DISCOVERY, EXPRESSION PROFILING, AND EVOLUTIONARY ANALYSIS OF *CYNODON* EXPRESSED SEQUENCE TAGS

by

CHANGSOO KIM

(Under the Direction of Andrew H. Paterson)

ABSTRACT

Bermudagrass (*Cynodon dactylon*) is a major turfgrass species for sports fields, lawns, parks, golf courses, and general utility turfs in tropical and subtropical regions. Despite its ecological importance, much of its study has been dependent upon classical approaches. Information about Bermudagrass at the molecular level has been deficient although molecular information for other plants has been accumulated for the last two dacades.

In the current study, we constructed a normalized cDNA library from leaf tissue of Bermudagrass in order to expand our knowledge of its transcriptome. We sequenced and annotated 15,588 expressed sequence tags (ESTs), which were deposited in the National Center for Biotechnology Information (NCBI) to be shared with other scientists.

We also conducted cDNA array hybridization (macroarray) to profile genes responding to drought stress. A total of 120 and 69 genes were identified as up- and down-regulated, respectively. BLASTX annotation suggested that up-regulated genes may be involved in osmotic adjustment, signal transduction pathways, protein repair systems, and removal of toxins, while down-regulated genes were mostly related to basic plant metabolism such as photosynthesis and glycolysis.

Using the cDNA sequences, we performed a comparative genomic study to gain new insight into the evolution of Bermudagrass. Results suggested that the common ancestor of the grass family experienced a whole genome duplication event at ca. 50.0 ~ 65.4 million years ago (MYA), before the divergence of the PACC and BEP clades at ca. 42.3 ~ 50.0 MYA. This evolutionary study also provided concrete evidence that the Chloridoideae and Panicoideae subfamilies diverged from a common ancestor at ca. 34.6 ~ 38.5 MYA. However, we were not able to find any evidence of a recent whole genome duplication event in Bermudagrass, possibly due to its autopolyploid genome structure.

INDEX WORDS: Bermudagrass, *Cynodon dactylon*, cDNA library, Expressed sequence tag, EST, Drought stress, Macroarray, Gene duplication, Genome evolution, Grass family, Poaceae, Synonymous substitution rate, Phylogenetic analysis

DISCOVERY, EXPRESSION PROFILING, AND EVOLUTIONARY ANALYSIS OF *CYNODON* EXPRESSED SEQUENCE TAGS

by

CHANGSOO KIM

B.S., Korea University, Seoul, Korea, 1997

M.S., Korea University, Seoul, Korea, 1999

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2007

© 2007

Changsoo Kim

All Rights Reserved

DISCOVERY, EXPRESSION PROFILING, AND EVOLUTIONARY ANALYSIS OF *CYNODON* EXPRESSED SEQUENCE TAGS

by

CHANGSOO KIM

Major Professor: Andrew Paterson

Committee: Paul Raymer
Robert Carrow
Russell Malmberg
Wayne Hanna

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2007

DEDICATION

I dedicate this dissertation to my respectful parents and to my loving wife, who supported me in completion of my challenge.

ACKNOWLEDGEMENTS

First of all, I would like to thank Dr. Andrew Paterson, my major professor, for his guidance, patience, and assistance in completion of my doctoral dissertation. Also, I would like to thank the rest of my committee members, Dr. Wayne Hanna, Dr. Robert Carrow, Dr. Russell Malmberg, and Dr. Paul Raymer, for serving and guidance. I am also grateful to all the members in the *Plant Genome Mapping Laboratory*. Without their support and friendship, I could not have done what I was able to do.

I would also like to express appreciation to my friend, Glenn Hawes. Over the past 4 years, Glenn has encouraged me in completion of my goal as a teacher and friend. Finally, I would like to acknowledge the *United States Golf Association/Turfgrass and Environmental Research Committee* for making my graduate years financially secure.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS v

LIST OF TABLES viii

LIST OF FIGURES x

CHAPTER 1

INTRODUCTION 1

CHAPTER 2

REVIEW OF LITERATURE 5

CHAPTER 3

CONSTRUCTION AND CHARACTERIZATION OF A NORMALIZED cDNA LIBRARY FROM *Cynodon dactylon* L. 55

ABSTRACT 56

INTRODUCTION 57

MATERIALS AND METHODS 59

RESULTS 63

DISCUSSION 68

REFERENCES 75

CHAPTER 4

PROFILE OF DROUGHT STRESS-RESPONSIVE GENES IN *Cynodon dactylon* L. 99

ABSTRACT 100

INTRODUCTION 101

MATERIALS AND METHODS 103

RESULTS 108

DISCUSSION 114

REFERENCES 125

CHAPTER 5

PHYLOGENETIC PREDATING OF GENE DUPLICATION EVENTS IN *Cynodon dactylon* L. AND MODEL GRASS SPECIES BY COMPARATIVE GENOMIC APPROACHES 156

ABSTRACT 157

INTRODUCTION 158

MATERIALS AND METHODS 159

RESULTS 165

DISCUSSION 170

REFERENCES 178

CHAPTER 6

SUMMARY AND CONCLUSIONS 198

LIST OF TABLES

Page

Table 2.1: A classification of the genus *Cynodon*. ····· 52

Table 3.1: Summary of the *Cynodon dactylon* normalized cDNA library. ····· 84

Table 3.2: Summary of annotation of 9,414 unigenes. ····· 88

Table 3.3: The 100 most frequent protein signatures identified by the InterProScan application using 9,414 *C. dactylon* unigenes. ····· 90

Table 3.4: 25 different protein motifs further annotated by InterProScan for 43 unigenes that were not significantly matched (No hits) in a BLASTX search. ····· 91

Table 3.5: Gene ontology mappings of 9,414 unigenes using GOblet's plants database. ····· 92

Table 4.1: Significantly up- or down-regulated genes during entire drought-stress treatment. ·· 136

Table 4.2: Significantly up-regulated genes in at least one treatment or time point. ····· 137

Table 4.3: Significantly down-regulated genes in at least one treatment or time point. ····· 141

Table 4.4: Gene ontology (GO) mappings of the 189 drought candidate genes using Goblet's plant database. ····· 144

Table 4.5: Summary of selected cis-acting regulatory elements highly represented in rice homologs of up-regulated *C. dactylon* genes. ····· 150

Table.4.6: Summary of comparison between the genes from each cluster and the corresponding GO terms. ····· 155

Table 5.1: Number of sequences and paralogs used to analyze the distribution of Ks values in eight tested grasses. ····· 184

Table 5.2: All possible secondary Ks peaks formed by paralogous pairs for the analyzed grasses. ····· 185

Table 5.3: Ks values representing speciation events of grass subfamilies by the analysis of orthologous pairs between grasses. ····· 188

Table 5.4: Phylogenetic dating of a genomic duplication in the rice lineage. ····· 192

Table 5.5: Comparison of average Ks values for each rice duplication block and its corresponding time in MYA. ····· 193

Table 5.6: Pairwise comparison of the frequency of internal trees for each duplicated block between two different data types. ····· 194

Table 5.7: Comparisons of frequencies of internal trees for Bermudagrass and tall fescue among different E-value thresholds of TBLASTN search. ····· 196

Table 5.8: Pairwise comparisons of internal tree ratio for all the blocks between each species tested. Each number indicates P-value based on the 99 % confidence limit of a binomial distribution. ····· 197

LIST OF FIGURES

Page

Figure 2.1: Functional demarcation of salt and drought stress signaling pathways. ····· 53

Figure 2.2: Phylogeny of the grass family based on combined data from various studies. ····· 54

Figure 3.1: Normalization protocol applied to this study. ····· 83

Figure 3.2: Length distribution of 9,414 unigenes. ····· 85

Figure 3.3: Number of contigs, singletons, and unigenes in 15,588 ESTs from *C. dactylon* normalized cDNA library. ····· 86

Figure 3.4: Redundancy of 15,588 ESTs from the *C. dactylon* normalized cDNA library. ····· 87

Figure 3.5: Taxonomic distribution of 6,268 unigenes showing significant BLASTX matches. ····· 89

Figure 3.6: Multiple sequence alignment of unigene 0263 with nucleotide sequences from *Panicum virgatum* (DN142518), *Sorghum bicolor* (CN136815), *Saccharum officinarum* (CA291314), and *Oryza sativa* (CT849368). ····· 98

Figure 4.1: Turfgrass density represented by percent coverage in each drought stress treatment. ····· 133

Figure 4.2: Turfgrass color as affected by three different intensities of drought stress treatment. ····· 134

Figure 4.3: Venn diagrams showing the classification of genes inducible in different PEG concentrations or at different sampling time points. ····· 135

Figure 4.4: Percentage representations of third-level annotations from GO molecular function category between up- and down-regulated genes. ····· 147

Figure 4.5: Percentage representations of third-level annotations from GO cellular component category between up- and down-regulated genes. ····· 148

Figure 4.6: Percentage representations of third-level annotations from GO biological process category between up- and down-regulated genes. ····· 149

Figure 4.7: Clustering of the 189 drought candidate genes. ····· 151

Figure 5.1: Step-by-step procedures and related bioinformatics tools used for this study. ····· 183

Figure 5.2: Age distribution of paralogous sequences from grass subfamilies included in the PACC clade. ····· 186

Figure 5.3: Age distribution of paralogous sequences from grass subfamilies included in the BEP clade. ····· 187

Figure 5.4: Age distribution of orthologous sequences from grass subfamilies included in the PACC clade. ····· 189

Figure 5.5: Age distribution of orthologous sequences from grass subfamilies included in the BEP clade. ····· 190

Figure 5.6: Age distribution of orthologous sequences between grass subfamilies included in the PACC and the BEP clade, respectively. ····· 191

Figure 5.7: The phylogeny of grass subfamilies analyzed in this study. ····· 195

CHAPTER 1

INTRODUCTION

Cultivated turfgrass is a ubiquitous feature of the urban landscape in the United States and many other developed regions of the world. According to Beard (1973), turfgrass provides at least three major benefits to human activities: functional, recreational, and ornamental. In addition, a very large industry has rapidly evolved to produce and deliver turfgrass products and services in the United States. Almost 30 different species are being used as turfgrass around the world (Beard, 1973). Bermudagrass [*Cynodon* (L.) Rich] is a major turfgrass species for sports fields, lawns, parks, golf courses, and general utility turfs in Australia, Africa, India, South America, and the Southern region of the United States. The genus *Cynodon* comprises nine species with *C. dactylon* being the most widespread. *C. dactylon* is a tetraploid with much genetic variability which may, in part, explain its widespread distribution. Other *Cynodon* species have a more limited natural distribution and are often restricted to a particular habitat.

Although bermudagrass has been the most prevalent turf species in the southern United States for a long time, so far there has been little molecular-level research regarding its reactions. More generally, turfgrass biotechnology lags behind the state of biotechnology in other grasses such as corn, barley, wheat, and rice (Caetano-Anolles, 1998; Chai and Sticklen, 1998). Advances in these turfgrass relatives can, however, be leveraged to accelerate understanding of turfgrass molecular and physiological biology. To date, the complete genomes of *Arabidopsis*,

rice and sorghum have been sequenced, and a large body of genomic and cDNA sequence is available in public databases. Furthermore, several genes and pathways involved in the expression of morphological traits and physiological parameters associated with water-stress tolerance have been identified. Given this large volume of information, a growing number of candidate genes that may be related to drought tolerance are identifiable in the literature and gene databases (Skriver and Mundy, 1990; Bray, 1993; Ingram and Bartels, 1996; Seki et al., 2001).

Genomic resources such as ESTs can also help to reveal the evolutionary history of bermudagrass, as valuable comparative genomic tools. The grass family (Poaceae) contains 10,000 species and 700 genera. Although other angiosperm families contain even more taxa, the Poaceae exceed all other families in one important trait: ecological dominance. Grasses are found throughout the globe and can dominate temperate and tropical habitats. Collectively, grasses cover more than 20% of the earth's land surface (Gaut, 2002). Given their ecological dominance, it is not surprising that grasses play a central role in the human endeavor. The grass family includes all the major cereals, such as wheat, maize, rice, sorghum, barley, and oats, and less familiar grains such as rye, common millet, finger millet, and many others. Also included are such economically important non-grain species such as sugarcane, as well as several under-appreciated crops. For example, turfgrasses are a major crop group; in 1992, they generated $600 million in seed sales in the United States, more than any other U.S. crop except corn (Ligon, 1993).

In this study, we constructed a normalized cDNA library as a starting point for bermudagrass genetic research. A large sample of clones from the library was sequenced, and genes were profiled using a macroarray technique at two different intensities of water stress and

three different time points. The unigene set from the library has also been utilized to unravel the evolutionary history of bermudagrass in the grass family.

The objectives of this study were to (a) construct and characterize a normalized cDNA library from bermudagrass, (b) identify genes that correlate with physiological responses under drought conditions, (c) develop and deploy tools for gene discovery and deciphering of gene function in bermudagrass and related species, (d) maximize two-directional flow of information between bermudagrass and other grass genomes about gene function, and (e) provide a foundation for advancing knowledge about *Cynodon* (Chloridoideae) evolution using comparative genomic approaches.

REFERENCES

Beard, J.B. 1973. Turfgrass: Science and culture. Prentice-Hall, Engle- Minner, D.D., P.H. Dernoeden, D.J. Wehner, and M.S. McIntosh. Wood Cliffs, NJ.

Bray, E.A. 1993. Molecular response to water deficit. Plant Physiologist. 103:1035-1040.

Caetano-Anolles, G. 1998. DNA analysis of turfgrass genetic diversity. Crop Sci. 38:1415-1424.

Chai, B. and M. Sticklen. 1998. Applications of biotechnology in turfgrass genetic improvement. Crop Sci. 38:1320-1328.

Gaut, B. 2002. Evolutionary dynamics of grass genomes. New Phytologist. 154:15-28.

Ingram, J. and D. Bartels. 1996. The molecular basis of dehydration tolerance in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47:377-403.

Ligon, P.C. 1993. Seeds of change. Dealer Progress Magazine. Nov-Dec:29-30.

Seki, M., M. Narusaka, H. Abe, M. Kasuga, K. Yamaguchi-Shinozaki, P. Carninci, Y. Hayashizaki, and K. Shinozaki. 2001. Monitoring the expression pattern of 1300

Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. Plant Cell. 13: 61-72.

Skriver, K. and J. Mundy. 1990. Gene expression in response to abscisic acid and osmotic stress. Plant Cell. 2:503-512.

CHAPTER 2

REVIEW OF LITERATURE

BERMUDAGRASS BIOLOGY, GENETICS AND BREEDING

Bermudagrass [*Cynodon* (L.) Rich] is one of the most prevalent turf species in the southern United States. The genus *Cynodon* (L.) Rich. belongs to the family Gramineae (Poaceae), subfamily Chloridoideae, tribe Cynodonteae, and subtribe Chloridinae (Casler and Duncan, 2003). Nine species and ten varieties of *Cynodon* have been recognized so far (Table 2.1), of which turf types are included only in the 2n = 36 *C. dactylon* (L.) Pers. var. *dactylon* ("common bermudagrass"), the 2n = 18 *C. transvaalensis* Burtt-Davy ("African bermudagrass"), and their 2n = 27 interspecific hybrid *C.* x *magenissii* Hurcombe (= *C. dactylon* x *C. transvaalensis*). This tripartite interpretation of turf species relationships in *Cynodon* is a simplified classification based on incomplete knowledge. Cytological evidence suggests that *C. transvaalensis* could be considered a botanical variety of *C. dactylon*, but that it is distinctive in geography, ecology, and morphology (Harlan et al., 1970). Most of all, *C. dactylon* var. *dactylon* contains enormously variable plant types ranging from the small, fine-textured plants used as turf to the large robust plants with high biomass production capability that are used for cultivated pasture (Casler and Duncan, 2003). Harlan and de Wet (1969), however, characterized the taxon as a ubiquitous, cosmopolitan weed of the world and gave a detailed description of its distributional patterns and variation. Its distribution extends in the warm zone from lat. 45°N to

45°S; it is thus adapted to a wide range of soils and climates from rainy tropics to arid land (Holm et al., 1977), but Taliaferro (1995) identified some cold-hardy varieties reaching farther north. According to *A Geographical Atlas of World Weeds* (Holm et al., 1979), bermudagrass is still classified as a severe, principal or common weed (decreasing degrees of damage) in most countries with a warm climate in Africa, America, Asia, Australasia and southern Europe although it has been widely used as turfgrass in many geographical areas.

