Pak. J. Bot., 46(2): 605-608, 2014.

SINGLE NUCLEOTIDE POLYMORPHISM ANALYSIS OF UBIQUITIN EXTENSION PROTEIN GENES (UBQ) OF *GOSSYPIUM ARBOREUM* AND *GOSSYPIUM HERBACEUM* IN COMPARISON WITH *ARABIDOPSIS THALIANA*

TAYYABA SHAHEEN[1, 2*], YUSUF ZAFAR[2] AND MEHBOOB-UR-RAHMAN[2]

[1]*Department of Bioinformatics and Biotechnology, Government College University Faisalabad, Pakistan*
[2]*Plant Genomics & Molecular Breeding Lab, National Institute for Biotechnology & Genetic Engineering (NIBGE), P.O. Box 577, Jhang Road Faisalabad, Pakistan*
[*]*Corresponding author e-mail: tayaba_pgmb@yahoo.com, mehboob_pbd@yahoo.com*

Abstract

Single nucleotide polymorphism analysis is an expedient way to study polymorphisms at genomic level. In the present study we have explored Ubiquitin extension protein gene of *G. arboreum* (A_2) and *G. herbaceum* (A_1) of cotton which is a multiple copy gene. We have found SNPs at 16 positions in 200 bp region within A genome of cotton indicating frequency of SNPs 1/13 bp. Both sequences from cotton have shown maximum similarity with UBQ5 and UBQ6 of *Arabidopsis thaliana*. Sequence obtained from *G. arboreum* has shown SNPs at 28 positions in comparison with each UBQ5 and UBQ6 of *Arabidopsis thaliana* while sequence obtained from *G. herbaceum* has shown SNPs at 31 positions in comparison with each UBQ5 and UBQ6 of *Arabidopsis thaliana*. In conclusion although during pace of evolution ubiquitin extension protein genes of both A genome species have got some mutations from nature but still most of their sequence is similar. Single nucleotide polymorphism study can prove a vital tool to identify gene type in case of Multicopy genes.

Introduction

SNPs are the most abundant molecular markers in plants and animals (Gupta *et al*., 2001; Bansal *et al*., 2010; Gomez-Uchida *et al*., 2011), and have been used in the genetic studies of a wide range of organisms (Qi *et al*., 2008). SNPs are a rich source of variations that can be used to saturate genetic maps (Ayeh, 2008). SNPs have also potential to be used for association mapping of interesting traits (Botstein & Risch, 2003; Kolkman *et al*., 2007; Aly *et al*., 2008). They are advantageous over other markers due to frequent occurrence, codominant nature and ease of automation (Ayeh, 2008).

In plants, SNPs have been detected through high-throughput analysis in *Arabidopsis* (Cho *et al*., 1999) and the same technique was used to find cis-acting elements in *Arabidopsis* (Narusaka *et al*., 1999 & 2003; Kidokoro *et al*., 2009). In *Arabidopsis* genome sequencing 25,274 SNPs and 14,041 Indels (small insertions/deletions) were found (Cho *et al*., 1999) which was very promising for detection of SNPs in plants. Although because of the expense of high-throughput technology for SNP discovery and detection, they have not been widely used in plant species.

Mining for SNPs in cotton is just in infancy (Shaheen *et al*., 2006; Ahmad *et al*., 2007). Genetic improvement of cotton will be enhanced by the availability of rapidly developing genetic resources and tools, including a high-density genetic map (Rahman *et al*., 2005; Rong *et al*., 2004; Lacape *et al*., 2005; Rahman *et al*., 2012). Cotton like other polyploid crops has the problem of huge genome size (Udall *et al*., 2006). As a remedy mining SNPs in diploid genomes first seems to be more feasible due to low level of complexity in diploid genomes (Wang *et al*., 2005). Coding regions which are conserved among many organisms are least prone to mutations and hence other marker systems can not be effective to explore these regions (Semagn *et al*., 2006).

Various markers are explored to help in markers assisted selection (Rabbani *et al*., 2010; Turi *et al*., 2012; Sarwar *et al.,* 2013); however, gene-mediated breeding instead of marker-assisted selection can prove a new system of breeding if SNP markers are combined with QTL data for phenotypic character (Lange & Whittacker, 2001; Masood *et al*., 2005).

