Exploiting a wheat EST database to assess genetic diversity

Ozge Karakas¹, Filiz Gurel² and Ahu Altinkut Uncuoglu¹

¹The Scientific and Technological Research Council of Turkey, Marmara Research Center, Genetic Engineering and Biotechnology Institute, Gebze-Kocaeli, Turkey. 
²Department of Molecular Biology and Genetics, Istanbul University, Istanbul, Turkey.

Abstract

Expressed sequence tag (EST) markers have been used to assess variety and genetic diversity in wheat (Triticum aestivum). In this study, 1549 ESTs from wheat infested with yellow rust were used to examine the genetic diversity of six susceptible and resistant wheat cultivars. The aim of using these cultivars was to improve the competitiveness of public wheat breeding programs through the intensive use of modern, particularly marker-assisted, selection technologies. The F₂ individuals derived from cultivar crosses were screened for resistance to yellow rust at the seedling stage in greenhouses and adult stage in the field to identify DNA markers genetically linked to resistance. Five hundred and sixty ESTs were assembled into 136 contigs and 989 singletons. BlastX search results showed that 39 (29%) contigs and 96 (10%) singletons were homologous to wheat genes. The database-matched contigs and singletons were assigned to eight functional groups related to protein synthesis, photosynthesis, metabolism and energy, stress proteins, transporter proteins, protein breakdown and recycling, cell growth and division and reactive oxygen scavengers. PCR analyses with primers based on the contigs and singletons showed that the most polymorphic functional categories were photosynthesis (contigs) and metabolism and energy (singletons). EST analysis revealed considerable genetic variability among the Turkish wheat cultivars resistant and susceptible to yellow rust disease and allowed calculation of the mean genetic distance between cultivars, with the greatest similarity (0.725) being between Harmankaya99 and Sönmez2001, and the lowest (0.622) between Aytin98 and Izgi01.

Key words: biodiversity, EST, genetic diversity, Triticum, yellow rust.

Received: October 29, 2009; Accepted: June 24, 2010.

Introduction

Wheat (Triticum aestivum L.) is one of the most important crops in the world and is grown in all agricultural regions of Turkey. The total area cultivated worldwide and in Turkey is 210 and 9.4 million ha, respectively (Zeybek and Yigit, 2004). The allohexaploid wheat genome (2n = 6x = 42) is one of the largest among crop species, with a haploid size of 16 billion bp (Bennett and Leitch, 1995), and its genetics and genome organization have been extensively studied using molecular markers (Yu et al., 2004; Ercan et al., 2010; Akfirat-Senturk et al., 2010).

PCR-based molecular markers such as simple sequence repeats (SSR) (Plaschke et al., 1995), restriction fragment length polymorphism (RFLP) (Nagaoka and Ogi- hara, 1997), amplified fragment length polymorphism (AFLP) (Gülbitti-Onarici et al., 2007), selective amplification of microsatellite polymorphic loci (SAMPL) (Altitas et al., 2008) and random amplified polymorphic DNA (RAPD) (Asif et al., 2005) are easy to use and show a high degree of polymorphism. A number of wheat genetic maps have been constructed using PCR based markers (Li et al., 2007).

In recent year, expressed-sequence tags (ESTs) have become a valuable tool for genomic analyses and are currently the most widely used approach for sequencing plant genomes, both in terms of the number of sequences and total nucleotide counts (Rudd, 2003). EST analysis provides a simple strategy for studying the transcribed regions of genomes, and renders complex, highly redundant genomes such as that of wheat amenable to large-scale analysis. The number of ESTs and cDNA sequences in public databases such as GenBank has increased exponentially in recent few years, and EST-based markers have been used to distinguish varieties and assess genetic diversity in wheat (Kantety et al., 2002; Leigh et al., 2003).

Yellow rust, a destructive disease of wheat triggered by the biotrophic fungus Puccinia striiformis f. sp. tritici (Chen 2005), is the most frequent and important cereal disease in Turkey, where it causes grain yield losses of 40%-60% and lowers the quality of cereal products (Zeybek and Yigit, 2004). In this study, an EST database for yellow rust-infested wheat was used, in conjunction with a multi-
vioximate statistical package (MVSP v.3.1), to assess the genetic diversity of yellow rust resistant and susceptible wheat genotypes. For this, EST sequences were assembled into longer contiguous sequences (contigs) using Vector NTI 10.0 software. Difficulties related to sequencing errors and the determination of orthology associated with the use of ESTs for systematics can be minimized by using several reads to assemble contigs and EST clusters for each region (Parkinson et al., 2002; Torre et al., 2006). The knowledge gained about the genetic constitution and relationships of genotypes using this approach should prove useful in the optimization of wheat breeding programs.

Materials and Methods

Plant material and evaluations

Six homozygous bread wheat genotypes (three yellow rust-resistant cultivars: PI178383, Izgi01, Sönmez2001, and three yellow rust-susceptible cultivars: Harmankaya99, ES14, Aytin98) were obtained from the Anatolian Agricultural Research Institute, Eskisehir, Turkey. The resistance of the parental cultivars and F2 generation was tested in greenhouses by applying uredospores. Two weeks after the inoculation the infection was scored on a scale of 0-9 (McNeal et al., 1971), with scores of 0-6 indicating a low infection and 7-9 indicating a high infection. The disease score for PI178383, Izgi01 and Sönmez2001 was 0 while that of Harmankaya99, ES14 and Aytin98 was 8, this confirming the resistance and susceptibility of the parental genotypes.

