

Large scale analysis of expressed genes in common wheat

Ogihara Y¹, Kawaura K¹, Mochida K²

¹Kihara Institute for Biological Research, Yokohama City University, Japan, ²RIKEN Plant Science Center, Japan

ABSTRACT

Because of wheat's huge genome size and polyploidy, we have concentrated on conducting large-scale analysis of expressed sequence tags (ESTs) in common wheat. Up to now, 46 cDNA libraries derived from tissues during the wheat life cycle as well as stress-treated tissues were constructed. Several thousands of colonies were randomly selected from each of these 46 cDNA libraries and sequenced from both ends. Sequence data of 628,226 ESTs are now available. These ESTs were grouped into about 90 thousand homoeologs and 36 thousands gene clusters with CAP3 and BLAST methods. These contigs were estimated to cover more than 90 % of the expressed wheat genes. Homology of these contigs was searched against the rice genome data base. About 30 % of those genes were plant specific. By computing abundantly expressed ESTs, correlated expression patterns of genes across the tissues (Virtual Display: VD) were monitored. Furthermore, the relationships between gene expression profiles among the stress-treated tissues were inferred using VD, and genes specifically induced and/or suppressed by stresses were able to be selected from the VD. These genes were annotated with BLAST searches. *In silico* selection of

In addition to EST data, we are also accumulating full length cDNA data of Chinese Spring wheat. We extracted total RNA from 17 tissues during the life cycle and/or stress-treatments, and constructed a full-length cDNA library from these pooled RNAs with the CAP-trapper method. We randomly selected about 20,000 clones from the library, and sequenced from both ends. These sequences were classified into 7,149 gene clusters. Finally, the complete sequences of the 6,162 independent clones were determined.

INTRODUCTION

Wheat is characterised by its polyploidy nature. Common wheat is allohexaploid and has three homoeologous genomes, designated as A, B and D. The huge genome size of the hexaploid (16 Gbp¹) along with the high content of repeat sequences makes it difficult to carry out the entire sequencing of the whole genome, although an effort to determine the DNA sequences of homoeologous chromosome 3 has been started on an international level². Taking these situations into accounts, we have concentrated on collecting expressed genes in common wheat for several years. At present, cDNA libraries from 46 tissues during the wheat life cycle and stress-treated tissues have been constructed. By accumulating comprehensive EST data, (1) we could classify these ESTs into contigs which are corresponding to each of their homoeologs, and group those contigs into gene clusters³. (2) Their gene expression patterns can be monitored *in silico* throughout the life cycle and stress-treatments⁴. (3) Homoeogenes whose expression patterns were separately traced, can be assigned to their homoeologous chromosomes⁵. (4) A DNA microarray harboring 38k gene probes is now available⁶. (5) Several thousands of full-length cDNA sequence data are also available⁷. These are important genome resources for wheat, and can make a great contribution to genome breeding of wheat.

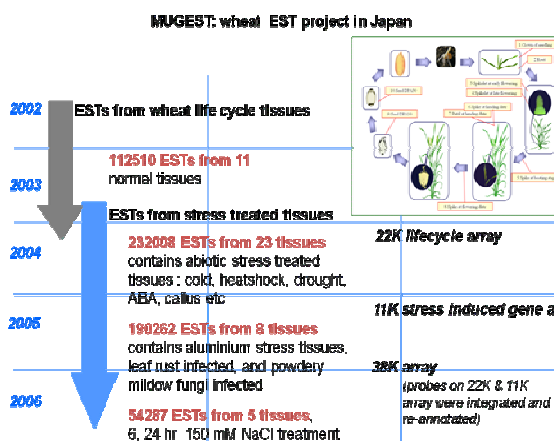


Fig. 1. Wheat EST project in Japan.

RNAs from various tissues during the wheat life cycle and stress-treated tissues were supplied for cDNA construction. One pass-sequencing of several thousands cDNA clones from both ends was carried out to supply EST data base.

screened genes from VD should provide a powerful tool for functional genomics of cereals.

