BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 11(1)

Analysis of expressed sequence tags from cDNA library of *Fusarium culmorum* infected barley (*Hordeum vulgare* L.) roots

Feyza Tufan[1], Cüneyt Uçarlı[1] & Filiz Gürel[1, 2*]

[1]Department of Molecular Biology and Genetics, Faculty of Science, Istanbul University, 34134 Vezneciler, Istanbul, Turkey; [2]Istanbul University, Research and Application Center for Biotechnology and Genetic Engineering, 34134 Vezneciler, Istanbul, Turkey; Filiz Gürel – Email: filiz@istanbul.edu.tr; *Corresponding author

Received January 09, 2015; Accepted January 10, 2015; Published January 30, 2015

Abstract:
Fusarium culmorum is one of the most common and globally important causal agent of root and crown rot diseases of cereals. These diseases cause grain yield loss and reduced grain quality in barley. In this study, we have analyzed an expressed sequence tag (EST) database derived from *F. culmorum* infected barley root tissues available at the National Center for Biotechnology Information (NCBI). The 2294 sequences were assembled into 1619 non-redundant sequences consisting of 359 contigs and 1260 singletons using the program CAP3. BLASTX analysis for these sequences was conducted in order to find similar sequences in all databases. Gene Ontology search, enzyme search, KEGG mapping and InterProScan search were done using Blast2GO 3.0.7 tool. By BLASTX analysis, 41.7%, 7.7%, 3.2% and 47.4% of ESTs were categorized as annotated, unannotated, not mapping and without blast hits, respectively. BLASTX analysis revealed that the majority of top hits were barley proteins (43.5%). Based on Gene Ontology classification, 38.3%, 31.3%, and 16% of ESTs were assigned to molecular function, biological process, and cellular component GO terms, respectively. Most abundant GO terms were as follows: 157 sequences were related to response to stress (biological process), 207 sequences were related to ion binding (molecular function), and 160 sequences were related to plastid (cellular component). Furthermore, based on KEGG mapping, 369 sequences could be assigned to 264 enzymes and 83 different KEGG pathways. According to Enzyme Commission (EC) distribution; 94 sequences were transferases (EC2) while 70 sequences were hydrolases (EC3).

Keywords: barley, ESTs, root rot and crown rot diseases, *Fusarium culmorum*

Background:
One of the most common and globally important problems of cereal breeding is *Fusarium* root rot (FRR) and *Fusarium* crown rot (FCR) diseases **[1-3].** In the development of these diseases, a fungi complex consisting of *Fusarium culmorum*, *Fusarium pseudograminearum* (Syn: *Fusarium graminearum*), *Microdochium nivale* (Syn: *Fusarium nivale*), *Fusarium avenaceum*, *Fusarium acuminatum*, *Bipolaris sorokiniana*, *Gaeumannomyces graminis* and other species plays a major role **[4]**. The surveys carried out in Turkey showed that *F. culmorum* was the most abundant among these fungi **[5]**. In order to reduce the use of pesticide in farming and struggling with the disease, it is crucial to investigate the mechanisms of natural resistance **[6]**. Investigation of resistance at the molecular level involves mapping of quantitative trait loci **[7]** and identification of resistance genes **[8]**. On the other hand, induction of genes from several pathways after pathogen infection strongly indicates their relation to the plant defense system **[9].**

Expressed Sequence Tags (ESTs) database is a collection of short-single sequences of cDNA copies of mRNA that are expressed under different conditions. ESTs represent part of the

transcribed portion of the genome **[10]**. ESTs are a robust resource (because of the relative simplicity and low expense for their production) in structural and functional genomics. ESTs can be used for gene discovery, genome annotations and comparative genomics. Various EST sequencing projects have been done to understand the transcriptome of genes associated with biotic stress **[11-13].**