Analysis of wheat yellow rust ESTs

ESTs from a yellow rust-infected wheat cDNA library (TA117G1X) were selected from the GrainGenes website and processed by means of VecScreen database searches to remove undesired vector fragments from the sequences. The Vector NTI 10.0 contig express program (InforMax, Bethesda, MD, USA) was used to construct contig tags from the EST sequences and the Contig Express module was used to assemble small fragments in text or chromatogram formats into contigs (Lu and Moriyama, 2004). Singletons were constructed from unassembled ESTs. The EST sequences were aligned and analyzed with ClustalW v.1.82 to identify conserved domains. Functional annotation was done using the BlastX algorithm of the Basic Alignment Search Tool (Altschul et al., 1990). PCR primers for the contigs and singletons selected for further characterization were designed with Primer Premier 5.0 and Primer 3.0 software (Figure 1). EST-derived contig and singleton primers were used to assess the genetic diversity of the six wheat genotypes.

PCR analyses of contigs and singletons

Total genomic DNA was extracted from the leaves of resistant and susceptible plants using the method of Weining and Langridge (1991) as modified by Song and Henry (1995). Genomic DNA amplifications with sense and antisense primers were done using a PTC-100 MJ thermocycler (MJ Research, Watertown, MA) in a 25 μL reaction volume. Each reaction contained 1X Taq buffer (MBI Fermentas, Germany), 2.5 mM MgCl2 (MBI Fermentas), 0.2 mM dNTP (MBI Fermentas), 400 nM of forward primer, 400 nM of reverse primer, 0.625 U of Taq polymerase/μL (MBI Fermentas) and 100 ng of genomic DNA. The thermal cycling parameters were: 3 min at 94 °C (initial denaturation), 37 cycles of 1 min at 94 °C, 1 min at 40-58 °C (depending on the annealing temperature) and 1 min at 72 °C, followed by a final extension at 72 °C for 10 min. PCR products were separated in 2% agarose gels, stained with ethidium bromide and examined under UV light.

Genetic similarity estimation and cluster analyses

Each contig and singleton band was scored as absent (0) or present (1) for the different cultivars and the data were entered into a binary matrix as discrete variables (‘1’ for presence and ‘0’ for absence of a homologous fragment). Only distinct, reproducible, well-resolved fragments were scored and the data were analyzed using MVSP 3.1 software (Kovach, 1999). This software package was also used to cal-
culation Jaccard (1908) similarity coefficients to construct a dendrogram by a neighbour-joining algorithm.

Results

Assembly of contigs and blast analysis

Table 1 summarizes the characteristics of the database used in this analysis. 1549 ESTs were selected from a yellow rust-infested wheat cDNA library (TA117G1X) and used to assemble 136 contigs. The number of individual ESTs belonging to each contig ranged from 2 to 57. Singletons were derived from unassembled ESTs and accounted for 72.63% of ESTs. Tables 2 and 3 show the results of the NCBI database searches done using the contig and singleton sequences. The BlastX searches revealed that 39 contigs (29%) were homologous to wheat genes (Figure 2). Contigs 3, 4, 11, 13, 16 and 112 did not match any organism. Contig 77 matched a sequence of unknown function (data not shown) while other contigs (71%) showed homology to genes of known function. The BlastX search also showed that 96 singletons (10%) were homologous to wheat genes (Figure 3), whereas 147 singletons (14%) did not match any organism and had no functional annotation (data not shown). The 39 contigs and 96 singletons that matched wheat proteins were assigned to eight functional groups that included protein synthesis, photosynthesis, metabolism and energy, stress proteins, transporter proteins, protein breakdown and recycling, cell growth and division and reactive oxygen scavengers. Photosynthesis was the major functional category of contigs, with nine proteins (22%), whereas cell growth and division was the smallest, with one protein (3%) (Figure 2). Metabolism was the major functional category of singletons, with 37 proteins (38%), whereas protein breakdown and recycling and cell growth and division were the smallest functional categories, with three proteins (3%) (Figure 3). Tables 4 and 5 show the sense and antisense primers used to assess the genetic diversity of wheat cultivars; these primers were designed

![Figure 2 - Classification of contigs homologous to proteins of known function.](image)

![Figure 3 - Classification of singletons homologous to proteins of known function.](image)

| Contig name | Blast hit number | Annotation | Accession number |
|-------------|------------------|------------|-----------------|
| Contig 1    | 100              | ribosomal protein L16 | NP_114295       |
| Contig 8    | 44               | ribosomal protein S7 | AAW50993        |
| Contig 9    | 101              | lipid transfer protein | ABB90546        |
| Contig 12   | 101              | chlorophyll a/b binding protein, chloroplast precursor (LHCII type I CAB) (LHCP) | P04784           |
| Contig 17   | 100              | ferredoxin, chloroplast precursor | P00228           |
| Contig 19   | 100              | triosephosphate-isomerase | CAC14917        |
| Contig 21   | 196              | putative glycine decarboxylase subunit | AAM92707        |
| Contig 22   | 281              | eukaryotic translation initiation factor 5A1 | AAZ95171        |
| Contig 24   | 100              | single-stranded nucleic acid binding protein | AAA75104        |
| Contig 30   | 100              | cytosolic glyceraldehyde-3-phosphate dehydrogenase | AAP83583        |
| Contig 33   | 294              | chlorophyll a/b-binding protein WCAB precursor [Triticum aestivum] | AAB18209        |
### Table 2 (cont.)

