populus tremula x Populus tremuloides]	3	TGG	21	7	Plus	Gossypium_hirsutum
88	CO100284 Aux/IAA protein [Populus tremula x Populus tremuloides]	3	TGG	18	6	Plus	Gossypium_raimondii
	DW480781 Auxin-responsive protein IAA11 [Arabidopsis thaliana (Mouse-ear cress)]	3	TGG	18	6	Plus	Gossypium_hirsutum
	DW488302	3	ACC	18	6	Minus	Gossypium_hirsutum
	DW488301	3	TGG	18	6	Plus	Gossypium_hirsutum
	TA26814_3635 Auxin-responsive protein IAA11 [Arabidopsis thaliana (Mouse-ear cress)]	3	TGG	18	6	Plus	Gossypium_hirsutum
89	CO083872 Putative cytosolic factor [Trifolium pratense (Red clover)]	3	TTA	18	6	Plus	Gossypium_raimondii
	TA26727_3635 Hypothetical protein [Cicer arietinum (Chickpea) (Garbanzo)]	3	TAT	18	6	Plus	Gossypium_hirsutum
90	CO089818 Hypothetical protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	18	6	Plus	Gossypium_raimondii
	TA28845_3635 At5g37660 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	18	6	Plus	Gossypium_hirsutum
91	CO082594 Nucellin-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	18	6	Plus	Gossypium_raimondii
	TA43902_3635 Nucellin-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	18	6	Plus	Gossypium_hirsutum
92	CO070235	3	TTC	21	7	Plus	Gossypium_raimondii
	TA39369_3635 Inducer of CBF expression 1 protein [Capsella bursa-pastoris (Shepherd's purse)]	3	TTC	21	7	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

93	TA16995_29730 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	21	7	Plus	Gossypium_raimondii
	AI731296 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	21	7	Plus	Gossypium_hirsutum
	TA28796_3635 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	21	7	Plus	Gossypium_hirsutum
94	TA12750_29730 Polypyrimidine tract-binding protein 1-like [Oryza sativa (japonica cultivar-group)]	4	ATCT	24	6	Plus	Gossypium_raimondii
	TA34023_3635 Polypyrimidine tract-binding protein 1-like [Oryza sativa (japonica cultivar-group)]	4	ATCT	24	6	Plus	Gossypium_hirsutum
95	TA9530_29730 Chromosome 10 SCAF12030, whole genome shotgun sequence [Tetraodon nigroviridis (Green puffer)]	4	ATGA	20	5	Plus	Gossypium_raimondii
	TA22504_3635 Chromosome 10 SCAF12030, whole genome shotgun sequence [Tetraodon nigroviridis (Green puffer)]	4	ATGA	20	5	Plus	Gossypium_hirsutum
86	TA11291_29730	4	CATA	24	6	Plus	Gossypium_raimondii
	TA24344_3635 Dual specificity phosphatase, catalytic domain, putative [Medicago truncatula (Barrel medic)]	4	GTAT	24	6	Minus	Gossypium_hirsutum
97	TA14342_29730 Putative MYB family transcription factor [Arabidopsis thaliana (Mouse-ear cress)]	4	TATG	20	5	Plus	Gossypium_raimondii
	DW511541	4	TATG	20	5	Plus	Gossypium_hirsutum
98	TA14305_29730 Mitogen-activated protein kinase 1 [Lycopersicon esculentum (Tomato)]	4	TATG	20	5	Plus	Gossypium_raimondii
	TA23495_3635 Mitogen-activated protein kinase 1 [Lycopersicon esculentum (Tomato)]	4	TATG	20	5	Plus	Gossypium_hirsutum
99	TA16726_29730 Mitogen-activated protein kinase 2 [Lycopersicon esculentum (Tomato)]	4	TATG	20	5	Plus	Gossypium_raimondii
	TA23495_3635 Mitogen-activated protein kinase 1 [Lycopersicon esculentum (Tomato)]	4	TATG	20	5	Plus	Gossypium_hirsutum
100	TA9059_29730 Root phototropism protein 2 [Arabidopsis thaliana (Mouse-ear cress)]	4	TTGT	20	5	Plus	Gossypium_raimondii
	CA993380 Phototropic response-like protein [Malus domestica (Apple) (Malus sylvestris)]	4	TTGT	20	5	Plus	Gossypium_hirsutum
101	TA9210_29730 antiporter/ glucose-6-phosphate transporter [Arabidopsis thaliana]	5	AAACC	25	5	Plus	Gossypium_raimondii
	TA28118_3635 antiporter/ glucose-6-phosphate transporter [Arabidopsis thaliana]	5	AAACC	25	5	Plus	Gossypium_hirsutum
102	TA10662_29730 Expressed protein [Arabidopsis thaliana (Mouse-ear cress)]	5	ATGCT	25	5	Plus	Gossypium_raimondii
	TA25857_3635 Expressed protein [Arabidopsis thaliana (Mouse-ear cress)]	5	ATGCT	25	5	Plus	Gossypium_hirsutum
103	CO102286 F2D10.13 [Arabidopsis thaliana (Mouse-ear cress)]	5	GAAGA	25	5	Plus	Gossypium_raimondii
	TA26724_3635 F2D10.13 [Arabidopsis thaliana (Mouse-ear cress)]	5	GAAGA	25	5	Plus	Gossypium_hirsutum
104	TA11200_29730 Eukaryotic peptide chain release factor subunit 1-3 [Arabidopsis thaliana (Mouse-ear cress)]	5	GTTTG	25	5	Plus	Gossypium_raimondii
	TA22251_3635 Putative eukaryotic peptide chain release factor subunit 1-3 [Oryza sativa (japonica cultivar-group)]	5	GTTTG	25	5	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

105	TA10378_29730 Hypothetical protein At1g70160 [Arabidopsis thaliana (Mouse-ear cress)]	6	AATCCC	24	4	Plus	Gossypium_raimondii
	DT559035 Hypothetical protein At1g70160 [Arabidopsis thaliana (Mouse-ear cress)]	6	CAATCC	24	4	Plus	Gossypium_hirsutum
	TA23974_3635 Hypothetical protein At1g70160 [Arabidopsis thaliana (Mouse-ear cress)]	6	CAATCC	24	4	Plus	Gossypium_hirsutum
106	TA9618_29730 Hypothetical protein [Arabidopsis thaliana (Mouse-ear cress)]	6	ACCAAA	30	5	Plus	Gossypium_raimondii
	DT564564	6	ACCAAA	30	5	Plus	Gossypium_hirsutum
	DT556816	6	ACCAAA	30	5	Plus	Gossypium_hirsutum
	DT558262	6	ACCAAA	30	5	Plus	Gossypium_hirsutum
	TA24430_3635 Hypothetical protein [Arabidopsis thaliana (Mouse-ear cress)]	6	ACCAAA	30	5	Plus	Gossypium_hirsutum
107	TA9545_29730 Alpha-1,4 glucan phosphorylase L isozyme, chloroplast precursor [Vicia faba (Broad bean)]	6	AGAAGC	24	4	Plus	Gossypium_raimondii
	TA28987_3635 Alpha-1,4 glucan phosphorylase L isozyme, chloroplast precursor [Vicia faba (Broad bean)]	6	AGAAGC	24	4	Plus	Gossypium_hirsutum
108	CO105571	6	CAACGC	24	4	Plus	Gossypium_raimondii
	TA34092_3635 Hypothetical protein [Arabidopsis thaliana (Mouse-ear cress)]	6	CCAACG	24	4	Plus	Gossypium_hirsutum
109	TA11832_29730 Hypothetical protein At4g32600 [Arabidopsis thaliana (Mouse-ear cress)]	6	CCAGCT	24	4	Plus	Gossypium_raimondii
	DW512264	6	CCAGCT	24	4	Plus	Gossypium_hirsutum
110	CO114031 unknown protein [Arabidopsis thaliana]	6	CCAGTA	24	4	Plus	Gossypium_raimondii
	TA25756_3635 unknown protein [Arabidopsis thaliana]	6	CCAGTA	24	4	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

111	CO118091	6	CCCATG	24	4	Plus	Gossypium_raimondii
	TA24455_3635 Hypothetical protein At2g43170 [Arabidopsis thaliana (Mouse-ear cress)]	6	GCATGG	24	4	Minus	Gossypium_hirsutum
112	DT555662 At1g14000/F7A19_9 [Arabidopsis thaliana (Mouse-ear cress)]	6	CCGACG	24	4	Plus	Gossypium_hirsutum
	TA16508_29730 At1g14000/F7A19_9 [Arabidopsis thaliana (Mouse-ear cress)]	6	GCCGAC	24	4	Plus	Gossypium_raimondii
113	TA17036_29730 MYB transcription factor MYB127 [Glycine max (Soybean)]	6	CCTCTC	30	5	Plus	Gossypium_raimondii
	DR041738 MYB transcription factor MYB127 [Glycine max (Soybean)]	6	CCTCTC	30	5	Plus	Gossypium_hirsutum
114	CO071668 MYB transcription factor MYB127 [Glycine max (Soybean)]	6	CCTCTC	30	5	Plus	Gossypium_raimondii
	DR041738 MYB transcription factor MYB127 [Glycine max (Soybean)]	6	CCTCTC	30	5	Plus	Gossypium_hirsutum
115	TA12532_29730 Putative ABC transporter ATPase; 10053-12032 [Arabidopsis thaliana (Mouse-ear cress)]	6	CCTCTC	24	4	Plus	Gossypium_raimondii
	TA25929_3635 Putative ABC transporter ATPase; 10053-12032 [Arabidopsis thaliana (Mouse-ear cress)]	6	CCTCTC	24	4	Plus	Gossypium_hirsutum
116	CO093803 Timing of CAB expression 1 protein [Castanea sativa (Sweet chestnut)]	6	CGGGAG	24	4	Plus	Gossypium_raimondii
	TA26774_3635 Two-component response regulator-like PRR37 [Oryza sativa (Rice)]	6	GAGCGG	24	4	Plus	Gossypium_hirsutum
117	TA13748_29730 Tubby; Di-trans-poly-cis-decaprenylcistransferase; Cyclin-like F- box [Medicago truncatula (Barrel medic)]	6	GGCGGA	24	4	Plus	Gossypium_raimondii
	AI732056 F12M16.22 [Arabidopsis thaliana (Mouse-ear cress)]	6	GGCGGA	24	4	Plus	Gossypium_hirsutum
118	TA10515_29730 Fructose-1,6-bisphosphatase [Pisum sativum (Garden pea)]	6	TCTCAC	30	5	Plus	Gossypium_raimondii
	DW483497 Fructose-1,6-bisphosphatase [Pisum sativum (Garden pea)]	6	TCTCAC	30	5	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

119	TA9736_29730 Transcription factor DRE-binding factor 2 [Gossypium hirsutum (Upland cotton)]	6	TCTCCT	24	4	Plus	Gossypium_raimondii
	DV848960 Transcription factor DRE-binding factor 2 [Gossypium hirsutum (Upland cotton)]	6	TCTCCT	24	4	Plus	Gossypium_hirsutum
	DT465897	6	TCTCCT	24	4	Plus	Gossypium_hirsutum
	DT572833	6	TCTCCT	24	4	Plus	Gossypium_hirsutum
	TA22759_3635 Transcription factor DRE-binding factor 2 [Gossypium hirsutum (Upland cotton)]	6	TCTCCT	24	4	Plus	Gossypium_hirsutum
120	TA13949_29730 Plastid ribosomal protein L19, putative [Brassica oleracea (Wild cabbage)]	6	TGAAGG	24	4	Plus	Gossypium_raimondii
	DR455677 Plastid ribosomal protein L19, putative [Brassica oleracea (Wild cabbage)]	6	TGAAGG	24	4	Plus	Gossypium_hirsutum
	CO499248 Plastid ribosomal protein L19, putative [Brassica oleracea (Wild cabbage)]	6	TGAAGG	24	4	Plus	Gossypium_hirsutum
121	TA15877_29730 At1g26300/F28B23_4 [Arabidopsis thaliana (Mouse-ear cress)]	6	TGATGC	24	4	Plus	Gossypium_raimondii
	TA28020_3635 At1g26300/F28B23_4 [Arabidopsis thaliana (Mouse-ear cress)]	6	TGATGC	24	4	Plus	Gossypium_hirsutum
122	CO126735 Protein kinase [Arabidopsis thaliana (Mouse-ear cress)]	6	TGGTGA	24	4	Plus	Gossypium_raimondii
	DW489261 Protein kinase [Arabidopsis thaliana (Mouse-ear cress)]	6	TGGTGA	24	4	Plus	Gossypium_hirsutum
123	TA13723_29730 Protein kinase [Arabidopsis thaliana (Mouse-ear cress)]	6	TGGTGA	24	4	Plus	Gossypium_raimondii
	DW489261 Protein kinase [Arabidopsis thaliana (Mouse-ear cress)]	6	TGGTGA	24	4	Plus	Gossypium_hirsutum
Group II							
HG	Sequence ID	Period Size	Period	Rep. Length	Copy Num.	Blast Strand	Species
1	CO074722 Gb|AAC24081.1 [Arabidopsis thaliana (Mouse-ear cress)]	2	AG	34	17	Plus	Gossypium_raimondii
	TA34307_3635 Gb|AAC24081.1 [Arabidopsis thaliana (Mouse-ear cress)]	2	AG	30	15	Plus	Gossypium_hirsutum
	DT574307 Gb|AAC24081.1 [Arabidopsis thaliana (Mouse-ear cress)]	2	AG	30	15	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