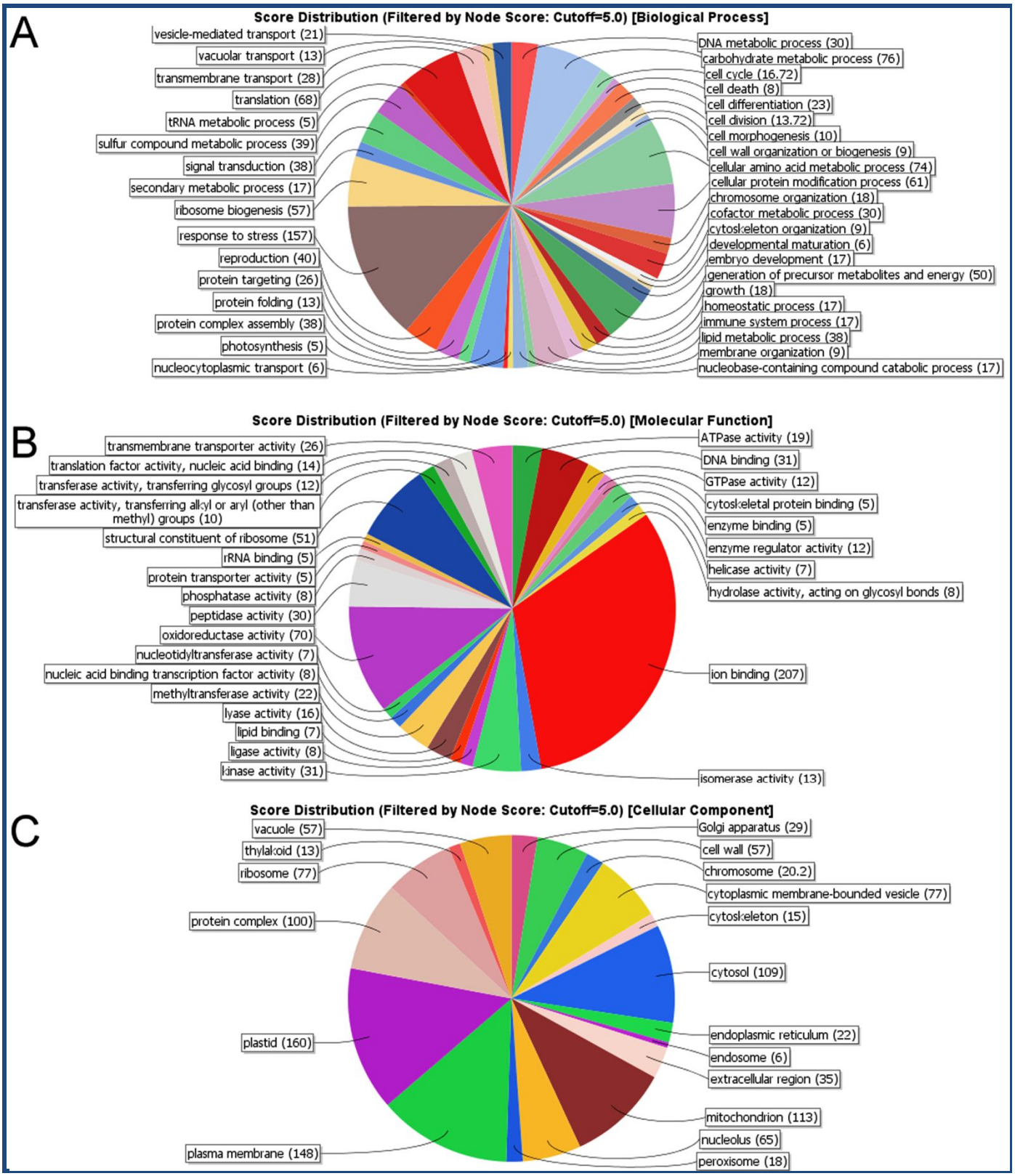

Figure 1: Gene Ontology annotation: Sequence distribution of ESTs regarding: **A)** Biological process; **B)** Molecular function and **C)** Cellular component.

