Turk J Biol
29 (2005) 7-13

Efficient Cloning of the Rare Wheat (*Triticum aestivum* L.) Transcripts

Muharrem DİLBİRLİĞİ
Central Research Institute for Field Crops, Ankara - TURKEY
Mustafa ERAYMAN
Mustafa Kemal University, Agricultural College Department of Crop Science, Hatay - TURKEY

Received: 15.09.2004

Abstract: Expressed sequence tag (EST) technology has spurred targeting the functional portion of genomes for several organisms. Constructed redundant complementary DNA (cDNA) libraries have targeted a majority of constitutive and over-expressed genes. However, the majority of lowly expressed and un-induced genes have not been able to be easily tagged and cloned. The objective of this study was to clone low and rarely expressed transcripts in wheat. Using a combination of a polyacrylamide gel and a motif-based PCR amplification, we cloned 144 bands containing 189 wheat fragments. Sequence analysis revealed 74 contigs, of which only 22 being present in database. The remaining 52 were unique, and were not present in any public EST or non-redundant database. Compared to 5% efficiency in the public EST database, the efficiency of cloning of unique sequences was about 28%. Therefore, the approach we describe here is highly promising for targeting the rarely and lowly expressed genes in wheat and, probably, various other genomes.

Key Words: EST, cDNA, Wheat, Functional Genomics, Cloning

Nadir Bugday (*Triticum aestivum* L.) Transkript'lerinin Etkin Klonlanması

Özet: EST (expressed sequence tag) teknolojisi çeşitli organizmalarda genomun işlevsel kısımlarının klonlanmasını hedeflemede büyük mesafe kat etmeyi sağlamıştır. Bol miktar traskriptlerden (redundant) oluşturulan cDNA kütüphaneleri genelde aralıksız (constitutive) ve yoğun (over-expressed) eksprese olan genleri içermektedir. Ancak, organ spesifik ve çok az eksprese olan genler kolaylıkla işaretlenip klonlanamamaktadır. Bu çalışma, ekmeklik buğdayda düşük seviyede ve nadir sıklıkla eksprese olan gen parçacıklarının klonlanmasını amaçlamaktadır. Poliakrilamid jel ve motife (farklı baz dizilimli genlerin aynı veya benzer sekansları) dayalı PCR (Polymerase Chain Reaction) amplifikasyonunun birlikte kullanımıyla, 189 buğday gen parçacığı taşıyan 144 bant klonlanmıştır. Gen baz dizilimlerinin analizi yalnızca 22'si gen bankasında kayıtlı olan 74 farklı kontig (contig) ile sonuçlanmıştır. Diğer 52 ise özgün (unique) ve herhangi bir EST veya 'non-redundant' gen bankasında rastlanmamıştır. EST gen bankasının % 5'lik özgün gen klonlama oranına karşın çalışmamızda bu oran yaklaşık % 28 olmuştur. Tanımladığımız bu yaklaşım, buğday da ve belki de farklı genomlarda nadir bulunan ve çok düşük seviyede eksprese edilen genlerin hedeflenmesi ve klonlanması konusunda oldukça ümit vericidir.

Anahtar Sözcükler: EST, cDNA, Buğday, Fonksiyonel Genomik, Klonlama

Introduction

The constructed redundant complementary DNA (cDNA) approach is a very important come out for targeting the expressed portion of all organisms irrespective of genome size (1-4). cDNA technology was first standardized with the human genome (5) and over 18 millions of cDNA and expressed sequence tags (ESTs) are currently available for hundreds of organisms in public databases (www.ncbi.nlm.nih.gov). The EST and cDNA sequences for more than 50 plant species are also available. Among plants, wheat has the largest amount of EST collection in the public database. Estimates of the number of wheat genes range from 75,000 to 150,000 (6). A total of 500,898 ESTs from 38 different libraries corresponding to about 23,000 uni-genes are present in the EST database for wheat (www.ncbi.nlm.nih.gov). A vast majority of wheat genes, therefore, are not represented in the current EST collection in the database.

The genome sequencing programs from diverse organisms have revealed gene families (7). The majority

of genes in a genome belong to various gene families that show a common structure and functional unity. About half of the genes even in diverged organisms such as fly, worm, and yeast are common (8). For example, a protein kinase family constitute about 2% of genomes ranging from 300 genes in *Drosophila* to 1100 genes in humans (8). Genome sequencing of monocot and dicot model plants, *Arabidopsis thaliana* and rice (*Oryza sativa*), revealed about 25,000 and 35,000-50,000 genes, respectively (9,10). Although the gene numbers vary, these genomes are categorized into about 11,000 gene families comprised of 2 to about 2,500 related gene sequences (9,10). The expression levels of genes belonging even to the same family may, however, greatly vary (10). ESTs for only about 65% of the annotated genes of both *Arabidopsis* and rice genomes have been able to be tagged (9,10). The remaining 35%, however, could not be tagged due to low expression and uninduction.