2	TA14574_29730 IRE homolog; protein kinase-like protein [Arabidopsis thaliana (Mouse-ear cress)]	2	AG	26	13	Plus	Gossypium_raimondii
	CO498746 IRE homolog; protein kinase-like protein [Arabidopsis thaliana (Mouse-ear cress)]	2	AG	20	10	Plus	Gossypium_hirsutum
3	TA10286_29730 Gamma hydroxybutyrate dehydrogenase-like protein [Oryza sativa (japonica cultivar-group)]	2	AG	28	14	Plus	Gossypium_raimondii
	DN804214 Gamma hydroxybutyrate dehydrogenase [Arabidopsis thaliana (Mouse-ear cress)]	2	AG	26	13	Plus	Gossypium_hirsutum
4	TA15654_29730 Hypothetical protein F23E12.210 [Arabidopsis thaliana (Mouse-ear cress)]	2	CA	22	11	Plus	Gossypium_raimondii
	TA38621_3635 Hypothetical protein F23E12.210 [Arabidopsis thaliana (Mouse-ear cress)]	2	CA	18	9	Plus	Gossypium_hirsutum
5	TA11234_29730 Putative fasciclin-like arabinogalactan protein FLA2 [Trifolium pratense (Red clover)]	2	CA	30	15	Plus	Gossypium_raimondii
	TA21226_3635 Putative fasciclin-like arabinogalactan protein [Salicornia europaea (Common glasswort)]	2	CA	20	10	Plus	Gossypium_hirsutum
6	CO114663 Arabidopsis thaliana gl1 homolog [Arabidopsis thaliana (Mouse-ear cress)]	2	CT	28	14	Plus	Gossypium_raimondii
	DT049270 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	TC	20	10	Plus	Gossypium_hirsutum
7	CO081596 NAM-like protein [Arabidopsis thaliana (Mouse-ear cress)]	2	CT	22	11	Plus	Gossypium_raimondii
	TA22929_3635 NAM-like protein [Arabidopsis thaliana (Mouse-ear cress)]	2	CT	18	9	Plus	Gossypium_hirsutum
	DT527598 Unknown protein [Lycopersicon esculentum (Tomato)]	2	CT	20	10	Plus	Gossypium_hirsutum
8	TA9287_29730 Arabidopsis thaliana gl1 homolog [Arabidopsis thaliana (Mouse-ear cress)]	2	CT	28	14	Plus	Gossypium_raimondii
	DT049270 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	TC	20	10	Plus	Gossypium_hirsutum
9	TA13307_29730 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	CT	26	13	Plus	Gossypium_raimondii
	DN802296	2	CT	20	10	Plus	Gossypium_hirsutum
10	TA14305_29730 Mitogen-activated protein kinase 1 [Lycopersicon esculentum (Tomato)]	2	TC	40	20	Plus	Gossypium_raimondii
	CA993935 Mitogen-activated protein kinase 1 [Lycopersicon esculentum (Tomato)]	2	TC	38	19	Plus	Gossypium_hirsutum
11	CO086021	2	TC	22	11	Plus	Gossypium_raimondii
	DR044199 Hypothetical protein OSJNBa0014O06.1 [Oryza sativa (japonica cultivar-group)]	2	TC	18	9	Plus	Gossypium_hirsutum
	TA31302_3635	2	TC	18	9	Plus	Gossypium_hirsutum
12	CO094396 Hypothetical protein OSJNBa0014O06.1 [Oryza sativa (japonica cultivar-group)]	2	TC	22	11	Plus	Gossypium_raimondii
	DR044199 Hypothetical protein OSJNBa0014O06.1 [Oryza sativa (japonica cultivar-group)]	2	TC	18	9	Plus	Gossypium_hirsutum
	TA31302_3635	2	TC	18	9	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

13	CO109699 Zinc finger protein [Pimpinella brachycarpa]	3	AAC	21	7	Plus	Gossypium_raimondii
	TA21928_3635 SPF1 protein [Ipomoea batatas (Sweet potato) (Batate)]	3	AAC	18	6	Plus	Gossypium_hirsutum
14	TA13731_29730 Protein phosphatase-2C-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AAG	21	7	Plus	Gossypium_raimondii
	DW513876	3	AAG	18	6	Plus	Gossypium_hirsutum
	DW513876	3	AAG	18	6	Plus	Gossypium_hirsutum
15	CO129071 Putative C2H2-type zinc finger protein [Arabidopsis thaliana (Mouse-ear cress)]	3	ACA	24	8	Plus	Gossypium_raimondii
	DN804675	3	ACA	21	7	Plus	Gossypium_hirsutum
16	TA9550_29730	3	ACC	21	7	Plus	Gossypium_raimondii
	TA42267_3635	3	TGG	18	6	Minus	Gossypium_hirsutum
	TA29583_3635	3	ACC	18	6	Plus	Gossypium_hirsutum
17	CO116035 Adenylate translocator (Brittle-1)-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	21	7	Plus	Gossypium_raimondii
	TA23448_3635 Adenylate translocator (Brittle-1)-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	18	6	Plus	Gossypium_hirsutum
18	CO129567 Adenylate translocator (Brittle-1)-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	21	7	Plus	Gossypium_raimondii
	TA23448_3635 Adenylate translocator (Brittle-1)-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	18	6	Plus	Gossypium_hirsutum
19	CO103751 Hypothetical protein At2g17970 [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	21	7	Plus	Gossypium_raimondii
	DT566790 Hypothetical protein At2g17970 [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	18	6	Plus	Gossypium_hirsutum
20	TA10409_29730 Hly-III related proteins [Medicago truncatula (Barrel medic)]	3	ATC	45	15	Plus	Gossypium_raimondii
	TA37119_3635 Hly-III related proteins [Medicago truncatula (Barrel medic)]	3	ATC	42	14	Plus	Gossypium_hirsutum
21	TA14749_29730 Hypothetical protein T16K5.250 [Arabidopsis thaliana (Mouse-ear cress)]	3	ATG	39	13	Plus	Gossypium_raimondii
	DT555093 Hypothetical protein T16K5.250 [Arabidopsis thaliana (Mouse-ear cress)]	3	ATG	21	7	Plus	Gossypium_hirsutum
	TA28739_3635 Hypothetical protein T16K5.250 [Arabidopsis thaliana (Mouse-ear cress)]	3	ATG	21	7	Plus	Gossypium_hirsutum
22	TA8867_29730 Wound-responsive protein, putative [Arabidopsis thaliana (Mouse-ear cress)]	3	ATT	36	12	Plus	Gossypium_raimondii
	TA27626_3635 Wound-responsive protein, putative [Arabidopsis thaliana (Mouse-ear cress)]	3	ATT	30	10	Plus	Gossypium_hirsutum
23	CO080595 Putative calcium-dependent protein kinase [Oryza sativa (japonica cultivar-group)]	3	CAA	21	7	Plus	Gossypium_raimondii
	TA41417_3635	3	CAA	18	6	Plus	Gossypium_hirsutum
24	CO097102 Putative zinc finger protein [Oryza sativa (japonica cultivar-group)]	3	CAA	24	8	Plus	Gossypium_raimondii

Supplementary Table 4. Contd.

	TA21598_3635 Zinc finger DNA-binding protein [Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle)]	3	CAA	21	7	Plus	Gossypium_hirsutum
	TA21596_3635 Cys2/His2-type zinc finger protein, putative [Arabidopsis thaliana (Mouse-ear cress)]	3	CAA	21	7	Plus	Gossypium_hirsutum
25	TA13961_29730 Zinc finger DNA-binding protein [Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle)]	3	CAA	24	8	Plus	Gossypium_raimondii
	TA21596_3635 Cys2/His2-type zinc finger protein, putative [Arabidopsis thaliana (Mouse-ear cress)]	3	CAA	21	7	Plus	Gossypium_hirsutum
	TA21598_3635 Zinc finger DNA-binding protein [Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle)]	3	CAA	21	7	Plus	Gossypium_hirsutum
26	TA10131_29730 Squamosa promoter binding protein-homologue 3 [Antirrhinum majus (Garden snapdragon)]	3	CCA	24	8	Plus	Gossypium_raimondii
	DT569996	3	CCA	18	6	Plus	Gossypium_hirsutum
	TA34843_3635	3	CCA	24	8	Plus	Gossypium_hirsutum
27	CO115407 Squamosa promoter binding protein-homologue 3 [Antirrhinum majus (Garden snapdragon)]	3	CCA	24	8	Plus	Gossypium_raimondii
	DT569996	3	CCA	18	6	Plus	Gossypium_hirsutum
	TA34843_3635	3	CCA	24	8	Plus	Gossypium_hirsutum
28	CO125017	3	CCA	24	8	Plus	Gossypium_raimondii
	DT569996	3	CCA	18	6	Plus	Gossypium_hirsutum
	TA34843_3635	3	CCA	24	8	Plus	Gossypium_hirsutum
29	TA11642_29730	3	CTT	24	8	Plus	Gossypium_raimondii
	DT458545	3	CTT	21	7	Plus	Gossypium_hirsutum
30	TA8814_29730 F22K20.9 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAA	30	10	Plus	Gossypium_raimondii
	DW486618	3	GAA	18	6	Plus	Gossypium_hirsutum
	DW486619	3	TTC	18	6	Minus	Gossypium_hirsutum
31	TA9084_29730 Calnexin [Glycine max (Soybean)]	3	GAA	27	9	Plus	Gossypium_raimondii
	TA22700_3635 Calnexin [Glycine max (Soybean)]	3	GAA	21	7	Plus	Gossypium_hirsutum
32	CO087702	3	GAC	21	7	Plus	Gossypium_raimondii
	DT455649	3	GAC	18	6	Plus	Gossypium_hirsutum
33	CO118175	3	GAT	21	7	Plus	Gossypium_raimondii
	TA24244_3635	3	GAT	18	6	Plus	Gossypium_hirsutum
34	CO081152	3	GAT	21	7	Plus	Gossypium_raimondii
	TA24244_3635	3	GAT	18	6	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

35	TA12772_29730 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	GAT	21	7	Plus	Gossypium_raimondii
	TA24244_3635	3	GAT	18	6	Plus	Gossypium_hirsutum
36	TA10501_29730 RNA recognition motif family protein, expressed [Oryza sativa (japonica cultivar-group)]	3	GAT	21	7	Plus	Gossypium_raimondii
	DW232867 RNA recognition motif family protein, expressed [Oryza sativa (japonica cultivar-group)]	3	GAT	18	6	Plus	Gossypium_hirsutum
	DW489296 Hypothetical protein At2g37340 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_hirsutum
	TA20824_3635 RNA recognition motif family protein, expressed [Oryza sativa (japonica cultivar-group)]	3	GAT	18	6	Plus	Gossypium_hirsutum
37	CO111128	3	GCA	27	9	Plus	Gossypium_raimondii
	DT565265	3	GCA	18	6	Plus	Gossypium_hirsutum
	DT463674 Auxin response factor 8 [Arabidopsis thaliana (Mouse-ear cress)]	3	GCA	27	9	Plus	Gossypium_hirsutum
38	TA9639_29730 Unnamed protein product-related [Medicago truncatula (Barrel medic)]	3	GCT	24	8	Plus	Gossypium_raimondii
	TA22416_3635 Translation initiation factor 3 [Oryza sativa (Rice)]	3	GCT	18	6	Plus	Gossypium_hirsutum
	TA22417_3635	3	GCT	18	6	Plus	Gossypium_hirsutum
	DN762148	3	GCT	24	8	Plus	Gossypium_hirsutum
39	CO126702 9G8-like SR protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GGT	21	7	Plus	Gossypium_raimondii
	DR458655 Putative RSZp22 splicing factor [Arabidopsis thaliana (Mouse-ear cress)]	3	GGT	18	6	Plus	Gossypium_hirsutum
40	CO128109	3	TCA	21	7	Plus	Gossypium_raimondii
	DW514675 Protein SCAR4 [Arabidopsis thaliana (Mouse-ear cress)]	3	TGA	18	6	Minus	Gossypium_hirsutum
	DT047243	3	TCA	21	7	Plus	Gossypium_hirsutum
41	TA15947_29730	3	TGA	21	7	Plus	Gossypium_raimondii
	TA27337_3635	3	TGA	18	6	Plus	Gossypium_hirsutum
42	TA16375_29730 Cytochome b5 [Olea europaea (Common olive)]	3	TGT	21	7	Plus	Gossypium_raimondii
	TA36324_3635 Cytochome b5 [Olea europaea (Common olive)]	3	TGT	18	6	Plus	Gossypium_hirsutum
43	TA9599_29730 Glutamine synthetase cytosolic isozyme 1 [Vitis vinifera (Grape)]	3	TTA	21	7	Plus	Gossypium_raimondii
	TA21971_3635 Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS(1)) [Alnus glutinosa (Alder)]	3	TTA	18	6	Plus	Gossypium_hirsutum
44	CO102079 Putative serine protease-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	27	9	Plus	Gossypium_raimondii
	TA23671_3635 Putative serine protease-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	21	7	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