In this study we have used SNPs for identification of gene type in multiple copy genes Ubiquitin extension protein gene. Ubiquitin is a highly conserved, 76-aa' protein that appears to be present in all eukaryotes. Several biological roles for Ubiquitin have been proposed; the best characterized is as a covalently bound recognition signal for proteolysis (Callis *et al*., 1990).

Biological function of the extension proteins has been partially elucidated with the observation that in yeast both extension proteins co-sediment with ribosomes (monosome and polysome fractions), indicating that they are constituents of mature active ribosomes (Callis *et al*., 1990). Ubiquitin extension protein genes are multiple copy genes in *Arabidopsis* (Callis *et al*., 1995).

The two Arabidopsis genes encoding ubiquitin 81-aa extension proteins (UBQ5 and UBQ6) have 90% nucleotide identity between them for both the ubiquitin and extension protein coding regions. The two 81-aa extension proteins are not identical they have 4-aa substitutions between them, of which 2 (at positions 32 and 80) are conservative substitutions. Unlike the 52-aa extension protein genes, the coding regions of UBQ5 and UBQ6 are not interrupted by introns. Arabidopsis extension proteins are highly conserved and we have used sequences of Arabidopsis extension proteins to identify the ubiquitin extension protein genes of cotton.

Materials and Methods

Plant material: Experimental material consisted of *G. herbaceum*. ESTs of *G. arboreum* (accession 8401) were obtained from cotton ESTdb.

Isolation of total genomic DNA: DNA was extracted from *G. herbaceum* plants according to the method used by Iqbal *et al*., (1997). After RNase treatment the DNA concentration was measured by flouremeter DyNA QuantTM 200. Running 25 ng DNA on 0.8% agarose gel checked the quality of DNA. The DNA samples, which were not showing a discrete band, were rejected. The total genomic DNA was diluted in double distilled water to a concentration of 15 ng/ul for PCR analysis.

Primer designining: Gene specific primers were designed based on ESTs showing homology with genes encoding for Ubiquitin extension protein gene of *G. arboreum* (accession 8401). Primers were designed using primer 3 software (Table 1). Polymerase chain reaction (PCR) was performed in a total volume of 20µl, using 2.5µl (15ng/µl) of cotton DNA, 10 x PCR buffer without $MgCl_2$ (10mM Tris-HCl, 50mM KCl, PH 8.3), 3mM $MgCl_2$, 0.1mM each of dATP, dGTP, dCTP and dTTP and 0.5 units of *Taq* DNA polymerase, 0.15 mM of each primer. *Taq* DNA polymerase together with 10 x PCR buffer, $MgCl_2$ and dNTPs were from MBI Fermentas. Polymerase chain reaction consisted of 35 cycles of 94°C for 1 min, 94°C for 30 sec, 50°C for 30 sec, 72°C extension for 1 min and final extension at 72°C for 10 min. PCR products were resolved on 2% agarose gel.

Table 1. EST used for primer designing, its homology and primer sequence.

EST	Best blast homology	Primer sequence
CON_002_01293	Ubiquitin extension protein	5´CGTCAAGATGCAGATCTTCG3´ 5´CTTCTTCCTCTTCTTGGCAC3´

Sequencing of PCR product: Sequencing of PCR product was done on ABI automated DNA sequencer. Sequence was edited manually. Sequences of Ubiquitin extension protein genes UBQ5 and UBQ6 of *Arabidopsis thaliana,* was obtained from GenBank. Maximum similarity of cotton gene sequences was searched with BLAST search tool in NCBI.

SNPs detection: DNASTAR (DNASTAR Inc., Madison, WI, USA) and Clustal v were used for sequence alignment (Fig. 1).

Results and Discussions

Distribution of SNPs: EST sequence of Ubiquitin extension protein gene of *G. arboreum* was used to design primers to amplify region from *G. herbaceum*. Ubiquitin extension protein gene sequence of *G. arboreum* and *G. herbaceum* was containing similarity with UBQ5 and UBQ6 genes of *Arabidopsis thaliana* (Fig. 1). On alignment of sequences obtained from *G. arboreum* and *G. herbaceum* along with UBQ5 and UBQ6 of Arabidopsis. SNPs between two cotton species were detected at 16 positions including 14 substitutions and 2 Indels (Table 2). Nucleotide variations of *G. arboreum* (accession 8401) with *Arabidopsis thaliana* UBQ5 were at 28 positions and with UBQ6 were also at 28 positions (Fig. 1). Nucleotide variations of *G. herbaceum* with *Arabidopsis thaliana* UBQ5 were at 31 positions and with UBQ6 were also at 31 positions. SNPs were detected at 18 positions between UBQ5 and UBQ6 of *Arabidopsis thaliana* (Table 3).