LARGE SCALE ANALYSES OF ESTs IN COMMON WHEAT

Wheat EST project in Japan

We had started the wheat EST project in Japan in 2002, as shown in Fig. 1. At first, we extracted RNAs from 11 tissues during the life cycle of Chinese Spring wheat. Using those RNAs, cDNA libraries were constructed. Several thousands of cDNA clones were selected from

each library, and sequenced from both ends of the inserts. 112,510 ESTs were obtained. Then, ESTs were collected from the stress-treated tissues along with successive EST collections from the tissues during the life cycle. Biotic and abiotic stresses were induced to get stress-response genes in common wheat. Consequently, 628,226 ESTs of common wheat were pooled in the genome resource of Japan (Fig. 1). These ESTs were grouped into 89,658 contigs with CAP3 and BLAST methods, and finally into 35,604 gene clusters with BLAST method ($E < 1e-120$). The gene number in wheat was estimated to cover more than 90%, since the gene number in rice is predicted to be ca. 38,000.

Tissue expression map

In the EST contigs, those abundantly expressed were selected to construct a tissue expression map. Correlated expression patterns of genes across the 46 tissues/stress treatments were monitored by computing abundantly expressed ESTs (designated as Virtual Display: VD). Furthermore, the relationships between gene expression profiles among the stress-induced tissues were able to be inferred from the gene expression patterns. Multi-dimensional analysis of EST data is analogous to microarray experiments. For example, genes specifically induced and/or suppressed by cold acclimation and heat-shock treatments were selected *in silico*. Selected genes were annotated with a BLAST search. Subsequently, gene ontology was conducted for selected genes with the InterPro search. Because genes regulated in response to temperature treatment were successfully selected, this method can be applied to other stress-treated tissues. Then, the methods were applied to screen genes in response to abiotic as well as biotic stresses. *In silico* selection of screened genes from virtual display should provide a powerful tool for functional wheat genomics.

Discrimination of homoeologous gene expression in common wheat

Computer analysis of a large number of ESTs from common wheat enabled us to assign each homoeolog expressed from the three genomes to its respective chromosome, and trace expression pattern during the life cycle and in stress-treated tissues⁵. Relatively abundant ESTs were grouped into three categories. (1) all genes were expressed from three genomes. (2) two of three genes were expressed, and the remaining one was completely shut down during the life cycle. (3) only one gene was expressed. These ESTs were assigned to each of the homoeologous chromosomes using a nullisomic/tetrasomic series of Chinese Spring wheat in combination with pyrosequencing. At present, about 150 genes from the three categories have been analysed. No preference for gene-silencing in particular genomes or chromosome was observed, suggesting that gene-silencing occurred after polyploidization, and at the gene level, not at the chromosome or genome level.

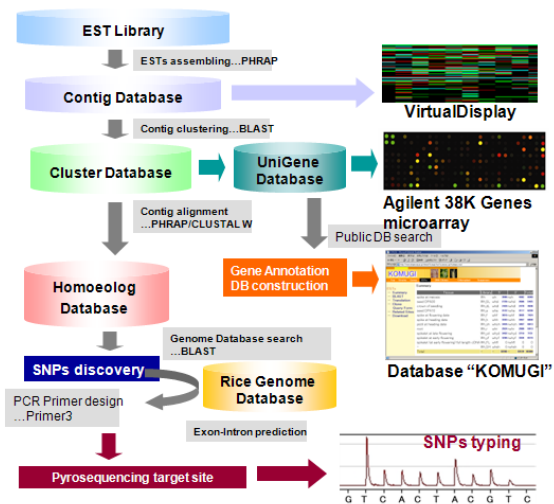


Fig. 2. Wheat ESTs clustering pipe line

Construction of DNA microarray and its application to screen salt-responsive genes

We constructed three kinds of oligo DNA microarray in collaboration with Agilent Co. Ltd. At first, we constructed a 22K wheat oligo-DNA microarray. A total of 148,676 ESTs of common wheat were collected in addition to our sequenced ones. These were grouped into 34,064 contigs, which were then used to design an oligonucleotide DNA microarray. Following a multistep selection of the sense strand, 21,939 60-mer oligo-DNA probes were selected for attachment on the microarray slide⁸. Secondly, we made an 11K oligo DNA microarray⁶ which harbors 10,263 gene probes independent of those of the 22K oligo DNA microarray⁸. Probes of the 11K microarray were designed from 232,008 wheat ESTs obtained from 23 wheat cDNA libraries⁴. These ESTs were grouped into gene clusters with the PHRAP and BLAST methods. The resultant gene clusters were compared with those of the 22K microarray by the BLAST to exclude redundant gene clusters. Steps for construction of oligo DNA microarray followed those of the 22K microarray. Finally, we constructed a 38K oligo DNA microarray by summation of gene probes on the 22K and the 11K microarrays. In the 38K microarray, actually 37,826 genes probes were attached.