The proteins in a gene family contain highly conserved small sequence units called "motifs". The size of the motifs ranges from a hexamer to a very large unit of 60 amino acids (11,12). P-loop (kinase-1a) is one of the most prevalent motifs shared by thousands of proteins (13). It is characterized by -a four- residue fragment (Gxxx) where glycine residue is always fixed, while the other three residues are variable among the subfamily members (13,14). With the objective to clone the rarely transcribed genes, it is possible to utilize primers for these conserved motifs to amplify the related sequences (15,16). Irrespective of expression level, the objective of this study was to clone the p-loop-containing genes of bread wheat.

Material and Methods

Plant material

The cultivar Thatcher of Hexaploid wheat (*Triticum aestivum* L.) leaves were used to extract the total RNA and mRNA. Three to four week old seedlings were collected and kept in liquid nitrogen (N2) until they were ground for total RNA extraction.

Poly (A)+ RNA Isolation and Amplification Conditions

Total RNA and the poly A^+ RNA were isolated from a few seedling leaves using the standard protocol (17). The first-strand cDNA synthesis PCR amplification reactions were carried out following the recommended protocols except that ^{35}S isotope was used instead of ^{33}P (17). A degenerate primer for the p-loop (GVGKTT; 5'-AAGAATTCGGNGTNGGNAAAACAAC-3') motif was used. Nine different 'T' primers for the 3' end were also used (CLONTECH Lab Inc.). The "T" primers had a common 19-bp sequence at the 5' end followed by 9 thymidine and 2 variable bases at the 3' end in all possible pairwise combinations of A, C, and G for the amplification of poly A^+ tails. The "T" primers are given in Table 1.

Table 1. 'T' primers used for amplification of PCR product.

Primer	Sequence (5' to 3')
T1	CATTATGCTGAGTGATATCTTTTTTTTTAA
T2	CATTATGCTGAGTGATATCTTTTTTTTTAC
T3	CATTATGCTGAGTGATATCTTTTTTTTTAG
T4	CATTATGCTGAGTGATATCTTTTTTTTTCA
T5	CATTATGCTGAGTGATATCTTTTTTTTTCC
T6	CATTATGCTGAGTGATATCTTTTTTTTTCG
T7	CATTATGCTGAGTGATATCTTTTTTTTTGA
T8	CATTATGCTGAGTGATATCTTTTTTTTTGC
T9	CATTATGCTGAGTGATATCTTTTTTTTTGG

The PCR reactions were performed in a total volume of 25 µl containing 5 µM of dNTPs, 0.4 µl of Advantage® cDNA polymerase mix (Promega), 20 µM of each primer, 50 ng of the first strand cDNA, 2 µl of 10X cDNA PCR reaction buffer and 0.2 µl of ^{35}S-dATP. The PCR conditions were: 1 cycle of 5 min at 94 °C, 5 min at 45 °C, 5 min at 68 °C; 2 cycles of 2 min at 94 °C, 5 min at 45 °C, and 5 min at 68 °C; 25 cycles of 1 min at 94 °C, 1 min at 60 °C, and 2 min at 68 °C, followed by 7 min at 68 °C.

The amplification products were size separated on a 0.4 mm denaturing 5% polyacrylamide/8M urea gel, following a standard sequencing gel protocol (17), and 1% agarose gel. After cutting out the bands, they were re-amplified with the corresponding primers and cloned in the vectors (pGEM-T easy) (Promega).

Sequence Analysis

The DNA sequencing was done commercially (http://geneseek.com). BLASTX and BLASTN analyses were performed in order to assign putative functions and the virtual northern for the sequences (www.ncbi.nlm.nih.gov). The "Old Distance" (GCG-

Genetic Computer Group, Wisconsin Package version 10.1, Madison, WI, USA) sequence comparison and analysis were performed to calculate the homology level among the sequences. Sequences and detailed bioinformatics analysis of the cloned sequences have been reported elsewhere (18).

Results

Cloning of Fragments

Using a p-loop primer and different "T" primer combinations, 9 different PCR reactions from the cDNA were set. A polyacrylamide gel was used to resolve the PCR product. To demonstrate the power of polyacrylamide gel, the same PCR products were also run on agarose gel (Figure 1a). Separation of amplified products on the agarose gel revealed a smear pattern ranging from 150 bp to 1630 bp along with a few visible bands, whereas, the same products on a 5% polyacrylamide-urea gel generated about 90 bands per p-loop/'T' primer combination (Figures 1a, b).

The size of the fragment bands ranged from 100 bp to 1300 bp (Figure 1). Fragments smaller than 100 bp-sized fragments were not noticeable on the polyacrylamide gel. One hundred forty-four bands that seemed unique on 9 different lanes were excised from the gel. In some cases, a few closest bands were counted as one and were excised together. The DNA from each gel piece was eluted, re-amplified, and cloned (Figures 1c, 2). At least 3 colonies were analyzed to single out each cut-band. Thirteen cut-bands contained more than 1 fragments. For example, the bands 18 (B18) and B21 revealed 2 different fragments when digested with *Eco*RI enzyme (Figure 2). Depending on the carried fragments in each cut-band, 1 to 3 clones per bands were sequenced and analyzed further.