Gene Ontology provides a structured and controlled vocabulary to describe gene products according to three ontologies, namely biological process, molecular function, and cellular component **[14]**. Enzyme commission (EC) numbers are hierarchical classification schemes for enzymes based on the reaction catalyzed. Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database resource that integrates genomic, chemical, and systemic functional information and has been widely used for pathway mapping **[15, 16]**. In this study, we analyzed two libraries of ESTs derived from barley roots infected with *Fusarium culmorum* KF 350 **[17]**. These libraries were established with Nickel and Carola cultivars grown for 7 days, then their roots were harvested 6, 24, 48, 72, and 96 hours post inoculation (HPI). Results may help to identify key factors in stress response in barley after infection by *F. culmorum* and provide additional information regarding their roles under biotic stress.

Methodology:

EST Source

A total of 2306 ESTs from two EST libraries LIBEST_016901 and LIBEST_016904 derived from *F. culmorum* KF 350 infected *Hordeum vulgare* cv. Nickel and Carola root tissues from GeneBank were used for bioinformatic analysis. Infection process was performed on 7 day-old seedlings and root samples were harvested in 6, 24, 48, 72, and 96 HPI. ESTs were downloaded in FASTA format for further analysis.

EST Processing

VecScreen tool (http: // www.ncbi.nlm.nih.gov/tools /vecscreen /) was used to find regions derived from vectors and vector contaminations were manually removed. The clean sequences were assembled into contigs and singletons with the CAP3 program **[18]**.

Functional Annotation of ESTs

Functional annotations were performed in 3 steps by using Blast2GO 3.0.7 tool **[19]**. Firstly, contigs and singletons were used as query in BLASTX searches against non-redundant protein database (Blast DB: nr, Number of Blast Hits: 20, Blast Expectation Value (E value): 1.0E-3) to find homologous sequences **[20]**. For further analysis of ESTs, mapping was used to retrieve GO terms as associated with BLASTX hits and finally annotation (E-Value Hit Filter:1.0E-6, Annotation CutOff:55, GO Weight:5, Hsp-Hit Coverage CutOff:0) was used to associate with queries reliable information from GO, Enzyme Codes, InterProScan, and KEGG databases.

Results & Discussion:

Analysis of EST libraries

A total of 2306 ESTs, 1168 and 1138 from LIBEST_016901 and LIBEST_016904 were downloaded in FASTA format. VecScreen tool was used to remove contaminant sequences and 12 sequences were manually removed. The trimmed 2294 ESTs were assembled into clusters by means of CAP3 program. 1034 ESTs were grouped into 359 contigs and 1260 ESTs represented singletons.

Statistics of BLASTX results

In order to assign a putative function, Blast2GO 3.0.7 tool was used. All of the contigs and singletons were subjected to BLASTX analysis for homology search. Afterwards, mapping and annotation were performed. Data distribution of 1619 ESTs were as follows: 52 sequences with blast hits, 125 sequences with mapping and 674 sequences with GO-Slim annotation. 768 sequences were without blast hits and removed because of lack of annotation. According to Top Hit species distribution of BLAST Top-Hits, the majority of top matches were *H. vulgare* 368 (43.5%) followed by *Aegilops tauschii* (121), *Triticum urartu* (96), *Brachypodium distachyon* (44), *Zea mays* (44), *Triticum aestivum* (38), and *Oryza sativa* (29).

Enzyme commision numbers are a hierarchical classification scheme for enzymes based on the reaction catalyzed. According to EC classification, 94 sequences were transferases, 70 sequences were hydrolases, 68 sequences were oxidoreductases, 15 sequences were lyases, 13 sequences were isomerases, and 2 sequences were ligases. KEGG is a database resource that integrates genomics, chemical and systemic functional information and has been widely used for pathway mapping. According to KEGG analysis, 369 sequences were belonging to 264 enzymes and mapped to 83 different KEGG pathways **Table 1 (see supplementary material).** KEGG pathways included cysteine and methionine metabolism (12 enzymes), amino sugar and nucleotide sugar metabolism (12 enzymes), and starch and sucrose metabolism (12 enzymes) and carbon fixation in photosynthetic organisms (7 enzymes). InterProScan tool was used to predict conserved domains in corresponding translated protein sequence. InterProScan of ESTs resulted in 1028 sequence of no InterPro, 591 sequences of InterPro and 223 sequences of GOs.