45	TA17231_29730 Protein At3g63460 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	21	7	Plus	Gossypium_raimondii
	TA26634_3635	3	TAA	18	6	Minus	Gossypium_hirsutum
	TA26635_3635 Protein At3g63460 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	21	7	Plus	Gossypium_hirsutum
	TA26636_3635	3	ATA	21	7	Minus	Gossypium_hirsutum
46	TA9513_29730 Putative serine protease-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	27	9	Plus	Gossypium_raimondii
	TA23671_3635 Putative serine protease-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	21	7	Plus	Gossypium_hirsutum
47	CO089244 Ser/Thr protein kinase [Lotus japonicus]	3	TTA	36	12	Plus	Gossypium_raimondii
	TA41161_3635 Ser/Thr protein kinase [Lotus japonicus]	3	TTA	33	11	Plus	Gossypium_hirsutum
48	TA15579_29730 TCP transcription factor [Medicago truncatula (Barrel medic)]	3	TTC	21	7	Plus	Gossypium_raimondii
	TA27906_3635 TCP transcription factor [Medicago truncatula (Barrel medic)]	3	TTC	18	6	Plus	Gossypium_hirsutum
49	CO117323 Adenylate kinase, chloroplast, putative, expressed [Oryza sativa (japonica cultivar-group)]	3	TTC	21	7	Plus	Gossypium_raimondii
	TA42874_3635 Probable adenylate kinase 1, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	18	6	Plus	Gossypium_hirsutum
50	CO115391 F7F22.5 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	24	8	Plus	Gossypium_raimondii
	DT552155 unknown protein [Arabidopsis thaliana]	3	CTT	21	7	Plus	Gossypium_hirsutum
	TA35637_3635 F7F22.5 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	24	8	Plus	Gossypium_hirsutum
51	TA10229_29730	3	GTA	27	9	Plus	Gossypium_raimondii
	TA34896_3635 Protein At4g19110 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTA	18	6	Minus	Gossypium_hirsutum
52	CO084206	4	AAAG	36	9	Plus	Gossypium_raimondii
	TA39376_3635 Hypothetical protein At4g35530 [Arabidopsis thaliana (Mouse-ear cress)]	4	AAAG	24	6	Plus	Gossypium_hirsutum
53	CO102888 Hypothetical protein T5J17.10 [Arabidopsis thaliana (Mouse-ear cress)]	4	AATT	24	6	Plus	Gossypium_raimondii
	DW235242	4	AATT	20	5	Plus	Gossypium_hirsutum
	DV850022 Hypothetical protein [Medicago truncatula (Barrel medic)]	4	AATT	20	5	Plus	Gossypium_hirsutum
54	CO078485 Hypothetical protein [Medicago truncatula (Barrel medic)]	4	AATT	24	6	Plus	Gossypium_raimondii
	DW235242	4	AATT	20	5	Plus	Gossypium_hirsutum
	DV850022 Hypothetical protein [Medicago truncatula (Barrel medic)]	4	AATT	20	5	Plus	Gossypium_hirsutum
55	TA11144_29730 NAM-like protein [Prunus persica (Peach)]	4	ATAC	24	6	Plus	Gossypium_raimondii
	TA29199_3635	4	ATAC	20	5	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

56	CO092319 Branched-chain-amino-acid aminotransferase 3, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	5	GTTGA	30	6	Plus	Gossypium_raimondii
	TA25329_3635 Branched-chain-amino-acid aminotransferase 3, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	5	GTTGA	25	5	Plus	Gossypium_hirsutum
57	TA11688_29730 Branched-chain-amino-acid aminotransferase 3, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	5	GTTGA	30	6	Plus	Gossypium_raimondii
	TA25329_3635 Branched-chain-amino-acid aminotransferase 3, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	5	GTTGA	25	5	Plus	Gossypium_hirsutum
58	CO078229 Branched-chain-amino-acid aminotransferase 3, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	5	GTTGA	30	6	Plus	Gossypium_raimondii
	TA25329_3635 Branched-chain-amino-acid aminotransferase 3, chloroplast precursor [Arabidopsis thaliana (Mouse-ear cress)]	5	GTTGA	25	5	Plus	Gossypium_hirsutum
59	CO105097 Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MYI13 [Arabidopsis thaliana (Mouse-ear cress)]	6	AAATGG	30	5	Plus	Gossypium_raimondii
	TA22883_3635 ZF-HD homeobox protein [Flaveria bidentis (Coastalplain yellowtops)]	6	ATGGAA	24	4	Plus	Gossypium_hirsutum
	DW224844 Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MYI13 [Arabidopsis thaliana (Mouse-ear cress)]	6	ATGGAA	24	4	Plus	Gossypium_hirsutum
60	TA14584_29730	6	AACCTC	30	5	Plus	Gossypium_raimondii
	AI726550	6	AACCTC	24	4	Plus	Gossypium_hirsutum
61	CO085294 Granule bound starch synthase II precursor [Manihot esculenta (Cassava) (Manioc)]	6	AAGATG	30	5	Plus	Gossypium_raimondii
	TA40470_3635	6	AAGATG	24	4	Plus	Gossypium_hirsutum
62	CO097080 Hypothetical protein [Medicago truncatula (Barrel medic)]	6	AATGAG	36	6	Plus	Gossypium_raimondii
	TA25429_3635	6	AATGAG	30	5	Plus	Gossypium_hirsutum
	DW228127	6	AATGAG	36	6	Plus	Gossypium_hirsutum
	TA25430_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	6	AATGAG	36	6	Plus	Gossypium_hirsutum
63	TA14464_29730 SWAP/Surp; Ubiquitin [Medicago truncatula (Barrel medic)]	6	CAGCTG	30	5	Plus	Gossypium_raimondii

Supplementary Table 4. Contd.

	DT571043 SWAP/Surp; Ubiquitin [Medicago truncatula (Barrel medic)]	6	CAGCTG	24	4	Plus	Gossypium_hirsutum
64	TA9415_29730 Two-component response regulator-like APRR7 [Arabidopsis thaliana (Mouse-ear cress)]	6	GAGCGG	30	5	Plus	Gossypium_raimondii
	TA26774_3635 Two-component response regulator-like PRR37 [Oryza sativa (Rice)]	6	GAGCGG	24	4	Plus	Gossypium_hirsutum
65	TA11166_29730 Zinc finger, RING-type; RINGv [Medicago truncatula (Barrel medic)]	6	GATAGG	30	5	Plus	Gossypium_raimondii
	TA37112_3635 Zinc finger, RING-type; RINGv [Medicago truncatula (Barrel medic)]	6	GATAGG	24	4	Plus	Gossypium_hirsutum
66	TA17050_29730 Unknown protein [Nicotiana plumbaginifolia (Leadwort-leaved tobacco)]	6	TATATC	30	5	Plus	Gossypium_raimondii
	TA28109_3635 ARIADNE-like protein ARI7 [Arabidopsis thaliana (Mouse-ear cress)]	6	TATATC	24	4	Plus	Gossypium_hirsutum
Group III							
HG	Sequence ID	Period Size	Period	Rep. Length	Copy Num.	Blast Strand	Species
1	TA9287_29730 Arabidopsis thaliana gl1 homolog [Arabidopsis thaliana (Mouse-ear cress)]	2	AT	20	10	Plus	Gossypium_raimondii
	DT049270 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	AT	24	12	Plus	Gossypium_hirsutum
2	CO114663 Arabidopsis thaliana gl1 homolog [Arabidopsis thaliana (Mouse-ear cress)]	2	AT	20	10	Plus	Gossypium_raimondii
	DT049270 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	AT	24	12	Plus	Gossypium_hirsutum
3	TA11468_29730 Hypothetical protein T11I11.9 [Arabidopsis thaliana (Mouse-ear cress)]	2	CT	20	10	Plus	Gossypium_raimondii
	TA23603_3635 Hypothetical protein T11I11.9 [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	22	11	Plus	Gossypium_hirsutum
4	CO096573 T2K10.11 protein [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	20	10	Plus	Gossypium_raimondii
	TA24566_3635 T2K10.11 protein [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	22	11	Plus	Gossypium_hirsutum
5	CO113228	2	TG	20	10	Plus	Gossypium_raimondii
	TA20238_3635	2	TG	28	14	Plus	Gossypium_hirsutum
6	CO092723	3	ACG	18	6	Plus	Gossypium_raimondii
	DT562671	3	CGA	24	8	Plus	Gossypium_hirsutum
	DR044179	3	GAC	24	8	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

	DR044172	3	GAC	27	9	Plus	Gossypium_hirsutum
7	CO121631	3	AGC	18	6	Plus	Gossypium_raimondii
	TA22417_3635	3	GCT	18	6	Minus	Gossypium_hirsutum
	DN762148	3	GCT	24	8	Minus	Gossypium_hirsutum
8	TA14807_29730 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	CAA	18	6	Plus	Gossypium_raimondii
	AI729014 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	CAA	21	7	Plus	Gossypium_hirsutum
	TA31876_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	CAA	21	7	Plus	Gossypium_hirsutum
9	TA16121_29730	3	CAG	18	6	Plus	Gossypium_raimondii
	DR457648	3	CAG	24	8	Plus	Gossypium_hirsutum
10	TA16599_29730	3	CAG	24	8	Plus	Gossypium_raimondii
	TA22510_3635	3	CAG	27	9	Plus	Gossypium_hirsutum
11	TA11781_29730 Hypothetical protein OSJNBa0003O19.4 [Oryza sativa (Rice)]	3	CGT	21	7	Plus	Gossypium_raimondii
	DR463779 Zinc finger Glo3-like protein [Medicago sativa (Alfalfa)]	3	CGT	27	9	Plus	Gossypium_hirsutum
	AI727625 Hypothetical protein At4g17885 [Arabidopsis thaliana (Mouse-ear cress)]	3	CGT	27	9	Plus	Gossypium_hirsutum
13	TA9638_29730 Eukaryotic translation initiation factor 3 subunit 10 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTG	18	6	Plus	Gossypium_raimondii
	TA22416_3635 Translation initiation factor 3 [Oryza sativa (Rice)]	3	GCT	18	6	Plus	Gossypium_hirsutum
	TA22417_3635	3	GCT	18	6	Plus	Gossypium_hirsutum
	DN762148	3	GCT	24	8	Plus	Gossypium_hirsutum
14	TA13992_29730 Biotin/lipoyl attachment [Medicago truncatula (Barrel medic)]	3	CTT	18	6	Plus	Gossypium_raimondii
	TA24677_3635 Biotin/lipoyl attachment [Medicago truncatula (Barrel medic)]	3	CTT	21	7	Plus	Gossypium_hirsutum
	DW518702 Biotin/lipoyl attachment [Medicago truncatula (Barrel medic)]	3	CTT	21	7	Plus	Gossypium_hirsutum
	TA24676_3635	3	CTT	24	8	Plus	Gossypium_hirsutum
15	TA14710_29730 Ultraviolet-B-repressible protein [Gossypium hirsutum (Upland cotton)]	3	CTT	21	7	Plus	Gossypium_raimondii
	TA39593_3635 Ultraviolet-B-repressible protein [Gossypium hirsutum (Upland cotton)]	3	CTT	24	8	Plus	Gossypium_hirsutum
16	TA11488_29730 Nucleosome/chromatin assembly factor group A [Zea mays (Maize)]	3	GAA	18	6	Plus	Gossypium_raimondii
	TA32508_3635 NAP1Ps [Pisum sativum (Garden pea)]	3	GAA	18	6	Plus	Gossypium_hirsutum
	TA32507_3635 Putative nucleosome assembly protein [Oryza sativa (japonica cultivar-group)]	3	GAA	21	7	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

17	CO082472	3	AAG	18	6	Plus	Gossypium_raimondii
	DN760252	3	GAA	21	7	Plus	Gossypium_hirsutum
	DT526895	3	AAG	18	6	Plus	Gossypium_hirsutum
	TA34933_3635	3	AAG	18	6	Plus	Gossypium_hirsutum
	TA34933_3635	3	AGA	18	6	Plus	Gossypium_hirsutum
18	CO130189	3	GAC	18	6	Plus	Gossypium_raimondii
	DT562671	3	GAC	18	6	Plus	Gossypium_hirsutum
	DR044179	3	GAC	24	8	Plus	Gossypium_hirsutum
	DT562671	3	CGA	24	8	Plus	Gossypium_hirsutum
	CO130189	3	GAC	24	8	Plus	Gossypium_raimondii
	DR044172	3	GAC	27	9	Plus	Gossypium_hirsutum
19	CO090142 F6A14.10 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_raimondii
	TA25557_3635 Putative SET protein, phospatase 2A inhibitor; 76220-74135 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	45	15	Plus	Gossypium_hirsutum
20	CO076869 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	GAT	18	6	Plus	Gossypium_raimondii
	DW225975 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	GAT	21	7	Plus	Gossypium_hirsutum
	DW499141 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	TA20630_3635 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
21	CO093670	3	GAT	24	8	Plus	Gossypium_raimondii
	TA33834_3635	3	GAT	27	9	Plus	Gossypium_hirsutum
22	CO086193 Putative SET protein, phospatase 2A inhibitor; 76220-74135 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_raimondii
	TA25557_3635 Putative SET protein, phospatase 2A inhibitor; 76220-74135 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	45	15	Plus	Gossypium_hirsutum
23	CO097964 Emb|CAB81597.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_raimondii
	TA34412_3635 Emb|CAB81597.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	21	7	Plus	Gossypium_hirsutum
24	TA9239_29730 F6A14.10 protein [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_raimondii
	TA25557_3635 Putative SET protein, phospatase 2A inhibitor; 76220-74135 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	45	15	Plus	Gossypium_hirsutum
25	CO120982 Type-B response regulator [Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle)]	3	GCA	21	7	Plus	Gossypium_raimondii
	DT548587 Type-B response regulator [Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle)]	3	GCA	24	8	Plus	Gossypium_hirsutum
26	CO113411 At2g36630/F1O11.26 [Arabidopsis thaliana (Mouse-ear cress)]	3	GGC	18	6	Plus	Gossypium_raimondii

Supplementary Table 4. Contd.