Sequence Analysis

A total of 189 wheat clones corresponding to 144 cut-band fragments were sequenced. A "ContigExpress" analysis among positive sequences resulted in 74 contigs. The longest sequences of each contig were used for analysis. Detailed sequence analysis and putative function

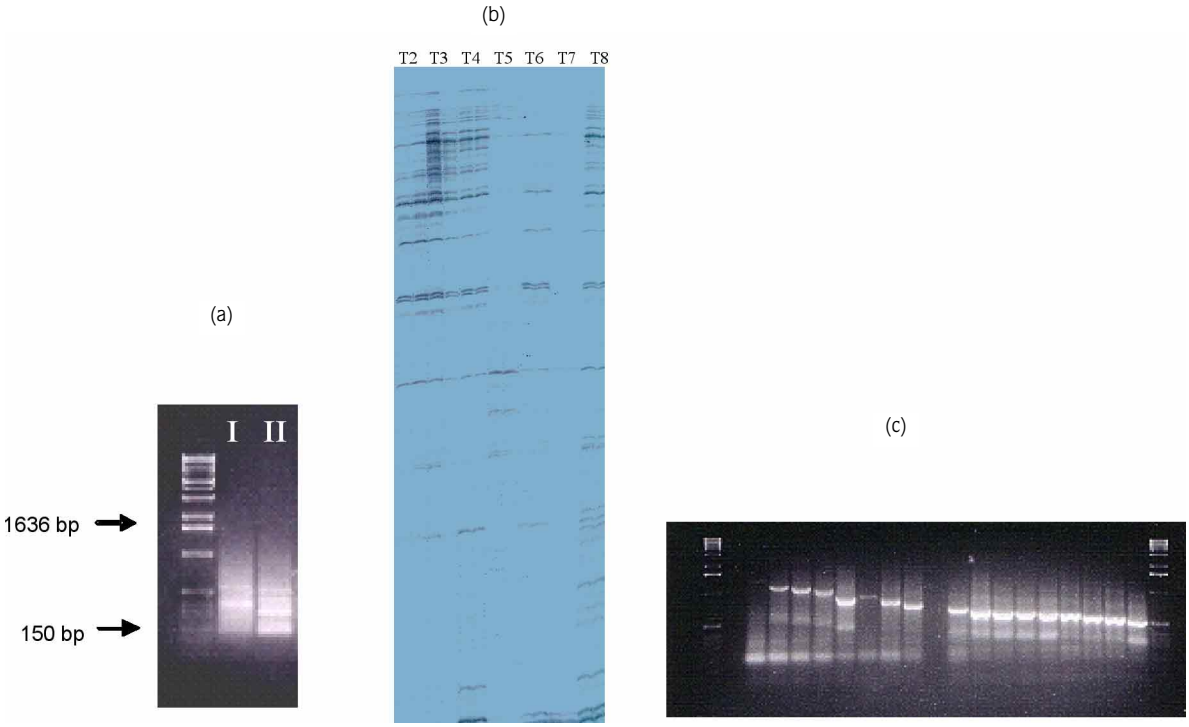

Figure 1. PCR amplification for genomic DNA and cDNA using "p-loop-T" primer set combinations. a) a p-loop-T1 primer set combination product run on 1% agarose gel for genomic DNA (I) and cDNA (II) for wheat. The lane on the far left shows a 1 Kb extension DNA ladder. Arrows indicate the sizes. b) PCR products run on a polyacrylamide gel as lines represent the primer sets from p-loop-T2 to p-loop-T8, respectively c) Re-amplification of the cloned unique fragments.