C. dactylon is generally characterized by a folded vernation, no auricle, and a fringe of hairs on the ligule. It is also rhizomatous and stoloniferous, and produces seed excessively on 4-5 digitate spikes raised above the leaf canopy. It has highly variable leaf characteristics, with color ranging from light to dark green and texture ranging from medium to coarse. During winter dormancy, leaves turn light brown and the species has little or no tolerance of shade or low temperatures. On the other hand, it tolerates drought, heat, and high foot traffic, and forms an aggressive sod on a wide range of soil types. The strong and creeping growth habit makes bermudagrass a weed in ornamental beds. Because of its versatility and specifically selected enhancements such as durability, drought tolerance, and uniformity of leaf texture, *C. dactylon* is suitable for use as a turfgrass (Beard, 1997).

Breeding and selection of bermudagrass has been almost continuous over the past century. Since *C. dactylon* has high genetic variability, the early stage of breeding was dependent upon selection. According to Casler and Duncan (2003), the search for superior turf bermudagrasses in the United States has been conducted since the early 1900s. Bermudagrasses used as turf at that time were relatively coarse-textured and on average probably little different from the “common” bermudagrasses presently ubiquitous to the region.

In 1977, Burton reported that the first bermudagrass putting greens in the South were seed-planted. The resulting heterogeneous populations tolerant of various biotic and abiotic stresses were screened and selected, as in the case of the 'U-3' cultivar, one of many fine strains selected by Superintendent D. Lester Hall from putting greens on a golf course in Savannah, GA (Casler and Duncan, 2003). Plants selected under such conditions, together with many foreign genetic resources, mainly originating from Africa, were accumulated in the germ plasm pool during the first half of the twentieth century. A *Cynodon* collection from all over the world was centered at Oklahoma State University in the 1960s by J. R. Harlan and colleagues for use in biosystematic investigations of the genus (Harlan and de Wet, 1969). The germ plasm pool has been developed over the past 40 years with accessions of domestic and foreign genetic resources.

According to Casler and Duncan (2003), the first documented turf bermudagrass breeding program was started in 1946 by Glenn Burton, USDA-ARS Geneticist at the Coastal Plains Experimental Station, Tifton, Georgia. This successful breeding program has bred Tifway and Tifgreen by selection and the program still continues under the direction of W. W. Hanna. Breeding for seed-propagated turf bermudagrass cultivars was performed by W. R. Kneebone at the University of Arizona in the 1960s and 1970s.

Busey and Dudeck (1997) describe both seeded and vegetatively propagated bermudagrass as "a complex of interbreeding species undergoing rapid evolution through natural and human intervention". Developments in seeded varieties have yielded grasses that are more cold tolerant, have lower fertilization requirements, are finer textured, and provide more uniform plant quality than the wild relatives. *C. dactylon* is currently the main source of seeded varieties in the United States (Taliaferro, 1995). Numex, Sahara, and Mirage are a few of the limited number of seeded bermudagrass varieties raised in the southern United States.

Hybrid bermudagrass is the result of the interspecific cross of tetraploid *C. dactylon* and diploid *C. transvaalensis* plants, which has been extensively used to produce sterile triploid clonal cultivars. Generally, hybrid bermudagrass provides uniformity of leaf texture, color, foot-traffic resistance, and winter color that are of great value to athletic fields and golf courses. For example, Tiffine and Tifgreen were produced by controlled crosses of *C. dactylon* and *C. transvaalensis* parents (Burton, 1991). Alderson and Sharp (1995) at Kansas State University introduced Midway, Midiron, Midlawn, and Midfield, which are triploid intraspecific hybrids of two taxa.

In addition, ionizing radiation has been used to induce variation among clonal propagules of bermudagrass plants as a means of cultivar improvement. Powell et al. (1974) found 71 mutant plants of Tifgreen and Tifdwarf resulting from treatment of rhizomes with 90 or 113 Gy of cobalt-60 g-radiation. This work produced the mutant clones released as Tifgreen II and Tifway II. Burton (1991) reported that Tifgreen II has lighter green color, greater cold tolerance, and lower maintenance requirements, but is coarser in texture than Tifgreen. Burton (1985, 1991) also reported that Tifway II was superior to Tifway in resistance to root knot, ring and sting nematodes, frost tolerance, and spring green up.

Hanna (1986) induced fine-textured mutations in cold hardy, but coarse textured, Midiron, and dwarf mutations in Tifway and Tifway II. Tift 94 (now renamed ‘TifSport’) is one of 66 finer-textured mutants induced with gamma irradiation from Midiron (Hanna et al., 1997). TifSport shows significantly higher turf quality than Midiron and moderate cold tolerance. TifEagle was also induced with the same method from Tifway II, which showed it to be superior to Tifdwarf in putting green quality under close mowing (3 to 4 mm) and capable of producing more stolons than Tifdwarf (Hanna and Elsner, 1999; Hanna, 1986).

Although bermudagrass is generally considered highly drought tolerant, Beard and Sifers (1997) confirmed that there is substantial genetic variation in bermudagrass for dehydration avoidance and drought tolerance. Genetic variation for water use was also reported by two different research groups (Kneebone and Pepper, 1982; Beard et al., 1992). Carrow (1996) showed the importance of selecting for rooting characteristics as a means of increasing drought avoidance in turfgrasses. Bermudagrass breeding, however, has been largely dependent upon selection from natural accessions, random mating, or mutations without any detailed information about of molecular mechanisms. Duncan and Carrow (2001) suggested that the genetic engineering of turfgrass offers an opportunity to target several components of the drought resistance mechanism. They also mentioned two different mechanisms governing dessication tolerance - cellular protection and cellular repair/recovery. Cellular protection includes membrane stabilization, water replacement, lipid modifications, compartmental stabilization, structural modification, antioxidants, osmotic adjustment, mRNA conservation, chromatin condensation, greater cell wall elasticity, cell-wall membrane interactions, sugar-protein-dehydrin-polyamine synthesis stabilization, and rate of water loss. Cellular repair/recovery includes water-loss induced damage repair, UV light-induced damage repair, cell structural and compartmental integrity, membrane and cytoskeletal reassembly, pH and ion balance maintenance, consistent electron transport, energy supply sufficiency, re-establishment of chromatin, DNA repair, lipid synthesis, protein synthesis, RNA synthesis, nutrient uptake, and metabolic re-establishment. Generally, dessication tolerant plants utilize a combination of some constitutive cellular protection strategy and a rehydration-induced recovery mechanism. Understanding molecular mechanisms regarding water stress will be valuable for integrating biotechnology-based approaches into efficient and systematic bermudagrass breeding programs.

TOOLS FOR DISSECTING PLANT PHYSIOLOGY IN TERMS OF MOLECULAR BIOLOGY

Since Howard Temin and David Baltimore discovered reverse transcriptase in 1970, cDNA cloning has been a crucial tool for gene discovery (Sambrook and Russell, 2002). In the 1980s and 1990s, large and stable cDNA libraries could be established from very small amounts of cDNA by virtue of a PCR technique and a variety of cloning vectors (Short et al., 1988; Tam et al., 1989; D'Alessio et al., 1992; Hu et al., 1992; Brady and Iscove, 1993; Rothstein et al., 1993; Revel et al., 1995). Constructing a cDNA (also known as EST; Expressed Sequence Tag) library laid a solid foundation for finding relevant genes and investigating their functions. The advantage of a cDNA library is that the introns are spliced out and the mRNA sequence can be used as a template to create DNA (via Reverse Transcriptase) to collect the preferred genes; thus, a cDNA library can be used not only to screen the target genes required, but also to express them. ESTs are usually generated by sequencing the 5' and/or the 3' ends of randomly picked clones from a cDNA library constructed from mRNA isolated at a particular developmental stage or tissue. One problem with this approach is the over-representation of abundant transcripts in the library. To overcome this difficulty, it would be desirable to be able to construct cDNA libraries containing equal amounts of cDNA from each gene expressed in a given cell, tissue, or organ (normalized cDNA libraries). Several approaches toward obtaining normalized cDNA libraries have been proposed. Weissman (1987) first reported cDNA library normalization by saturation hybridization to genomic DNA; however, this approach is impractical, since it would be extremely difficult to provide saturating amounts of the rarer cDNA species to the hybridization reaction. The alternative is the use of reassociation kinetics: Assuming that cDNA reannealing follows second-order kinetics, rarer species will anneal less rapidly than abundant species, and the remaining single-stranded fraction of cDNA will become progressively normalized during

the course of the reaction (Soares et al., 1994; Bonaldo et al., 1996). These protocols rely heavily on the reassociation of the nucleic acids in amplified plasmid libraries. However, plasmid libraries are associated with a cDNA-size cloning bias that manifests as an increased cloning efficiency of short cDNAs. In addition, during library amplification before normalization, the growth of cDNA clones varies with plasmid length; therefore, long clones are underrepresented after bulk amplification of the library. This discrepancy would lead to underrepresentation of long cDNA and difficulty in cloning long and rare cDNAs. To avoid problems related to amplification of libraries, Carninci et al. (2000) developed a technique to normalize cDNA before cloning (Figure 3.1). In addition, Zhulidov et al. (2004) recently described a simple cDNA normalization method (termed duplex-specific nuclease [DSN] normalization) that utilizes the DSN from kamchatka crab (*Paralithodes camtschaticus*). This technique provides a very simple and efficient way to normalize cDNA samples enriched with full-length sequences, does not include laborious physical separation procedures, and requires minimal hands-on time.

As mentioned, expressed sequence tags are created by sequencing the 5' and/or 3' ends of randomly isolated gene transcripts that have been converted into cDNAs. Although a typical EST represents only a part (approximately 200-900 base pairs) of a coding sequence, this partial sequence data is substantially useful. For example, EST collections are a relatively quick and inexpensive route for discovering new genes, confirming coding regions in genomic sequence (Adams et al., 1991), providing opportunities to elucidate phylogenetic relationships (Nishiyama et al., 2003), facilitating the construction of genome maps (Paterson et al., 2000), and providing the basis for development of expression arrays also known as DNA chips (Schena et al., 1995; Shalon et al., 1996; DeRisi et al., 1996; Chen et al., 1998). Currently, there are nearly 20 million ESTs in the NCBI public collection, more than 4 million of which derive from plants

(http://ww.ncbi.nlm.nih.gov/dbEST/). In order to deal with the large amount of sequence data, new computer-based tools have been developed for systematic collection, organization, storage, access, analysis, and visualization of this data. Alba et al. (2004) listed the details of bioinformatics resources that exist for these purposes.

A variety of methods have also been developed for quantifying mRNA abundance in plant tissues, which have enabled researchers to find genes related to particular traits. Although the reliable method of RNA gel-blot analysis (also known as 'Northern' blot) can be quite sensitive and allows for the accurate quantification of specific transcripts (Hauser et al., 1997), this method is not readily applied to genome wide analysis. Liang and Pardee (1992) reported that the differential display method, which uses low stringency PCR, a combinatorial primer set, and gel electrophoresis to amplify and to visualize large populations of cDNAs has significant advantages over scale-limited approaches such as RNA gel-blot analysis; however, this technique gives non-quantitative output.

Bachem et al. (1996) applied the principles of AFLP (Amplified Fragment Length Polymorphism) to cDNA templates (i.e. cDNA-AFLP) and this approach has been used to identify differentially expressed genes involved in a variety of plant processes (Qin et al., 2000; Durrant et al., 2000). The most important advantage of this method is that poorly characterized genomes can be investigated in a high-throughput manner (Bachem et al., 1996). Additionally, this method can be applied to a wide variety of tissue types, developmental stages, or time points to be compared sequentially; however, the sensitivity of cDNA-AFLP is largely limited by the ability of cDNA libraries to capture low-abundance transcripts. Therefore, the method is less sensitive to low-abundance transcripts as well as requiring substantial resources for cloning and sequencing. Another method, Serial Analysis of Gene Expression (i.e. SAGE; Velculescu et al.,

1995) combines differential display and cDNA sequencing approaches, and is quantifiable; however, SAGE is laborious, requires a foundation of sequence information as a prerequisite, and is also of limited power for detecting low-abundance transcripts just like cDNA-AFLP.

Assessment of transcription at the genomic scale has been mostly achieved with DNA microarrays. Microarrays take advantage of existing EST collections and genome sequence data, robotic instrumentation for miniaturization, and fluorescent dyes for simultaneously detecting nucleic acid abundance in RNA populations derived from multiple samples. The idea of depositing multiple DNA spots representing different genes onto a solid surface was used to investigate *Escherichia coli* gene expression on membranes (macroarrays) as long ago as 1993 (Chuang et al., 1993). Commercially available macroarrays have continued to produce useful data and should be considered before recourse to costly microarrays (Tao et al., 1999), because macroarrays are less expensive and still give researchers reliable data. Nonetheless, the recent application of microarray-robotics to achieve high spotting densities of DNA on glass slides was innovative and facilitated the construction of microarrays containing up to 50,000 DNA sequences (typically PCR products, cDNAs, or oligonucleotides) on a single microscope slide (DeRisi et al., 1996; Shalon et al., 1996). The development that has facilitated the reproducible comparison of gene expression between two samples, and hence between experiments, is dual fluorescent labeling (Schena et al., 1995). Simultaneous hybridization of two populations of DNA sequences labeled with the fluorescent dyes Cy3 and Cy5 allows accurate assessment of relative levels of gene expression, which is unaffected by hybridization variability or the differences between individual microarrays.

WATER STRESS RESEARCH AT THE LEVEL OF MOLECULAR BIOLOGY

Abiotic stresses such as drought, extreme temperatures, or high and fluctuating salinity can severely impair plant growth and performance causing significant reductions in yield. Thus, for decades the responses of plants to various stresses have been the focus of physiological studies. More recently, this research has been extended by molecular and reverse genetic studies and transgenic experiments (Grover, 1999; Bajaj et al., 2000; Hasegawa et al., 2000; Zhang et al., 2000). In particular, drought stress has long been a central issue in plant science because good quality water is an especially limited natural resource that is essential for plants to survive. Therefore, understanding how plants respond to water stress and using that understanding to improve drought tolerance should be an integral component of water conservation. In fact, little has been done concerning turfgrass biotechnology; however, a large and increasing number of genes, transcripts, and proteins have been revealed in drought stress pathways in other plant species by virtue of many tools and approaches for gene discovery.

Ingram and Bartels (1996) published the first systematic review of the molecular basis of dehydration tolerance in plants. At that time, knowledge of the regulatory network governing drought-stress responses was fragmentary, with almost no information on signal perception. The drought-stress responsive genes were largely classified into five categories in this review: (a) Metabolism, (b) Osmotic Adjustment, (c) Structural Adjustment, (d) Degradation and Repair including removal of toxins, and (e) Late-Embryogenesis-Abundant Proteins.

Water stress leads to increased expression of genes involved in general carbon metabolism. For example, the increased expression of glyceraldehyde-3-phosphate dehydrogenase indicates that enhanced carbon flux is needed to produce energy and/or carbon skeletons for osmolyte biosynthesis (Velasco et al., 1994). In addition, the level of mRNA

encoding a specific isoform of phosphoenolpyruvate carboxylase (PEPCase) increases during water deficit-stress in *Mesembryanthemum crystallinum* (Vernon et al., 1993).

Drought stress-responsive genes cloned from a variety of plants encode enzymes involved in the synthesis of osmoprotectants such as glycine betaine, proline, and various sugar alcohols (Weretilnyk and Hanson, 1990; Vernon and Bohnert, 1992; Ishitani et al, 1995; Yoshiba et al., 1995). The expression of these genes is induced by water stress and salt stress. For example, a gene encoding an enzyme involved in the synthesis of pinitol (myo-inositol O-methyl transferase) was isolated from *M. crystallinum* and is induced by salt stress (Vernon and Bohnert, 1992). The accumulation of proline has been extensively studied and it has been demonstrated that alteration of proline content through genetic engineering can improve stress tolerance (Kavi Kishor et al., 1995). In plants, Δ'-pyrroline-5-carboxylate synthetase (P5CS) catalyzes the conversion from L-glutamate to glutamic-γ-semialdehyde, which spontaneously forms Δ'-pyrroline-5-carboxylate, which is converted to proline by P5C reductase (P5CR) (Yoshiba et al., 1995). The level of P5CS mRNA is increased in response to drought in *Arabidopsis*, but P5CR mRNA levels are not altered. Upon rehydration, the transcript level of proline dehydrogenase, an enzyme involved in proline degradation is increased in *Arabidopsis* (Taylor, 1996; Yoshiba et al., 1997), and it has been shown that water deficit stress reciprocally controls the expression of the genes encoding P5CS and proline dehydrogenase to ensure that proline accumulates (Yoshiba et al., 1995).