	DW482895	3	CGC	21	7	Minus	Gossypium_hirsutum
	DW482894	3	GGC	21	7	Plus	Gossypium_hirsutum
27	TA9284_29730 At2g36630/F1O11.26 [Arabidopsis thaliana (Mouse-ear cress)]	3	GGC	18	6	Plus	Gossypium_raimondii
	DW482895	3	CGC	21	7	Minus	Gossypium_hirsutum
	DW482894	3	GGC	21	7	Plus	Gossypium_hirsutum
28	CO071933 At2g36630/F1O11.26 [Arabidopsis thaliana (Mouse-ear cress)]	3	GGC	18	6	Plus	Gossypium_raimondii
	DW482895	3	CGC	21	7	Minus	Gossypium_hirsutum
	DW482894	3	GGC	21	7	Plus	Gossypium_hirsutum
29	TA10085_29730	3	TCA	21	7	Plus	Gossypium_raimondii
	TA31716_3635	3	TCA	24	8	Plus	Gossypium_hirsutum
30	TA17016_29730	3	TCA	21	7	Plus	Gossypium_raimondii
	TA28467_3635 Putative bHLH transcription factor [Arabidopsis thaliana (Mouse-ear cress)]	3	CAT	18	6	Plus	Gossypium_hirsutum
	TA28466_3635	3	CAT	18	6	Plus	Gossypium_hirsutum
	TA17016_29730	3	CAT	21	7	Plus	Gossypium_raimondii
	TA28466_3635	3	CAT	27	9	Plus	Gossypium_hirsutum
31	CO116980	3	TCA	18	6	Plus	Gossypium_raimondii
	DT549374	3	TCA	24	8	Plus	Gossypium_hirsutum
32	CO128567 Putative arabinogalactan-protein [Oryza sativa (japonica cultivar-group)]	3	TCT	21	7	Plus	Gossypium_raimondii
	DR044171 Gb|AAF02137.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	27	9	Plus	Gossypium_hirsutum
	TA21781_3635 Gb|AAF02137.1 [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	27	9	Plus	Gossypium_hirsutum
	TA21780_3635 At5g06390 [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	27	9	Plus	Gossypium_hirsutum
33	TA11292_29730 Hypothetical protein At2g35680; T20F21.13 [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	21	7	Plus	Gossypium_raimondii
	TA24343_3635 Hypothetical protein At2g35680; T20F21.13 [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	24	8	Plus	Gossypium_hirsutum
34	CO085995 Putative PTEN-like phosphatase [Oryza sativa (japonica cultivar-group)]	3	TCT	21	7	Plus	Gossypium_raimondii
	TA24343_3635 Hypothetical protein At2g35680; T20F21.13 [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	24	8	Plus	Gossypium_hirsutum
35	CO095090 Putative PTEN-like phosphatase [Oryza sativa (japonica cultivar-group)]	3	TCT	21	7	Plus	Gossypium_raimondii
	TA24343_3635 Hypothetical protein At2g35680; T20F21.13 [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	24	8	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

36	TA12487_29730 Putative leucine-rich repeat disease resistance protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	21	7	Plus	Gossypium_raimondii
	TA35487_3635 Putative leucine-rich repeat disease resistance protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TCT	24	8	Plus	Gossypium_hirsutum
37	CO125503 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	18	6	Plus	Gossypium_raimondii
	DT047070	3	TCA	21	7	Minus	Gossypium_hirsutum
	AJ513319 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DW499141 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	TA20630_3635 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
38	CO123383 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	21	7	Plus	Gossypium_raimondii
	DT047070	3	TCA	21	7	Minus	Gossypium_hirsutum
	TA20630_3635 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	AJ513319 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DW499141 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DW225975 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	GAT	21	7	Plus	Gossypium_hirsutum
	TA42103_3635	3	TGA	24	8	Plus	Gossypium_hirsutum
	CO493420 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	24	8	Plus	Gossypium_hirsutum
39	TA9193_29730 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	21	7	Plus	Gossypium_raimondii
	AJ513319 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DT047070	3	TCA	21	7	Minus	Gossypium_hirsutum
	TA20630_3635 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DW225975 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	GAT	21	7	Plus	Gossypium_hirsutum
	DW499141 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	TA42103_3635	3	TGA	24	8	Plus	Gossypium_hirsutum
	CO493420 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	24	8	Plus	Gossypium_hirsutum
40	TA10749_29730 Phloroglucinol O-methyltransferase [Rosa chinensis var. spontanea]	3	TGA	21	7	Plus	Gossypium_raimondii
	TA35930_3635 Phloroglucinol O-methyltransferase [Rosa chinensis var. spontanea]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DT048525	3	TCA	24	8	Minus	Gossypium_hirsutum
41	TA9659_29730 MADS box protein GHMADS-1 [Gossypium hirsutum (Upland cotton)]	3	TTA	18	6	Plus	Gossypium_raimondii
	DR457971	3	AAT	18	6	Minus	Gossypium_hirsutum
	DW231765 MADS box protein GHMADS-1 [Gossypium hirsutum (Upland cotton)]	3	TTA	18	6	Plus	Gossypium_hirsutum
	DT048252 MADS box protein GHMADS-1 [Gossypium hirsutum (Upland cotton)]	3	TTA	18	6	Plus	Gossypium_hirsutum
	DT457922 MADS box protein GHMADS-1 [Gossypium hirsutum (Upland cotton)]	3	TTA	21	7	Plus	Gossypium_hirsutum
	TA24540_3635 MADS box protein GHMADS-1 [Gossypium hirsutum (Upland cotton)]	3	TTA	21	7	Plus	Gossypium_hirsutum

Supplementary Table 4. Contd.

42	TA13626_29730 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	21	7	Plus	Gossypium_raimondii
	DR459816 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	27	9	Plus	Gossypium_hirsutum
	TA26423_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	27	9	Plus	Gossypium_hirsutum
43	TA13856_29730 Thylakoid lumenal protein-like [Oryza sativa (japonica cultivar-group)]	3	TTC	18	6	Plus	Gossypium_raimondii
	TA26484_3635 Thylakoid lumenal protein-like [Oryza sativa (japonica cultivar-group)]	3	TTC	18	6	Plus	Gossypium_hirsutum
	DN803954 Hypothetical protein At1g12250 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	21	7	Plus	Gossypium_hirsutum
44	TA14465_29730 Acid phosphatase-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	24	8	Plus	Gossypium_raimondii
	AI054761	3	TTC	27	9	Plus	Gossypium_hirsutum
45	CO105461 CBL-interacting serine/threonine-protein kinase 11 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	18	6	Plus	Gossypium_raimondii
	TA42588_3635 CBL-interacting serine/threonine-protein kinase 11 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	24	8	Plus	Gossypium_hirsutum
46	CO074478 Thylakoid lumenal protein-like [Oryza sativa (japonica cultivar-group)]	3	TTC	18	6	Plus	Gossypium_raimondii
	TA26484_3635 Thylakoid lumenal protein-like [Oryza sativa (japonica cultivar-group)]	3	TTC	18	6	Plus	Gossypium_hirsutum
	DN803954 Hypothetical protein At1g12250 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTC	21	7	Plus	Gossypium_hirsutum
47	CO123080	3	TTC	18	6	Plus	Gossypium_raimondii
	DN802775	3	TTC	21	7	Plus	Gossypium_hirsutum
48	CO069614 Acid phosphatase 1 precursor (EC 3.1.3.2) (Apase-1(1)) [Lycopersicon esculentum (Tomato)]	3	TTC	24	8	Plus	Gossypium_raimondii
	AI054761	3	TTC	27	9	Plus	Gossypium_hirsutum
49	TA13320_29730	4	GACA	20	5	Plus	Gossypium_raimondii
	TA28576_3635	4	GACA	24	6	Plus	Gossypium_hirsutum
50	CO075206	4	GATA	20	5	Plus	Gossypium_raimondii
	TA26759_3635	4	TATC	24	6	Minus	Gossypium_hirsutum
51	TA12438_29730	3	TGG	21	7	Plus	Gossypium_raimondii
	TA22248_3635	3	CAC	39	13	Minus	Gossypium_hirsutum
	TA22249_3635	3	CAC	18	6	Minus	Gossypium_hirsutum
	DN801177	3	CAC	39	13	Minus	Gossypium_hirsutum
52	TA10237_29730 Expressed protein [Arabidopsis thaliana (Mouse-ear cress)]	4	TTTA	20	5	Plus	Gossypium_raimondii

Supplementary Table 4. Contd.

	TA22189_3635 Expressed protein [Arabidopsis thaliana (Mouse-ear cress)]	4	TTTA	24	6	Plus	Gossypium_hirsutum
	TA22190_3635 Expressed protein [Arabidopsis thaliana (Mouse-ear cress)]	4	TTTA	24	6	Plus	Gossypium_hirsutum
	DT569576 Expressed protein [Arabidopsis thaliana (Mouse-ear cress)]	4	TTTA	24	6	Plus	Gossypium_hirsutum
53	TA10238_29730 Hypothetical protein [Medicago truncatula (Barrel medic)]	4	TTTA	20	5	Plus	Gossypium_raimondii
	TA22190_3635 Expressed protein [Arabidopsis thaliana (Mouse-ear cress)]	4	TTTA	24	6	Plus	Gossypium_hirsutum
	DT569576 Expressed protein [Arabidopsis thaliana (Mouse-ear cress)]	4	TTTA	24	6	Plus	Gossypium_hirsutum
	TA22189_3635 Expressed protein [Arabidopsis thaliana (Mouse-ear cress)]	4	TTTA	24	6	Plus	Gossypium_hirsutum
54	TA14521_29730 Protein At2g38820 [Arabidopsis thaliana (Mouse-ear cress)]	6	CCTTCA	24	4	Plus	Gossypium_raimondii
	TA39319_3635 Protein At2g38820 [Arabidopsis thaliana (Mouse-ear cress)]	6	CCTTCA	30	5	Plus	Gossypium_hirsutum
55	CO116889 Fas-associated factor 1-like protein [Capsicum annuum (Bell pepper)]	6	GCCTCC	30	5	Plus	Gossypium_raimondii
	TA24449_3635 Fas-associated factor 1-like protein [Capsicum annuum (Bell pepper)]	6	GCCTCC	36	6	Plus	Gossypium_hirsutum
56	TA9155_29730 Splicing factor SR1D [Arabidopsis thaliana (Mouse-ear cress)]	6	GGAGCA	24	4	Plus	Gossypium_raimondii
	TA22897_3635 Splicing factor SR1D [Arabidopsis thaliana (Mouse-ear cress)]	6	GGAGCA	30	5	Plus	Gossypium_hirsutum
57	TA17292_29730 Hypothetical protein OJ1150_A11.18 [Oryza sativa (japonica cultivar-group)]	6	TCACCA	24	4	Plus	Gossypium_raimondii
	TA32583_3635 Hypothetical protein OJ1150_A11.18 [Oryza sativa (japonica cultivar-group)]	6	TCACCA	30	5	Plus	Gossypium_hirsutum
58	CO123990	6	TGCTCC	24	4	Plus	Gossypium_raimondii
	TA22898_3635	6	GGAGCA	36	6	Minus	Gossypium_hirsutum
59	TA9156_29730	6	TGCTCC	24	4	Plus	Gossypium_raimondii
	TA22898_3635	6	GGAGCA	36	6	Minus	Gossypium_hirsutum
60	CO110189	6	TGCTCC	24	4	Plus	Gossypium_raimondii
	TA22898_3635	6	GGAGCA	36	6	Minus	Gossypium_hirsutum
61	CO086499	6	TGCTCC	24	4	Plus	Gossypium_raimondii
	TA22898_3635	6	GGAGCA	36	6	Minus	Gossypium_hirsutum

Supplementary Table 5.