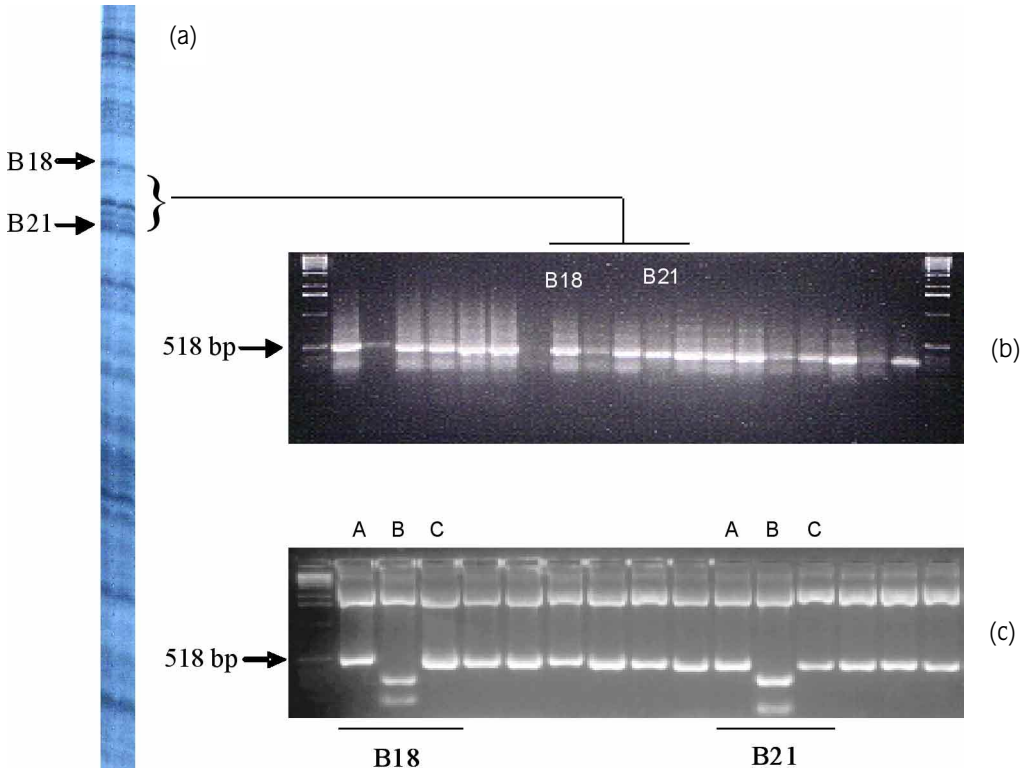

Figure 2. Size separation of co-migrating fragments a) Amplified PCR product run on polyacrylamide gel, b) Re-amplification of the same cut-bands and run on agarose gel and c) Digestion of cloned fragment re-amplified fragments with *Eco*RI restriction digestion enzyme.

assignment of the sequences have been explained elsewhere (18). Twenty-two of the sequences partially or fully matched the sequences in the database (www.ncbi.nlm.nih.gov). Only 7 sequences revealed more than an 80% sequence similarity, of which 2 sequences were previously cloned. Fifteen sequences also showed a 35% to 80% sequence similarity with the cloned plant genes ($E \leq 10^{-40}$ to $E \leq 10^{-8}$). The remaining 52 sequences, however, were not present in the database. In general the 22 cloned sequences matched the p-loop-containing putatively-annotated sequences such as kinases, NTP-binding, ATP-dependent, pathogen related proteins, proteases, transcription factors and resistance genes.

'Gap' (GCG) analysis of 74 sequences showed 21% to 98% sequence identity. Excluding the 24 sequences that were 90%-99% identical to each other, the "Old Distance" (GCG) analyses among 50 sequences are given in Table 2. Only about 27% of the sequences seemed closely related, and the sequence similarity for the remaining 73% was less than 50% (Table 2). The majority of the cloned sequences belonged to the diverged p-loop containing sequence.

Expression Profiles of Unique Sequences

BLASTN search was performed among wheat ESTs and the 22 unique sequences matching to previously cloned plant sequences in the database (www.ncbi.nlm.nih.gov). The majority of these 22 sequences matched more than 1 EST. The 12 unique sequences had only 1 EST in the database. The 10 unique sequences, however, had more than a single EST, ranging from 2 to 7. Two sequences were represented by more than 35 ESTs. The 52 sequences, however, had no EST hit in the database.

Discussion

We particularly used the wheat (*Triticum aestivum* L.) genome because it has the largest EST collection in the database. Furthermore, wheat has a large genome, of which about 95% is non-transcribing (19). Most of the

Table 2. Sequence homology levels among the cloned wheat sequences.