According to GO Level distribution of ESTs, GO levels varied from 1 to 11 and total number of annotations were 4448 (Term Filter Mode: by Sequence Count, Term Filter Value:5). Based on the gene ontology category in biological process, 157 sequences were related to response to stress (GO: 0006950) while 76 sequences were associated with carbohydrate metabolism (GO:0005975), and 74 sequences were related to cellular amino acid metabolism (GO:0006520). In molecular function, 207 sequences were related to ion binding (GO: 0043167) and 70 sequences were related to oxidoreductase activity (GO: 0016491). With respect to cellular component ontology, the subcellular localization of 160 sequences was plastid (GO: 0009536), for 148 sequences was plasma membrane (GO: 0005886), and for 113 sequences was mitochondrion (GO: 0005739). A Multi-Level Pie Chart representation was used to display GO graphs **(Figure 1).**

Conclusion:

Root and crown rot disease is one of the most serious problems for cereal breeding. Despite this fact, very little information is available regarding the molecular mechanisms of such diseases. In this study, we performed EST analysis in order to investigate transcriptome during the early stages of colonization of barley roots by *F. culmorum*. After evaluation of 2294 ESTs in a *F.culmorum*-infected barley cDNA library, 157 response to stress-related genes were identified. 264 enzymes were mapped to 83 different KEGG pathways. These results allow to identify a pool of stress and/or defense-related candidate genes. Primer design against those genes will allow to assess comparatively expression patterns of those genes in *F. culmorum*-resistant and *F. culmorum*-susceptible barley cultivars.

Acknowledgement:
This work was supported by the Scientific Research Projects Coordination Unit of Istanbul University, Project No. 27149. We thank to Dr. Paolo Bagnaresi for his valuable recommendations.

References:

[1] Beccari G *et al. Plant Pathol*. 2011 **60:** 671
[2] Paulitz TC *et al. Can J Plant Pathol*. 2002 **24**: 416
[3] Scherm B *et al. Mol Plant Pathol*. 2013 **14:** 323 [PMID: 23279114]
[4] Moya-Elizondo EA *et al. Plant Dis*. 2011 **95**: 1099
[5] Bentley AR *et al. Sydowia* 2006 **58**: 163
[6] Boutigny L *et al. Eur J Plant Pathol*. 2008 **121**: 411
[7] Chen GD *et al. Theor Appl Genet*. 2013 **126**: 2511 [PMID: 23824201]
[8] Li HB *et al. Theor Appl Genet*. 2010 **121**: 941 [PMID: 20535443]
[9] Desmond OJ *et al. Physiol Mol Plant Pathol*. 2005 **67**: 171
[10] Wolfsberg TG & Landsman D, *Methods Biochem Anal*. 2001 **43**: 283 [PMID: 11449729]
[11] Naganeeswaran SA *et al. Bioinformation* 2012 **8**: 65 [PMID: 22359437]
[12] Feng J *et al. Ann Appl Biol*. 2012 **161:** 214
[13] Karakaş Ö *et al. Genet Mol Biol*. 2010 **33**: 719 [PMID: 21637582]
[14] Ashburner M *et al. Nat Genet*. 2000 **25**: 25 [PMID: 10802651]
[15] Kanehisa M & Goto S, *Nucleic Acids Res*. 2000 **28**: 27 [PMID: 10592173]
[16] Kanehisa M *et al. Nucleic Acids Res*. 2014 **42**: D199 [PMID: 24214961]
[17] Eichmann R *et al. J Plant Physiol*. 2006 **163**: 657 [PMID: 16545999]
[18] Huang X & Madan A, *Genome Res*. 1999 **9:** 868 [PMID: 10508846]
[19] http://www.blast2go.com/
[20] Altschul SF *et al. J Mol Biol*. 1990 **215:** 403 [PMID: 2231712]

Edited by P Kangueane
Citation: **Tufan *et al.*** Bioinformation 11(1): 034-038 (2015)

BIOINFORMATION

open access

Supplementary material:

Table 1: Number of sequences and enzymes involved into reconstruction of KEGG pathway.