Group I

HG	Sequence ID	Period Size	Period	Rep. Length	Copy Num.	Blast Strand	Species
1	BQ405858	2	GA	18	9	Plus	Gossypium_arboreum
	CO103483 F26K24.14 protein [Arabidopsis thaliana (Mouse-ear cress)]	2	GA	18	9	Plus	Gossypium_raimondii
2	BQ402652 Eukaryotic translation initiation factor 2 subunit beta [Triticum aestivum (Wheat)]	3	AAG	18	6	Plus	Gossypium_arboreum
	CO088220 Eukaryotic translation initiation factor 2 subunit beta [Triticum aestivum (Wheat)]	3	AAG	18	6	Plus	Gossypium_raimondii
	TA14327_29730 Eukaryotic translation initiation factor 2 beta subunit-like [Solanum tuberosum (Potato)]	3	AAG	18	6	Plus	Gossypium_raimondii
3	BE052321 Putative ubiquinol-cytochrome-c reductase cytochrome c1 [Oryza sativa (japonica cultivar-group)]	3	ACC	18	6	Plus	Gossypium_arboreum
	TA9532_29730 Putative ubiquinol-cytochrome-c reductase cytochrome c1 [Oryza sativa (japonica cultivar-group)]	3	ACC	18	6	Plus	Gossypium_raimondii
4	BG440558 Putative ubiquinol-cytochrome-c reductase cytochrome c1 [Oryza sativa (japonica cultivar-group)]	3	ACC	18	6	Plus	Gossypium_arboreum
	TA9532_29730 Putative ubiquinol-cytochrome-c reductase cytochrome c1 [Oryza sativa (japonica cultivar-group)]	3	ACC	18	6	Plus	Gossypium_raimondii
5	BG443450	3	ACC	18	6	Plus	Gossypium_arboreum
	TA9532_29730 Putative ubiquinol-cytochrome-c reductase cytochrome c1 [Oryza sativa (japonica cultivar-group)]	3	ACC	18	6	Plus	Gossypium_raimondii
6	TA5263_29729 Cytochrome c1, heme protein, mitochondrial precursor [Solanum tuberosum (Potato)]	3	ACC	18	6	Plus	Gossypium_arboreum
	TA9532_29730 Putative ubiquinol-cytochrome-c reductase cytochrome c1 [Oryza sativa (japonica cultivar-group)]	3	ACC	18	6	Plus	Gossypium_raimondii
7	BM360664 Cytochrome c1 [Arabidopsis thaliana (Mouse-ear cress)]	3	ACC	18	6	Plus	Gossypium_arboreum
	TA9532_29730 Putative ubiquinol-cytochrome-c reductase cytochrome c1 [Oryza sativa (japonica cultivar-group)]	3	ACC	18	6	Plus	Gossypium_raimondii
8	BQ411088 4-alpha-glucanotransferase [Solanum tuberosum (Potato)]	3	ATT	51	17	Plus	Gossypium_arboreum
	CO075512	3	AAT	51	17	Minus	Gossypium_raimondii
	TA9504_29730 Putative 4-alpha-glucanotransferase [Oryza sativa (japonica cultivar-group)]	3	ATT	51	17	Plus	Gossypium_raimondii
9	BG443488 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_arboreum
	CO092245 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_raimondii
10	BF276058 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_arboreum
	CO092245 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_raimondii
11	BQ403789 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_arboreum
	CO092245 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_raimondii

Supplementary Table 5. Contd.

12	BG446135	3	CTT	18	6	Plus	Gossypium_arboreum
	TA14664_29730 Zinc finger, RING-type; RINGv [Medicago truncatula (Barrel medic)]	3	CTT	18	6	Plus	Gossypium_raimondii
13	BE052283 Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K14A3 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	18	6	Plus	Gossypium_arboreum
	CO080640 Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K14A3 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	18	6	Plus	Gossypium_raimondii
14	BQ405354 Ultraviolet-B-repressible protein [Gossypium hirsutum (Upland cotton)]	3	CTT	21	7	Plus	Gossypium_arboreum
	TA14710_29730 Ultraviolet-B-repressible protein [Gossypium hirsutum (Upland cotton)]	3	CTT	21	7	Plus	Gossypium_raimondii
15	BG445086	3	GAT	18	6	Plus	Gossypium_arboreum
	TA15485_29730 Hypothetical protein At3g20260 [Arabidopsis thaliana (Mouse-ear cress)]	3	GAT	18	6	Plus	Gossypium_raimondii
16	BG442229 Von Willebrand factor, type A; Ubiquitin interacting motif [Medicago truncatula (Barrel medic)]	3	TGC	18	6	Plus	Gossypium_arboreum
	CO088783 Von Willebrand factor, type A; Ubiquitin interacting motif [Medicago truncatula (Barrel medic)]	3	TGC	18	6	Plus	Gossypium_raimondii
	CO105343 Von Willebrand factor, type A; Ubiquitin interacting motif [Medicago truncatula (Barrel medic)]	3	TGC	18	6	Plus	Gossypium_raimondii
17	BQ404857 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	21	7	Plus	Gossypium_arboreum
	CO099414 Hypothetical protein OJ1734_E02.2 [Oryza sativa (japonica cultivar-group)]	3	TTC	21	7	Plus	Gossypium_raimondii
	TA13626_29730 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	21	7	Plus	Gossypium_raimondii
18	TA7048_29729 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	21	7	Plus	Gossypium_arboreum
	CO099414 Hypothetical protein OJ1734_E02.2 [Oryza sativa (japonica cultivar-group)]	3	TTC	21	7	Plus	Gossypium_raimondii
	TA13626_29730 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	21	7	Plus	Gossypium_raimondii
19	BM357948 Mitogen-activated protein kinase [Euphorbia esula (Leafy spurge)]	4	TATG	20	5	Plus	Gossypium_arboreum
	TA14305_29730 Mitogen-activated protein kinase 1 [Lycopersicon esculentum (Tomato)]	4	TATG	20	5	Plus	Gossypium_raimondii
	TA16726_29730 Mitogen-activated protein kinase 2 [Lycopersicon esculentum (Tomato)]	4	TATG	20	5	Plus	Gossypium_raimondii
20	TA7673_29729 Transthyretin-like protein [Arabidopsis thaliana (Mouse-ear cress)]	5	AGAAG	25	5	Plus	Gossypium_arboreum
	CO107385 Transthyretin-like protein [Arabidopsis thaliana (Mouse-ear cress)]	5	AGAAG	25	5	Plus	Gossypium_raimondii
	TA10548_29730 Transthyretin-like protein [Arabidopsis thaliana (Mouse-ear cress)]	5	AGAAG	25	5	Plus	Gossypium_raimondii
	CO106914 Transthyretin-like protein [Arabidopsis thaliana (Mouse-ear cress)]	5	AGAAG	25	5	Plus	Gossypium_raimondii
	CO071413 Transthyretin-like protein [Arabidopsis thaliana (Mouse-ear cress)]	5	AGAAG	25	5	Plus	Gossypium_raimondii
21	BG444055 DNA-binding protein-related-like [Oryza sativa (japonica cultivar-group)]	5	AGAGA	30	6	Plus	Gossypium_arboreum

Supplementary Table 5. Contd.

	TA13177_29730 DNA-binding protein-related-like [Oryza sativa (japonica cultivar-group)]	5	AGAGA	30	6	Plus	Gossypium_raimondii
22	BQ412345	5	CAAAC	25	5	Plus	Gossypium_arboreum
	CO107215	5	CAAAC	25	5	Plus	Gossypium_raimondii

Group II

HG	Sequence ID	Period Size	Period	Rep. Length	Copy Num.	Blast Strand	Species
1	BF269259	3	CAG	21	7	Plus	Gossypium_arboreum
	CO086279	3	CTG	18	6	Minus	Gossypium_raimondii
2	BQ413494	3	CAT	24	8	Plus	Gossypium_arboreum
	CO119751	3	CAT	21	7	Plus	Gossypium_raimondii
3	BQ415217 CBS1 [Hyacinthus orientalis (Common hyacinth)]	3	GAT	24	8	Plus	Gossypium_arboreum
	TA14678_29730 CBS1 [Hyacinthus orientalis (Common hyacinth)]	3	GAT	21	7	Plus	Gossypium_raimondii
4	TA7800_29729 CBS1 [Hyacinthus orientalis (Common hyacinth)]	3	GAT	24	8	Plus	Gossypium_arboreum
	TA14678_29730 CBS1 [Hyacinthus orientalis (Common hyacinth)]	3	GAT	21	7	Plus	Gossypium_raimondii
5	BQ405592 unknown protein [Arabidopsis thaliana]	3	GTG	21	7	Plus	Gossypium_arboreum
	TA10731_29730 unknown protein [Arabidopsis thaliana]	3	GTG	18	6	Plus	Gossypium_raimondii
6	BG444426	3	TCT	24	8	Plus	Gossypium_arboreum
	CO128567 Putative arabinogalactan-protein [Oryza sativa (japonica cultivar-group)]	3	TCT	21	7	Plus	Gossypium_raimondii
7	CK640498 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	24	8	Plus	Gossypium_arboreum
	CO076869 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	GAT	18	6	Plus	Gossypium_raimondii
	CO125503 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	18	6	Plus	Gossypium_raimondii
	CO123383 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	21	7	Plus	Gossypium_raimondii
	TA9193_29730 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	21	7	Plus	Gossypium_raimondii
8	BQ410226	3	TGA	24	8	Plus	Gossypium_arboreum
	CO121651	3	TGA	21	7	Plus	Gossypium_raimondii
9	BF277771 Alanine aminotransferase [Capsicum annuum (Bell pepper)]	3	TTC	21	7	Plus	Gossypium_arboreum
	TA10152_29730 Alanine aminotransferase [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	18	6	Plus	Gossypium_raimondii

Supplementary Table 5. Contd.

Group III

HG	Sequence ID	Period Size	Period	Rep. Length	Copy Num.	Blast Strand	Species
1	BF268601	2	TA	24	12	Plus	Gossypium_arboreum
	CO102551	2	AT	28	14	Minus	Gossypium_raimondii
2	BF270284 Pentaxin [Medicago truncatula (Barrel medic)]	2	TC	18	9	Plus	Gossypium_arboreum
	CO108840 F6I1.14 protein [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	20	10	Plus	Gossypium_raimondii
3	BG444624	3	AGA	18	6	Plus	Gossypium_arboreum
	CO116035 Adenylate translocator (Brittle-1)-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	21	7	Plus	Gossypium_raimondii
	CO129567 Adenylate translocator (Brittle-1)-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	21	7	Plus	Gossypium_raimondii
4	BE055500 MYC2 [Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle)]	3	CTC	18	6	Plus	Gossypium_arboreum
	TA17138_29730 MYC2 [Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle)]	3	CTC	21	7	Plus	Gossypium_raimondii
5	BE054572 Calnexin [Glycine max (Soybean)]	3	GAA	21	7	Plus	Gossypium_arboreum
	TA9085_29730	3	TCT	27	9	Minus	Gossypium_raimondii
	TA9084_29730 Calnexin [Glycine max (Soybean)]	3	GAA	27	9	Plus	Gossypium_raimondii
6	BM359886 Cleft lip and palate associated transmembrane protein-like [Arabidopsis thaliana (Mouse-ear cress)]	3	GGA	21	7	Plus	Gossypium_arboreum
	CO085114 Cleft lip and palate associated transmembrane protein-like [Arabidopsis thaliana (Mouse-ear cress)]	3	GGA	24	8	Plus	Gossypium_raimondii
7	BQ402342	3	TCT	21	7	Plus	Gossypium_arboreum
	TA9085_29730	3	TCT	27	9	Plus	Gossypium_raimondii
8	TA5514_29729 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	18	6	Plus	Gossypium_arboreum
	CO076869 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	GAT	18	6	Plus	Gossypium_raimondii
	CO125503 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	18	6	Plus	Gossypium_raimondii
	CO123383 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	21	7	Plus	Gossypium_raimondii
	TA9193_29730 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	21	7	Plus	Gossypium_raimondii
9	TA5985_29729 Protein At3g63460 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	18	6	Plus	Gossypium_arboreum
	TA17231_29730 Protein At3g63460 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	21	7	Plus	Gossypium_raimondii
10	BQ405038 Hypothetical protein T5J17.10 [Arabidopsis thaliana (Mouse-ear cress)]	4	AATT	20	5	Plus	Gossypium_arboreum

Supplementary Table 5. Contd.

	CO102888 Hypothetical protein T5J17.10 [Arabidopsis thaliana (Mouse-ear cress)]	4	AATT	24	6	Plus	Gossypium_raimondii
	CO078485 Hypothetical protein [Medicago truncatula (Barrel medic)]	4	AATT	24	6	Plus	Gossypium_raimondii
11	BG445815 T5A14.7 protein [Arabidopsis thaliana (Mouse-ear cress)]	6	GATAGG	24	4	Plus	Gossypium_arboreum
	TA11166_29730 Zinc finger, RING-type; RINGv [Medicago truncatula (Barrel medic)]	6	GATAGG	30	5	Plus	Gossypium_raimondii
12	BG445177	6	GATAGG	24	4	Plus	Gossypium_arboreum
	TA11166_29730 Zinc finger, RING-type; RINGv [Medicago truncatula (Barrel medic)]	6	GATAGG	30	5	Plus	Gossypium_raimondii

Supplementary Table 6.