Drought stress has been shown to cause alterations in the chemical composition and physical properties of the cell wall (e.g. wall extensibility), and such changes may involve the genes encoding S-adenosylmethionine synthetase (Espartero et al., 1994). Under nonstressful conditions, increased expression of S-adenosyl-L-methionine synthetase genes correlates with

areas where lignification is occurring (Peleman et al., 1989). Thus, the increased expression in drought-stressed tissue could also be due to lignification in the cell wall.

A number of genes expressed in drought-stressed plants encode products that may play a role in cellular repair and general defense responses. These proteins include heat shock proteins (Borkird et al.. 1991; Almoguera and Jordano, 1992; Ouvrad et al., 1996), ubiquitin (Borkird et al., 1991), protease inhibitors (Downing et al., l992), proteases (Guerrero et al., 1990; Koizumi et al., 1993; Williams et al., 1994), endochitinases (Chen et al., 1994), lipoxygenases (Bell and Muilet, 1991), peroxidase (Mittler and Zilinskas, 1994), superoxide dismutase (Perl-Treves and Galun, 1991; White and Zilinskas, 1991) and adenosylmethionine (AdoMet) synthetases (Espartero et al., 1994). In addition, the genes encoding late-embryogenesis-abundant (LEA) proteins are consistently represented in differential screens for transcripts with increased levels during drought stress: however, the exact function of LEA proteins is still being studied (Ingram and Bartels, 1996).

Signal transduction via ABA and the promoter modules of several responsive genes began to be elucidated from the early 1990s. Zhu (2002) reviewed salt and drought stress signal transduction in plants based on scattered candidate gene studies from the 1990s, which are briefly summarized in figure 2.1. Xiong et al. (2002) also divided the signal transduction pathways in plants under environmental stresses into three major types: (a) osmotic/oxidative stress signaling that makes use of mitogen activated protein kinase (MAPK) modules; (b) Ca^{2+}-dependent signaling that leads to activation of LEA-type genes such as the dehydration responsive element (DRE) class of genes, and (c) Ca^{2+}-dependent salt overly sensitive (SOS) signaling that results in ion homeostasis. Various signal pathways can operate independently of each other or they may positively or negatively modulate other pathways. Different signaling

pathways may also share components and second messengers to achieve their objectives. This interdependence of various pathways on each other is called cross-talk among themselves. As a result, many signals could interact in a cooperative fashion with each other (Knight and Knight, 2001).

For the last decade, powerful experimental tools such as microarray, reverse genetics approaches, and bioinformatics have made it possible for researchers to find new clues to understand molecular mechanisms regulating gene expression in response to various abiotic stresses. Study of drought stress mechanisms in plants at the molecular level has been significantly accelerated by analyzing *cis*- and *trans*-acting elements that function in gene expression during stresses in *Arabidopsis* (Yamaguchi-Shinozaki and Shinozaki, 2005). Recently, Yamaguchi-Shinozaki and Shinozaki (2006) extensively reviewed transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. In this review, regulation of gene expression under drought and cold stresses was divided into ABA-independent, and ABA-dependent gene expression pathways. In fact, several reports have described genes that are induced by dehydration and cold stresses but that do not respond to exogenous ABA treatment (Zhu, 2002; Shinozaki et al., 2003). This suggests the existence of ABA-independent, as well as ABA-dependent, signal transduction cascades between the initial stress signal and the expression of specific genes. For example, Yamaguchi-Shinozaki and Shinozaki (1992) reported that the *Arabidopsis RD29A/COR78/LTI78* gene is induced by drought, cold, and ABA. Analysis of this promoter has shown that a 9-bp conserved sequence, TACCGACAT, named the DRE (Dehydration-responsive element) , is an essential *cis*-element for regulating *RD29A* induction in the ABA-independent response to dehydration and cold (Yamaguchi-Shinozaki and Shinozaki, 1994). DRE is also found in the promoter regions of

many drought- and cold-inducible genes (Thomashow, 1999; Yamaguchi-Shinozaki and Shinozaki, 2000). Similar *cis*-acting elements, named C-repeat (CRT) and low-temperature-responsive element (LTRE), both containing an A/GCCGAC motif that forms the core of the DRE sequence, regulate cold-inducible promoters (Jiang et al., 1996; Stockinger et al., 1997). Liu et al. (1998) isolated the cDNAs encoding DRE-/CRT-binding proteins, CBF/DREB1 (C-repeat Binding Factor/DRE Binding protein 1), and DREB2 using yeast one-hybrid screening. These proteins contained the conserved DNA-binding domain found in the ERF (ethylene-responsive element-binding factor) and AP2 proteins. These proteins specifically bind to the DRE/CRT sequence and activate the transcription of genes driven by the DRE/CRT sequence. In *Arabidopsis*, three genes encoding DREB1/CBF lie in tandem on chromosome 4 in the following order: DREB1B/CBF1, DREB1A/CBF3, and DREB1C/CBF2. Furihata et al. (2006) identified that there are two DREB2 proteins, DREB2A and DREB2B (Yamaguchi-Shinozaki and Shinozaki, 2006). Expression of the DREB1/CBF gene is induced by cold, but not by dehydration and high-salinity stresses (Liu et al., 1998; Shinwari et al., 1998). In contrast, expression of the DREB2 genes is induced by dehydration and high-salinity stresses but not by cold stress (Liu et al., 1998; Nakashima et al., 2000). Later, Sakuma et al. (2002) reported three novel *DREB1/CBF*-related genes and six novel *DREB2*-related genes that were not expressed at high levels under various stress conditions. The three DREB1 proteins are probably major transcription factors involved in cold-induced gene expression, and the DREB2A and DREB2B proteins are involved in high-salinity- and drought-induced gene expression. However, the expression of one of the *CBF/DREB1* genes, *CBF4/DREB1D*, is induced by osmotic stress (Haake et al., 2002) and the other two *CBF/DREB1* genes, *DDF/DREB1F* and *DDF2/DREB1E*, are induced by high salinity stress (Magome et al., 2004), suggesting the existence of cross-talk

between the CBF/DREB1 and the DREB2 pathways. There are also several dehydration-inducible genes that do not respond to either cold or ABA treatment, suggesting the existence of another ABA-independent pathway in the dehydration stress response. These genes include early response to dehydration1 (*ERD1*), which encodes a Clp protease regulatory subunit, ClpD (Nakashima et al., 1997). Simpson et al. (2003) reported that *ERD1* is not only induced by dehydration, but also upregulated during natural senescence and dark-induced senescence. Promoter analysis of *ERD1* in transgenic plants indicates that the *cis*-acting elements responsible for gene expression during dehydration and etiolation are separately located in two discrete portions of the *ERD1* promoter. Moreover, two different novel *cis*-acting elements, a MYC like sequence (CATGTG) and a 14-bp rps1 site 1-like sequence, are involved in induction by dehydration stress (Simpson et al., 2003). Recently, three cDNAs encoding MYC-like sequence-binding proteins – *ANAC019, ANAC055*, and *ANAC072* – were isolated by the yeast one-hybrid screening method (Tran et al., 2004). They also revealed that several stress-inducible genes were significantly upregulated in the transgenic plants overexpressing *ANAC019, ANAC055,* or *ANAC072*, and the plants showed significantly increased drought tolerance.

ABA plays an important role in the adaptation of vegetative tissues to abiotic stresses such as drought and high salinity (Bray et al., 2000). ABA promotes stomatal closure in guard cells, mediated by solute efflux, and regulates the expression of many genes that may function in dehydration tolerance in both vegetative tissues and seeds (Himmelbach et al., 2003). Guiltinan et al. (1990) first identified the *cis*-acting element, named ABRE (ABA-responsive element; PyACGTGGC), involved in ABA-regulated gene expression in the wheat *Em* gene, which functions mainly in seeds during late embryogenesis. Mundy et al. (1990) also identified ABRE in the promoter region of rice the *RAB16* gene, which is expressed in both dehydrated vegetative

tissues and maturing seeds. However, a single copy of ABRE is not sufficient for ABA-responsive transcription. ABRE and coupling elements such as CE1 and CE3 constitute an ABA-responsive complex in the regulation of wheat *HVA1* and *HVA22* genes (Shen et al., 1995, 1996). Uno et al. (2000) also reported that two ABRE sequences are necessary for the expression of *Arabidopsis RD29B* in seeds and for the ABA-responsive expression of *RD29B* in vegetative tissue. One of these ABRE sequences might function as a coupling element. Most of the known coupling elements have similarity with ABREs and contain an A/GCGT motif (Hobo et al., 1999). *Arabidopsis* cDNAs encoding the bZIP transcription factors referred to as ABRE-binding (AREB) proteins of ABRE-binding factors (ABFs) were isolated using one-hybrid screening method (Choi et al., 2000; Uno et al., 2000). Uno et al. (2000) also showed that expression of AREB1/ABF2, AREB2/ABF4, and ABF3 was upregulated by ABA, dehydration, and high-salinity stresses among those AREB/ABF proteins. Their activities were reduced in the ABA-deficient *aba2* mutant and in the ABA-insensitive *abi1* mutant, but were enhanced in the ABA hypersensitive *era1* mutant (Koornneef et al., 1984; Koornneef et al., 1992; Uno et al., 2000). Kang et al. (2002) reported that overexpression of ABF3 and ABF4/AREB2 resulted in ABA-hypersensitive phenotypes in germination and seedling growth stages in *Arabidopsis*. These transgenic plants also showed improvement of drought stress tolerance and the expression of some ABA-responsive genes such as LEA class genes (*RD29B, rab18*), cell cycle regulator genes (*ICK1*), and protein phosphatase 2C genes (*ABI1* and *ABI2*), suggesting that AREB/ABF proteins are involved in ABA response and stress tolerance in plants. Moreover, Kim et al., (2004) reported that ABF2/AREB1 was shown to be an essential component of glucose signaling, and its overexpression also improved tolerance of drought.

There have been a number of reports regarding ABA-dependent gene expression other

than AREBs/ABFs. For example, induction of the dehydration-responsive *RD22* is mediated by ABA and requires protein biosynthesis for its ABA dependent expression (Abe et al., 1997). MYC and MYB recognition sites in the *RD22* promoter function as *cis*-acting elements in the dehydration inducible expression of *RD22* (Abe et al., 2003). A MYC transcription factor, AtMYC2 (*rd22BP1*), and a MYB transcription factor, AtMYB22, bind these *cis*-elements in the *RD22* promoter and cooperatively activate the expression of *RD22*. These two transcription factors are synthesized after the accumulation of endogenous ABA, indicating that they play roles in a late stage of the plant's response to different stresses. Abe et al. (2003) also constructed transgenic plants overproducing MYC and MYB which had higher sensitivity to ABA and showed osmotic stress tolerance. Recently, AtMYC2 was also reported as a transcription factor that functions in jasmonic acid (JA) and JA-ethylene-regulated defense responses in *Arabidopsis*, which reflects cross-talk between ABA- and JA-responsive gene expression at the MYC recognition sites in the promoters (Anderson et al., 2004; Boter et al., 2004; Lorenzo et al., 2004). Recently, Fujita et al. (2004) constructed transgenic plants either overexpressing or repressing *RD26* in *Arabidopsis*. *Arabidopsis RD26* encodes a NAC protein and is induced not only by dehydration but also by ABA. The transgenic plants overexpressing *RD26* were highly sensitive to ABA, whereas *RD26*-repressed plants were insensitive. Microarray analysis showed that ABA- and stress-inducible genes were upregulated in *RD26*-overexpressing plants and repressed in *RD26*-repressed plants, indicating that a *cis*-regulatory factor, the NAC recognition site, may function in ABA-dependent gene expression under stress conditions.

Based on the prior knowledge of the gene expression regulatory pathway in *Arabidopsis*, a number of stress-responsive gene homologs have been identified in other plant species and genetic engineering has allowed the introduction of new pathways for the biosynthesis of various

compatible solutes into plants, resulting in the production of transgenic plants with improved stress tolerance (Chen and Murata, 2002). In addition, recent transgenic plants include regulatory proteins like transcription factors, which make it possible to induce expression of a group of genes that show tolerance to various abiotic stresses (Cherian et al., 2006). For example, introduction of dehydration responsive element binding protein (DREB) family genes under the control of different promoters were reported in *Oryza sativa* L. (Dubouzet et al., 2003). Transgenic wheat plants expressing *DREB1A* gene under the control of rd29A promoter also showed substantial resistance to water stress in comparison with controls (Pellegrineschi et al., 2004). The stress inducible expression of this gene had minimal effects on plant growth and provided greater tolerance of stress conditions than genes driven by the 35S promoter.

GENOMIC DUPLICATION (POLYPLOIDY) IN PLANT EVOLUTION

Polyploidy means more than two sets of chromosomes per nucleus. Polyploidy permeates virtually all of angiosperm biology (Paterson, 2005) and has long been recognized to be an important process in the evolutionary history of plants (Mable, 2003). Polyploidy was traditionally proposed to have occurred in the lineage of at least 70% of angiosperms (Masterson, 1994) and in 95% of pteridophytes (Soltis and Soltis, 1999). Traditionally, estimates of the number of polyploids have been based on chromosome number alone or in conjunction with chromosome pairing analysis (Leitch et al., 2004).

Polyploids can be largely classified into allopolyploids and autopolyploids. Autopolyploids derive from the multiplication of a single genome or genomes of the same species. Allopolyploids combine two or more genomes from distinct species (Osborn et al., 2003). Ramsey and Schemske (1998) suggested that the formation of allopolyploids might be

more common in nature than that of autopolyploids because of heterosis and homeostasis conferred by permanent hybridity in allopolyploids, which is lacking in autopolyploids. Autopolyploids often show reduced fertility due to meiotic irregularities. Despite the disadvantages of autopolyploids, recent molecular data have demonstrated an autopolyploid origin for an increasing number of polyploids and have stressed that autopolyploidy is of significant evolutionary importance (Mahy et al., 2000).

Technological advances in the analysis of genome structure and function have enabled researchers to study the genetic consequences of genome duplication at an unprecedented level. Recent studies have shown that polyploidy is not just the simple merger and subsequent collaboration of two genomes, but rather involves various molecular and physiological changes (Adams and Wendel, 2005).

The recent advent of whole genome sequencing, comparative genome mapping, micro-colinearity studies, and analyses of EST collections has shown that polyploidy events occurred at many times in the evolutionary history of *Arabidopsis* (Grant et al., 2000; Ku et al., 2000; Vision et al., 2000; Blanc et al., 2003; Bowers et al., 2003; Ermolaeva et al., 2003; Raes et al., 2003; Ziolkowske et al., 2003). Although a variety of results were deduced from the studies because of methodological differences, strong evidence points to one round of genomic duplication after the monocot-eudicot divergence and a second polyploidization event sometime after the divergence of *Arabidopsis* and *Brassica* from the Malvaceae. Paterson et al. (2004) provided the evidence of an ancient genome-doubling event in the common ancestor of the modern grasses. Although Wendel et al. (1986) showed duplicated chromosomal segments based on isozyme study and Ahn and Tanksley (1993) reported polyploidy associated with extensive chromosomal restructuring with comparative molecular maps, a more recent polyploidy have been discussed in

the maize lineage using comparative genomics (Messing et al., 2004; Lai et al., 2004). Other examples of ancient polyploidization events were described in the ancestor of the solanaceous crops tomato and potato, in the legumes *Glycine* (soybean) and *Medicago truncatula* (Schlueter et al., 2004), and in a common ancestor of the cotton (*Gossypium*) genus (Blanc et al., 2004; Rong et al., 2004).