Development of new SNPs by re-sequencing of PCR amplicons with or without pre-screening has been reported in previous studies (Rafalski, 2002; Ayeh, 2008; Hina *et al.,* 2012; Safdar *et al.,* 2012). In this study we observed 16 SNPs including 14 substitutions and 2 indels in 200 bp amplified Ubiquitin extension protein gene region in *G. arboreum* (accession 8401) and *G. herbaceum* (Table 2). Maximum type of variations are transitions (T/C and A/G) it is comparable with the previous results with data published for human genome and sugar beet where transitions were more frequent than transversions (Schneider *et al*., 2001). The frequency of SNPs and Indels observed in this study is higher as expected from previous intraspecific studies in cotton (Shaheen *et al*., 2006; Ahmad *et al*., 2007). High frequency of occurrence of SNPs makes them potent markers to develop a dense genetic map of cotton. Molecular markers being used in cotton like RAPD, SSRs, AFLP, and SCARs have limitations of low frequency (Ayeh *et al*., 2008).

Ubiquitin genes are highly conserved in nature. In *Arabidopsis* two genes with 90% nucleotide identity in their exons encode ubiquitin and identical 52-amino acid (aa) extension proteins with 85 and 79% aa identity to 52-aa extension proteins from humans and yeast, respectively. Two other genes with 90% nucleotide identity encode ubiquitin and 81-aa extension proteins that differ by 4 amino acids from each other and are approximately 70% identical to the 76- and the 80-aa extension proteins from yeast and humans, respectively (Callis *et al*., 1995).

Cotton and *Arabidopsis* are close relatives (Rong *et al*., 2005) and in this study we have used *Arabidopsis* genes to identify gene type in cotton. Gene sequences obtained from cotton shown maximum similarity with UBQ5 and UBQ6 of *Arabidopsis thaliana*. On alignment of sequences obtained from *G. arboreum* and *G. herbaceum* along with UBQ5 and UBQ6 of *Arabidopsis*. SNPs between two cotton species were detected at 16 positions including 14 substitutions and 2 Indels (Table 2). Nucleotide variations of *G. arboreum* (accession 8401) with both protein sequences of *Arabidopsis thaliana* UBQ5 and UBQ6 were at 28 positions but at different sites (Fig. 1). Nucleotide variations of *G. herbaceum* with both protein sequences of *Arabidopsis thaliana* UBQ5 and UBQ6 were also at 31 positions but at different locations. SNPs were detected at 18 positions between UBQ5 and UBQ6 of *Arabidopsis thaliana* (Table 3). Presence of SNPs between both species and their different number of SNPs (31 and 28) for *Arabidopsis* protein genes sequences show effect of evolutionary processes which have caused mutations but not at very high ratio, as reported Stupar *et al*., (2006) who studied structural diversity of patatin Multicopy gene family.