HG with identical repeat number							
HG	Sequence ID	Period Size	Period	Rep. Length	Copy Num.	Blast Strand	Species
1	BQ402652 Eukaryotic translation initiation factor 2 subunit beta [Triticum aestivum (Wheat)]	3	AAG	18	6	Plus	Gossypium_arboreum
	CO088220 Eukaryotic translation initiation factor 2 subunit beta [Triticum aestivum (Wheat)]	3	AAG	18	6	Plus	Gossypium_raimondii
	TA14327_29730 Eukaryotic translation initiation factor 2 beta subunit-like [Solanum tuberosum (Potato)]	3	AAG	18	6	Plus	Gossypium_raimondii
	TA28888_3635 Eukaryotic translation initiation factor 2 beta subunit-like [Solanum tuberosum (Potato)]	3	AAG	18	6	Plus	Gossypium_hirsutum
2	BQ403789 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_arboreum
	DN800615 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_hirsutum
	CO092245 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_raimondii
	TA23709_3635 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_hirsutum
3	BF276058 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_arboreum
	DN800615 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_hirsutum
	CO092245 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_raimondii
	TA23709_3635 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_hirsutum
4	BG443488 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_arboreum
	DN800615 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_hirsutum
	CO092245 RAR1 [Arabidopsis thaliana (Mouse-ear cress)]	3	CCA	21	7	Plus	Gossypium_raimondii

Supplementary Table 6. Contd.

	TA23709_3635 Putative Rar1 protein [Solanum demissum (Wild potato)]	3	CCA	21	7	Plus	Gossypium_hirsutum
5	BE052283 Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K14A3 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	18	6	Plus	Gossypium_arboreum
	DW496704 Hypothetical protein [Brassica oleracea (Wild cabbage)]	3	CTT	18	6	Plus	Gossypium_hirsutum
	CO080640 Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K14A3 [Arabidopsis thaliana (Mouse-ear cress)]	3	CTT	18	6	Plus	Gossypium_raimondii
6	BM357948 Mitogen-activated protein kinase [Euphorbia esula (Leafy spurge)]	4	TATG	20	5	Plus	Gossypium_arboreum
	TA14305_29730 Mitogen-activated protein kinase 1 [Lycopersicon esculentum (Tomato)]	4	TATG	20	5	Plus	Gossypium_raimondii
	TA23495_3635 Mitogen-activated protein kinase 1 [Lycopersicon esculentum (Tomato)]	4	TATG	20	5	Plus	Gossypium_hirsutum
	TA16726_29730 Mitogen-activated protein kinase 2 [Lycopersicon esculentum (Tomato)]	4	TATG	20	5	Plus	Gossypium_raimondii

HG with different repeat number

HG	Sequence ID	Period Size	Period	Rep. Length	Copy Num.	Blast Strand	Species
1	BF268204	2	TC	28	14	Plus	Gossypium_arboreum
	TA13307_29730 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	CT	26	13	Plus	Gossypium_raimondii
	DN802296	2	CT	20	10	Plus	Gossypium_hirsutum
	TA35543_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	TC	22	11	Plus	Gossypium_hirsutum
2	TA5890_29729 LIM-domain SF3 protein [Nicotiana tabacum (Common tobacco)]	2	CT	30	15	Plus	Gossypium_arboreum
	CO094613 LIM-domain SF3 protein [Nicotiana tabacum (Common tobacco)]	2	CT	20	10	Plus	Gossypium_raimondii
	DR458537 Putative LIM domain protein PLIM-2 [Oryza sativa (japonica cultivar-group)]	2	CT	22	11	Plus	Gossypium_hirsutum
	AI054962 Putative LIM domain protein PLIM-2 [Oryza sativa (japonica cultivar-group)]	2	CT	22	11	Plus	Gossypium_hirsutum
3	TA7156_29729 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	TC	28	14	Plus	Gossypium_arboreum
	TA13307_29730 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	CT	26	13	Plus	Gossypium_raimondii
	DN802296	2	CT	20	10	Plus	Gossypium_hirsutum
	TA35543_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	2	TC	22	11	Plus	Gossypium_hirsutum
4	BF272268 OSJNBb0039L24.7 protein [Oryza sativa (japonica cultivar-group)]	2	CT	30	15	Plus	Gossypium_arboreum
	CO094613 LIM-domain SF3 protein [Nicotiana tabacum (Common tobacco)]	2	CT	20	10	Plus	Gossypium_raimondii
	DR458537 Putative LIM domain protein PLIM-2 [Oryza sativa (japonica cultivar-group)]	2	CT	22	11	Plus	Gossypium_hirsutum
	AI054962 Putative LIM domain protein PLIM-2 [Oryza sativa (japonica cultivar-group)]	2	CT	22	11	Plus	Gossypium_hirsutum

Supplementary Table 6. Contd.

5	BE052113	2	CT	26	13	Plus	Gossypium_arboreum
	TA12822_29730	2	CT	22	11	Plus	Gossypium_raimondii
	TA30762_3635 Targeting protein for Xklp2 containing protein, expressed [Oryza sativa (japonica cultivar-group)]	2	TC	22	11	Plus	Gossypium_hirsutum
6	BF270284 Pentaxin [Medicago truncatula (Barrel medic)]	2	TC	18	9	Plus	Gossypium_arboreum
	CO108840 F6I1.14 protein [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	20	10	Plus	Gossypium_raimondii
	TA44152_3635 F6I1.14 protein [Arabidopsis thaliana (Mouse-ear cress)]	2	TC	22	11	Plus	Gossypium_hirsutum
7	BF276199 Putative early nodule-specific-like protein ENOD8 [Oryza sativa (japonica cultivar-group)]	3	AAG	21	7	Plus	Gossypium_arboreum
	CO086777 Putative early nodule-specific-like protein ENOD8 [Oryza sativa (japonica cultivar-group)]	3	AAG	18	6	Plus	Gossypium_raimondii
	TA35757_3635 Putative early nodule-specific-like protein ENOD8 [Oryza sativa (japonica cultivar-group)]	3	AAG	24	8	Plus	Gossypium_hirsutum
8	TA5263_29729 Cytochrome c1, heme protein, mitochondrial precursor [Solanum tuberosum (Potato)]	3	ACC	18	6	Plus	Gossypium_arboreum
	TA9532_29730 Putative ubiquinol-cytochrome-c reductase cytochrome c1 [Oryza sativa (japonica cultivar-group)]	3	ACC	18	6	Plus	Gossypium_raimondii
	TA24730_3635 Putative ubiquinol-cytochrome-c reductase cytochrome c1 [Oryza sativa (japonica cultivar-group)]	3	ACC	24	8	Plus	Gossypium_hirsutum
9	BG444624	3	AGA	18	6	Plus	Gossypium_arboreum
	CO116035 Adenylate translocator (Brittle-1)-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	21	7	Plus	Gossypium_raimondii
	CO129567 Adenylate translocator (Brittle-1)-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	21	7	Plus	Gossypium_raimondii
	TA23448_3635 Adenylate translocator (Brittle-1)-like protein [Arabidopsis thaliana (Mouse-ear cress)]	3	AGA	18	6	Plus	Gossypium_hirsutum
10	BQ411088 4-alpha-glucanotransferase [Solanum tuberosum (Potato)]	3	ATT	51	17	Plus	Gossypium_arboreum
	CO075512	3	AAT	51	17	Minus	Gossypium_raimondii
	TA9504_29730 Putative 4-alpha-glucanotransferase [Oryza sativa (japonica cultivar-group)]	3	ATT	51	17	Plus	Gossypium_raimondii
	TA38929_3635 4-alpha-glucanotransferase [Solanum tuberosum (Potato)]	3	ATT	33	11	Plus	Gossypium_hirsutum
11	TA6708_29729	3	CAG	33	11	Plus	Gossypium_arboreum
	TA16599_29730	3	CAG	24	8	Plus	Gossypium_raimondii
	TA22510_3635	3	CAG	27	9	Plus	Gossypium_hirsutum
12	BF269259	3	CAG	21	7	Plus	Gossypium_arboreum
	CO086279	3	CTG	18	6	Minus	Gossypium_raimondii
	TA37795_3635	3	TGC	18	6	Minus	Gossypium_hirsutum

Supplementary Table 6. Contd.

13	BE055500 MYC2 [Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle)]	3	CTC	18	6	Plus	Gossypium_arboreum
	TA17138_29730 MYC2 [Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle)]	3	CTC	21	7	Plus	Gossypium_raimondii
	TA21503_3635	3	CTC	21	7	Plus	Gossypium_hirsutum
14	BQ405354 Ultraviolet-B-repressible protein [Gossypium hirsutum (Upland cotton)]	3	CTT	21	7	Plus	Gossypium_arboreum
	TA14710_29730 Ultraviolet-B-repressible protein [Gossypium hirsutum (Upland cotton)]	3	CTT	21	7	Plus	Gossypium_raimondii
	TA39593_3635 Ultraviolet-B-repressible protein [Gossypium hirsutum (Upland cotton)]	3	CTT	24	8	Plus	Gossypium_hirsutum
15	BE054572 Calnexin [Glycine max (Soybean)]	3	GAA	21	7	Plus	Gossypium_arboreum
	TA9085_29730	3	TCT	27	9	Minus	Gossypium_raimondii
	TA9084_29730 Calnexin [Glycine max (Soybean)]	3	GAA	27	9	Plus	Gossypium_raimondii
	TA22700_3635 Calnexin [Glycine max (Soybean)]	3	GAA	21	7	Plus	Gossypium_hirsutum
16	BM359886 Cleft lip and palate associated transmembrane protein-like [Arabidopsis thaliana (Mouse-ear cress)]	3	GGA	21	7	Plus	Gossypium_arboreum
	CO085114 Cleft lip and palate associated transmembrane protein-like [Arabidopsis thaliana (Mouse-ear cress)]	3	GGA	24	8	Plus	Gossypium_raimondii
	TA28051_3635 Cleft lip and palate associated transmembrane protein-like [Arabidopsis thaliana (Mouse-ear cress)]	3	GGA	24	8	Plus	Gossypium_hirsutum
	DT559190 Cleft lip and palate associated transmembrane protein-like [Arabidopsis thaliana (Mouse-ear cress)]	3	GGA	27	9	Plus	Gossypium_hirsutum
17	BQ405592 unknown protein [Arabidopsis thaliana]	3	GTG	21	7	Plus	Gossypium_arboreum
	TA10731_29730 unknown protein [Arabidopsis thaliana]	3	GTG	18	6	Plus	Gossypium_raimondii
	TA38018_3635 unknown protein [Arabidopsis thaliana]	3	GTG	18	6	Plus	Gossypium_hirsutum
18	TA5514_29729 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	18	6	Plus	Gossypium_arboreum
	CO125503 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	18	6	Plus	Gossypium_raimondii
	CO123383 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	21	7	Plus	Gossypium_raimondii
	TA9193_29730 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	21	7	Plus	Gossypium_raimondii
	AJ513319 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DT047070	3	TCA	21	7	Minus	Gossypium_hirsutum
	DW225975 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	GAT	21	7	Plus	Gossypium_hirsutum
	DW499141 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	TA20630_3635 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	CO493420 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	24	8	Plus	Gossypium_hirsutum
19	CK640498 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	24	8	Plus	Gossypium_arboreum

Supplementary Table 6. Contd.

	CO125503 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	18	6	Plus	Gossypium_raimondii
	CO076869 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	GAT	18	6	Plus	Gossypium_raimondii
	CO123383 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	21	7	Plus	Gossypium_raimondii
	TA9193_29730 Salt tolerance protein [Sesuvium portulacastrum (Shoreline sea purslane)]	3	TGA	21	7	Plus	Gossypium_raimondii
	DT047070	3	TCA	21	7	Minus	Gossypium_hirsutum
	AJ513319 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DW499141 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	DW225975 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	GAT	21	7	Plus	Gossypium_hirsutum
	TA20630_3635 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	21	7	Plus	Gossypium_hirsutum
	CO493420 Late embryogenesis-like protein [Prunus armeniaca (Apricot)]	3	TGA	24	8	Plus	Gossypium_hirsutum
20	TA5985_29729 Protein At3g63460 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	18	6	Plus	Gossypium_arboreum
	TA17231_29730 Protein At3g63460 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	21	7	Plus	Gossypium_raimondii
	TA26634_3635	3	TAA	18	6	Minus	Gossypium_hirsutum
	TA26636_3635	3	ATA	21	7	Minus	Gossypium_hirsutum
	TA26635_3635 Protein At3g63460 [Arabidopsis thaliana (Mouse-ear cress)]	3	TTA	21	7	Plus	Gossypium_hirsutum
21	BQ404857 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	21	7	Plus	Gossypium_arboreum
	TA13626_29730 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	21	7	Plus	Gossypium_raimondii
	CO099414 Hypothetical protein OJ1734_E02.2 [Oryza sativa (japonica cultivar-group)]	3	TTC	21	7	Plus	Gossypium_raimondii
	TA26423_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	27	9	Plus	Gossypium_hirsutum
	DR459816 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	27	9	Plus	Gossypium_hirsutum
22	TA7048_29729 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	21	7	Plus	Gossypium_arboreum
	TA13626_29730 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	21	7	Plus	Gossypium_raimondii
	CO099414 Hypothetical protein OJ1734_E02.2 [Oryza sativa (japonica cultivar-group)]	3	TTC	21	7	Plus	Gossypium_raimondii
	TA26423_3635 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	27	9	Plus	Gossypium_hirsutum
	DR459816 Hypothetical protein [Medicago truncatula (Barrel medic)]	3	TTC	27	9	Plus	Gossypium_hirsutum
23	BQ405038 Hypothetical protein T5J17.10 [Arabidopsis thaliana (Mouse-ear cress)]	4	AATT	20	5	Plus	Gossypium_arboreum
	CO078485 Hypothetical protein [Medicago truncatula (Barrel medic)]	4	AATT	24	6	Plus	Gossypium_raimondii
	CO102888 Hypothetical protein T5J17.10 [Arabidopsis thaliana (Mouse-ear cress)]	4	AATT	24	6	Plus	Gossypium_raimondii
	DV850022 Hypothetical protein [Medicago truncatula (Barrel medic)]	4	AATT	20	5	Plus	Gossypium_hirsutum
	DW235242	4	AATT	20	5	Plus	Gossypium_hirsutum

Supplementary Table 6. Contd.