Classical ideas (Stephens, 1951; Ohno, 1970) have suggested that the genomic redundancy resulting from polyploidy might free extra copies of genes to mutate and diverge, adopting new functions without compromising essential functions. Song et al. (1995) reported that synthetic polyploid populations from *Brassica* showed extensive genomic rearrangements. Levy and Feldman (2004) demonstrated that extensive genomic arrangements, including exchanges between genomes and gene loss, often arise with the onset of polyploidization in the wheat genome. In the long term, ancient polyploids effectively become "diploidized" with single dosages restored for many genes. The prominence of polyploidy in flowering plants implies that it has some adaptive significance. Polyploids often show novel phenotypes that are not present in their diploid progenitors or exceed the range of the contributing species (Ramsey and Schemske, 2002). Some of these phenotypes, such as increased drought tolerance, apomixis (asexual seed production), pest resistance, flowering time, organ size, and biomass, enabled polyploids to adapt themselves to new environments or enhance their chances of being selected for use in agriculture. The mechanisms by which polyploidy contributes to novel variation are not well understood, but Osborn et al. (2003) suggested that duplicate genes have relaxed constraints on their function, and thus can create new phenotypes in polyploids (neofunctionalization) or diverge in only part of a gene function, such as tissue specificity (subfunctionalization). Mena et al. (1996) found diversification of C-function activity in maize flower development, which strongly supports

subfunctionalization after polyploidization. In contrast, however, several authors (Hughes and Hughes, 1993; Moore and Purugganan, 2003; Chapman et al., 2006) have shown duplicated genes to evolve more slowly than singletons. Some evidence also supports the suggestion that the retention of duplicate genes is non-random (Kellis et al, 2004; Blanc and Wolfe, 2004; Seoighe and Gehring, 2004; Chapman et al., 2006). Blanc and Wolfe (2004) conducted a comprehensive analysis of the functional divergence of genes that have been duplicated by polyploidy during the evolutionary history of *Arabidopsis*. They showed that more than half of the gene pairs formed by the most recent polyploidy have significantly different expression patterns. These authors also provide evidence that 62% of recently duplicated gene pairs have undergone functional diversification. The most interesting finding reported in this publication was that some duplicated gene pairs have diverged in concert, forming two parallel networks expressed in different cell types or under different environmental conditions, which they termed *concerted divergence.* This finding might have important implications for divergence in metabolic pathways and, hence, evolutionary diversification.

Genomic doubling also significantly affects gene expression, resulting in rapid epigenetically induced gene silencing (Osborn et al., 2003; Liu and Wendel, 2003). Since novel phenotypes are known to emerge from polyploidy formation, including some with high visibility to natural selection such as organ size and flowering time, one can speculate that such short term changes might somehow play a role in the formation of new species (Abbott and Lowe, 2004). An interesting recent revelation is that the silencing of some duplicated genes often accompanies the onset of allopolyploidy, as shown by studies of newly created synthetic polyploids (Wang et al., 2004; Kashkush et al., 2002; He et al., 2003; Adams et al., 2003, 2004). Wang et al. (2004) observed that silencing occurs as early as the first generation after polyploidy, although some

genes are not silenced until later generations. Adams et al. (2004) also observed that some duplicated genes are silenced immediately upon polyploidization in some organs of the plant but remain expressed in other organs at varying levels. Simultaneously with the aforementioned studies, researchers have also posed three different questions regarding gene silencing in polyploidization – the cause of gene silencing, the direction of gene silencing (random or nonrandom), and the reason for gene silencing in polyploids.

Silencing arising immediately upon polyploidization, in the absence of gene deletion, must be epigenetically induced because there is insufficient time for point mutations to accumulate. Changes in cytosine methylation, histone modifications (such as deacetylation and methylation), and positional effects from higher order changes in chromatin structure have been known to be the general causes of silencing in the polyploidization event (Osborn et al., 2003; Liu and Wendel, 2003). However, Comai et al. (2003) proposed models for homoeologous gene silencing, involving repeats and long terminal repeats (LTRs) of retroelements. Kashkush et al. (2003) also suggested that polyploidy in synthetic wheat could be related to antisense transcripts generated by readout transcription of a retrotransposon caused silencing of an adjacent gene

An important and controversial question is whether gene silencing is a random process or not. Some recent studies have shown synchronous silencing of the same duplicate gene in multiple polyploid genotypes or lines (Adams et al., 2003), suggesting that silencing is a nonrandom process due to dosage requirements. In contrast, silencing of other genes appeared to be random (Adams et al., 2004).

Understanding patterns of gene silencing may help to reveal why duplicated genes are silenced in polyploids. Osborn et al. (2003) summarized the maintenance of appropriate gene dosage and altered regulatory networks as the most probable explanations. According to the

authors, polyploidy has the general effect of increasing gene expression levels on a per cell basis in proportion to the gene dosage conferred by ploidy level; therefore, gene silencing is inevitable in order to eliminate this genetic redundancy. Also, the expression of most genes is dependent upon networks of regulators, such as transcription factors, that are organized into hierarchies. The numbers of regulators in diploid networks is high, but in polyploids they can be expanded several fold, which, in turn, may be able to affect the intensity of gene silencing.

THE GRASS FAMILY AND ITS EVOLUTION

The grass family (Poaceae or Gramineae) has been of particular interest to humans because most people in the world rely on grasses, including rice, wheat, and maize, for a major portion of their diet. Domestic animals are also raised on diets partly or wholly of grasses. In addition, grasses are an important part of the urban and suburban landscape. The family contains approximately 10,000 species and 700 genera, and covers approximately 20% of the earth's land surface, which reflect "ecological dominance" (Gaut, 2002). Owing to the importance of the family, grasses have been the subject of intense phylogenetic, ecological, agronomic, and molecular study. In the last few decades, a clear picture has been formed regarding the evolutionary history of the grass family by virtue of the advent of high-throughput molecular biology (Kellogg, 2001).

Most taxonomic studies of the grasses have recognized six or seven major subfamilies, with several smaller subfamilies. Initial grass classifications were based on morphological structures such as the spikelet, leaf blade, and embryo, but morphology alone failed to clearly resolve systematic relationships. As a result, molecular markers have been employed to construct grass phylogenies. Initially, chloroplast markers such as the *rbcL* and *ndhF* genes (Clark et al.,

1995; Duvall and Morton, 1996) have been used for molecular studies; however, more recently, phylogenetic studies have been performed on the basis of nuclear markers such as internal transcript spacer (ITS) (Hsiao et al., 1998), *waxy* (Mason-Gamer et al., 1998), and *phyB* (Mathews et al., 2000). Although some of these studies have been hampered by small sample size or insufficient numbers of variable bases, all have reached similar conclusions about the order of events in the evolution of the grasses (Kellogg, 2001). Some of these molecular and morphological studies have been combined by the Grass Phylogeny Working Group (GPWG) into a robust phylogeny of the family (Figure 2.2, Grass Phylogeny Working Group, 2001).

The phylogenetic approaches of the GPWG have revealed unexpected information regarding the evolutionary history of the grasses. For example, before molecular phylogenetic analyses, the Anomochlooideae were considered members of subfamily Bambusoideae (Clark et al., 1995); however, it is now known that the Anomochlooideae and Bambusoideae represent divergent grass lineages (figure 2.2), with the anomochlooids representing the most early diverged grass lineage. In contrast, bambusoids fall within a monophyletic group known as the 'BEP' clade because they contain Ehrhartoideae and Pooideae. The latter two subfamilies include the economically important cereals, such as rice, wheat, barley and oats (Kellogg, 2000, 2001).

The remaining major grass subfamilies fall into a second monophyletic clade known as the 'PACC' clade. The PACC clade contains subfamilies Panicoideae, Arundinoideae, Centothecoideae, and Chloridoideae. According to Kellogg (2000), the PACC clade is of evolutionary importance in that all C4 species fall within the PACC clade. Kellogg also found that the distribution of C4 plants in the PACC clade suggests that C4 photosynthesis originated at

least four times, which suggests that regulation of C4 photosynthesis may differ among species with independent origins of C4 photosynthesis.

Gaut (2002) put the divergence of key grass taxa into a temporal framework to better understand and discuss grass genome evolution. He provided a phylogeny and divergence times among eight economically important grasses – rice, oats, barley, wheat, foxtail millet, pearl millet, sorghum, and maize - along with a basal grass (*Anomochloa*) and an outgroup (*Joinvillea*) based on the GPWG phylogeny. Stebbins (1981) and Wolfe et al. (1987) assumed that maize and rice diverged 50 million years ago, respectively, which was confirmed by Gaut's analysis with the nonparametric rate smoothing method of Sanderson (Sanderson, 1997). Gaut (2002) also suggested that the divergence time estimated in the study may be an improvement over some previously published estimates because Sanderson's method does not assume a molecular clock. The estimates suggest the grass family originated roughly 77 million years ago (MYA). The age of the grass family has previously been reported to be 55-70 MYA based on fossil evidence (Linder, 1987; Jacobs et al., 1999). More recently, Prasad et al. (2005) reported their analysis of phytoliths in coprolites of titanosaurid sauropods that lived in central India about 65 to 71 MYA. Their data indicated that those dinosaurs ate grasses and provided the first unambiguous evidence that the grass family originated and had already diversified during the Cretaceous. All those fossil data confirmed that the higher estimate of 77 MYA is reasonable (80-85 MYA).

The collaborative work of the GPWG has resolved the broad outline of grass phylogeny, and we now have considerable confidence about which species are most closely related. Despite rapid progress in the last decade, our understanding of grass genomes is still quite limited. For example, Hilu and Alice (2001) mentioned that Chloridoideae has been a poorly studied group although it is of significant importance in agriculture. Morphologically,

Chloridoideae is not a well-defined group despite its large number of species. Representatives of the Chloridoideae were included in only a few studies (Barker et al., 1995; Clark et al., 1995; Duvall and Morton, 1996; Hilu and Alice, 2001). The chloridoid species were always grouped together as a single group and appeared to be related to Arundinoideae. The monophyletic origin of Chloridoideae based on DNA sequence is not supported by previous numerical and immunological studies (Hilu and Esen, 1988). Paterson et al. (2004) also addressed the lack of DNA sequence information as a major limitation in order to study the evolution of Chloridoideae and Arundinoideae.

REFERENCES

Abbott, R.J. and A.J. Lowe. 2004. Origins, establishment and evolution of new polyploidy species: *Senecio cambrensis* and *S. eboracensis* in the British Isles. Biol. J. Linn. Soc. 82:467-474.

Abe, H., K. Yamaguchi-Shinozaki, T. Urao, T. Iwasaki, D. Hosokawa, and K. Shinoazki. 1997. Role of *Arabidopsis* MYC and MYB homologs in drought- and abscisic acid-regulated gene expression. Plant Cell. 9:1859-1868.

Abe, H., T. Urao, T. Ito, M. Seki, K. Shinozaki, and K. Yamaguchi-Shinozaki. 2003. *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. Plant Cell. 15:63-78.

Adams, K.L. and J.F. Wendel. 2005. Polyploidy and genome evolution. Curr. Opin. Plant Biol. 8:135-141.

Adams, K.L., R. Cronn, R. Percifield, and J.F. Wendel. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc. Natl. Acad. Sci. USA. 100:4649-4654.

Adams, K.L., R. Percifield, and J.F. Wendel. 2004. Organ-specific silencing of duplicated genes in a newly synthesized cotton allotetraploid. Genetics. 168:2217-2226.

Adams, M., J. Kelley, J. Gocayne, M. Dubnick, M.H. Polymeropoulos, H. Xiao, C.R. Merril, A. Wu, B. Olde, R.F. Moreno, A.R. Kerlavage, W.R. McCombie, and J.C. Venter. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science, 252:1651–1656.

Ahn, S. and S.D. Tanksley. 1993. Comparative linkage maps of rice and maize genomes. Proc. Natl. Acad. Sci. USA 90:7980-7984.

Alba, R., Z. Fei, P. Payton, Y. Liu, S.L. Moore, P. Debbie, J. Cohn, M. D'Ascenzo, J.S. Gordon, J.K.C. Rose, G. Martin, S.D. Tanksley, M. Bouzayen, M.M. Jahn, and J. Giovannoni. 2004. ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development. Plant J. 38:696-714.

Alderson, J. and W.C. Sharp. 1995. Grass Varieties in the United States, CRC Lewis Publishers, Boca Raton, FL.

Almoguera, C. and J. Jordano. 1992. Developmental and environmental concurrent expression of sunflower dry-seed-stored low-molecular-weight heat-shock protein and LEA messenger RNAs. Plant Mol. Biol. 19:781-792.

Anderson, J.P., E. Badruzsaufari, P.M. Schenk, J.M. Manners, O.J. Desmond, C. Ehlert, D.J. Maclean, P.R. Ebert, and K. Kazan. 2004. Antagonistic interaction between abscisic acid

and jasmonate-ethylene signaling pathways modulates defense gene expression and disease resistance in *Arabidopsis*. Plant Cell. 16:3460-3479.

Bachem, C., R. van der Hoeven, S. de Bruijn, D. Vhil, M. Zabeau, and R. Visser. 1996. Visualisation of differential gene expression using a novel method of RNA finger-printing base on AFLP: analysis of gene expression during potato tuber development. Plant J. 9:745-753.

Bajaj, S., J. Targolli, L.F. Liu, T.H.D. Ho, and R. Wu. 2000. Transgenic approaches to increase dehydration-stress tolerance in plants. Mol. Breed. 5:493-503.

Barker, N.P., H.P. Linder, and E.H. Harley. 1995. Polyphyly of Arundinoideae (Poaceae): Evidence from *rbcL* sequence data. Syst. Bot. 20:423-435.

Beard J.B. and S.I. Sifers. 1997. Genetic diversity in dehydration avoidance and drought resistance within the *Cynodon* and *Zoysia* species. Intl. Turfgrass Soc. Res. J. 8:603-610.

Beard, J.B., R.L. Green, and S.I. Sifers. 1992. Evapotranspiration and leaf extension rates of 24 well-watered turf-type *Cynodon* genotypes. HortSci. 27:986-988.

Bell, E. and J.E. Mullet. 1991. Lipoxygenase gene expression is modulated in plants by water deficit, wounding and methyl jasmonate. Mol. Gen. Genetics. 230:456-462.

Blanc, G. and K.H. Wolfe. 2004. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell. 16:1679-1691.

Blanc, G., K. Hokamp, and K.H. Wolfe. 2003. A recent polyploidy superimposed on older large-scale duplications in the *Arabidopsis* genome. Genome Res. 13:137-144.

Bonaldo, M., G. Lennon, and M. Soares. 1996. Normalization and subtraction: Two approaches to facilitate gene discovery. Genome Res. 6:791-806.

Borkird, C., C. Sirnoens, T. Villarroel, and M. Van Montagu. 1991. Gene expression associated with water-stress adaptation of rice cells and identification of two genes as hsp 70 and ubiquitin. Physiologia Plantarum 82:449-447.

Boter, M., O. Ruiz-Rivero, A. Abdeen, and S. Prat. 2004. Conserved MYC transcription factors play a key role in jasmonate signaling both in tomato and *Arabidopsis*. Genes Dev. 18:1577-1591.

Bowers, J.E., B.A. Chapman, J. Rong, and A.H. Paterson. 2003. Unravelling angiosperm genome evolution by phlogenetic analysis of chromosomal duplication events. Nature. 422:433-438.

Brady, G. and N.N. Iscove. 1993. Construction of cDNA libraries from single cells. Methods Enzymol. 225:611-623.

Bray, E., J. Bailey-Serres, and E. Weretilnyk. 2000. Responses to abiotic stresses. In Biochemistry and Molecular Biology of Plants, ed. B.B. Buchanan,W. Gruissem, R.L. Jones. pp. 1158-1203. Amer. Soc. Plant Physiol. Rockville, MD.

Burton, G.W. 1985. Registration of Tifway II bermudagrass. Crop Sci. 25:364.

Burton, G.W. 1991. A history of turf research at Tifton. Green Section Record 29(3):12-14. United States Golf Association, Far Hills, NJ.

Busey, P. and A.E. Dudeck. 1999. Bermudagrass Varieties. In Unruh J., Elliot M. (Eds), Best Management Practices for Florida Golf Courses, 2nd edition. pp. 97-101. Institute of Food and Agricultural Sciences, Gainesville FL.

Carninci, P., Y. Shibata, N. Hayatsu, Y. Sugahara, K. Shibata, M. Itoh, H. Konno, Y. Okazaki, M. Muramatsu, and Y. Hayashizaki. 2000. Normalization and Subtraction of Cap-

Trapper-Selected cDNAs to Prepare Full-Length cDNA Libraries for Rapid Discovery of New Genes. Genome Res. 10:1617-1630.

Carrow, R.N. 1996. Drought resistance aspects of turfgrasses in the southeast: root-shoot responses. Crop Sci. 36:687-694.

Casler, M.D. and R.R. Duncan. 2003. Turfgrass Biology, Genetics, and Breeding. pp 235-256. John Wiley & Sons, Inc. Hoboken, NJ.

Chapman, B.A., J.E. Bowers, F.A. Feltus, and A.H. Paterson. 2006. Buffering of crucial functions by paleologous duplicated genes may contribute cyclicality to angiosperm genome duplication. Proc. Natl. Acad. Sci. USA. 103:2730-2735.

Chen, J., R. Wu, P.-C. Yang, J.Y. Huang, Y.P. Sher, M.H. Han, W.C. Kao, P.J. Lee, T.F. Chiu, F. Chang, Y.W. Chu, C.W. Wu, and K. Peck. 1998. Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. Genomics. 51:313–324.