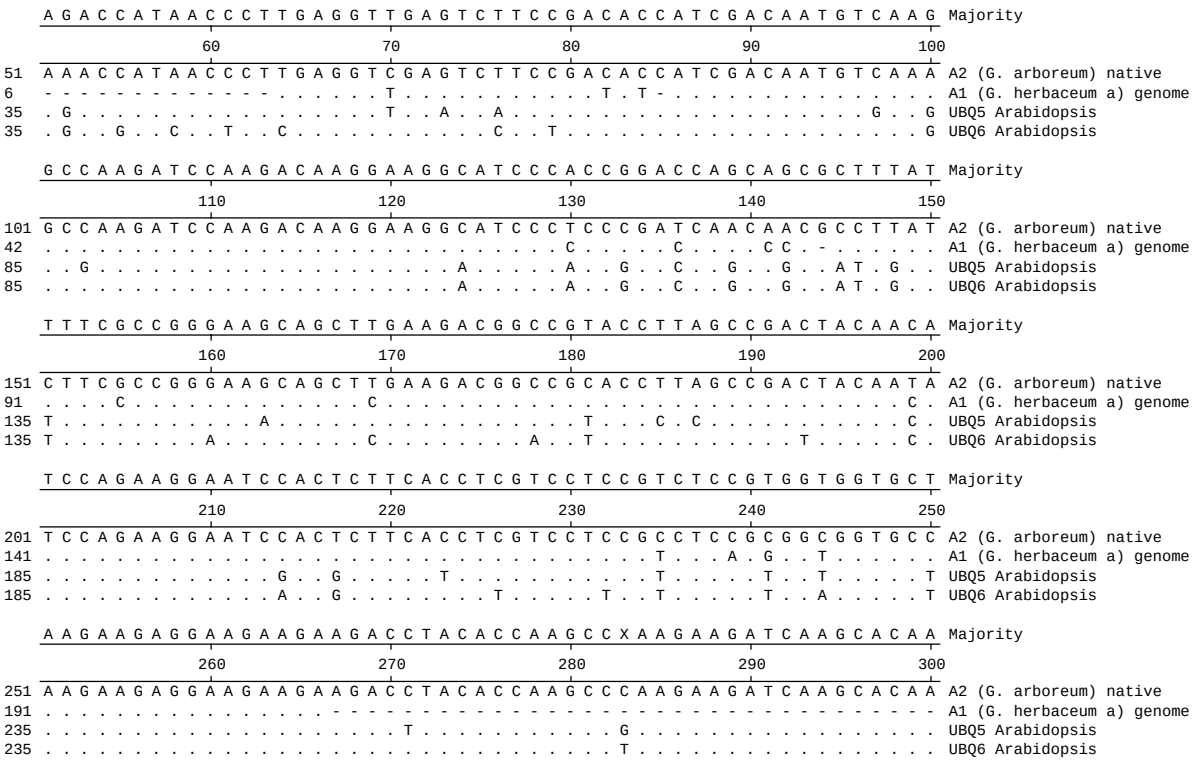

Fig. 1. Alignment of *G. arboreum*, *G. herbaceum* and UBQ5 and UBQ6 of *Arabidopsis thaliana* showing polymorphisms.

Table 2. SNPs identified in cotton (*G. arboreum* and *G. herbaceum*).

EST	Amplicons size	No. of SNPs/ Indels	Type of mutation
Ubiquitin extension protein gene	200	16	*T→C*
			C→T
			C→T
			-C
			T→C
			T→C
			A→C
			A→C
			-G
			G→C
			T→C
			T→C
			C→T
			C→A
			C→G
			C→T

Table 3. Number of SNPs identified in cotton sequences, UBQ5 and UBQ6 sequences of *Arabidopsis* and in cotton and *Arabidopsis* sequences.

Comparison of sequences	No. of SNPs
G. arboreum and *G.herbaceum*	16
G. arboreum and UBQ5 of *Arabidopsis*	28
G. herbaceum and UBQ5 of *Arabidopsis*	31
G. arboreum and UBQ6 of *Arabidopsis*	28
G. herbaceum and UBQ6 of *Arabidopsis*	31
UBQ5 of *Arabidopsis* and UBQ6 of *Arabidopsis*	18

Both varieties separated from each other 1-4 MYA (Wendel, 1989). More numbers of SNPs were detected in *G. herbaceum* when aligned with A. thaliana as compared to *G. arboreum* which indicate that *G. herbaceum* was separated from A. thaliana before than *G. arboreum*.

In conclusion SNPs are an effective tool for whole genome survey and are potent markers to survey conserved regions where other markers may not prove very effective. They can prove very effective in study of multiple copy genes.

Acknowledgements

We are thankful to Higher Education Commission Pakistan for providing funds for the present studies through the HEC Project "Finding Single Nucleotide Polymorphisms in cotton genome." under Presidential Young Innovative Programme and indigenous PhD fellowship scheme.

References

Ahmad, S., M. Ashraf, T. Zhang, N. Islam, T. Shaheen and M. Rahman. 2007. Identifying genetic variation in Gossypium L. based on single nucleotide polymorphism. *Pak. J. Bot.,* 39: 1245-1250.

Aly, T.A., E.E. Baschal, M.M. Jahromi, M.S. Fernando, S.R. Babu, T.E. Fingerlin, A. Kretowski, H.A. Erlich, P.R. Fain, M.J. Rewers, and G.S. Eisenbarth. 2008. Analysis of single nucleotide polymorphisms identifies major type 1A diabetes locus telomeric of the major histocompatibility complex. *Diabetes,* 57: 770-776.

Ayeh, K.O. 2008. Expressed sequence tags (ESTs) and single nucleotide polymorphisms (SNPs): Emerging molecular

marker tools for improving agronomic traits in plant biotechnology. *Afr. J. Biotech.*, 7(4): 331-341.