24	BG445815 T5A14.7 protein [Arabidopsis thaliana (Mouse-ear cress)]	6	GATAGG	24	4	Plus	Gossypium_arboreum
	TA11166_29730 Zinc finger, RING-type; RINGv [Medicago truncatula (Barrel medic)]	6	GATAGG	30	5	Plus	Gossypium_raimondii
	TA37112_3635 Zinc finger, RING-type; RINGv [Medicago truncatula (Barrel medic)]	6	GATAGG	24	4	Plus	Gossypium_hirsutum
25	BG445177	6	GATAGG	24	4	Plus	Gossypium_arboreum
	TA11166_29730 Zinc finger, RING-type; RINGv [Medicago truncatula (Barrel medic)]	6	GATAGG	30	5	Plus	Gossypium_raimondii
	TA37112_3635 Zinc finger, RING-type; RINGv [Medicago truncatula (Barrel medic)]	6	GATAGG	24	4	Plus	Gossypium_hirsutum

Supplementary Table 7.

	GO Term	All SSR-ESTs	Dinucleotides	Trinucleotides	AC motif	AG motif	AT motif	AAC motif	AAG motif	AAT motif	ACC motif
Ga	**Gh**	**Gr**	**Ga**	**Gh**	**Gr**	**Ga**	**Gh**	**Gr**	**Ga**	**Gh**	**Gr**
585	**1680**	**937**	**175**	**267**	**289**	**283**	**993**	**476**	**10**	**37**	**20**
Biological process	Biological regulation	27(4.62)	113(6.73)	98(10.46)	9(5.14)	25(9.36)	31(10.73)	11(3.89)	68(6.85)	43(9.03)	2(20.00)
	Cellular component organization	49(8.38)	60(3.57)	38(4.06)	9(5.14)	13(4.87)	7(2.42)	35(12.37)	30(3.02)	23(4.83)	1(10.00)
	Cellular process	219(37.44)	591(35.18)	429(45.78)	72(41.14)	107(40.07)	141(48.79)	109(38.52)	346(34.84)	212(44.54)	8(80.00)
	Death	6(1.03)	12(0.71)	8(0.85)	1(0.57)	0(0.00)	2(0.69)	3(1.06)	9(0.91)	2(0.42)	0(0.00)
	Developmental process	45(7.69)	119(7.08)	80(8.54)	22(12.57)	22(8.24)	25(8.65)	17(6.01)	67(6.75)	41(8.61)	0(0.00)
	Growth	9(1.54)	24(1.43)	16(1.71)	6(3.43)	8(3.00)	2(0.69)	1(0.35)	15(1.51)	9(1.89)	0(0.00)
	Localization	36(6.15)	112(6.67)	73(7.79)	12(6.86)	21(7.87)	26(9.00)	14(4.95)	57(5.74)	31(6.51)	0(0.00)
	Metabolic process	189(32.31)	545(32.44)	394(42.05)	65(37.14)	107(40.07)	138(47.75)	95(33.57)	311(31.32)	190(39.92)	6(60.00)
	Multicellular organismal process	39(6.67)	115(6.85)	81(8.64)	18(10.29)	21(7.87)	25(8.65)	16(5.65)	63(6.34)	42(8.82)	0(0.00)
	Multi-organism process	1(0.17)	11(0.65)	5(0.53)	0(0.00)	3(1.12)	1(0.35)	0(0.00)	6(0.60)	3(0.63)	0(0.00)
	Reproduction	27(4.62)	69(4.11)	43(4.59)	15(8.57)	13(4.87)	15(5.19)	9(3.18)	35(3.52)	18(3.78)	0(0.00)
	Response to stimulus	77(13.16)	241(14.35)	161(17.18)	31(17.71)	35(13.11)	55(19.03)	38(13.43)	153(15.41)	69(14.50)	2(20.00)
	Signaling	23(3.93)	75(4.46)	84(8.96)	7(4.00)	18(6.74)	27(9.34)	10(3.53)	40(4.03)	35(7.35)	0(0.00)
Cellular component	Cell	298(50.94)	747(44.46)	517(55.18)	88(50.29)	124(46.44)	145(50.17)	140(49.47)	429(43.20)	279(58.61)	7(70.00)
	Extracellular region	43(7.35)	26(1.55)	8(0.85)	10(5.71)	3(1.12)	4(1.38)	11(3.89)	8(0.81)	2(0.42)	0(0.00)
	Macromolecular complex	17(2.91)	26(1.55)	16(1.71)	5(2.86)	4(1.50)	5(1.73)	0(0.00)	11(1.11)	8(1.68)	2(20.00)
	Membrane-enclosed lumen	42(7.18)	36(2.14)	64(6.83)	12(6.86)	4(1.50)	19(6.57)	23(8.13)	23(2.32)	39(8.19)	1(10.00)
	Organelle	193(32.99)	509(30.30)	353(37.67)	53(30.29)	86(32.21)	98(33.91)	116(40.99)	298(30.01)	197(41.39)	4(40.00)
	Binding	218(37.26)	655(38.99)	438(46.74)	67(38.29)	109(40.82)	132(45.67)	110(38.87)	397(39.98)	227(47.69)	5(50.00)

Supplementary Table 7.

	Catalytic activity	201(34.36)	409(24.35)	304(32.44)	54(30.86)	83(31.09)	119(41.18)	89(31.45)	218(21.95)	125(26.26)	4(40.00)
	Enzyme regulator activity	3(0.51)	12(0.71)	10(1.07)	3(1.71)	3(1.12)	1(0.35)	4(1.41)	9(0.91)	7(1.47)	0(0.00)
	Molecular transducer activity	10(1.71)	42(2.50)	29(3.09)	2(1.14)	5(1.87)	9(3.11)	0(0.00)	25(2.52)	11(2.31)	0(0.00)
	Structural molecule activity	35(5.98)	33(1.96)	14(1.49)	2(1.14)	8(3.00)	4(1.38)	30(10.60)	12(1.21)	8(1.68)	1(10.00)
	Transcription regulator activity	35(5.98)	140(8.33)	72(7.68)	11(6.29)	27(10.11)	20(6.92)	19(6.71)	90(9.06)	42(8.82)	1(10.00)
	Transporter activity	14(2.39)	65(3.87)	42(4.48)	2(1.14)	13(4.87)	17(5.88)	9(3.18)	36(3.63)	19(3.99)	0(0.00)
	GO Term	**ACG motif**	**ACT motif**	**AGC motif**	**AGG motif**	**ATC motif**	**CCG motif**				
Ga	**Gh**	**Ga**	**Gh**	**Gr**	**Ga**	**Gh**	**Gr**	**Ga**	**Gh**	**Gr**	**Ga**
585	**1680**	**77**	**184**	**81**	**91**	**51**	**196**	**11**	**73**	**34**	**87**
Biological process	Biological regulation	2(5.41)	1(5.00)	3(3.90)	18(9.78)	7(8.64)	4(4.40)	6(11.76)	24(12.24)	0(0.00)	3(4.11)
	Cellular component organization	0(0.00)	0(0.00)	5(6.49)	11(5.98)	4(4.94)	3(3.30)	2(3.92)	3(1.53)	3(27.27)	1(1.37)
	Cellular process	17(45.95)	9(45.00)	32(41.56)	74(40.22)	46(56.79)	34(37.36)	19(37.25)	92(46.94)	7(63.64)	26(35.62)
	Death	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	1(1.10)	0(0.00)	2(1.02)	0(0.00)	1(1.37)
	Developmental process	2(5.41)	0(0.00)	11(14.29)	18(9.78)	11(13.58)	11(12.09)	2(3.92)	15(7.65)	1(9.09)	4(5.48)
	Growth	0(0.00)	0(0.00)	5(6.49)	7(3.80)	2(2.47)	1(1.10)	1(1.96)	0(0.00)	0(0.00)	1(1.37)
	Localization	2(5.41)	0(0.00)	7(9.09)	12(6.52)	3(3.70)	5(5.49)	7(13.73)	24(12.24)	2(18.18)	4(5.48)
	Metabolic process	15(40.54)	8(40.00)	27(35.06)	78(42.39)	51(62.96)	34(37.36)	17(33.33)	85(43.37)	5(45.45)	22(30.14)
	Multicellular organismal process	2(5.41)	0(0.00)	8(10.39)	18(9.78)	11(13.58)	10(10.99)	1(1.96)	15(7.65)	1(9.09)	4(5.48)
	Multi-organism process	0(0.00)	0(0.00)	0(0.00)	3(1.63)	1(1.23)	0(0.00)	0(0.00)	1(0.51)	0(0.00)	0(0.00)
	Reproduction	1(2.70)	0(0.00)	7(9.09)	11(5.98)	6(7.41)	8(8.79)	1(1.96)	11(5.61)	0(0.00)	5(6.85)
	Response to stimulus	4(10.81)	2(10.00)	13(16.88)	24(13.04)	13(16.05)	16(17.58)	8(15.69)	42(21.43)	1(9.09)	16(21.92)
	Signaling	1(2.70)	1(5.00)	3(3.90)	16(8.70)	6(7.41)	4(4.40)	2(3.92)	21(10.71)	0(0.00)	3(4.11)
cellular component	Cell	14(37.84)	10(50.00)	40(51.95)	86(46.74)	44(54.32)	45(49.45)	27(52.94)	96(48.98)	8(72.73)	28(38.36)
	Extracellular region	0(0.00)	0(0.00)	1(1.30)	1(0.54)	1(1.23)	9(9.89)	2(3.92)	4(2.04)	0(0.00)	1(1.37)
	Macromolecular complex	1(2.70)	1(5.00)	3(3.90)	3(1.63)	3(3.70)	1(1.10)	0(0.00)	1(0.51)	1(9.09)	0(0.00)
	Membrane-enclosed lumen	0(0.00)	2(10.00)	2(2.60)	3(1.63)	5(6.17)	9(9.89)	1(1.96)	12(6.12)	4(36.36)	1(1.37)
	Organelle	10(27.03)	7(35.00)	27(35.06)	62(33.70)	36(44.44)	23(25.27)	16(31.37)	60(30.61)	7(63.64)	17(23.29)
Molecular function	Binding	13(35.14)	7(35.00)	26(33.77)	80(43.48)	45(55.56)	37(40.66)	18(35.29)	84(42.86)	8(72.73)	43(58.90)
	Catalytic activity	8(21.62)	6(30.00)	26(33.77)	57(30.98)	42(51.85)	26(28.57)	21(41.18)	76(38.78)	1(9.09)	14(19.18)
	Enzyme regulator activity	0(0.00)	0(0.00)	0(0.00)	2(1.09)	0(0.00)	3(3.30)	1(1.96)	1(0.51)	0(0.00)	0(0.00)
	Molecular transducer activity	1(2.70)	1(5.00)	0(0.00)	4(2.17)	3(3.70)	2(2.20)	1(1.96)	6(3.06)	0(0.00)	1(1.37)
	Structural molecule activity	1(2.70)	1(5.00)	1(1.30)	7(3.80)	0(0.00)	0(0.00)	0(0.00)	3(1.53)	0(0.00)	0(0.00)
	Transcription regulator activity	6(16.22)	2(10.00)	2(2.60)	20(10.87)	5(6.17)	8(8.79)	1(1.96)	13(6.63)	3(27.27)	8(10.96)
	Transporter activity	2(5.41)	0(0.00)	2(2.60)	7(3.80)	2(2.47)	0(0.00)	4(7.84)	15(7.65)	2(18.18)	1(1.37)
	GO Term										

Supplementary Table 7. Contd.