| Contig name | Blast hit number | Annotation | Accession number |
|-------------|------------------|------------|-----------------|
| Contig 34   | 65               | jasmonate-induced protein | AAR20919 |
| Contig 35   | 44               | oxygen-evolving enhancer protein 2, chloroplast precursor (OEE2) | Q00434 |
| Contig 39   | 100              | geranylgeranyl hydrogenase | AAZ67145 |
| Contig 40   | 100              | chlorophyll a/b-binding protein WCAB precursor | AAB18209 |
| Contig 46   | 102              | chlorophyll a/b-binding protein WCAB precursor | AAB18209 |
| Contig 49   | 31               | oxygen-evolving complex precursor | AAP80632 |
| Contig 52   | 9                | metallothionein-like protein 1 (MT-1) | P43400 |
| Contig 55   | 198              | glycine-rich RNA-binding protein | BAF30986 |
| Contig 57   | 100              | type 1 non-specific lipid transfer protein precursor | CAH04983 |
| Contig 58   | 33               | RUB1-conjugating enzyme | AAP80608 |
| Contig 63   | 103              | oxygen-evolving enhancer protein 1, chloroplast precursor (OEE1) (33 kDa subunit of oxygen evolving system of photosystem II) (OEC 33 kDa subunit) (33 kDa thylakoid membrane protein) | P27665 |
| Contig 65   | 101              | acidic ribosomal protein P2 | AAP80619 |
| Contig 66   | 199              | cyclophilin A-1 | AAK49426 |
| Contig 73   | 190              | dehydroascorbate reductase | AAL71854 |
| Contig 75   | 63               | metallothionein | AAP80616 |
| Contig 80   | 33               | wali7 | AAC37416 |
| Contig 90   | 52               | putative membrane protein | ABB90549 |
| Contig 91   | 100              | cold shock protein-1 | BAB78536 |
| Contig 93   | 155              | Ps16 protein | BAA22411 |
| Contig 96   | 109              | elongation factor 1-alpha (EF-1-alpha) | Q03033 |
| Contig 99   | 72               | histone H1 WH1A.2 | AAD41006 |
| Contig 105  | 131              | ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain precursor (clone pWS4.3) - wheat | RKWTS |
| Contig 110  | 82               | cytochrome b6-f complex iron-sulfur subunit, chloroplast precursor (Rieske iron-sulfur protein) (plastohydroquinone:plastocyanin oxidoreductase iron-sulfur protein) (ISP) (RISP) | Q7X9A6 |
| Contig 113  | 103              | lipid transfer protein 3 | AAP23941 |
| Contig 122  | 163              | ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit | BAB19814 |
| Contig 133  | 100              | ribosomal protein L36 | AAW50980 |
| Contig 135  | 100              | 60s ribosomal protein L21 | AAP80636 |
| Contig 136  | 100              | histone H2A.2.1 | P02276 |

### Table 3 - Singletons showing homology to genes with proteins matching *Triticum aestivum* identified in a BlastX search of the NCBI database.

| Singleton name | Blast hit number | Annotation | Accession number |
|----------------|------------------|------------|-----------------|
| CA599282       | 199              | ATP synthase CF1 alpha subunit | NP_114256 |
| CA599218       | 88               | ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit | BAB19811 |
| CA598725       | 191              | ribosomal protein L14 | NP_114294 |
| CA597765       | 119              | RuBisCO large subunit-binding protein subunit alpha, chloroplast precursor (60 kDa chaperonin subunit alpha) (CPN-60 alpha) | P08823 |
| CA597760       | 100              | type 1 non-specific lipid transfer protein precursor | CAH69210 |
| CA597766       | 3                |aintegumenta-like protein | ABB90555 |
| CA597808       | 116              | geranylgeranyl hydrogenase | AAZ67145 |
| CA597830       | 100              | 14-3-3 protein | AAR89812 |
| CA597851       | 49               | plastid glutamine synthetase isoform GS2c | AAZ30062 |
| CA597983       | 100              | GRAB2 protein | CAA09372 |
| CA598020       | 103              | protein H2A.5 (weH2A-2) | Q43213 |
| CA598034       | 100              | histone deacetylase | AUA82113 |
| CA598102       | 22               | WIR1A protein | Q01482 |
| Singleton name | Blast hit number | Annotation | Accession number |
|----------------|------------------|------------|-----------------|
| CA598128       | 100              | probable light-induced protein | AAP80856 |
| CA598130       | 100              | tubulin beta-2 chain (beta-2 tubulin) | Q9ZRB1 |
| CA598143       | 172              | thioredoxin M-type, chloroplast precursor (TRX-M) | Q9ZP21 |
| CA598151       | 100              | lipid transfer protein precursor | AAG27707 |
| CA598174       | 200              | S28 ribosomal protein | AAP80664 |
| CA598181       | 110              | pathogenesis-related protein 1.2 | CAA07474 |
| CA598182       | 2                | pathogenesis-related protein 1.2 | CAA07474 |
| CA598187       | 98               | VER2 | BAA32786 |
| CA598196       | 1                | putative cytochrome c oxidase subunit | AAM92706 |
| CA598235       | 100              | plasma membrane intrinsic protein 1 | AAF16463 |
| CA598239       | 151              | triosephosphate translocator | AAK01174 |
| CA598244       | 14               | glycosyltransferase | CAI30070 |
| CA598256       | 100              | heat shock protein 80 | AAD11549 |
| CA598258       | 22               | fasciclin-like protein FLA26 | ABO5416 |
| CA598266       | 80               | elongation factor 1-beta (EF-1-beta) | P29546 |
| CA598296       | 106              | beta-1,3-glucanase precursor | AAD28734 |
| CA598314       | 11               | oxygen-evolving enhancer protein 2, chloroplast precursor (OEE2) | Q00434 |
| CA598347       | 114              | putative ribosomal protein S18 | AAM92708 |
| CA598359       | 198              | sucrose synthase type I | CAA04543 |
| CA598366       | 105              | receptor-like kinase protein | AAS93629 |
| CA598421       | 121              | ribulose-bisphosphate carboxylase (EC 4.1.1.39) small chain precursor (clone pWS4.3) | RKWTS |
| CA598422       | 75               | wali5 | AAA50850 |
| CA598432       | 99               | ribosomal protein P1 | AAW50990 |
| CA598476       | 100              | LRR19 | AAK20736 |
| CA598485       | 100              | ribulose-bisphosphate carboxylase | CAA25058 |
| CA598489       | 64               | histone H2A | AAB00193 |
| CA598518       | 157              | phosphoribulokinase; ribulose-5-phosphate kinase | CAA41020 |
| CA598523       | 100              | ribosomal protein L19 | AAP80588 |
| CA598557       | 79               | type 2 non-specific lipid transfer protein precursor | CAH69201 |
| CA598577       | 252              | ferredoxin, chloroplast precursor | P00228 |
| CA598584       | 258              | putative fructose 1-6-biphosphate aldolase | CAD12665 |
| CA598630       | 101              | translationally-controlled tumor protein homolog (TCTP) | Q8LMR8 |
| CA598637       | 100              | histone H2A | AAB00193 |
| CA598672       | 100              | lipid transfer protein | ABB90546 |
| CA598674       | 100              | glutathione transferase F6 | CAD29479 |
| CA598677       | 100              | ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit | BAB19812 |
| CA598687       | 55               | wali6 | AAC37417 |
| CA598691       | 100              | type 1 non-specific lipid transfer protein precursor | CAH69483 |
| CA598694       | 45               | cold-responsive LEA/RAB-related COR protein | AF25053 |
| CA598700       | 195              | fructan 1-exohydrolase | CAD48199 |
| CA598719       | 24               | 50S ribosomal protein L9, chloroplast precursor (CL9) | Q8L803 |
| CA598755       | 100              | type 1 non-specific lipid transfer protein precursor | CAH69190 |
| CA598762       | 95               | cysteine synthase (O-acetylserylserine sulphydrylase) (O-acetylserylserine (thiol)-lyase) (CSase A) (OAS-TL A) | P38076 |
| CA598818       | 100              | putative fructose 1,6-biphosphate aldolase | CAD12665 |
| CA598837       | 126              | glutathione S-transferase | AAD56395 |
| CA598848       | 167              | glyceraldehyde-3-phosphate dehydrogenase | AAW68026 |
based on the contig and singleton sequences that were homologous to wheat genes.