	B17	B16	B51	B31	B30	B27	B52	B15	B61	B38	B44	B7	B5	B40	B50	B36	B28	B37	B6	B19	B42	B8	B20	B41	B35
B17	1.00	0.93	0.86	0.23	0.25	0.24	0.28	0.25	0.23	0.24	0.27	0.18	0.22	0.25	0.31	0.30	0.26	0.27	0.28	0.28	0.24	0.26	0.27	0.29	0.25
B16		1.00	0.89	0.27	0.27	0.28	0.31	0.25	0.23	0.25	0.27	0.18	0.24	0.28	0.33	0.32	0.29	0.32	0.29	0.30	0.30	0.27	0.30	0.32	0.25
B51			1.00	0.25	0.24	0.26	0.26	0.30	0.23	0.23	0.27	0.25	0.25	0.26	0.30	0.28	0.26	0.28	0.30	0.31	0.26	0.21	0.27	0.32	0.26
B31				1.00	0.41	0.26	0.27	0.27	0.25	0.30	0.27	0.19	0.26	0.26	0.28	0.29	0.27	0.28	0.27	0.28	0.30	0.27	0.35	0.30	0.26
B30					1.00	0.25	0.26	0.27	0.24	0.31	0.26	0.19	0.28	0.30	0.29	0.29	0.27	0.28	0.24	0.26	0.32	0.25	0.29	0.29	0.27
B27						1.00	0.37	0.30	0.21	0.25	0.29	0.19	0.24	0.28	0.32	0.28	0.26	0.29	0.28	0.30	0.28	0.29	0.29	0.31	0.27
B52							1.00	0.29	0.29	0.26	0.23	0.22	0.27	0.32	0.37	0.30	0.34	0.38	0.27	0.27	0.28	0.29	0.26	0.29	0.31
B15								1.00	0.22	0.25	0.27	0.16	0.21	0.26	0.28	0.22	0.21	0.23	0.20	0.21	0.24	0.25	0.25	0.30	0.24
B61									1.00	0.38	0.31	0.32	0.22	0.30	0.24	0.29	0.21	0.25	0.29	0.29	0.23	0.26	0.25	0.27	0.26
B38										1.00	0.36	0.32	0.27	0.30	0.31	0.33	0.33	0.33	0.27	0.31	0.30	0.33	0.33	0.29	0.27
B44											1.00	0.25	0.29	0.30	0.25	0.23	0.29	0.26	0.26	0.26	0.28	0.25	0.29	0.27	0.30
B7												1.00	0.27	0.26	0.23	0.18	0.22	0.25	0.14	0.15	0.19	0.20	0.19	0.18	0.22
B5													1.00	0.36	0.23	0.26	0.30	0.30	0.22	0.22	0.23	0.27	0.25	0.27	0.27
B40														1.00	0.30	0.32	0.24	0.27	0.27	0.28	0.26	0.22	0.27	0.30	0.28
B50															1.00	0.36	0.30	0.32	0.26	0.27	0.31	0.31	0.29	0.30	0.26
B36																1.00	0.26	0.33	0.30	0.32	0.34	0.30	0.26	0.28	0.26
B28																	1.00	0.73	0.29	0.31	0.30	0.34	0.29	0.29	0.27
B37																		1.00	0.28	0.31	0.27	0.33	0.27	0.30	0.30
B6																			1.00	0.85	0.34	0.31	0.30	0.28	0.26
B19																				1.00	0.36	0.32	0.31	0.29	0.25
B42																					1.00	0.37	0.29	0.24	0.26
B8																						1.00	0.25	0.24	0.25
B20																							1.00	0.39	0.31
B41																								1.00	0.30
B35																									1.00