Bansal, V., O. Harismendy, R. Tewhey, S.S. Murray, N.J. Schork, E.J. Topol, K.A. Frazer. 2010. Accurate detection and genotyping of SNPs utilizing population sequencing data Published in Advance. *Genome Res*., 20: 537-545.

Botstein, D. and N. Risch. 2003. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. *Nat. Genet*., 33: 228-237.

Callis, J., J.A. Raasch and R.D. Vierstras. 1990. Ubiquitin extension proteins of *Arabidopsis thaliana*. *J. Bio. Chem*., 21(265): 12466-12493.

Callis, J., T. Carpenter, C. Sun and R.D Vierstra. 1995. Structure and evolution of genes encoding polyubiquitin and ubiquitin-like proteins in *Arabidopsis thaliana* ecotype Columbia. *Genetics*, 159: 921-939.

Cho, R.J., M. Mindrinos, D.R. Richards, R.J. Sapolsky, M. Anderson, E. Drenkard, J. Dewdney, T.L. Reuber, M. Stammers, N. Federspiel, A. Theologis, W.H. Yang, E. Hubbell, M. Au, E.Y. Chung, D. Lashkari, B. Lemieux, C. Dean, R.J. Lipshutz, F.M. Ausubel, R.W. Davis and P.J. Oefner. 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat. Genet*., 23: 203-207.

Gomez-Uchida, D., J.E. Seeb, M.J Smith, C. Habicht, T.P. Quinn and L.W. Seeb. 2011. Single nucleotide polymorphisms unravel hierarchical divergence and signatures of selection among Alaskan sockeye salmon (*Oncorhynchus nerka*) populations. *BMC Evol. Bio*., 11: 48.

Gupta, P.K., J.K. Roy and M. Prasad. 2001. Single nucleotide polymorphisms: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Curr. Sci.*, 80: 524-535.

Hina, S. M. A. Javed, S. Haider and M. Saleem. 2012. Isolation and sequence analysis of cotton infecting begomovir uses. *Pak. J. Bot.,* 44: 223-230.

Iqbal, M.J., N. Aziz, N.A. Saeed, Y. Zafar and K.A. Mailk. 1997. Genetic diversity of some elite cotton varieties by RAPD analysis. *Theor. Appl. Genet*., 94: 139-144.

Kidokoro S., K. Nakashima, Z.K. Shinwari, K. Shinozaki and K. Yamaguchi-Shinozaki. 2009. The Phytochrome-Interacting Factor PIF7 Negatively Regulates *DREB1* Expression under Circadian Control in *Arabidopsis. Plant Physiol.,* 151(4): 2046-2057.

Kolkman, J.M., S.T. Berry, A.J. Leon, M.B. Slabaugh, S. Tang, W. Gao, D.K. Shintani, J.M. Burke and S.J. Knapp. 2007. Single nucleotide polymorphisms and linkage disequilibrium in sunflower. *Genetics*, 177: 457-468.

Lacape, J.M., T.B. Nguyen, B. Courtois, J.L. Belot, M. Giband, J.P. Gourlot, G. Gawryziak, S. Roques and B. Hau. 2005. QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum, Gossypium barbadense* backcross generations. *Crop Sci*., 45: 123-140.

Lange, C. and J.C. Wittacher. 2001. On prediction of genetic values in marker-assisted selection. *Genetics*, 159: 1375-1381.

Masood, S., Y. Seiji, Z. K. Shinwari and R. Anwar. 2005. Mapping quantitative trait loci (QTLs) for salt tolerance in rice (*Oryza sativa*) using RFLPs. *Pak. J. Bot*., 36(4): 825-834.

Narusaka, Y., K. Nakashima, Z.K. Shinwari, Y. Sakuma T. Furihata, H. Abe, M. Narusaka K. Shinozaki and K.Y. Shinozaki. 2003. Interaction between two cis-acting elements, ABRE and DRE, in ABA-dependent expression of *Arabidopsis* rd29A gene in response to dehydration and high salinity stresses. *The Plant J.,* 34(2): 137-149.

Narusaka, Y., Z.K. Shinwari K. Nakashima, K. Yamaguchi-Shinozaki and K. Shinozaki. 1999. The roles of the two cis-acting elements, DRE and ABRE in the dehydration, high salt and low temperature responsive expression of the Rd29a gene in *Arabidopsis thaliana*. *Plant and Cell Physiol*., 40: 91.