Chen, R.D., L.X. Yu, A.F. Greer, H. Cheriti, and Z. Tabaeizadeh. 1994. Isolation of an osmotic stress- and abscisic acid-induced gene encoding an acidic endochitinase from *Lycopersicon chilense*. Mol. Gen. Genetics 245:195-202.

Chen, T.H.H. and N. Murata. 2002. Enhancement of tolerance of abiotic stress by metabolic engineering of betaines and other compatible solutes. Curr. Opin. Plant Biol. 5:250-257.

Cherian, S., M.P. Reddy, and R.B. Ferreira. 2006. Transgenic plants with improved dehydration-stress tolerance: progress and future prospects. Biologia Plantarum. 50:481-495.

Choi, H., J. Hong, J. Ha, J. Kang, and S.Y. Kim. 2000. ABFs, a family of ABA-responsive element binding factors. J. Biol. Chem. 275:1723-1730.

Chuang, S.E., D.L. Daniels, and F.R. Blattner. 1993. Global regulation of gene expression in *Escherichia coli*. J. Bacteriol. 175:2026-2036.

Clark, L.G., W. Zhang, and J.F. Wendel. 1995. A phylogeny of the Grass family (Poaceae) based on *ndhF* sequence data. Syst. Bot. 20:436-360.

Comai, L., A. Madlung, C. Joseffson, and A. Tyagi. 2003. Do the different parental ‘‘heteromes’’ cause genomic shock in newly formed allopolyploids? Philos. Trans. R. Soc. Lond. B. Biol. Sci. 358:1149-1155.

D'Alessio, J.M., R. Bebee, J.L. Hartley, M.C. Noon, and D. Polayes. 1992. λZipLox: Automatic subcloning of cDNA. Focus (Life Technologies). 14:76-79.

DeRisi, J., V. Iyer, and P. Brown.1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science. 278:680–686.

Dowing, W.L., F. Mauxion, M.O. Fauvarque, M.P.Reviron, D. de Vieen, N. Vartanian, and J. Giraudat. 1992. A *Brassica napus* transcript encoding a protein related to the kunitz protease inhibitor family accumulates upon water stress in leaves, not in seeds. Plant J. 2:685-693.

Dubouzet, J.G., Y. Sakuma, Y. Ito, M. Kasuga, E.G. Dubouzet, S. Miura, M. Seki, K. Shinozaki, and K. Yamaguchi-Shinozaki. 2003. *OsDREB* genes in rice, *Oryza sativa* L. Encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant J. 33:751-763.

Duncan, R.R. and R.N. Carrow. 2001. Molecular breeding for tolerance to abiotic/edaphic stresses in forage and turfgrass. Mol. Breeding of Forage Crops. 2:251-260.

Durrant, W., O. Rowland, P. Piedras, K. Hammond-Kosack, and J. Jones. 2000. cDNA-AFLP reveals a striking overlap in race specific resistance and wound response gene expression profiles. Plant Cell. 12:963-977.

Duvall, M.R. and B.R. Morton. 1996. Molecular phylogenetics of Poaceae: an expanded analysis of *rbcL* sequence data. Mol. Phylogen. Evol. 5:352-358.

Ermolaeva, M.D., M. Wu, J.A. Eisen, and S.L. Salzberg. 2003. The age of the *Arabidopsis thaliana* genome duplication. Plant Mol. Biol. 51:859-866.

Espartero, J., J.A. Pintor-Toro, and J.M. Pardo. 1994. Differential accumulation of S-adenosylmethionine synthetase transcripts in response to salt stress. Plant Mol. Biol. 25:217-227.

Fujita, M., Y. Fujita, K. Maruyama, M. Seki, K. Hiratsu, M. Ohme-Takagi, L.P. Tran, K. Yamaguchi-Shinozaki, and K. Shinozaki. 2004. A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. Plant J. 39:863-876.

Furihata, T., K. Maruyama, Y. Fujita, T. Umezawa, R. Yoshida, K. Shinozaki, and K. Yamaguchi-Shinozaki. 2006. ABA-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. Proc. Natl. Acad. Sci. USA. 103:1988-1993.

Gaut, B. 2002. Evolutionary dynamics of grass genomes. New Phytologist. 154:15-28.

Grant, D., P. Cregan, and R.C. Shoemaker. 2000. Genome organization in dicots: genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*. Proc. Natl. Acad. Sci. USA. 97:4168-4173.

Grass Phylogeny Working Group. 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). Ann. Missouri Bot. Gard. 88:373-457.

Grover, A. 1999. A novel approach for raising salt tolerant transgenic plants based on altering stress signaling through Ca^{++}/calmodulin-dependent protein phosphatase calcineurin. Curr. Sci. 76:136-137.

Guerrero, F.D., J.T. Jones, and J.E. Mullet. 1990. Turgor-responsive gene transcription and RNA levels increase rapidly when pea shoots are wilted: sequence and expression of three inducible genes. Plant Mol. Biol. 15:11-26.

Guiltinan, M.J., W.R. Marcotte, and R.S. Quatrano. 1990. A plant leucine zipper protein that recognizes an abscisic acid response element. Science. 250:267-271.

Haake, V., D. Cook, J.L. Riechmann, O. Pineda, M.F. Thomashow, and J.Z. Zhang. 2002. Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. Plant Physiol. 130:639-648.

Hanna, W.W. 1986. Induced mutations in Midiron and Tifway bermudagrasses. p.175. Agronomy Abstracts. Amer. Soc. Agron., Madison, WI.

Hanna, W.W. and J.E. Elsner. 1999. Registration of 'TifEagle' bermudagrass. Crop.Sci. 39:1258.

Hanna, W.W., R.N. Carrow, and A.J. Powell. 1997. Registration of 'Tift 94' bermudagrass. Crop Sci. 37:1012.

Harlan J.R. and J.M.J. de Wet. 1969. Sources of variation in *Cynodon dactylon* (L.) Pers. Crop Sci. 9:774-778.

Harlan, J.R., J.M.J. de Wet, K.M. Rawal, M.R. Felder, and W.L. Richardson. 1970. Cytogenetic studies in *Cynodon* L. C. Rich. (*Gramineae*). Crop Sci. 10:288-291.

Hasegawa, P.M., R.A. Bressan, J.K. Zhu, and H.J. Bohnert. 2000. Molecular biology of salinity stress responses in higher plants. Annu. Rev. Plant Physiol Plant Mol. Biol. 32:525-529.

Hauser, B., L. Pratt, and M.-M. Cordonnier-Pratt. 1997. Absolute quantification of five phytochrome transcripts in seedlings and mature plants of tomato (*Solanum lycopersicum* L.). Planta. 201:379-387.

He, P., B. Friebe, B. Gill, and J.-M. Zhou. 2003. Allopolyploidy alters gene expression in the highly stable hexaploid wheat. Plant Mol. Biol. 52:401-414.

Hilu, K.W. and L.A. Alice. 2001. A Phylogeny of Chloridoideae (Poaceae) Based on *matK* Sequences. Syst. Bot. 26: 386-405.

Hilu, K.W. and A. Esen. 1988. Prolamin size diversity in the Poaceae. Biochemical Syst. Ecol. 16:457-465.

Himmelbach, A., Y. Yang, and E. Grill. 2003. Relay and control of abscisic acid signaling. Curr. Opin. Plant Biol. 6:470-479

Hobo, T., M. Asada, Y. Kowyama, and T. Hattori. 1999. ACGT-containing abscisic acid response element (ABRE) and coupling element 3 (CE3) are functionally equivalent. Plant J. 19:679-689.

Holm, L., D.L. Plucknett, J.V. Pancho, and J.P. Herberger. 1977. The World's Worst Weeds. Ed. East-West Center, Honolulu, HI, USA.

Holm, L., J.V Pancho, J.P. Herberger, and D.L. Plucknett. 1979. A Geographical Atlas of World Weeds. Wiley-Interscience Publ., New York, NY.

Hsiao, C., S.W.L. Jacobs, N.P. Barker, and J.J. Chatterton. 1998. A molecular phylogeny of the subfamily Arundinoideae (Poaceae) based on sequences of rDNA. Australian Syst. Bot.11:41-52.

Hu, W.N., W. Kopachik, and R.N. Band. 1992. A simple, efficient method to create a cDNA library. Bio Techniques. 13:862-864.

Hughes, M.K. and A.L. Hughes. 1993. Evolution of duplicate genes in a tetraploid animal, Xenopus laevis. Mol. Biol. Evol. 10:1360-1369.

Ingram, J. and D. Bartels. 1996. The molecular basis of dehydration tolerance in plants. Annu. Rev. Plant Physiol. Plant Mol. Biol. 44:377-403.

Ishitani, M., T. Nakamura, S.Y. Han, and T. Takabe. 1995. Expression of the betaine aldehyde dehydrogenase gene in barley in response to osmotic stress and abscisic acid. Plant Mol. Biol. 27:307-315.

Jacobs, B.F., J.D. Kingston, and L.L. Jacobs. 1999. The origin of grass-dominated ecosystems. Annu. Rev. Ecol. Syst. 86:590-643.

Jiang, C., B. Iu, and J. Singh. 1996. Requirement of a CCGAC *cis*-acting element for cold induction of the *BN115* gene from winter *Brassica napus*. Plant Mol. Biol. 30:679-684.

Kang, J.Y., H.I. Choi, M.Y. Im, and S.Y. Kim. 2002. *Arabidopsis* basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling. Plant Cell. 14:343-357.

Kashkush, K., M. Feldman, and A.A. Levy. 2002. Gene loss, silencing, and activation in a newly synthesized wheat allotetraploid. Genetics. 160:1651-1659.

Kashkush, K., M. Feldman, and A.A. Levy. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat. Genet. 33:102-106.

Kavi Kishor, P.B., Z. Hong, G.H. Miao, C.A.A. Hu, and D.P.S. Venna. 1995. Overexpression of Δ'-pyrroline-5-carboxylate synthetase increases proline production and confers osmotolerance in transgenic plants. Plant Physiol. 108:1387- 1394.

Kellis, M., B.W. Birren, and E.S. Lander. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cereviseae*. Nature. 428:617-624.

Kellogg, E.A. 2000. The grasses: a case study of macroevolution. Annu. Rev. Ecol. Syst. 31:217-238.

Kellogg, E.A. 2001. Evolutionary history of the grasses. Plant Physiol. 125:1198-1205.

Kim, S., J.Y. Kang, D.I. Cho, J.H. Park, and S.Y. Kim. 2004. ABF2, an ABRE-binding bZIP factor, is an essential component of glucose signaling and its overexpression affects multiple stress tolerance. Plant J. 40:75-87.

Kneebone, W.R. and I.L. Pepper. 1982. Consumptive water use by sub-irrigated turfgrasses under desert conditions. Agron. J. 74:419-423.

Knight, H. and M.R. Knight. 2001. Abiotic stress signalling pathways: specificity and cross-talk. Trends Plant Sci. 6: 262-267.

Koizumi, M., K. Yamaguchi-Shinozaki, H. Tsuji, and K. Shinozaki. 1993. Structure and expression of two genes that encode distinct drought-inducible cysteine proteinases in *Arabidopsis thaliana*. Gene. 129:175-182.

Koornneef, M., G. Reuling, and C.M. Karssen. 1984. The isolation and characterization of abscisic acid-insensitive mutants of *Arabidopsis thaliana*. Physiol. Plant. 61:377-383.

Koornneef, M., M.L. Jorna, D.L.C. Brinkhorst-van der Swan, and C.M. Karssen. 1992. The isolation of abscisic acid (ABA)-deficient mutants by selection of induced revertants in nongerminating gibberellin sensitive lines of *Arabidopsis thaliana* (L.) Heynh. Theor. Appl. Genet. 61:385-393.

Ku, H.M., T. Vision, J. Liu, and S.D. Tanksley. 2000. Comparing sequenced segments of tomato and *Arabidopsis* genomes: large scale duplication followed by selective gene loss creates a network of synteny. Proc. Natl. Acad. Sci. USA. 97:9121-9126.

Lai, J., J. Ma, Z. Swigonova, W. Ramakrishna, E. Linton, V. Llaca, B. Tanyolac, Y.-J. Park, O.-Y. Jeong, J.L. Bennetzen, and J. Messing. 2004. Gene loss and movement in the maize genome. Genome Res. 14:1924-1931.

Leitch, A.R., Soltis D.E., Soltis P.S. Leitch I.J., and J.C. Pires. 2004. Genome downsizing in polyploidy plants. Biol. J. Linn. Soc. 82:651-663.

Liang, P. and A. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science. 257. 967-971.

Linder, H.P. 1987. The evolutionary history of the Poales/Restionales: a hypothesis. Kew Bulletin 42: 297-318.

Liu, B. and J.F. Wendel. 2003. Epigenetic phenomena and the evolution of plant allopolyploids. Mol. Phylogenet. Evol. 29:365-379.

Liu, Q., M. Kasuga, Y. Sakuma, H. Abe, S. Miura, K. Yamaguchi-Shinozaki, and K. Shinozaki. 1998. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. Plant Cell. 10:1391-1406.

Lorenzo, O., J.M. Chico, J.J. Sanchez-Serrano, and R. Solano. 2004. JASMONATE INSENSITIVE1 encodes a MYC transcription factor essential to discriminate between different jasmonate-regulated defense responses in *Arabidopsis*. Plant Cell. 16:1938-1950.

Mable, M.B. 2003. Breaking down taxonomic barriers in polyploidy research. Trends in Plant Sci. 8:582-590.

Magome H., S. Yamaguchi, A. Hanada, Y. Kamiya, and K. Oda. 2004. Dwarf and delayed-flowering 1, a novel *Arabidopsis* mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. Plant J. 37:720-729.

Mahy, G., L.P. Bruederle, B. Connors, M. van Hofwegen, and N. Vorsa. 2000. Allozyme evidence for genetic autopolyploidy and high genetic diversity in tetraploid cranberry, *Vaccinium oxycoccos* (Ericaceae). Am. J. Bot. 87:1882-1889.

Mason-Gamer, R.J., C.F. Weil, and E.A. Kellogg. 1998. Granule-bound starch synthase: structure, function and phylogenetic utility. Mol. Biol. Evol. 15:1658-1673.

Masterson, J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science. 264:421-423.

Mathews, S., R.C. Tsai, and E.A. Kellogg. 2000. Phylogenetic structure in the grass family (Poaceae): Evidence from the nuclear gene phytochrome B. Am. J. Bot. 87:96-107.

Mena, M., A. Ambrose, R.B. Meeley, S.P. Briggs, M.F. Yanofsky, and R.J. Schmidt. 1996. Diversification of C-function activity in maize flower development. Science. 274:1537-1540.

Messing, J., A.K. Bharti, W.M. Karlowski, H. Gundlach, H.R. Kim, T, Yu, F. Wei, G. Fuks, C.A. Soderlund, K.F.X. Mayer, and R.A. Wing. 2004. Sequence composition and genome organization of maize. Proc. Natl. Acad. Sci. USA. 101:14349-14354.

Mittler, R. and B.A. Zilinskas. 1994. Regulation of pea cytosolic ascorbate peroxidase and other antioxidant enzymes during the progression of drought stress and following recovery from drought. Plant J. 5:397-405.

Moore, R.C. and M.D. Purugganan. 2003. The early stages of duplicate gene evolution. Proc. Natl. Acad. Sci. USA. 100:15682-15687.

Mundy, J., K. Yamaguchi-Shinozaki, and N.H. Chua. 1990. Nuclear proteins bind conserved elements in the abscisic acid-responsive promoter of a rice *rab* gene. Proc. Natl. Acad. Sci. USA. 87:1406-1410.

Nakashima, K., T. Kiyosue, K. Yamaguchi-Shinozaki, and K. Shinozaki. 1997. A nuclear gene, *erd1*, encoding a chloroplast-targeted Clp protease regulatory subunit homolog is not only induced by water stress but also developmentally upregulated during senescence in *Arabidopsis thaliana*. Plant J. 12:851-861.

Nakashima, K., Z.K. Shinwari, Y. Sakuma, M. Seki, S. Miura, K. Yamaguchi-Shinozaki, and K. Shinozaki. 2000. Organization and expression of two *Arabidopsis DREB2* genes encoding DRE-binding proteins involved in dehydration and high-salinity-responsive gene expression. Plant Mol. Biol. 42:657-665.

Nishiyama, T., T. Fujita, T. Shin-I et al. 2003. Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. Proc. Natl Acad. Sci. USA, 100:8007–8012.

Ohno, S. 1970. Evolution by Gene Duplication. George Allen and Unwin. London, UK.