**EST-derived contig and singleton polymorphisms**

PCR analyses with the contig and singleton primers showed that the most polymorphic functional categories were photosynthesis (30%) and metabolism and energy (46%) for contigs and singletons, respectively (Figures 4 and 5). Of the 39 contig and 92 singleton primers used to characterize the genetic diversity of the six wheat genotypes, 14 contig and 48 singleton primers were polymorphic in susceptible and resistant wheat cultivars. Table 6

| Singleton name | Blast hit number | Annotation | Accession number |
|----------------|------------------|------------|-----------------|
| CA599850       | 42               | putative proteinase inhibitor-related protein | AAS49905         |
| CA599859       | 43               | ferredoxin-NADP(H) oxidoreductase | CAD30024         |
| CA599166       | 137              | cold acclimation induced protein 2-1 | AAY16797         |
| CA599172       | 135              | stress responsive protein | AAY44603         |
| CA599235       | 100              | beta-expansin TaEXPB3 | AAT9294         |
| CA599238       | 77               | oxygen-evolving enhancer protein 2, chloroplast precursor (OEE2) (23 kDa subunit of oxygen evolving system of photosystem II) (OEC 23 kDa subunit) (23 kDa thylakoid membrane protein) | Q00434 |
| CA599257       | 101              | glyceraldehyde-3-phosphate dehydrogenase | AAW68026         |
| CA599262       | 196              | histone H2A.2.1 | P02276         |
| CA599265       | 2                | phosphoglycerate kinase, chloroplast precursor | P12782 |
| CA599271       | 100              | ribosomal protein L18 | AAW50985 |
| CA599273       | 68               | outer mitochondrial membrane protein porin (voltage-dependent anion-selective channel protein) (VDAC) | P46274 |
| CA599277       | 103              | putative SKP1 protein | CAE53885         |
| CA599285       | 154              | putative lipid transfer protein | ABB90547         |
| CA599802       | 100              | ribosomal protein L11 | AAW50983         |
| CA599830       | 100              | thioredoxin h | CAB96931         |
| CA599840       | 199              | cyc07 | AAP80855         |
| CA599841       | 298              | calcium-dependent protein kinase | ABY59005         |
| CA599849       | 100              | putative 40S ribosomal protein S3 | AAM29710         |
| CA599861       | 100              | ribosomal protein L13a | AAW50984         |
| CA599862       | 57               | reversibly glycosylated polypeptide | CAA77237         |
| CA599866       | 282              | MAP kinase | ABS11090         |
| CA599875       | 105              | (1,3;1,4) beta glucanase | CAA80493         |
| CA599880       | 31               | minichromosomal maintenance factor | AAS68103         |
| CA599013       | 100              | D1 protease-like protein precursor | AAL99044         |
| CA599015       | 17               | putative beta-expansin | BAD06319         |
| CA599032       | 114              | tonoplast intrinsic protein | AB96817         |
| CA599049       | 41               | porphobilinogen deaminase | AAL12221         |
| CA599099       | 100              | gamma-type tonoplast intrinsic protein | AAD10494         |
| CA599101       | 100              | small GTP-binding protein | AAD28731         |
| CA599103       | 19               | pre-mRNA processing factor | AAY84871         |
| CA599107       | 82               | sedoheptulose-1,7-bisphosphatase, chloroplast precursor (sedoheptulose bisphosphatase) (SBPase) (SED(1,7)P2ase) | P46285 |
| CA599110       | 199              | ribulose bisphosphate carboxylase small chain PWS4.3, chloroplast precursor (RubisCO small subunit PWS4.3) | P00871 |
| CA599114       | 3                | metallothionein-like protein 1 (MT-1) | P43400 |
| CA599115       | 176              | type 1 non-specific lipid transfer protein precursor | CAH69199 |
| CA599119       | 5                | putative high mobility group protein | CAI64395         |
| CA599121       | 51               | putative proteinase inhibitor-related protein | AAS49905         |
| CA599135       | 257              | putative cellulose synthase | BAD06322         |
Table 4 - Contig primers used for genomic amplifications.