Ga	Gh	Gh	Gr	Ga	Gh	Gr	Ga	Gh	Gr	Ga	Gh
585	**1680**	**286**	**149**	**34**	**114**	**83**	**32**	**116**	**41**	**8**	**30**
Biological process	Biological regulation	2(5.88)	4(4.60)	17(5.94)	7(4.70)	1(2.94)	3(2.63)	3(3.61)	3(9.38)	13(11.21)	4(9.76)
	Cellular component organization	1(2.94)	23(26.44)	7(2.45)	3(2.01)	0(0.00)	4(3.51)	4(4.82)	4(12.50)	3(2.59)	0(0.00)
	Cellular process	15(44.12)	52(59.77)	116(40.56)	71(47.65)	11(32.35)	46(40.35)	31(37.35)	15(46.88)	42(36.21)	19(46.34)
	Death	0(0.00)	3(3.45)	6(2.10)	2(1.34)	0(0.00)	1(0.88)	0(0.00)	0(0.00)	0(0.00)	0(0.00)
	Developmental process	2(5.88)	3(3.45)	18(6.29)	7(4.70)	2(5.88)	7(6.14)	2(2.41)	3(9.38)	3(2.59)	3(7.32)
	Growth	0(0.00)	0(0.00)	3(1.05)	1(0.67)	0(0.00)	0(0.00)	1(1.20)	0(0.00)	1(0.86)	0(0.00)
	Localization	1(2.94)	5(5.75)	17(5.94)	11(7.38)	1(2.94)	10(8.77)	8(9.64)	1(3.13)	7(6.03)	3(7.32)
	Metabolic process	15(44.12)	28(32.18)	98(34.27)	70(46.98)	11(32.35)	43(37.72)	29(34.94)	21(65.63)	38(32.76)	14(34.15)
	Multicellular organismal process	2(5.88)	3(3.45)	17(5.94)	7(4.70)	2(5.88)	7(6.14)	3(3.61)	2(6.25)	3(2.59)	3(7.32)
	Multi-organism process	0(0.00)	0(0.00)	2(0.70)	0(0.00)	0(0.00)	1(0.88)	0(0.00)	0(0.00)	0(0.00)	0(0.00)
	Reproduction	2(5.88)	1(1.15)	9(3.15)	3(2.01)	2(5.88)	4(3.51)	2(2.41)	2(6.25)	0(0.00)	0(0.00)
	Response to stimulus	3(8.82)	12(13.79)	37(12.94)	19(12.75)	1(2.94)	10(8.77)	8(9.64)	8(25.00)	26(22.41)	8(19.51)
	Signaling	1(2.94)	3(3.45)	5(1.75)	6(4.03)	1(2.94)	2(1.75)	3(3.61)	3(9.38)	10(8.62)	4(9.76)
Cellular component	Cell	19(55.88)	53(60.92)	129(45.10)	83(55.70)	11(32.35)	52(45.61)	49(59.04)	22(68.75)	52(44.83)	26(63.41)
	Extracellular region	0(0.00)	0(0.00)	2(0.70)	2(1.34)	0(0.00)	3(2.63)	0(0.00)	0(0.00)	1(0.86)	0(0.00)
	Macromolecular complex	0(0.00)	2(2.30)	0(0.00)	3(2.01)	1(2.94)	1(0.88)	0(0.00)	4(12.50)	4(3.45)	1(2.44)
	Membrane-enclosed lumen	7(20.59)	4(4.60)	0(0.00)	9(6.04)	2(5.88)	2(1.75)	4(4.82)	4(12.50)	3(2.59)	1(2.44)
	Organelle	13(38.24)	46(52.87)	85(29.72)	58(38.93)	6(17.65)	35(30.70)	30(36.14)	17(53.13)	36(31.03)	18(43.90)
Molecular function	Binding	19(55.88)	45(51.72)	115(40.21)	76(51.01)	10(29.41)	40(35.09)	29(34.94)	20(62.50)	50(43.10)	16(39.02)
	Catalytic activity	4(11.76)	40(45.98)	72(25.17)	51(34.23)	11(32.35)	43(37.72)	31(37.35)	14(43.75)	12(10.34)	5(12.20)
	Enzyme regulator activity	0(0.00)	0(0.00)	0(0.00)	1(0.67)	0(0.00)	0(0.00)	2(2.41)	0(0.00)	0(0.00)	0(0.00)
	Molecular transducer activity	1(2.94)	1(1.15)	12(4.20)	3(2.01)	0(0.00)	1(0.88)	1(1.20)	2(6.25)	1(0.86)	0(0.00)
	Structural molecule activity	0(0.00)	22(25.29)	1(0.35)	1(0.67)	1(2.94)	1(0.88)	1(1.20)	4(12.50)	5(4.31)	3(7.32)
	Transcription regulator activity	7(20.59)	5(5.75)	25(8.74)	9(6.04)	2(5.88)	8(7.02)	5(6.02)	4(12.50)	11(9.48)	2(4.88)
	Transporter activity	0(0.00)	3(3.45)	7(2.45)	4(2.68)	1(2.94)	10(8.77)	7(8.43)	0(0.00)	2(1.72)	2(4.88)

	GO Term										
Ga	**Gh**	**Gr**	**Ga**	**Gh**	**Gr**	**Ga**	**Gh**	**Gr**	**Ga**	**Gh**	**Gr**
585	**1680**	**9**	**3**	**12**	**11**	**31**	**100**	**44**	**14**	**43**	**14**
biological process	Biological regulation	0(0.00)	2(6.67)	1(11.11)	0(0.00)	0(0.00)	0(0.00)	1(3.23)	7(7.00)	6(13.64)	1(7.14)
	Cellular component organization	0(0.00)	0(0.00)	0(0.00)	1(33.33)	3(25.00)	1(9.09)	2(6.45)	3(3.00)	2(4.55)	1(7.14)
	Cellular process	0(0.00)	3(10.00)	6(66.67)	1(33.33)	6(50.00)	6(54.55)	7(22.58)	31(31.00)	20(45.45)	4(28.57)
	Death	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)
	Developmental process	0(0.00)	2(6.67)	1(11.11)	0(0.00)	3(25.00)	2(18.18)	3(9.68)	14(14.00)	9(20.45)	1(7.14)
	Growth	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	1(3.23)	3(3.00)	3(6.82)	0(0.00)

Supplementary Table 7. Contd.

	Localization	2(25.00)	0(0.00)	0(0.00)	0(0.00)	1(8.33)	2(18.18)	1(3.23)	7(7.00)	2(4.55)	1(7.14)
	Metabolic process	0(0.00)	1(3.33)	5(55.56)	1(33.33)	7(58.33)	4(36.36)	6(19.35)	33(33.00)	20(45.45)	4(28.57)
	Multicellular organismal process	0(0.00)	2(6.67)	1(11.11)	0(0.00)	3(25.00)	2(18.18)	3(9.68)	14(14.00)	9(20.45)	1(7.14)
	Multi-organism process	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)
	Reproduction	0(0.00)	2(6.67)	1(11.11)	0(0.00)	0(0.00)	0(0.00)	1(3.23)	9(9.00)	5(11.36)	0(0.00)
	Response to stimulus	0(0.00)	2(6.67)	0(0.00)	0(0.00)	4(33.33)	2(18.18)	1(3.23)	15(15.00)	9(20.45)	2(14.29)
	Signaling	0(0.00)	2(6.67)	1(11.11)	0(0.00)	0(0.00)	0(0.00)	1(3.23)	4(4.00)	6(13.64)	1(7.14)
Cellular component	Cell	3(37.50)	6(20.00)	6(66.67)	1(33.33)	5(41.67)	5(45.45)	12(38.71)	40(40.00)	27(61.36)	11(78.57)
	Extracellular region	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	1(1.00)	0(0.00)	0(0.00)
	Macromolecular complex	0(0.00)	1(3.33)	1(11.11)	1(33.33)	0(0.00)	0(0.00)	1(3.23)	3(3.00)	2(4.55)	0(0.00)
	Membrane-enclosed lumen	0(0.00)	0(0.00)	2(22.22)	0(0.00)	0(0.00)	0(0.00)	5(16.13)	6(6.00)	7(15.91)	2(14.29)
	Organelle	3(37.50)	5(16.67)	6(66.67)	1(33.33)	5(41.67)	4(36.36)	11(35.48)	33(33.00)	23(52.27)	9(64.29)
Molecular function	Binding	0(0.00)	3(10.00)	4(44.44)	1(33.33)	4(33.33)	3(27.27)	7(22.58)	43(43.00)	27(61.36)	4(28.57)
	Catalytic activity	2(25.00)	7(23.33)	0(0.00)	0(0.00)	7(58.33)	5(45.45)	4(12.90)	10(10.00)	4(9.09)	1(7.14)
	Enzyme regulator activity	0(0.00)	2(6.67)	1(11.11)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)
	Molecular transducer activity	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	0(0.00)	1(3.23)	3(3.00)	2(4.55)	0(0.00)
	Structural molecule activity	0(0.00)	0(0.00)	1(11.11)	0(0.00)	0(0.00)	0(0.00)	1(3.23)	3(3.00)	0(0.00)	1(7.14)
	Transcription regulator activity	0(0.00)	0(0.00)	2(22.22)	0(0.00)	1(8.33)	1(9.09)	0(0.00)	14(14.00)	7(15.91)	3(21.43)
	Transporter activity	2(25.00)	6(20.00)	0(0.00)	0(0.00)	1(8.33)	2(18.18)	0(0.00)	1(1.00)	2(4.55)	1(7.14)

	GO Term								
Ga	**Gh**	**Ga**	**Gh**	**Gr**	**Ga**	**Gh**	**Gr**		
585	**1680**	**55**	**191**	**96**	**12**	**52**	**13**		
Biological process	Biological regulation	3(6.98)	1(7.14)	1(1.82)	13(6.81)	18(18.75)	0(0.00)	9(17.31)	2(15.38)
	Cellular component organization	0(0.00)	0(0.00)	1(1.82)	9(4.71)	12(12.50)	0(0.00)	0(0.00)	0(0.00)
	Cellular process	11(25.58)	7(50.00)	11(20.00)	54(28.27)	39(40.63)	3(25.00)	17(32.69)	4(30.77)
	Death	0(0.00)	0(0.00)	1(1.82)	1(0.52)	0(0.00)	0(0.00)	0(0.00)	0(0.00)
	Developmental process	1(2.33)	0(0.00)	5(9.09)	14(7.33)	15(15.63)	0(0.00)	1(1.92)	0(0.00)
	Growth	0(0.00)	0(0.00)	0(0.00)	5(2.62)	4(4.17)	0(0.00)	2(3.85)	0(0.00)
	Localization	2(4.65)	0(0.00)	0(0.00)	5(2.62)	2(2.08)	1(8.33)	7(13.46)	2(15.38)
	Metabolic process	12(27.91)	7(50.00)	18(32.73)	49(25.65)	30(31.25)	3(25.00)	12(23.08)	2(15.38)
	Multicellular organismal process	1(2.33)	0(0.00)	5(9.09)	11(5.76)	15(15.63)	0(0.00)	1(1.92)	0(0.00)
	Multi-organism process	0(0.00)	0(0.00)	0(0.00)	3(1.57)	3(3.13)	0(0.00)	0(0.00)	0(0.00)
	Reproduction	0(0.00)	0(0.00)	4(7.27)	6(3.14)	5(5.21)	0(0.00)	0(0.00)	0(0.00)
	Response to stimulus	3(6.98)	1(7.14)	15(27.27)	35(18.32)	19(19.79)	0(0.00)	6(11.54)	1(7.69)
	Signaling	3(6.98)	1(7.14)	1(1.82)	8(4.19)	14(14.58)	0(0.00)	3(5.77)	0(0.00)
	Cell	17(39.53)	10(71.43)	15(27.27)	73(38.22)	52(54.17)	6(50.00)	39(75.00)	9(69.23)

Supplementary Table 7. Contd.

	Extracellular region	0(0.00)	0(0.00)	0(0.00)	1(0.52)	0(0.00)	0(0.00)	0(0.00)	0(0.00)
	Macromolecular complex	0(0.00)	0(0.00)	0(0.00)	0(0.00)	1(1.04)	1(8.33)	2(3.85)	0(0.00)
	Membrane-enclosed lumen	0(0.00)	3(21.43)	2(3.64)	8(4.19)	7(7.29)	0(0.00)	3(5.77)	1(7.69)
	Organelle	13(30.23)	6(42.86)	14(25.45)	57(29.84)	41(42.71)	4(33.33)	17(32.69)	4(30.77)
Molecular function	Binding	19(44.19)	8(57.14)	12(21.82)	64(33.51)	47(48.96)	4(33.33)	24(46.15)	4(30.77)
	Catalytic activity	11(25.58)	6(42.86)	15(27.27)	32(16.75)	23(23.96)	2(16.67)	11(21.15)	1(7.69)
	Enzyme regulator activity	0(0.00)	2(14.29)	0(0.00)	6(3.14)	1(1.04)	0(0.00)	1(1.92)	0(0.00)
	Molecular transducer activity	0(0.00)	0(0.00)	0(0.00)	4(2.09)	4(4.17)	0(0.00)	3(5.77)	0(0.00)
	Structural molecule activity	0(0.00)	0(0.00)	0(0.00)	0(0.00)	2(2.08)	1(8.33)	2(3.85)	0(0.00)
	Transcription regulator activity	3(6.98)	3(21.43)	2(3.64)	20(10.47)	8(8.33)	0(0.00)	2(3.85)	0(0.00)
	Transporter activity	1(2.33)	0(0.00)	0(0.00)	5(2.62)	2(2.08)	0(0.00)	2(3.85)	0(0.00)