Osborn, T.C., J.C. Pires, J.A. Birchler, D.L. Auger, Z.J. Chen, H.S. Lee, L. Comai, A. madlung, R.W. Doerge, V. Colot, and R.A. Martienssen. 2003. Understanding mechanisms of novel gene expression in polyploids. Trends Genet. 19:141-147.

Ouvrard, O., F. Cellier, K. Ferrare, D. Tousch, T. Lamm, J.M. Dupuis, and F. Casse-Delbart. 1996. Identification and expression of water-stress- and abscisic acid-regulated genes in a drought-tolerant sunflower genotype. Plant Mol. Biol. 31:819-829.

Paterson, A., J. Bowers, M. Burow, X. Draye, C. G. Elsik, C. Jiang, C.S. Katsar, T. Lan, Y. Lin, R. Ming, and R.J. Wright**.** 2000. Comparative genomics of plant chromosomes. Plant Cell. 12:1523–1540.

Paterson, A.H. 2005. Polyploidy, evolutionary opportunity, and crop adaptation. Genetica. 123:191-196.

Paterson, A.H., J.E. Bowers, and B.A. Chapman. 2004. Ancient polyploidization predating divergence of the cereal, and its consequences for comparative genomics. Proc. Natl. Acad. Soc. USA. 101:9903-9908.

Peleman, J., W. Boerjan, G. Engler, J. Seurinck, J. Botterman, T. Alliotte, M. Van Montagu, and D. Inzé. 1989. Strong cellular preference in the expression of a house keeping gene of *Arabidopsis thaliana* encoding S-adenosylmethionine synthetase. Plant Cell. 1:81-93.

Pellegrineschi, A., M. Reynolds, M. Paceco, R.M. Brito, R. Almeraya, K. Yamaguchi-Shinozaki, and D. Hoisington. 2004. Stress induced expression in wheat of the *Arabidopsis thaliana DREB1A* gene delays water stress symptoms under greenhouse conditions. Genome. 47:493-500.

Perl-Treves, R. and E. Galun. 1991. The tomato Cu, Zn superoxide dismutase genes are developmentally regulated and respond to light and stress. Plant Mol. Biol. 17:745-760.

Powell, J.B., G.W. Burton, and J.R. Young. 1974. Mutations induced in vegetatively propagated turf bermudagrasses by gamma radiation. Crop Sci. 14:327-330.

Prasad, V., C.A.E. Stromberg, H. Alimohammadian, and A. Sahni. 2005. Dinosaur coprolites and the early evolution of grasses and grazers. Science. 310:1177-1180.

Qin, L., H. Overmars, J. Helder, H. Popeijus, J. van der Voort, W. Groenink, P. van Koert, A. Schots, J. Bakker, and G. Smant. 2000. An efficient cDNA-AFLP-based strategy for the

identification of putative pathogenicity factors from the potato cyst nematode *Globodera rostochiensis*. Mol. Plant Microbe Interact. 13:830-836.

Raes, J., K. Vandepoele, C. Simillion, Y. Saeys, Y. Van de Peer. 2003. Investigating ancient duplication events in the *Arabidopsis* genome. J. Struct. Funct. Genomics. 3:117-129.

Ramsey, J. and D.W. Schemske. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu. Rev. Ecol. Syst. 29:467-501.

Ramsey, J. and D.W. Schemske. 2002. Neopolyploidy in flowering plants. Annu. Rev. Ecol. Syst. 33:589-639.

Revel, F., J.P. Renard, and V. Duranthon. 1995. PCR-generated cDNA libraries from reduced numbers of mouse oocytes. Zygote. 3:241-250.

Rong, J., C. Abbey, J.E. Bowers, C.L. Brubaker, C. Chang, P.W. Chee, T.A. Delmonte, X. Ding, J.J. Garza, B.S. Marler et al. 2004. A 3347-locus genetic recombination map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics. 166:389-417.

Rothstein J.L., D. Johnson, J. Jessee, J. Skowronski, J.A. Deloia, D. Solter, and B.B. Knowles. 1993. Construction of primary and subtracted cDNA libraries from early embryos. Methods Enzymol. 225:587-610.

Sakuma, Y., Q. Liu, J.G. Dubouzet, H. Abe, K. Shinozaki, and K. Yamaguchi-Shinozaki. 2002. DNA-binding specificity of the ERF/AP2 domain of *Arabidopsis* DREBs, transcription factors involved in dehydration- and cold-inducible gene expression. Biochem. Biophys. Res. Commun. 290:998-1009.

Sambrook, J. and D.W. Russell. 2002. Molecular Cloning: A Laboratory Manual, 3rd Ed. Cold Spring Harbor Lab Press, p.11.8.

Sanderson, M.J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218-1231.

Schena, M., D. Shalon, R. Davis, and P. Brown. 1995. Quantitative monitoring of gene expression patterns with a complimentary DNA microarray. Science. 270:467–470.

Schlueter, J., P. Dixon, C. Granger, D. Grant, L. Clark, J.J. Doyle, and R.C. Shoemaker. 2004. Mining EST databases to resolve evolutionary events in major crop species. Genome. 47:868-876.

Seoighe, C. and C. Gehring. 2004. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. Trends Genet. 20:461-464.

Shalon, D., S. Smith. and P. Brown. 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res. 6:639–645.

Shen, Q. and T.H. Ho. 1995. Functional dissection of an abscisic acid (ABA)-inducible gene reveals two independent ABA-responsive complexes each containing a G-box and a novel *cis*-acting element. Plant Cell. 7:295-307.

Shen, Q., P. Zhang, and T.H. Ho. 1996. Modular nature of abscisic acid (ABA) response complexes: composite promoter units that are necessary and sufficient for ABA induction of gene expression in barley. Plant Cell. 8:1107-1119.

Shinozaki, K, K. Yamaguchi-Shinozaki, and M. Seki. 2003. Regulatory network of gene expression in the drought and cold stress responses. Curr. Opin. Plant Biol. 6:410-417.

Shinozaki, K. and K. Yamaguchi-Shinozaki. 2000. Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways. Curr. Opin. Plant Biol. 3:217-223.

Shinwari, Z.K., K. Nakashima, S. Miura, M. Kasuga, M. Seki et al. 1998. An *Arabidopsis* gene family encoding DRE/CRT binding proteins involved in low-temperature-responsive gene expression. Biochem. Biophys. Res. Commun. 250:161-170.

Short, J.M., J.M. Fernandez, J.A. Sorge, and W.D. Huse. 1988. λZAP: A bacteriophage λ expression vector with in vivo excision properties. Nucleic Acid Res. 16:7583-7600.

Simpson, S.D., K. Nakashima, Y. Narusaka, M. Seki, K. Shinozaki et al. 2003. Two different novel *cis*-acting elements of *erd1*, a clpA homologous *Arabidopsis* gene function in induction by dehydration stress and dark-induced senescence. Plant J. 33:259-270.

Soares, M., M. Bonaldo, P. Jelene, L. Su, L. Lawton, and A. Efstratiadis. 1994. Construction and characterization of normalized cDNA library. Proc. Natl. Acad. Sci. USA. 91:9228-9232.

Soltis, D.E. and P.S. Soltis. 1999. Polyploidy: recurrent formation and genome evolution. TREE. 14:348-352.

Song, K., P. Lu, K. Tank, and T.C. Osborn. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. Proc. Natl. Acad. Sci., USA. 92:7719-7723.

Stebbins, G.L. 1981. Coevolution of grasses and herbivores. Ann. Missouri Bot. Gard. 68: 75-86.

Stephens, S.G. 1951. Possible significance of duplication in evolution. Adv. Genet. 4:247-265.

Stockinger, E.J., S.J. Gilmour and M.F. Thomashow. 1997. *Arabidopsis thaliana CBF1* encodes an AP2 domain-containing transcription activator that binds to the C-repeat/DRE, a *cis*-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. Proc. Natl. Acad. Sci. USA. 94:1035-1040.

Taliaferro, C.M. 1995. Diversity and vulnerability of bermuda turfgrass species. Crop Sci. 35:327-332.

Tam, A.W., M.M. Smith, K.E. Fry, and J.W. Larrick. 1989. Construction of cDNA libraries from small numbers of cells using sequence independent primers. Nucleic Acids Res. 17:1269.

Tao, H., C. Bausch, C. Richmond, F.R. Blattner, and T. Conway. 1999. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. J. Bacteriol. 181:6425-6440

Taylor, C.B. 1996. Proline and Water Deficit: Ups, Downs, Ins, and Outs. Plant Cell. 8:1221-1224.

Thomashow, M.F. 1999. Plant cold acclimation: freezing tolerance genes and regulatory mechanisms Annu. Rev. Plant Physiol. Plant Mol. Biol. 50:571-599.

Tran, L.S.P., K. Nakashima, Y. Sakuma, S.D. Simpson, Y. Fujita et al. 2004. Functional analysis of *Arabidopsis* NAC transcription factors controlling expression of *erd1* gene under drought stress. Plant Cell. 16:2482.98.

Uno, Y., T. Furihata, H. Abe, R. Yoshida, K. Shinozaki, and K. Yamaguchi-Shinozaki. 2000. *Arabidopsis* basic leucine zipper transcription factors involved in an abscisic acid-dependent signal transduction pathway under drought and high-salinity conditions. Proc. Natl. Acad. Sci. USA. 97:11632-11637.

Velasco, R., F. Salamini, and D.I. Bartek. 1994. Dehydration and ABA increase mRNA levels and enzyme activity of cytosolic GAPDH in the resurrection plant *Craterostigma plantagimeum*. Plant Mol. Biol. 26:541-546.

Velculescu, V., L. Zhang, B. Vogelstein, and K. Kinzler. 1995. Serial analysis of gene expression. Science. 270:484-487

Vernon, D.M. and H.J. Bohnert. 1992. A novel methyltransferase induced by osmotic stress in the facultative halophyte *Mesernbryanthemum ctystallinum*. EMBO J. 11:2077-2985.

Vernon, D.M., and J.A. Ostrem, H.J. Bohnert. 1993. Stress perception and response in a facultative halophyte: the regulation of salinity-induced genes in *Mesernbryanthemum crystallinum* . Plant Cell Environ. 16:437-444.

Vision, T.J., D.G. Brown, and S.D. Tanksley. 2000. The origins of genomic duplications in *Arabidopsis*. Science. 290:2114-2117.

Wang, J., L. Tian, A. Madlung, H.S. Lee, M. Chen, J.J. Lee, B. Watson, T. Kagochi, L. Comai, and Z.J. Chen.2004. Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. Genetics. 167:1961-1973.

Weissman, S.M. 1987. Molecular genetic techniques for mapping the human genome. Mol. Biol. Med. 4:133-143.

Wendel, J.F., C.W. Stuber, M.D. Edwards, and M.M. Goodman. 1986. Duplicated chromosome segments in maize (*Zea mays* L.): Further evidence from hexokinase isozymes. Theor. Appl. Genet. 72:178-185.

Weretilnyk, E.A. and A.D. Hanson. 1990. Molecular cloning of a plant betaine-aldehyde dehydrogenase, an enzyme implicated in adaptation to salinity and drought. Proc. Natl Acad. Sci. USA. 87:2745-2749.

White, D.A. and B.A. Zilinskas. 1991. Nucleotide sequence of a complementary DNA encoding pea cytosolic copper/zinc superoxide dismutase. Plant Physiol. 96:1391-1392.

Williams, R.J., M. Bulman, A. Huttly, A. Phillips, and S. Neill. 1994. Characterization of a cDNA from *Arabidopsis thaliana* encoding a potential thiol protease whose expression is induced independently by wilting and abscisic acid. Plant Mol. Biol. 25:259-270.

Wolfe, K.H., W.-H. Li, and P.M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNAs. Proc. Natl. Acad. Sci. USA. 84:9054-9058.

Xiong, L., K.S. Schumaker, and J.-K. Zhu. 2002. Cell signaling during Cold, Drought, and Salt Stress. Plant Cell. S165-S183.

Yamaguchi-Shinozaki, K and K. Shinozaki. 1992. A novel *Arabidopsis* DNA binding protein contains the conserved motif of HMG-box proteins. Nucleic Acids Res. 20:6737.

Yamaguchi-Shinozaki, K and K. Shinozaki. 1994. A novel *cis*-acting element in an *Arabidopsis* gene is involved in responsiveness to drought, low temperature, or high-salt stress. Plant Cell. 6:251-264.

Yamaguchi-Shinozaki, K. and K. Shinozaki. 2005. Organization of *cis*-acting regulatory elements in osmotic- and cold-stress-responsive promoters. Trends Plant Sci. 10:88-94.

Yamaguchi-Shinozaki, K. and K. Shinozaki. 2006. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. Annu. Rev. Plant Biol. 57:781-803.

Yoshiba, Y, T. Kiyosue, T. Katagari, H. Ueda, T. Mizoguchi, K. Yamaguchi-Shinozaki, K. Wada, Y. Harada, and K. Shinozaki. 1995. Correlation between the induction of a gene for of Δ'-pyrroline-5carboxylate synthetase and accumulation of proline in *Arabidopsis thaliana* under osmotic stress. Plant J. 7:751-760.

Yoshiba, Y., T. Kiyosue, K. Nakashima, K. Yamaguchi-Shinozaki, and K. Shinozaki. 1997. Regulation of proline levels as an osmolyte in plants under water stress. Plant Cell Physiol. 38:1095- 1102.

Zhang, J.X., N.Y. Klueva, Z. Wang, R. Wu, T.H. Ho, H.T. Nguyen, and T.H.D. Ho. 2000. Genetic engineering for abiotic stress resistance in crop plants. In Vitro Cell Devel. Biol. Plant. 36:108-114.

Zhu, J.-K. 2002. Salt and drought stress signal transduction in plants. Annu. Rev. Plant Biol. 53:247-273.

Zhulidov, P.A., E.A. Bogdanova, A.S. Shcheglov, L.L. Vagner, G.L. Khaspekov, V.B. Kozhemyako, M.V. Matz, E. Meleshkevitch, L.L. Moroz, S.A. Lukyanov, and D.A. Shagin. 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acid Res. 32(3):e37.

Ziolkowski, P., G. Blanc, and J. Sadowski J. 2003. Structural divergence of chromosomal segments that arose from successive duplication events in the *Arabidopsis* genome. Nucleic Acids Res. 31:1339-1350.