No	Pathways	Seqs	Enzs	No	Pathways	Seqs	Enzs
1	Cysteine and methionine metabolism	16	12	43	β-Alanine metabolism	3	3
2	Carbon fixation in photosynthetic organisms	15	7	44	Betalain biosynthesis	3	1
3	Amino sugar and nucleotide sugar metabolism	14	12	45	Lysine degradation	3	3
4	Starch and sucrose metabolism	14	12	46	Tropane, piperidine and pyridine alkaloid biosynthesis	3	4
5	Glutathione metabolism	13	6	47	Fatty acid degradation	3	3
6	Phenylpropanoid biosynthesis	13	6	48	Glycerolipid metabolism	3	2
7	Purine metabolism	12	8	49	Novobiocin biosynthesis	3	4
8	Phenylalanine metabolism	12	8	50	Cyanoamino acid metabolism	3	4
9	Pyruvate metabolism	11	9	51	Isoquinoline alkaloid biosynthesis	3	3
10	Glycolysis / Gluconeogenesis	11	8	52	Lysine biosynthesis	3	1
11	Oxidative phosphorylation	9	3	53	Alanine, aspartate and glutamate metabolism	3	1
12	Sulfur metabolism	8	5	54	Fructose and mannose metabolism	3	3
13	Phenylalanine, tyrosine and tryptophan biosynthesis	7	8	55	Aminoacyl-tRNA biosynthesis	2	2
14	Drug metabolism - cytochrome P450	6	1	56	Tryptophan metabolism	2	2
15	Metabolism of xenobiotics by cytochrome P450	6	1	57	Geraniol degradation	2	2
16	Carbon fixation pathways in prokaryotes	6	3	58	Pantothenate and CoA biosynthesis	2	2
17	Methane metabolism	6	4	59	Other glycan degradation	2	2
18	Selenocompound metabolism	6	5	60	Limonene and pinene degradation	2	2
19	Tyrosine metabolism	6	5	61	Glycosphingolipid biosynthesis - ganglio series	2	1
20	Ascorbate and aldarate metabolism	6	4	62	Benzoate degradation	2	2
21	Galactose metabolism	6	5	63	Butanoate metabolism	2	2
22	Pyrimidine metabolism	6	3	64	Riboflavin metabolism	2	1
23	alpha-Linolenic acid metabolism	5	4	65	Propanoate metabolism	2	2
24	Arginine and proline metabolism	5	3	66	Glycosaminoglycan degradation	2	1
25	Pentose phosphate pathway	5	5	67	Linoleic acid metabolism	1	1
26	One carbon pool by folate	5	3	68	Steroid degradation	1	1
27	Glycerophospholipid metabolism	5	4	69	Drug metabolism - other enzymes	1	1
28	Glyoxylate and dicarboxylate metabolism	5	3	70	Chloroalkane and chloroalkene degradation	1	1
29	Pentose and glucuronate interconversions	5	4	71	Photosynthesis	1	1
30	Aminobenzoate degradation	4	3	72	Streptomycin biosynthesis	1	1
31	Valine, leucine and isoleucine degradation	4	4	73	Nitrogen metabolism	1	1
32	Citrate cycle (TCA cycle)	4	2	74	Glucosinolate biosynthesis	1	1
33	Sphingolipid metabolism	4	2	75	C5-Branched dibasic acid metabolism	1	1
34	Ether lipid metabolism	4	3	76	Vitamin B6 metabolism	1	1
35	Glycine, serine and threonine metabolism	4	3	77	Steroid biosynthesis	1	1
36	Fatty acid elongation	4	3	78	Fatty acid biosynthesis	1	1
37	Histidine metabolism	4	2	79	Ethylbenzene degradation	1	1
38	Steroid hormone biosynthesis	4	2	80	Biosynthesis of ansamycins	1	1
39	Biosynthesis of unsaturated fatty acids	3	3	81	Caprolactam degradation	1	1
40	Arachidonic acid metabolism	3	2	82	mTOR signaling pathway	1	1
41	Valine, leudne and isoleudne biosynthesis	3	3	83	Thiamine metabolism	1	1
42	Ubiquinone and other terpenoid-quinone biosynthesis	3	1				