These oligo DNA microarrays were supplied for global expression analysis of salt-responsive genes in wheat^{6,8}. Two weeks old seedling of wheat were treated with 150 mM NaCl for 1, 6, and 24 h, and their roots and shoots were subjected to analyses. Several thousand genes showed changes in expression of more than twofold and were classified into 12 groups by their expression patterns. Many of the genes with unknown function were clustered together with genes known to be involved in response to salt stress. Thus, analysis of gene expression patterns combined with gene ontology should help identify the function of the unknown genes. Also, functional analysis of these wheat genes should provide new insight into the response to salt stress.

These activities of wheat genome consortium of Japan is summarized in Fig. 2.

FULL LENGTH cDNA SEQUENCE DATA OF CHINESE SPRING WHEAT

We used the CAP-trapper method⁹ to construct the full length cDNA library from the pooled RNAs that were derived from the 17 tissues during the wheat life cycle and the abiotic stressed tissues of Chinese Spring wheat. There are 19,968 randomly selected recombinant clones from the full length cDNA library that have been sequenced from both ends. By grouping these one-path sequences, it was possible to obtain 7,149 independent gene clusters into which homoeogenes expressed from each of these three genomes of common wheat could be grouped. The largest clones, which were representative of each gene cluster, were rearranged, leading to the determination of the full length sequences of 6,162 clones. A detailed description of the sequence data for the full length cDNA clones and size distribution of the inserted DNAs are shown in Fig. 3. The data for the wheat full length cDNAs was used to carry out a blast homology search against the public DNA database. About ten percent of the clones were determined to be wheat-specific genes without any counterparts in the DNA database. These sequence data of wheat full length cDNAs should be an important genome resource of wheat and can be supplied for annotation of the wheat genome.

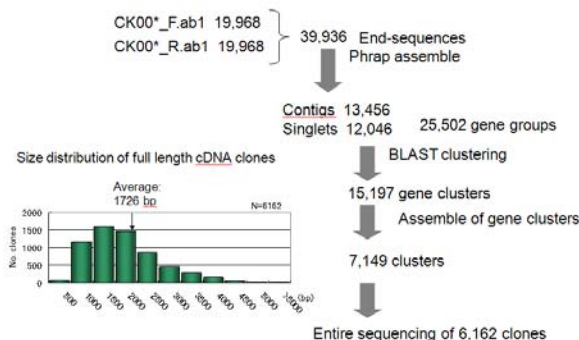


Fig. 3 Procedure for selection and determination of cDNA clones from the full-length cDNA library

ACKNOWLEDGEMENTS

The work was supported by Grants-in-Aid for Scientific Research on Priority Area "Comparative Genomics" and National BioResource Project from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

REFERENCES

1. Bennet MD, Leitch IJ (1995) Nuclear DNA amounts in angiosperms. *Ann Bot* 76: 113-176
2. Gill BS, et al. (2004) A workshop report on wheat genome sequencing: international genome research on wheat consortium. *Genetics* 168: 1087-1096
3. Ogihara Y, et al. (2003) Correlated clustering and virtual display of gene expression patterns in the wheat life cycle by large-scale statistical analyses of expressed sequence tags. *Plant J* 33: 1001-1011
4. Mochida K, et al. (2006) Tissue expression map of comprehensive expressed sequence tags and its application to *in silico* screening of stress response genes in common wheat. *Mol. Genet. Genomics* 276: 304-312
5. Mochida K, et al. (2003) Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol Genet. Genom.* 270: 371-377
6. Kawaura K, et al. (2008) Genome-wide analysis for identification of salt-responsive genes in common wheat. *Funct Integr Genomics* [Epub ahead of print] PMID 18326864
7. Kawaura K, et al. (2008) Assessment of adaptive evolution between wheat and rice as deduced from the full-length cDNA sequence data of common wheat and their tissue expression map. Submitted
8. Kawaura K, et al. (2006) Transcriptome analysis of salinity stress responses in common wheat using a 22K oligo-DNA microarray. *Funct Integr Genomics* 6: 132-142
9. Ogihara Y, et al. (2004) Construction of a full-length cDNA library from young spikelets of hexaploid wheat and its characterization by large-scale sequencing of expressed sequence tags. *Genes Genet. Syst.* 79: 227-232