Table 2.1. A classification of the genus *Cynodon* (Carlson and Duncan, 2003)

Epithet	Chromosome Number	Distribution
C. aethiopicus Clayton et Harlan	18, 36	East Africa: Ethiopia to Transvaal
C. arcuatus J. S. Presl ex C. B. Presl	36	Malagasy, India, S.E. Asia, S. Pacific to Australia
C. barberi Rang. Et Tad.	18	India
C. dactylon (L.) Pers.		
var. *dactylon**	36	Cosmopolitan
var. *afghanicus* Harlan et de Wet	18, 36	Afghanistan
var. *aridus* Harlan et de Wet	18	South Africa northward to Palestine, East to South India
var. *coursii* Harlan et de Wet	36	
var. *elegans* Rendle	36	Madagascar
var. *polevansii* (Stent) Harlan et de Wet	36	Southern Africa, south of lat. 13°S. near Baberspan S. Africa
C. incompletes Nees		
var. *incompletes*	18	Transvaal to Cape
var. *hirsutus* (Stent) de Wet et Harlan	18, 36	Transvaal to Cape
C. nlemfuensis Vanderyst		
var. *nlemfuensis*	18, 36	Tropical Africa
var. *robustus* Clayton et Harlan	18, 36	East Tropical Africa
C. plectostachyus (K. Schum.) Pilger	18	East Tropical Africa
C. transvaalensis Burtt-Davy*	18	South Africa
C. x *magennisii* Hurcombe*	27	South Africa

(*: species being used for turfgrass)

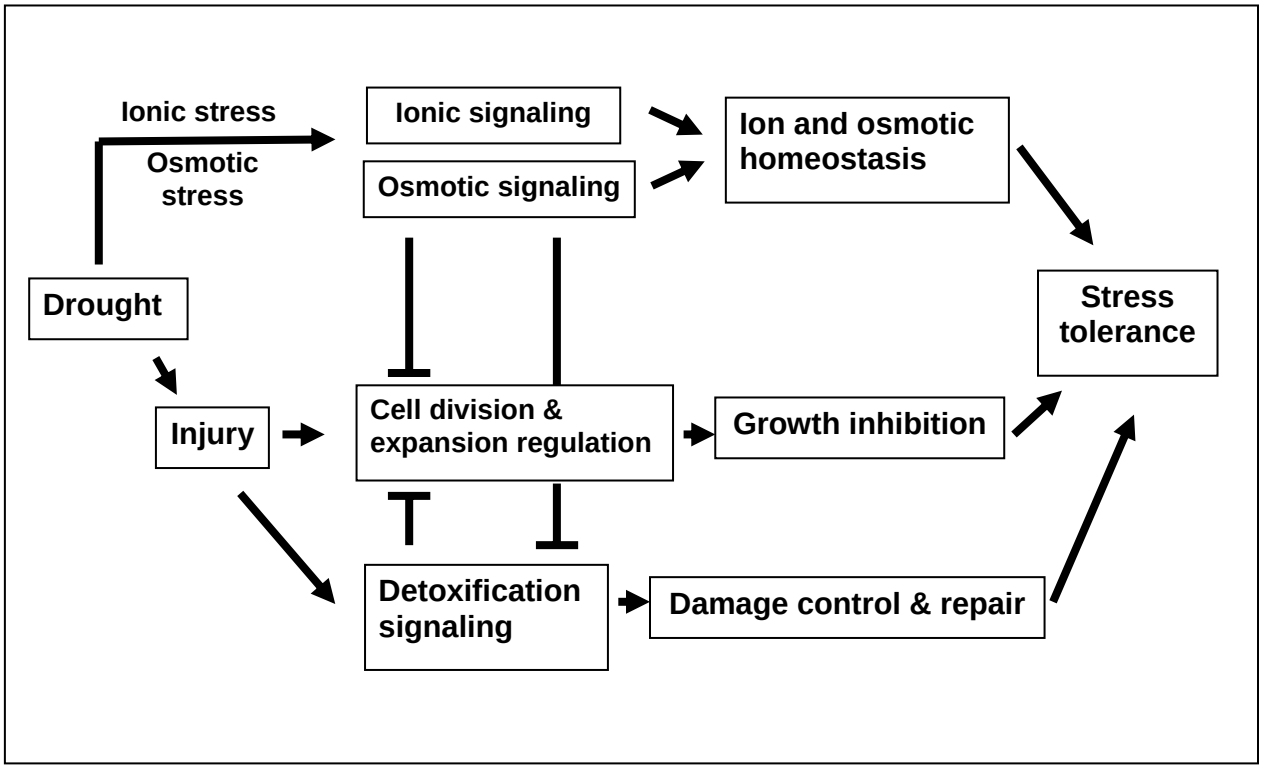


Figure 2.1. Functional demarcation of salt and drought stress signaling pathways. The inputs for ionic and osmotic signaling pathways are ionic (excess Na^+) and osmotic (e.g., turgor) changes. The output of ionic and osmotic signaling is cellular and plant homeostasis. Direct input signals for detoxification signaling are derived stresses (i.e. injury, and the signaling output is damage control and repair (e.g., activation of dehydration tolerance genes). Interactions between the homeostasis, growth regulation, and detoxification pathways are indicated (Zhu, 2002).

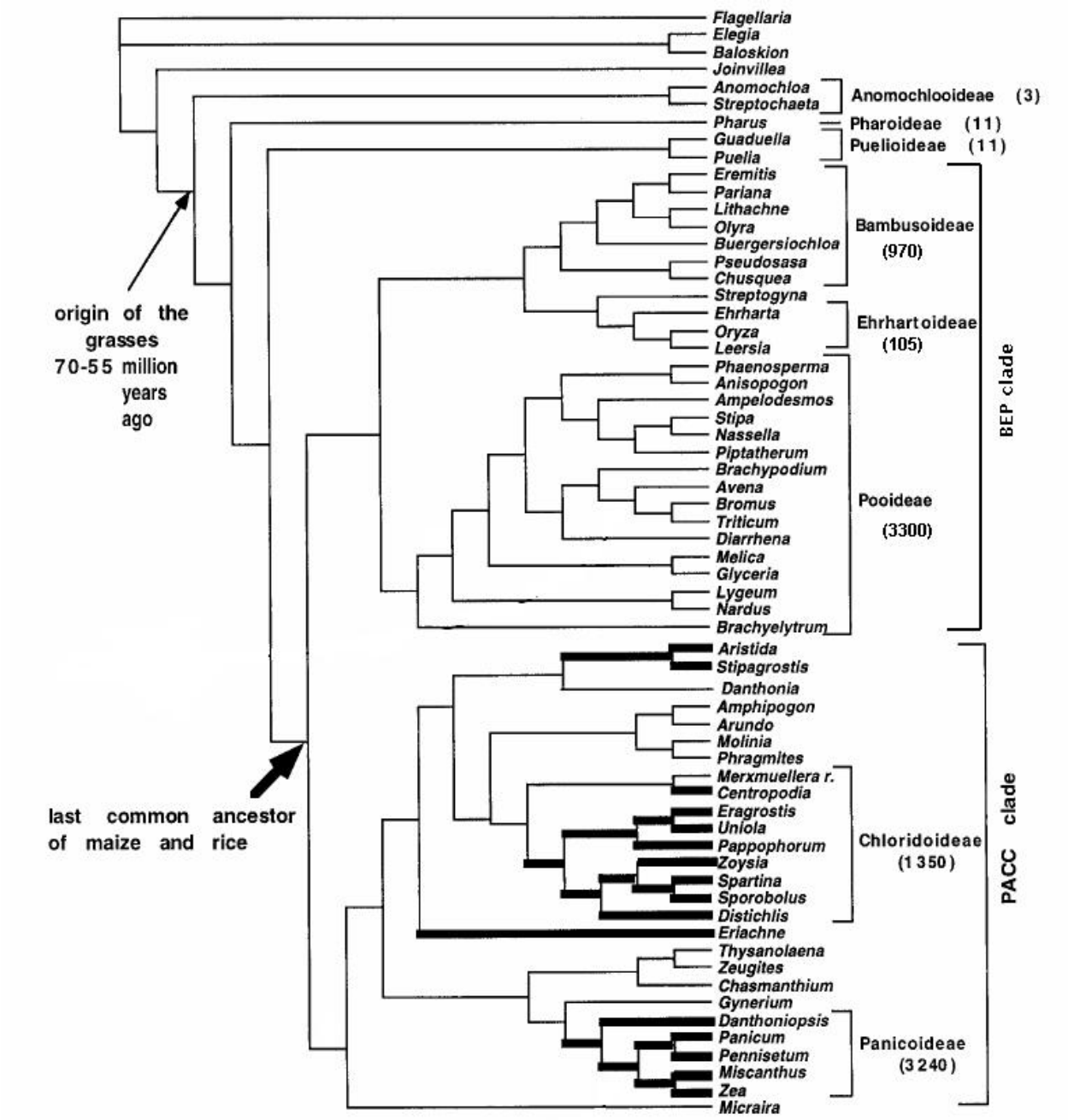


Figure 2.2. Phylogeny of the grass family based on combined data from various studies (Grass Phylogeny Working Group, 2001). Heavy lines indicate C4 photosynthesis; Numbers in parentheses indicate approximate numbers of species.

CHAPTER 3

CONSTRUCTION AND CHARACTERIZATION OF A NORMALIZED cDNA LIBRARY FROM *Cynodon dactylon* L.[1]

[1]Kim C, Kamps TL, Jang CS, Robertson JS, Feltus FA, and Paterson AH. To be submitted to *Theoretical and Applied Genetics.*

ABSTRACT

A normalized cDNA library was constructed from Bermudagrass in order to gain insight into the transcriptome of *Cynodon dactylon*. A total of 15,588 high-quality expressed sequence tags (ESTs) from the cDNA library were subjected to TIGR Gene Indices clustering tools (TGICL) to produce a unigene set. A total of 9,414 unigenes were obtained from the high-quality ESTs, including 2,467 contigs and 6,947 singletons. A total of only 39.6% of the high-quality ESTs were redundant, indicating that the normalization procedure was effective. A large-scale comparative genomic analysis of the unigenes was performed using publicly available tools such as BLAST, InterProScan, and Gene Ontology (GO). A total of 6,489 unigenes (69%) had sequence similarity with the NCBI's non-redundant (nr) plant protein database, using the BLASTX algorithm. Using the GO hierarchy, 4,608 of 6,268 BLASTX-annotated unigenes were further annotated and grouped based on their inferred functions, specific activities, and cellular localizations. Based on an InterProScan search to find known protein signatures, a total of 9,534 redundant protein motifs were assigned to 3,673 unigenes, 43 of which were not annotated during the sequence comparison against the NCBI nr database. As a result, 2,882 unigenes (31%) still remain unannotated. The unigenes were subjected to a search for EST-derived simple sequence repeats (EST-SSRs) and conserved-intron scanning primers (CISPs) which are candidate DNA markers for Bermudagrass. A total of 143 unigenes were identified to have repeat motifs and 95 primer pairs flanking the repeat motifs were designed. A total of 6,014 candidate CISP sets were also designed based on BLASTN results against the rice genome sequence. Although the candidate EST-SSRs and CISPs need to be further verified, they are expected to be useful as DNA markers for many purposes, including comparative genomic study of grass species by virtue of their significant similarities to EST sequences from other grasses.

Thus, knowledge of *Cynodon* ESTs will empower turfgrass research by providing *Cynodon* homologs for genes that are thought to confer important functions in other plants.

INTRODUCTION

Bermudagrass (*Cynodon* spp.) is one of the most prevalent turfgrasses in tropical and subtropical regions. It has been used for sports fields, lawns, parks, golf courses, and as general utility turf. Common Bermudagrass, *C. dactylon*, naturalized throughout the warmer regions of the United States, was introduced into this country during the colonial period from Africa or India. The earliest introductions are not recorded, but Mease's *Geological Account of the United States,* published in 1807, already listed Bermudagrass as one of the principal grasses in the Southern States.

C. dactylon is tetraploid, with the presence of up to four alleles per locus thought to confer advantages such as adaptation to a broader range of environments such as drought and heat conditions, or a wider range of pests, than otherwise would be possible. Its C4 photosynthetic pathway together with stoloniferous and rhizomatous growth habit enables it to produce considerable biomass, so it is often used as a forage crop. Its strong and creeping growth habit also makes *C. dactylon* a highly-competitive weed. Although some varieties have degrees of cold tolerance, Bermudagrass is generally cold susceptible. At the onset of low temperatures in the fall and winter, Bermudagrass begins to discolor, protein fractions change in composition, and reserve carbohydrates increase in the stems and rhizomes (winter dormancy). After the first killing frost, leaves and stems of Bermudagrass remain dormant until average daily temperatures rise above 50°F for several days. In warm frost-free climates, Bermudagrass remains green

throughout the year, but growth is significantly reduced with the onset of cool nights. The species exhibits its best growth where average daily temperatures are above 75°F (Beard 1973).

Turfgrass became a big business in the United States along with increased demand for recreational spaces. According to Ligon's report (1993), turfgrass seed sales in the United States were more than $600 million in 1992, higher than for any other crop except maize. If sod or sprig sales were included, the total would increase several times. Despite its economic significance, turfgrass biotechnology lags far behind that of other cereals such as corn, barley, wheat, and rice (Caetano-Anolles 1998; Chai and Sticklen 1998). Turfgrass improvement traditionally has relied on conventional breeding methods, in which the accessible genetic material is restricted by sexual reproduction. Plant biotechnologies such as DNA markers, *in vitro* tissue culture, and genetic transformation offer attractive complements. Of course, acquisition of basic molecular information must precede further study.

Since Howard Temin and David Baltimore discovered reverse transcriptase in 1970 (Baltimore 1970; Temin and Mizutani 1970), cDNA library production has been one of the best ways to gather basic molecular information, by sampling genes expressed from a particular tissue or developmental stage. One advantage of a cDNA library is that introns are spliced out and the mRNA sequence can be used as a template to create DNA, which means it contains only the coding region of a genome. However, a cDNA library tends to provide a biased sample of transcripts, favoring those of short length and high abundance. In order to overcome this drawback, it would be desirable to be able to construct cDNA libraries containing equal amounts of cDNA from each gene expressed in a given cell, tissue, or organ (normalized cDNA libraries). Since Patanjali et al. (1991) constructed a uniform-abundance ('normalized') cDNA library using

DNA reassociation kinetics (C_0t), various approaches to construct normalized cDNA library has been attempted (Soares et al. 1994; Carninci et al. 2000; Zhulidov et al. 2004).

Although Bermudagrass has high agricultural importance, there are only 4,560 ESTs available in the public domain as of March 1, 2007 (NCBI Taxonomy Browser). We constructed and sequenced a normalized cDNA library to accelerate cataloging of the expressed genes of *C. dactylon*. This EST database will serve as a foundation for gene expression and regulation studies in Bermudagrass, as well as for providing new insight into the evolutionary history of Chloridoids.

MATERIALS AND METHODS

Plant material

C. dactylon genotype T89 (PI 290869) sod was provided by Dr. Wayne Hanna at the Coastal Plain Experiment Station, Tifton, GA. The sod was grown in greenhouse conditions at temperatures of 25°C/30°C (night/day). In order to construct a normalized cDNA library, fresh leaves were harvested and stored at -80°C until use.

RNA extraction and construction of a normalized cDNA library

Total RNA was isolated from fresh leaf tissues using TRIZOL (Invitrogen, Carlsbad, CA) based on the manufacturer's instructions. Poly (A)+ RNA was extracted from total RNA using PolyATtract® mRNA Isolation Systems (Promega, Madison, WI). 1 μg mRNA was taken from the initially extracted mRNA (5 μg total amount) and subjected to first strand cDNA synthesis using the PowerScript Reverse Transcriptase (Clontech, Palo Alto, CA).

In order to minimize abundant transcripts, the first strand cDNA was normalized according to Carninci et al. (2000). The first strand cDNA as a tester and aliquots of mRNA (1 μg) as a driver were mixed and hybridized at R_0t 2.5, which has been optimized in *Arabidopsis* and *Oryza*. Under this condition, abundant transcripts such as Rubisco are expected to form RNA-DNA duplexes and rare transcripts remain single stranded. The duplexes were captured by magnetic beads and discarded. As a result, rare transcripts were used to construct the normalized cDNA library. Detailed steps are described in Figure 3.1.

After normalizing the first strand cDNA, the second strand cDNA was synthesized by primer extension. The second strand cDNA was digested by the restriction enzyme SfiIA-SfiIB and cDNA size fractionation was performed in order to eliminate cDNAs less than 500 bp in length using a CHROMA SPIN-400 column (Clontech). The digested and size-fractionated second strand cDNA was cloned directionally in the 3' to 5' orientation into the SfiIA-SfiIB restriction sites of the Lambda TriplEx2 vector (Clontech).

Mass excision, colony picking, and sequencing

The primary library was amplified once, and recombinant clones were excised in the phagemid vector λTriplEx2 by mass excision in the *Escherichia coli* SOLR strain (Stratagene, La Jolla, CA). The mass excision protocol used in this study was modified according to Eriksson et al. (2004) because the original protocol provided by Clontech does not support blue/white screening. The modified protocol significantly increased the efficiency of experimental procedures by screening false positives.

A total of 36,864 clones from the cDNA library were picked, stored in 96 384-well microtiter plates using the Qbot (Genetix, New Milton, Hampshire, UK) and archived as a

permanent community resource. A total of 21,504 clones representing the first 56 384-well plates were subjected to sequencing. The clones selected for sequencing were grown at 37°C overnight in 96-deep-well culture plates with 1.5 mL LB with 50μg/mL ampicillin per well. Plasmid DNAs were prepared by alkaline lysis with modifications for the 96-well plate format.

Sequencing reactions were carried out using ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kits Version 3.1 (Applied Biosystems, Foster City, CA). Reactions were set up in 96-well PCR plates using a fortriII (forward sequencing primer for pTriplex II, 5-AAGCGCGCCATTGTGTTGGTACCC-3) sequencing primer to generate 5' EST sequences. Cycle sequencing was performed in a PTC-100 thermocycler (MJ Research, Waltham, MA) with the folowing program: Preheat at 94°C for 5 minutes followed by 75 or 40 cycles of 94°C denaturation for 20 seconds, 50°C annealing for 5 seconds, 60°C extension for 4 minutes, and then held at 4°C. Of the 56 384-well plates sequenced, 24 plates were done based on 40 amplification cycles and 32 plates were done based on 75 cycles. Reactions were filtered through Sephadex filter plates (Krakowski et al. 1995), and then transferred directly into MicroAmp 96-well reaction plates (Applied Biosystems). High-throughput sequencing was performed on an ABI PRISM 3730xl (Applied Biosystems).