	B66	B9	B69	B10	B63	B11	B22	B4	B56	B12	B34	B33	B3	B58	B2	B1	B13	B68	B14	B64	B18	B43	B26	B65	B60
B17	0.20	0.19	0.26	0.29	0.25	0.27	0.25	0.23	0.24	0.27	0.30	0.31	0.26	0.23	0.26	0.28	0.23	0.26	0.27	0.27	0.25	0.22	0.23	0.30	0.27
B16	0.23	0.20	0.32	0.33	0.33	0.27	0.26	0.24	0.30	0.29	0.33	0.33	0.28	0.25	0.27	0.31	0.24	0.32	0.30	0.34	0.25	0.22	0.24	0.33	0.33
B51	0.15	0.23	0.25	0.29	0.26	0.27	0.23	0.26	0.24	0.25	0.31	0.32	0.27	0.21	0.25	0.26	0.22	0.25	0.26	0.29	0.27	0.23	0.25	0.28	0.26
B31	0.21	0.20	0.28	0.33	0.36	0.28	0.21	0.25	0.33	0.34	0.27	0.29	0.26	0.36	0.28	0.30	0.29	0.32	0.28	0.31	0.31	0.24	0.31	0.38	0.34
B30	0.23	0.23	0.29	0.29	0.33	0.25	0.18	0.27	0.32	0.32	0.27	0.27	0.27	0.27	0.31	0.33	0.32	0.33	0.28	0.30	0.31	0.21	0.28	0.32	0.41
B27	0.23	0.26	0.30	0.28	0.38	0.29	0.21	0.22	0.33	0.30	0.36	0.34	0.29	0.31	0.30	0.31	0.29	0.39	0.28	0.25	0.25	0.24	0.29	0.28	0.37
B52	0.24	0.25	0.35	0.29	0.34	0.26	0.26	0.27	0.30	0.32	0.34	0.36	0.31	0.39	0.31	0.31	0.29	0.28	0.27	0.30	0.26	0.25	0.26	0.35	0.33
B15	0.25	0.17	0.30	0.23	0.33	0.24	0.26	0.23	0.29	0.26	0.36	0.33	0.24	0.31	0.28	0.24	0.26	0.34	0.25	0.29	0.28	0.26	0.26	0.28	0.27
B61	0.06	0.22	0.26	0.19	0.12	0.22	0.24	0.27	0.18	0.25	0.23	0.24	0.27	0.23	0.29	0.20	0.19	0.16	0.23	0.29	0.28	0.20	0.28	0.26	0.22
B38	0.13	0.30	0.29	0.29	0.30	0.28	0.25	0.27	0.28	0.28	0.26	0.25	0.25	0.33	0.25	0.37	0.31	0.31	0.34	0.35	0.26	0.24	0.30	0.30	0.31
B44	0.26	0.23	0.25	0.26	0.22	0.26	0.25	0.22	0.26	0.23	0.25	0.26	0.23	0.29	0.26	0.26	0.23	0.22	0.25	0.28	0.26	0.24	0.28	0.23	0.32
B7	0.15	0.24	0.29	0.17	0.22	0.23	0.23	0.25	0.24	0.19	0.20	0.21	0.16	0.30	0.16	0.21	0.19	0.18	0.19	0.26	0.18	0.26	0.23	0.20	0.22
B5	0.28	0.23	0.28	0.30	0.33	0.32	0.29	0.25	0.31	0.24	0.29	0.29	0.23	0.28	0.26	0.28	0.23	0.29	0.23	0.28	0.25	0.27	0.25	0.27	0.30
B40	0.07	0.18	0.28	0.31	0.35	0.29	0.20	0.28	0.29	0.28	0.31	0.30	0.30	0.31	0.33	0.29	0.28	0.35	0.25	0.30	0.34	0.20	0.31	0.26	0.31
B50	0.18	0.25	0.28	0.27	0.35	0.28	0.26	0.26	0.30	0.30	0.30	0.29	0.23	0.37	0.30	0.31	0.28	0.31	0.29	0.31	0.27	0.22	0.29	0.36	0.35
B36	0.25	0.27	0.30	0.25	0.37	0.22	0.20	0.23	0.31	0.28	0.31	0.30	0.28	0.31	0.26	0.36	0.27	0.37	0.28	0.32	0.28	0.30	0.30	0.30	0.32
B28	0.28	0.23	0.37	0.27	0.32	0.25	0.33	0.24	0.27	0.29	0.29	0.29	0.28	0.32	0.30	0.27	0.24	0.24	0.28	0.27	0.26	0.22	0.30	0.33	0.29
B37	0.15	0.26	0.40	0.24	0.31	0.29	0.29	0.23	0.28	0.26	0.35	0.34	0.33	0.34	0.29	0.28	0.26	0.25	0.27	0.30	0.26	0.22	0.32	0.35	0.28
B6	0.25	0.20	0.23	0.26	0.35	0.22	0.21	0.20	0.24	0.30	0.30	0.32	0.25	0.22	0.28	0.25	0.23	0.30	0.25	0.28	0.22	0.28	0.26	0.27	0.29
B19	0.25	0.20	0.27	0.28	0.41	0.23	0.20	0.20	0.26	0.31	0.33	0.34	0.26	0.26	0.29	0.29	0.24	0.34	0.26	0.30	0.22	0.28	0.28	0.29	0.32
B42	0.08	0.24	0.35	0.27	0.34	0.29	0.19	0.29	0.34	0.33	0.30	0.30	0.29	0.35	0.30	0.28	0.23	0.33	0.31	0.32	0.30	0.20	0.31	0.34	0.33
B8	0.32	0.20	0.34	0.31	0.33	0.24	0.19	0.26	0.34	0.28	0.30	0.27	0.25	0.34	0.27	0.34	0.29	0.29	0.27	0.35	0.24	0.29	0.33	0.31	0.34
B20	0.25	0.22	0.29	0.28	0.33	0.25	0.24	0.22	0.30	0.26	0.31	0.31	0.31	0.34	0.26	0.26	0.25	0.33	0.28	0.31	0.25	0.25	0.29	0.33	0.32
B41	0.30	0.28	0.30	0.27	0.33	0.24	0.23	0.23	0.29	0.28	0.33	0.33	0.29	0.30	0.27	0.35	0.29	0.33	0.28	0.28	0.28	0.24	0.25	0.28	0.30
B35	0.41	0.33	0.29	0.24	0.23	0.27	0.22	0.24	0.28	0.25	0.29	0.28	0.21	0.31	0.23	0.31	0.24	0.19	0.22	0.30	0.27	0.31	0.25	0.30	0.27
B66	1.00	0.27	0.00	0.25	0.00	0.23	0.14	0.28	0.00	0.25	0.05	0.04	0.25	0.00	0.21	0.00	0.28	0.00	0.27	0.04	0.29	0.16	0.35	0.00	0.00
B9		1.00	0.15	0.23	0.15	0.22	0.29	0.25	0.19	0.19	0.23	0.22	0.16	0.25	0.24	0.25	0.20	0.23	0.24	0.25	0.20	0.24	0.26	0.25	0.23
B69			1.00	0.40	0.33	0.26	0.23	0.34	0.30	0.30	0.33	0.34	0.33	0.30	0.28	0.32	0.35	0.30	0.30	0.36	0.26	0.14	0.36	0.32	0.34
B10				1.00	0.39	0.27	0.29	0.34	0.35	0.30	0.29	0.31	0.28	0.34	0.27	0.37	0.27	0.34	0.27	0.32	0.30	0.27	0.29	0.27	0.41
B63					1.00	0.41	0.28	0.27	0.40	0.30	0.36	0.35	0.30	0.30	0.36	0.32	0.38	0.32	0.33	0.35	0.29	0.13	0.30	0.27	0.38
B11						1.00	0.34	0.25	0.37	0.24	0.32	0.29	0.24	0.31	0.26	0.26	0.25	0.31	0.26	0.26	0.26	0.27	0.26	0.27	0.30
B22							1.00	0.26	0.24	0.21	0.23	0.24	0.26	0.23	0.31	0.23	0.29	0.20	0.28	0.21	0.23	0.22	0.24	0.15	0.23
B4								1.00	0.29	0.22	0.24	0.26	0.20	0.26	0.25	0.28	0.21	0.26	0.25	0.30	0.23	0.26	0.29	0.26	0.22
B56									1.00	0.46	0.41	0.40	0.32	0.34	0.35	0.35	0.34	0.36	0.30	0.34	0.26	0.22	0.25	0.27	0.31
B12										1.00	0.37	0.36	0.29	0.38	0.30	0.31	0.26	0.30	0.26	0.31	0.22	0.29	0.25	0.36	0.31
B34											1.00	0.90	0.46	0.39	0.33	0.33	0.31	0.33	0.39	0.29	0.25	0.21	0.33	0.33	0.32
B33												1.00	0.42	0.36	0.35	0.32	0.28	0.30	0.37	0.29	0.23	0.21	0.30	0.32	0.33
B3													1.00	0.40	0.28	0.29	0.26	0.39	0.28	0.28	0.23	0.25	0.29	0.39	0.41
B58														1.00	0.41	0.29	0.38	0.33	0.33	0.31	0.26	0.17	0.34	0.37	0.36
B2															1.00	0.29	0.33	0.35	0.28	0.35	0.26	0.22	0.27	0.31	0.41
B1																1.00	0.41	0.39	0.30	0.38	0.28	0.17	0.27	0.29	0.35
B13																	1.00	0.34	0.27	0.27	0.20	0.22	0.30	0.39	0.36
B68																		1.00	0.45	0.42	0.26	0.14	0.36	0.30	0.36
B14																			1.00	0.35	0.27	0.21	0.34	0.37	0.41
B64																				1.00	0.42	0.16	0.29	0.33	0.33
B18																					1.00	0.25	0.25	0.23	0.23
B43																						1.00	0.27	0.23	0.19
B26																							1.00	0.44	0.39
B65																								1.00	0.35
B60																									1.00