Qi, H., X.L. Lu and G. Zhang. 2008. Characterization of 12 single nucleotide polymorphisms (SNPs) in Pacific abalone, *Haliotis discus hannai. Molecular Ecology Resources*, 8: 974-976.

Rabbani, M.A., M.S. Masood, Z.K. Shinwari and K. Y. Shinozaki. 2010. Genetic analysis of basmati and non-basmati Pakistani rice (*Oryza sativa* L.) cultivars using microsatellite markers. *Pak. J. Bot*., 42(4): 2551-2564.

Rafalski, A. 2002. Applications of single nucleotide polymorphism in crop genetics. *Curr. Opin. Plant Biol.,* 5: 94-100.

Rahman, M., M. Asif, I. Ullah, K.A. Malik and Y. Zafar. 2005. Overview of cotton genomic studies in Pakistan. *Plant & Animal Genome Conference XIII*. San Diego, CA. USA.

Rahman, M., T. Shaheen, N. Tabbasam, M.A. Iqbal, Y. Zafar and A.H. Paterson. 2012. Cotton genetic resources. A review. *Agron. Sustain. Dev*., 32: 419-432.

Rong, J., J.E. Bowers, S.R. Schulze, V.N. Waghmare, C.J. Rogers, G.J. Pierce, H. Zhang, J.C. Estill and A.H. Paterson. 2005. Comparative genomics of *Gossypium* and *Arabidopsis*: Unraveling the consequences of both ancient and recent polyploidy. *Genome Res.,* 15: 1198-1210.

Rong, J.K., C. Abbey, J.E. Bowers, C.L. Brubaker, C. Chang, P.W. Chee, T.A. Delmonte, X. Ding, J.J. Garza, B.S. Marler, C. Park, G.J. Pierce, K.M. Rainey, V. Rastogi, K. Schulze, N.L. Tronlinde, J.F. Wendel, T.A. Wilkins, R.A. Wing, R.J. Wright, X. Zhao, L. Zhu and A.H. Paterson. 2004. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). *Genetics*, 166: 389-417.

Safdar, W.H. Majeed, B. Ali and I. Naveed. 2012. Molecular evolution and diversity of small heat shock proteins genes in plants. *Pak. J. Bot.,* 44: 211-218.

Sarwar, M.K.S., M.Y. Ashraf, M. Rahman and Y. Zafar. 2012. Genetic variability in different biochemical traits and their relationship with yield and yield parameters of cotton cultivars grown under water stress conditions. *Pak. J. Bot.,* 44: 515-520.

Schneider, K., B. Weisshaar, D.C. Borchardt and F. Salamini. 2001. SNP frequency and allelic haplotype structure of *Beta vulgaris* expressed genes. *Mol. Breed*., 8: 63-74.

Semagn, K., A. Bjornstad and M.N. Ndjiondjop. 2006. An overview of molecular marker methods for plants. *Afr. J. Biotech*., 5(25): 2540-2568.

Shaheen, T., M. Rahman and Y. Zafar. 2006. Chloroplast RPS_8 gene of cotton reveals the conserved nature through out plant taxa. *Pak. J. Bot*., 38:1467-1476.

Stupar, R.M., K.A. Beaubien, W. Jin, J. Song, M-K. Lee, C. Wu, H-B. Zhang, B. Han and J. Jiang. 2006. Structural diversity and differential transcription of the patatin multicopy gene family during potato tuber development. *Genetics*, 172: 1263-1275.

Turi, N., A. Farhatullah, M.A. Rabbani and Z.K. Shinwari. 2012. Genetic diversity in the locally collected *Brassica* species of Pakistan based on microsatellite markers. *Pak. J. Bot*., 44(3): 1029-1035.

Udall, J.A., J.M. Swanson, D. Nettleton, R.J. Percifield and J.F. Wendel. 2006. A novel approach for characterizing expression levels of genes duplicated by polyploidy. *Genetics*, 173(3): 1823-1827.

Wang, R-S, L-Y. Wu, Z-P. Li and X-S Zhang. 2005. Haplotype reconstruction from SNP fragments by minimum error correction. *Bioinformatics*, 21(10): 2456-2462.

Wendel, J.F. 1989. New world tetraploid cotton contains old world cytoplasm. *Proc. Natl. Acad. Sci.* USA., 86: 4132-4136.

(Received for publication 7 September 2012)