| Primer | Sequence (5’-3’) | T<sub>o</sub>C | Product size (bp) | Primer | Sequence (5’-3’) | T<sub>o</sub>C | Product size (bp) |
|--------|-----------------|-------------|-----------------|--------|-----------------|-------------|-----------------|
| Contig 1F | ACA gAT AgA AgC Agg ACg AA | 50 | 370 | Contig 58F | ggg CAa gAA gAA gAA gAA AgA gA | 50 | 267 |
| Contig 1R | AAg ggg TgA Agg AAT TAT TgT gA | 53 | 168 | Contig 58R | TgA ggg TTA ggg AAg ggg gAA gA | 53 | 168 |
| Contig 8F | gCT CAC TgA gTA ggA TgA gAC | 50 | 340 | Contig 65F | gTT gCC TAT CgT gT gT gAC | 48 | 898 |
| Contig 8R | CTg TgA TTT gCA TAT TgT gAC | 50 | 340 | Contig 65R | gTT gCC TAT CgT gT gT gAC | 48 | 898 |
| Contig 12F | Acg CAC ATc gga CAC gC | 53 | 336 | Contig 73F | gTT gAT gAC TgA gAC TgA gAC | 46 | 417 |
| Contig 12R | CAg CTC CCC ggT TTT gT | 49 | 366 | Contig 73R | gTT gAT gAC TgA gAC TgA gAC | 46 | 417 |
| Contig 17F | gCC ACC TTC TCA gCC ACA | 49 | 366 | Contig 75F | gTT gAT gAC TgA gAC TgA gAC | 46 | 417 |
| Contig 17R | gCC gCC gCC gCC gCA gAC | 49 | 366 | Contig 75R | gTT gAT gAC TgA gAC TgA gAC | 46 | 417 |
| Contig 19F | gAg gCA ATc ggA AAT g | 50 | 350 | Contig 75F | gTT gAT gAC TgA gAC TgA gAC | 46 | 417 |
| Contig 19R | gAg gCA ATc ggA AAT g | 50 | 350 | Contig 75R | gTT gAT gAC TgA gAC TgA gAC | 46 | 417 |
| Contig 21F | gCC CTC AAC ATT TCA gAgC Ag | 50 | 516 | Contig 80F | gTT gAT gAC TgA gAC TgA gAC | 50 | 499 |
| Contig 21R | gCC CTC AAC ATT TCA gAgC Ag | 50 | 516 | Contig 80R | gTT gAT gAC TgA gAC TgA gAC | 50 | 499 |
| Contig 22F | gAg TcC gAg CAg gAg gAg T | 48 | 363 | Contig 90F | gCC ATc CAT gCa AAg gAa gAa | 50 | 412 |
| Contig 22R | gAg TcC gAg CAg gAg gAg T | 48 | 363 | Contig 90R | gCC ATc CAT gCa AAg gAa gAa | 50 | 412 |
| Contig 24F | gTg Tgg CTc cTc cTc CcT Tc | 51 | 331 | Contig 91F | gCC ATc CAT gCa AAg gAa gAa | 50 | 412 |
| Contig 24R | gTg Tgg CTc cTc cTc CcT Tc | 51 | 331 | Contig 91R | gCC ATc CAT gCa AAg gAa gAa | 50 | 412 |
| Contig 30F | gAg gTg AcT gAg gAc gCc gTT gTc | 44 | 450 | Contig 93F | gCC ATc CAT gCa AAg gAa gAa | 50 | 412 |
| Contig 30R | gAg gTg AcT gAg gAc gCc gTT gTc | 44 | 450 | Contig 93R | gCC ATc CAT gCa AAg gAa gAa | 50 | 412 |
| Contig 33F | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90F | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 33R | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90R | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 34F | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90F | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 34R | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90R | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 35F | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90F | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 35R | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90R | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 39F | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90F | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 39R | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90R | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 46F | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90F | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 46R | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90R | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 49F | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90F | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 49R | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90R | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 52F | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90F | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 52R | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90R | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 55F | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90F | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
| Contig 55R | gAg gAg gAc gAg gAg gAc gTg | 48 | 363 | Contig 90R | gAT gAg gAg gAg gAg gAg gAg gAg | 49 | 357 |
Table 5 - Singleton primers used for genomic amplifications.

| Primer | Sequence (5'→3') | T₀°C | Product size (bp) | Primer | Sequence (5'→3') | T₀°C | Product size (bp) |
|--------|------------------|------|------------------|--------|------------------|------|------------------|
| CA598034F | CA598034R | 38   | 418              | CA598930F | CA598930R | 46   | 366              |
| CA598174F | CA598174R | 43   | 263              | CA598143F | CA598143R | 50   | 279              |
| CA598286F | CA598286R | 42   | 278              | CA599110F | CA599110R | 58   | 138              |
| CA598347F | CA598347R | 54   | 326              | CA598930F | CA598930R | 51   | 241              |
| CA598432F | CA598432R | 47   | 122              | CA599218F | CA599218R | 44   | 490              |
| CA598523F | CA598523R | 50   | 262              | CA599271F | CA599271R | 49   | 483              |
| CA598719F | CA598719R | 52   | 266              | CA598802F | CA598802R | 47   | 461              |
| CA598762F | CA598762R | 44   | 224              | CA598949F | CA598949R | 45   | 209              |
| CA598930F | CA598930R | 50   | 262              | CA598577F | CA598577R | 55   | 363              |
| CA598949F | CA598949R | 51   | 234              | CA599238F | CA599238R | 50   | 426              |
| CA599103F | CA599103R | 48   | 272              | CA597808F | CA597808R | 48   | 308              |