wheat genes are present in clusters spanning small physical regions of varying gene densities and numbers (20). The genic regions are spaced by varying sized blocks of nested retrotransposons and duplicated genes (21, 22, 23). Due to its size, sequencing of the total wheat genome is not an option for the immediate future. Therefore, approaches targeting the expressed part of this genome are desirable.

We studied 2 different contexts using the motif-based primer along with "T" primers for amplifications; the cloning of distinct p-loop containing sequences and a well-normalized cDNA cloning approach. Combination of a polyacrylamide gel and cDNA technology allowed us to isolate the gene fragments varying 35-fold in expression. Expression of about 18% of the sequences (represented by 1 and 2 ESTs) was very low as no EST was present for about 70% of the sequences. Two sequences contained a high copy of EST matches indicating the over expression. A random EST cloning method would mainly represent several of these 2 sequences, due to a higher expression level, but none of the rare transcripts. In this study, however, the fragments with rare transcripts were targeted by labeling the PCR amplification products with ^{35}S and these were size separated on denaturing polyacrylamide gels (Figure 1). Therefore, the bands representing rare transcripts were as visible as the highly expressed fragment bands. Furthermore, additional sequencing of the bands containing more than 1 fragment ensured identification of even co-migrating sequences (Figure 2).

Estimates of the number of wheat genes range from 75,000 to 150,000 (6). However, over 500,000 wheat ESTs are currently available in database representing only about 23,000 wheat unigenes (www.ncbi.nlm.nih.gov). This translates to about 5% efficiency and an average of 20 transcripts per unigene. Random EST libraries are sequenced from both 5' and 3' ends, so that the actual gene number represented by these unigenes may even be lower. The main drawback with this low rate is, probably, because most of the wheat transcripts are single or few copies that will be out-competed by multiple copy sequences. In this study, however, by sequencing 189 clones we identified 74 unique sequences that translate to about 39% efficiency. Furthermore, 61% (52) of the unique sequences were not even present in any plant species.