Kandilas et al.
Table 5 (cont.)

| Primer | Sequence (5’-3’) | $T_m^\circ C$ | Product size (bp) | Primer | Sequence (5’-3’) | $T_m^\circ C$ | Product size (bp) |
|--------|------------------|---------------|-------------------|--------|------------------|---------------|-------------------|
| CA59891F | TAC TgA TTC Ttg TgT CTT A | 41 | 107 | CA598837F | gAg gA gTg gAg gA TgA gAA gA | 44 | 436 |
| CA59891R | CAC CCT TTA TCT ACT TTT A | 41 | 107 | CA598837R | AAA gGA TTA ggg ATT gAg TA | 44 | 436 |
| CA59884F | CCA gAT TTT CTT CCC CAT | 47 | 300 | CA5988850F | AAC CCC CgC CCT CAC gA gA | 51 | 189 |
| CA59884R | CAg CAC CAg CAg CAg CCC | 47 | 300 | CA5988850R | AAC gAC CCA CAC AAg A | 51 | 189 |
| CA59925F | TgT TCT CAA CTT CCC CTC C | 50 | 343 | CA599262F | CCA ACC CgC TTT CCT AAC ACC CT | 56 | 266 |
| CA59925R | CAA CgT ACT CgC CACC g | 50 | 343 | CA599262R | CgC gCC CgC TAC gCC ACC T | 56 | 266 |
| CA59785F | TTT gAg gC gC CAg gTA | 49 | 258 | CA598020F | gTC ACAA TCA TCT TCT CCA CCA gT | 47 | 185 |
| CA59785R | CTT gTg gAA ggg gCg Tg | 49 | 258 | CA598020R | CCA ACC TgT TTT CCC gT | 47 | 185 |
| CA598130F | ACT TTT TTT gTT gAg ggg AA | 46 | 482 | CA598235F | AAg gAg AAg AAC CCC AAA | 49 | 618 |
| CA598130R | CTC TCC CCC CTT CCC CAg g | 46 | 482 | CA598235R | AAg gAg AAg AAC CCA CCC Tg | 49 | 618 |
| CA598258F | gAg TCC ACC CCC gCC Cg | 57 | 338 | CA598359F | CAg TgC TgA AAT CAT TgT | 44 | 350 |
| CA598258R | CAg TgC TgA AAT CAT TgT | 57 | 338 | CA598359R | CAg TgC TgA AAT CAT TgT | 44 | 350 |
| CA598637F | CAC CTC gTg AgT CCT CgT Cg | 52 | 266 | CA598674F | AAg gAg CAg gAT TTT CCT TgC TgA | 47 | 230 |
| CA598637R | CTT gTT TTT TgT CTT CCC CTC | 52 | 266 | CA598674R | AAg gAg CAg gAT TTT CCT TgC TgA | 47 | 230 |
| CA599135F | AAg gCg AAg AAg CCA gAg TT | 53 | 292 | CA599114F | CgT TgC TCC TgC gCg Tg | 55 | 334 |
| CA599135R | CgT gAg AAC AgC gAA gCg | 53 | 292 | CA599114R | CgT TgC TCC TgC gCg Tg | 55 | 334 |
| CA599099F | CTC gAg gCg TgA gAg gA AAA T | 52 | 397 | CA599049F | ATT CgT CgT CTT CgT TCC | 51 | 278 |
| CA599099R | gAC CCC CcC gTT gAg AAg Cg | 52 | 397 | CA599049R | ATT CgT CgT CTT CgT TCC | 51 | 278 |
| CA599032F | gCC gAT CCA TTC ACC CCC gA | 56 | 375 | CA599013F | TgA ACA AAg gAg ACA Cgg TgT | 45 | 235 |
| CA599032R | gAg TgC TCC CCC CCC CAg Tg | 56 | 375 | CA599013R | TgA ACA AAg gAg ACA Cgg TgT | 45 | 235 |
| CA598962F | CAg gAg CgT gA gAg TCT CgT Cg | 51 | 225 | CA598940F | gAa TTT gTT TgC TCC CAA TgA | 47 | 601 |
| CA598962R | CAg gAg CgT gA gAg TCT CgT Cg | 51 | 225 | CA598940R | gAa TTT gTT TgC TCC CAA TgA | 47 | 601 |
| CA599166F | AAg gCT CCA TgT CTT TgC g | 54 | 311 | CA599172F | gCg TgC ACC CCC gCg TgT | 53 | 359 |
| CA599166R | gAg TCT gTT CTT CgC CAg TgT | 54 | 311 | CA599172R | gCg TgC ACC CCC gCg TgT | 53 | 359 |
| CA597830F | CgT gAg AAg CAg gAA gCg | 54 | 331 | CA597983F | gAa TTT gTT TgC TCC CAA TgA | 52 | 208 |
| CA597830R | gAg ATg TgA gAg CAg gAA gCg | 54 | 331 | CA597983R | gAa TTT gTT TgC TCC CAA TgA | 52 | 208 |
| CA598102F | gAg gAC CCT AAC CAC | 55 | 362 | CA598118F | CAC CCC gCA gAg TCC TgT CTT TTT ATT CTT TCC AgT TgA TgA | 36 | 382 |
| CA598102R | gAg gAC CCT AAC CAC | 55 | 362 | CA598118R | CAC CCC gCA gAg TCC TgT CTT TTT ATT CTT TCC AgT TgA TgA | 36 | 382 |
| CA598187F | CAg TCT TCT CCC CgC CAC | 36 | 450 | CA598128F | gCCT TCT TgA AAT CAC gTgT CTT | 49 | 451 |
| CA598187R | CAg TCT TCT CCC CgC CAC | 36 | 450 | CA598128R | gCCT TCT TgA AAT CAC gTgT CTT | 49 | 451 |
| CA598256F | gAg TGT CTA CgA tCA TgA | 52 | 135 | CA598366F | CCA ACC CCC AAg gAtg TA | 54 | 347 |
| CA598256R | gAg TGT CTA CgA tCA TgA | 52 | 135 | CA598366R | CCA ACC CCC AAg gAtg TA | 54 | 347 |
| CA598422F | CAC gAg TgA gAg gAg CgC | 38 | 356 | CA598476F | ATT TCC CgA gAg TgA gAg gAg | 52 | 160 |
| CA598422R | CAC gAg TgA gAg gAg CgC | 38 | 356 | CA598476R | ATT TCC CgA gAg TgA gAg gAg | 52 | 160 |
| CA598630F | CAA gAg CAA TCC CAC CAA CgC | 52 | 383 | CA598687F | gAg CAA CTT TgAG gAg CgC | 53 | 285 |
| CA598630R | CAA gAg CAA TCC CAC CAA CgC | 52 | 383 | CA598687R | gAg CAA CTT TgAG gAg CgC | 53 | 285 |
| CA598694F | gAg TgC TgA gAg TgC TA | 52 | 352 | CA599121F | AAA CCA CCC TgA AAg ACA CCA CAg | 48 | 370 |
| CA598694R | gAg TgC TgA gAg TgC TA | 52 | 352 | CA599121R | AAA CCA CCC TgA AAg ACA CCA CAg | 48 | 370 |
summarizes the mean genetic distance and genetic identity between the cultivars as determined by MVSP 3.1. Pair-wise within-group distances ranged from 0 to 0.725, with the highest similarity (0.725) occurring between Harmankaya99 and Sönmez2001 and the lowest (0.622) between Aytin98 and Izgi01.