The current study used the p-loop motif as a primer site. The p-loop (kinase-1a) motif frequently occurs and is shared by several protein families (24). It contains 12 different subfamilies representing a different conserved pattern of a unique structure (13). Taking a few variant amino acids within a subfamily also into account, it is possible to clone about 1000 unique sequences using only the 12 different p-loop motif primers. Furthermore, using different gene family specific primers for the 3'site along with a different p-loop primer for the 5' site can also increase the number of unique clones. In conclusion, the use of different motif sequences found in diverse protein families may increase the number of unique transcripts. For the cloning of the transcribed part of the genome, the approach described here is applicable to any other plant species, particularly, for those that cannot be completely sequenced.

Acknowledgment

We gratefully thank Dr. Kulvinder Gill, the Vogel Endowed Chair for Wheat Breeding at Washington State University, for allowing us to use his facility during the study and for his guidance during our academic careers.

Corresponding author:
Muharrem DİLBİRLİĞİ
Central Research Institute for Field Crops,
Pk: 226, 0642 Ulus/Ankara, Turkey
E-mail: mdilbirligi@tagem.gov.tr

References

1. Covitz PA, Smith LS, Long SR. Expressed sequence tags from a root-hair-enriched medicago truncatula cDNA library. Plant Physiol 117: 1325-1332, 1998.
2. Ewing RM, Kahla AB, Poirot O et al. Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. Genome Res 9: 950-959, 1999.
3. Fernandes J, Brendel V, Gai X et al. Comparison of RNA expression profiles based on maize expressed sequence tag frequency analysis and micro-array hybridization. Plant Physiol 128: 896-910, 2002.
4. Hillier LD, Lennon G, Becker M et al. Generation and analysis of 280,000 human expressed sequence tags. Genome Res 6: 807-828, 1996.

5. Adams MD, Kelly JM, Gocayne JD et al. Complementary DNA sequencing: Expressed sequence tags and human genome project. Science 252: 1651-1656, 1991.

6. Sandhu D, Gill KS. Gene-containing regions of wheat and the other grass genomes. Plant Physiol 128: 803-811, 2002.

7. Henikoff S, Greene EA, Pietrokovski S et al. Gene families: the taxonomy of protein paralogs and chimeras. Science 278: 609-614, 1997.

8. Rubin GM, Yandell MD, Wortman JR et al. Comparative genomics of the eukaryotes. Science 287: 2204-2215, 2000.

9. Goff SA, Ricke D, Lan TH et al. A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296: 92-100, 2002.

10. TAGI. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408: 796-815, 2000.

11. Raetz CR, Roderick SL. A left-handed parallel beta helix in the structure of UDP-N-acetylglucosamine acyltransferase. Science 270: 997-1000, 1995.

12. Zeng C, Pinsonneault J, Gellon G et al. Deformed protein binding sites and cofactor binding sites are required for the function of a small segment-specific regulatory element in Drosophila embryos. Embo J 13: 2362-2377, 1994.

13. Kinoshita K, Sadanami K, Kidera A et al. Structural motif of phosphate-binding site common to various protein super families: all-against-all structural comparison of protein-mononucleotide complexes. Protein Eng 12: 11-14, 1999.

14. Fujihashi M, Zhang YW, Higuchi Y et al. Crystal structure of cis-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase. Proc Natl Acad Sci USA 98: 4337-4342, 2001.

15. Collins NC, Webb CA, Seah S et al. The isolation and mapping of disease resistance gene analogs in maize. Mol Plant Microbe Interact 11: 968-978, 1998.

16. Shen KA, Meyers BC, Islam-Faridi MN et al. Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. Mol Plant Microbe Interact 11: 815-823, 1998.

17. Sambrook J, Fristch EF, Maniatis T."Molecular Cloning: A Laboratory Manual," Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY, 1989.

18. Dilbirligi M, Gill KS. Identification and analysis of expressed resistance gene sequences in wheat. Plant Mol Biol 53: 771-787, 2004.

19. Arumuganathan K, Earle ED. Nuclear DNA Content of Some Important Plant Species. Plant Molecular Biology Reporter 9: 208-218, 1991.

20. Erayman M, Sandhu D, Sidhu D, Dilbirligi M, Baenziger PS, Gill KS. Demarcating the gene-rich regions of wheat genome. Nuc Acids Res 32: 3546-3565, 2004.

21. Barakat A, Carels N, Bernardi G. The distribution of genes in the genomes of Graminae. Proc Natl Acad Sci USA 94: 6857-6861, 1997.

22. Wendel JF. Genome evolution in polyploids. Plant Mol Biol 42: 225-249, 2000.

23. Wicker T, Stein N, Albar L et al. Analysis of a contiguous 211 kb sequence in diploid wheat (Triticum monococcum L.) reveals multiple mechanisms of genome evolution. Plant J 26: 307-316, 2001.

24. Traut TW. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. Eur J Biochem 222: 9-19, 1994.