Figure 6 shows the dendrogram based on the similarity index (Jaccard’s coefficient) of the six cultivars. Two main clusters were observed, the first of which included cultivars Aytin98 and ES14 while the second was divided into two subclusters, the first of which comprised PI178383 while the second contained Izgi01, Sönmez2001 and Harmankaya99. The latter subcluster consisted a group containing Izgi01 and another containing Sönmez2001 and Harmankaya99. The construction of this dendrogram dem-
Table 6 - Similarity index (Jaccard's coefficient) between *Triticum aestivum* cultivars.

| Population ID | PI178383 | Izgi01 | Sönmez2001 | Harmankaya99 | ES14 | Aytin98 |
|---------------|----------|-------|-------------|--------------|------|---------|
| PI178383      | 1.000    |       |             |              |      |         |
| Izgi01        | 0.680*   | 1.000 |             |              |      |         |
| Sönmez2001    | 0.656*   | 0.692* | 1.000       |              |      |         |
| Harmankaya99  | 0.692*   | 0.680* | 0.725*      | 1.000        |      |         |
| ES14          | 0.682*   | 0.655* | 0.686*      | 0.712*       | 1.000|         |
| Aytin98       | 0.655*   | 0.622* | 0.628*      | 0.655*       | 0.703*| 1.000  |

*Genetically similar.

The low level of genetic diversity expected between self-pollinating plants means that EST databases can be useful tools for genetic studies in wheat and related species. Our results indicate that EST-derived primers were good tools for assessing the genetic diversity in wheat cultivars. A relatively high level of polymorphism (58.61% of loci were polymorphic) was observed with 39 contig and 92 singleton primers across the six wheat genotypes, despite the fact that all of them were local cultivars from geographically close locations. Several other studies have reported polymorphism in self-pollinating plants, including tef (4%) (Bai et al., 1999), azuki (18%) (Yee et al., 1999), rice (22%) (Maheswaran et al., 1997), sugar beet (50%) (Schondelmaier et al., 1996) and wild barley (76%) (Pakniyat et al., 1997). In a work similar to that reported here, Wei et al. (2005) used microsatellite markers to assess the polymorphic divergence in wheat landraces highly resistant to *Fusarium* head blight (FHB). The level of polymorphism observed among 20 wheat landraces resistant to FHB and four wheat landraces susceptible to FHB was 97.5% with a mean genetic similarity index among the 24 genotypes of 0.419 (range: 0.103 to 0.673).

In conclusion, we have used an EST database to examine the genetic diversity among Turkish wheat cultivars resistant and susceptible to yellow rust disease. Our results indicate that EST databases can be used to assess genetic diversity and identify suitable parents in populational studies designed to detect genes related to disease resistance.

Acknowledgments

The authors thank Dr. Selma Onarici for her helpful comments, Dr. Necmettin Bolat for providing plant material and Central Research Institute of Field Crops for performing pathogenity tests. This study is a part of a PhD thesis (“Investigation of yellow rust disease resistance in winter-type bread wheat (*Triticum aestivum* L.) using biotechnological methods”) by Ozge Karakas done at the Institute of Sciences and Research Foundation (project no. 1832), Istanbul University. This work was supported by TUBITAK KAMAG (project no. 105G075).

References

Akfirat-Senturk F, Aydn Y, Ertugrul F, Hasancebi S, Budak H, Akan K, Mert Z, Bolat N and Uncuoglu-Altnkut A (2010) A microsatellite marker for yellow rust resistance in wheat. Cereal Research Communications 38:203-210.

Alnta S, Toklu F, Kafkas S, Kilian B, Brandolini A and Ozkan H (2008) Estimating genetic diversity in drum and bread wheat cultivars from Turkey using AFLP and SAMPL markers. Plant Breed 127:9-14.

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

Asf M, Rahman MU and Zafar Y (2005) DNA fingerprinting studies of some wheat (*Triticum aestivum* L.) genotypes using Random Amplified Polymorphic DNA (RAPD) analysis. Pak J Bot 37:271-277.

Bai G, Tefera H, Ayele M and Nguyen HT (1999) A genetic linkage map of tef [Eragrostis tef (Zucc.) Trotter] based on amplified fragment length polymorphism. Theor Appl Genet 99:599-604.

Bennett MD and Leitch IJ (1995) Nuclear DNA amounts in angiosperms. Ann Bot 76:113-176.

Chen XM (2005) Epidemiology and control of stripe rust (*Puccinia striiformis* f. sp. *tritici*) on wheat. Can J Plant Pathol 27:314-337.

Ercan S, Ertugrul F, Aydin Y, Akfirat-Senturk F, Hasancebi S, Akan K, Mert Z, Bolat N, Yorgancilar O and Uncuoglu-Altnkut AA (2010) An EST-SSR marker linked with yellow rust resistance in wheat. Biol Plant (in press).
Gülbitti-Onarici S, Sümer S and Özcan S (2007) Determination of phylogenetic relationships between some wild wheat species using amplified fragment length polymorphism (AFLP) markers. Bot J Linn Soc 153:67-72.

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223-270.

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

Kovach WL (1999) MVS P – A Multivariate Statistical Package for Windows, v. 3.1. Kovach Computing Services, Pentraeth, 133 pp.

Leigh F, Lea V, Law J, Wolters P, Powell W and Donini P (2003) Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. Euphytica 133:359-366.

Li S, Jia J, Wei X, Zhang X, Li L, Chen H, Fan Y, Sun H, Zhao X, Lei T et al. (2007) An intervarietal genetic map and QTL analysis for yield traits in wheat. Mol Breed 20:167-178.

Lu G and Moriyama NE (2004) Vector NTI, a balanced all-in-one sequence analysis suite. Brief Bioinform 5:378-388.

Maheswaran M, Subudhi PK, Nandi S, Xu JC, Parco A, Yang DC and Huang N (1997) Polymorphism, distribution and segregation of AFLP markers in doubled haploid rice population. Theor Appl Genet 94:39-45.

McNeal FM, Conzak CF, Smith EP, Tade WS and Russell TS (1971) A uniform system for recording and processing cereal research data. US Agricultural Research Service 42:34-121.

Nagaoka T and Ogihara Y (1997) Applicability of inter-simple sequence repeat polymorphisms in wheat for use as DNA markers in comparison to RFLP and RAPD markers. Theor Appl Genet 94:597-602.

Pakniyat H, Powell W, Baird E, Handley LL, Robinson D, Scrimgeour CM, Nevo E, Hackett CA, Caligari PDS and Forster BP (1997) AFLP variation in wild barley (Hordeum spontaneum C. Kock) with reference to salt tolerance and associated ecogeography. Genome 40:332-341.

Parkinson J, Guiliano DB and Blaxter M (2002) Making sense of EST sequences by CLOBBing them. BMC Bioinformatics 3:31.

Pashley CH, Ellis JR, McCauley DE and Burke JM (2006) EST databases as a source for molecular markers: Lessons from Helianthus. J Hered 97:381-388.

Plaschke J, Ganal MW and Röder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91:1001-1007.

Rudd S (2003) Expressed sequence tags: Alternative or complement to whole genome sequences? Trends Plant Sci 8:321-328.

Schonelmaier J, Streinrucken G and Jung C (1996) Integration of AFLP markers into a linkage map of sugar beet (Beta vulgaris L.). Plant Breed 115:231-237.

Song W and Henry RJ (1995) Molecular analysis of the DNA polymorphism of wild barley (Hordeum spontaneum) germplasm using the polymerase chain reaction. Genet Resour Crop Evol 42:273-280.

Torre J, Egan MG, Katari MS, Brenner ED, Stevenson DW, Coruzzi GM and DeSalle R (2006) ESTimating plant phylogeny: Lessons from partitioning. BMC Evol Biol 6:48.

Wei YM, Hou YC, Yan ZH, Wu W, Zhang ZQ and Liu DC (2005) Microsatellite DNA polymorphism divergence in Chinese wheat (Triticum aestivum L.) landraces highly resistant to Fusarium head blight. J Appl Genet 46:3-9.

Weining S and Langridge P (1991) Identification and mapping of polymorphisms in cereals based on the polymerase chain reaction. Theor Appl Genet 82:209-216.

Yee E, Kidwell KK, Sills GR and Lumpkin TA (1999) Diversity among selected Vigna angularis (Azuki) accessions on the basis of RAPD and AFLP markers. Crop Sci 39:268-275.

Yu JK, Dake TM, Singh S, Benschcer D, Li W, Gill B and Sorrels ME (2004) Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome 47:805-818.

Zeybek A and Yigit F (2004) Determination of virulence genes frequencies in wheat stripe rust (Puccinia striliformis f. sp. tritici) populations during natural epidemics in the regions of southern Aegean and western Mediterranean in Turkey. Pak J Biol Sci 7:1967-1971.

Zhang W, Qu LJ, Gu H, Gao W and Liu M (2002) Studies on the origin and evolution of tetraploid wheats based on the internal transcribed spacer (ITS) sequences of nuclear ribosomal DNA. Theor Appl Genet 104:1099-1106.

Internet Resources

GrainGenes. http://wheat.pw.usda.gov/cgi-bin/westsql/est_lib.cgi (August 15, 2007).

VecScreen database. http://www.ncbi.nlm.nih.gov/VecScreen/VectScreen.html (September 20, 2007).

Associate Editor: Luciano da Fontoura Costa

License Information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.