Sequence processing and annotation

All sequence outputs obtained were initially processed by a custom Perl Script that includes Phred for quality base calling (Ewing et al. 1998) and Crossmatch for trimming vector sequences (Ewing and Green 1998). The first unigene set of *C. dactylon* was established by TIGR Gene Indices Clustering tools (TGICL; Pertea et al. 2003). All sequenced EST data are

publicly available through the National Center for Biotechnology Information (NCBI, USA; GenBank dbEST accession numbers ES291835 – ES307422).

Functional characterization of the unigene data set was composed of pairwise comparison of both the high quality clones (singletons) and the consensus sequences (contigs) against public databases. All unigenes from the TGICL assembly were queried against the NCBI non-redundant (nr) plant protein reference library (638,576 proteins, as of June 20, 2007), using the NCBI standalone blastall program and the BLASTX algorithm at default settings (Altschul et al. 1997). Within each unigene, the existence of a top high-scoring pair (HSP) with an E-value below 10^{-5} was taken as indicative of significant similarity.

Protein fingerprint and motif analysis of each unigene was performed using an InterProScan wrapper program. The program was run locally to search translated unigenes versus all InterPro protein domains (as of April 1, 2007; Zdobnov and Apwiler 2001). Pfam (Finn et al. 2006), PROSITE (Hulo et al. 2006), ProDom (Bru et al. 2005), PRINTS (Attwood et al. 1998), TIGRFAMS (Haft et al. 2003) and SMART (Letunic et al. 2006) search components of the InterPro database were used with default parameters. Matches with a cutoff E-value of 10^{-5} were regarded as significant protein motifs.

Unigenes were also assigned functions according to Gene Ontology terms (Harris et al. 2004) based on BLAST definitions with GOblet's plant database (271,009 plant proteins with 154,140 GO Identifiers; http://goblet.molgen.mpg.de; Groth et al. 2004) using an E-value threshold of 10^{-10}.

Candidate DNA marker search

SSRs from the unigenes were detected with RepeatMasker software (http://www.

repeatmasker.org/) using default settings, with "DNA source:grasses," and the "only mask simple sequence repeat and low complexity DNA" option selected. EST-SSRs with at least five di-, tri-, tetra-, or penta-repeats were selected. The unigenes containing SSRs were selected using an in-house python script, and Primer3 software was used to design primers flanking SSRs (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi; Rozen and Skaletsky 2000). Sequence information was provided in FASTA format, with the "target" option used to ensure primers were designed within regions flanking the repeat motif. Default settings were used from Primer3, with "optimum primer size: 20 bp" and "optimum temperature: 60 °C." When using Primer3 software, GC % was restricted from 20 % to 80 %. Primers were designed to amplify products of size 150-500 bp.

In order to find candidate conserved-intron scanning primers (CISPs; Feltus et al. 2006), the unigenes were aligned to targeted segments of rice chromosomes 1 to 12 [The Institute for Genomic Research (TIGR) v5.0, http://www.tigr.org] using BLASTN. Perfectly conserved HSP (high-scoring segment pair) fragments that were at least 20 bp in length and no more than 2000 bp apart were extracted from the BLAST report and primers from the conserved segments were designed with a target melting temperature of 60 °C using an in-house perl script.

RESULTS

Normalized cDNA library characterization

General characteristics of the cDNA library were summarized in Table 3.1. Average insert size determined by gel electrophoresis was 1200 basepairs. A total of 21,504 (56 384-well plates) ESTs were sequenced and 15,588 passed Phred and Crossmatch quality limitations, which reflects a 73% sequencing success rate, somewhat below the 85% success rate expected

by our group. In fact, two different sequencing reaction cycles were applied to Bermudagrass EST sequencing as described in the previous section. Of the 21,504 clones sequenced, a total of 9,216 clones were done by 40-cycle-reaction whereas a total of 12,288 clones were done by 75-cycle-reaction. For unknown reasons, 40- and 75-cycle-reaction showed success rates of 87% and 61%, respectively.

The high-quality ESTs were subjected to TIGR Gene Indices clustering tools (TGICL, Pertea et al. 2003) in order to generate the first unigene set of *C. dactylon*. TGICL produced 2,467 contigs from 2,306 different clusters, and 6,947 singletons. Figure 3.2 shows the length distribution of the 9,414 unigenes. The length of contigs (667 bp) was longer than that of singletons (597 bp) (Table 3.1). Approximately 6,816 unigenes (72.4 %) were longer than 500 basepairs, while 106 unigenes (1.1 %) were shorter than 100 basepairs.

In order to reduce abundant transcripts and to increase the possibility of discovering rare transcripts, the cDNA library was normalized according to slightly modified protocol of Carninci et al. (2000) as shown in Figure 3.1. Redundancy of the normalized cDNA library was calculated to evaluate the efficiency of the normalizing process as follows:

$$\text{Redundancy (\%)} = [1 - (\text{number of unigenes}/\text{number of ESTs before TGICL assembly})] \times 100$$

The redundancy of 15,588 ESTs from the normalized cDNA library was 39.6%. For further evaluation of the efficiency, redundancies in every 1,000 ESTs added up were calculated. Figure 3.3 and 3.4 show the increasing numbers of unigenes, contigs, and singletons and the changes of redundancies in every 1,000 ESTs accumulated, respectively. There were no sharp

increases in any range but instead constant increases, which indicated that the normalization was quite efficient.

Annotation of 9,414 unigenes: BLASTX search against NCBI nr database

Sequence comparison against the NCBI nr database allowed for putative annotation of 6,489 unigenes representing 69 % of the 9,414 unigenes (Table 3.2). Among these 6,489 BLASTX-annotated unigenes, 5,539, 390, and 59 putative annotations were derived from *Oryza sativa*, *Zea mays*, and *Arabidopsis thaliana*, respectively. The taxonomic distribution of BLASTX-annotated unigenes summarized in Figure 3.5. As shown in the pie chart, BLASTX best hits for 109 unigenes failed to assign taxonomic origins and 78 different species with small numbers of best hits were combined as one group. Of the 9,414 unigenes, 2,925 unigenes did not exhibit significant similarity (E-value > 10^{-5}) to genes in the nr database.

Annotation of 9,414 unigenes: InterProScan search

Additional information on the unigenes was obtained by protein-signature scanning. InterProScan was used for sequence comparison to the InterPro database (Zdobnov and Apwiler 2001). Submission of each unigene to an InterProScan analysis facilitated the annotation of 3,673 unigenes (39 % of 9,414 unigenes) with significant protein signatures and associated InterPro numbers (E-value < 10^{-5}). A total of 9,534 redundant protein motifs composed of 1,657 unique motifs were assigned to 3,673 InterProScan-annotated unigenes. The 100 most frequently occurring domain signatures with the InterProScan identification numbers (e.g. IPR005819) are listed in Table 3.3. More than 100 occurrences of Protein kinase (IPR000719), GTPase (IPR001806), RNA-binding region RNP-1 (IPR000504), tyrosine protein kinase (IPR001245),

G-protein beta WD-40 repeat (IPR001680), and ubiquitin-conjugating enzyme (IPR000608) motifs were found in the unigene set.

In addition, the InterProScan search annotated 43 unigenes which did not show significant similarity in BLASTX search against the NCBI nr database. A total of 25 different protein motifs found in InterProScan for those 43 unigenes are listed in Table 3.4. As a result, 2,882 unigenes have not yet been annotated in any way.

Annotation of 9,414 unigenes: Gene Ontology (GO) classification

Unigenes were further annotated according to Gene Ontology terms (Harris et al. 2004) based on BLASTX definitions with GOblet's plant database (271,009 plant proteins with 154,140 GO Identifiers; http://goblet.molgen.mpg.de; Groth et al. 2004) using an E-value threshold of 10^{-10}. The Gene Ontology (GO) has three basic level annotations: molecular function, cellular component, and biological process. GO gives more detailed descriptions of genes as the level of annotation increases since GO provides a controlled vocabulary to describe gene and gene product attributes in any organism and has hierarchical descriptions of genes based on their specific activities, locations, and functions. The 9,414 unigenes were allowed for the GOblet search and 4,608 of them were further classified. Table 3.5 summarizes the GO search results. Initially, the results were divided into three basic level classifications and the percentage of each category was calculated based on the total number of annotations (with GO identification numbers) involved in each basic level classification. Since one particular unigene can have multiple GO identification numbers, 4,608 unigenes further classified by GO have 9,169 redundant GO identification numbers. A total of 1,881 of the 6,489 BLASTX-annotated unigenes were not classified by GO terms.

In molecular function (GO:0003674), binding (GO:0005488, 59.8%) and catalytic activity (GO:0003824, 56.2%) were categories containing the largest numbers of Bermudagrass unigenes. Nucleotide binding (GO:0000166, 21.7%), nucleic acid binding (GO:0003676, 19%), protein binding (GO:0005515, 11.8%), and ion binding (GO:0043167, 17.6%) functional groups comprised most of the binding category. Transferase activity (GO:0016740, 21%) and hydrolase activity (GO:0016787, 18.8%) were the most frequent functional groups in the catalytic activity.

Cellular component (GO:0005575) annotations indirectly showed the locations of annotated genes. Cell (GO:0005623, 97.3%) and organelle (GO:0043226, 62.8%) categories occupied most of the cellular component. In the cell category, intracellular (GO:0005622) and membrane (GO:0016020) component groups showed 80.2% and 35.5%, respectively. On the other hand, membrane-bound organelle (GO: 0043227) and intracellular organelle (GO:0043229) component groups occupied 49.3% and 62% of the organelle category, respectively.

The two most frequent categories in the biological process (GO:0008150) were physiological process (GO:0007582, 96.2%) and cellular process (GO:0009987, 89.3%). Metabolism (GO:0008152, 96.2%) and cellular physiological process (GO:0050875, 89.3%) groups comprised the two most significant parts of the physiological process category.

Candidate DNA markers

The 9,414 *Cynodon* unigenes were screened to identify SSRs with a minimum of five repeats of di-, tri-, tetra-, and pentanucleotides. A total of 143 non-redundant EST-SSRs were identified (1.5 %). For the 143 EST-SSRs, 41 % of the motifs were dinucluotides and 48 % were trinucleotides. The remaining 11 % were tetra- and penta-nucleotides. The most frequent

dinuclotides were GA and TC, representing 38 % and 45 % of all dinucleotides, respectively. The predominant trinucleotide motif was CCG/CGG (30 %). Primers were developed targeting EST-SSR regions for only 95 of the 143 unigenes because of insufficient flanking sequences. In addition, the 143 unigenes containing candidate EST-SSRs were searched using BLASTN against ESTs from grasses, some of which showed significant similarities (data not shown).

Conserved-intron scanning primers (CISPs) within relatively conserved exons located near exon-intron boundaries are used to scan introns for variation suitable for DNA-marker identification such as single nucleotide polymorphisms (SNPs) and insertion-deletions (INDELs) (Feltus et al. 2006). The *Cynodon* unigenes could be used to find candidate CISPs by virtue of the full genome sequence of rice (Goff et al. 2002; Yu et al. 2002). Only perfect conservation (no mismatch) of an exon was considered for CISP design since it would increase the likelihood that these primers worked in additional grasses. A total of 6,014 redundant candidate CISP sets were designed from 1,387 unigenes based on the alignment between *Cynodon* unigenes and rice chromosomes 1 to 12 using BLASTN algorithm.

DISCUSSION

Normalized cDNA library from C. dactylon

Because of the general lack of *Cynodon* ESTs detailed above, the cDNA library in this study was constructed to discover as many genes as possible in *C. dactylon*. However, a cDNA library tends to have high levels of redundant sequences such as house-keeping genes unless it is normalized or subtracted. In other words, the expression patterns of different genes in a given tissue yield mRNAs that differ in abundance, making it difficult to capture rare mRNAs from cDNA libraries. This problem also leads to redundant sequencing of multiple clones representing

the same expressed gene, thereby affecting the efficiency and cost effectiveness of the EST approach (Bonaldo et al. 1996).

In order to reduce redundant sequences and accordingly increase the discovery of different *Cynodon* genes, the cDNA library was normalized as shown in Figure 3.1. The redundancy of 15,588 ESTs generated in this study was 39.6% (9,414 unigenes), reflecting very low redundancy in that normalized cDNA libraries from other plant species reach this level of redundancy at much smaller numbers of ESTs (ca. 10,000, data not shown). For example, Cho et al. (2004) reported 5,211 leaf ESTs from a non-normalized cDNA library of wild rice (*Oryza minuta*), which resulted in 34.7% redundancy. Vogel et al. (2006) produced 8,777 unigenes from 20,440 ESTs from a non-normalized library of *Brachypodium distachyon*, which represents 57.1% redundancy. Considering the redundancy of the *C. dactylon* cDNA library graphed in Figure 3.4, the probability of finding rare transcripts from the cDNA library generated in the current study is significantly increased. Indeed, further sequencing of the library is justifiable.

Annotation of 9,414 unigenes

The 9,414 unigenes were searched using the BLASTX algorithm against the non-redundant database of GenBank (NCBI's nr database). BLAST searches indicated that 2,925 (31%) unigenes lacked similarity to an entry in the nr protein database whereas 6,489 (69%) of the unigenes were significantly matched at the threshold E-value of 10^{-5}. This is not in accordance with prior studies in many C3 grass species, in that the percentage of BLASTX-annotated genes in *C. dactylon* was less than that in C3 grass species. For example, Zhang et al. (2004) constructed and sequenced cDNA libraries in wheat (*Triticum aestivum*), which resulted in 49,963 unigenes. Primarily, they used BLASTX against NCBI nr database and found that

16.5% of the wheat unigenes were not matched to the nr protein database. In addition, Brautigam et al. (2005) reported 2,800 unigenes from a cold acclimated oat (*Avena sativa*) cDNA library. A BLASTX search against NCBI nr database was also used in the same manner. Consequently, 427 (15.3%) unigenes did not show significant similarity. In contrast, the BLASTX result of C4 grass species was slightly different from that of C3 grass species and was more similar to that of *C. dactylon*. For example, Vettore et al. (2007) reported that 27,833 (65%) of the 43,131 sugarcane unigenes were similar to known protein sequences present in the NCBI nr database at the cutoff E-value of 10^{-5}, which was not different from the BLASTX results of the *C. dactylon* cDNA library. Brendel et al. (2002) also reported that 66% of 27,294 distinct maize ESTs were significantly matched to the database at the same cutoff E-value. The discrepancy of BLASTX results between C3 and C4 grass species may be explained by the number of proteins deposited in the NCBI nr protein database and their evolution. Generally, BLASTX results with 'no hits found' are considered 'novel proteins' or 'non-functional proteins'. Since the rice genome has been fully sequenced, rice proteins may be the major component of the database. Evolutionarily, C4 grass species are only involved in the PACC (Panicoideae, Arundinoideae, Chloridoideae, and Centothecoideae) clade separated from the BEP (Bambusoideae, Ehrhartoideae, and Pooideae) clade in which rice is involved. Therefore, it can be assumed that a large part of 'no hits' in C4 grasses including *C. dactylon* may result from C4 specific novel proteins.

In addition, Figure 3.5 shows the taxonomic distribution of 6,489 unigenes with significant BLASTX matches. A total of 85.4% of the BLASTX best hits were rice proteins. Interestingly, the best hits from sorghum and sugarcane comprised parts of the pie chart although only a small number of proteins from those species are deposited in the NCBI nr protein database. This may also result from the fact that they are evolutionarily close to *C. dactylon*.

C. dactylon unigenes were also subjected to InterProScan searches of the InterPro databases in order to find known protein domains (Table 3.3). The profile of protein signatures in *A. thaliana* (Austin et al. 2004) is not remarkably different from that in *C. dactylon*: however, haem peroxidase (IPR002016) and plant peroxidase (IPR000823) motifs were highly represented in *C. dactylon*. Zhang and Kirkham (1996) reported the activities of antioxidant enzymes in the leaves of sorghum (C4) and sunflower (C3) under either watered or dry conditions. The activities of most enzymes were elevated under the dry conditions and the activities in sorghum were higher than in sunflower. Most notably, ascorbate peroxidase activity in sorghum was significantly higher both under watered and dry conditions than in sunflower. Porubleva et al. (2001) also investigated the proteome of maize (C4) leaves using 2-D electrophoresis and mass spectrometry, which revealed that a high level of peroxidase was detected in maize leaves when compared with *A. thaliana* (C3). The elevated activity of peroxidase under dry conditions might be explained by defense mechanisms of plants, but the higher activity or contents of peroxidase in C4 plants under normal conditions still need to be elucidated.

Furthermore, an InterProScan search made it possible to annotate 43 additional unigenes that were not annotated by the BLASTX search (Table 3.4). As mentioned, the BLASTX search was mostly dependent upon rice proteins due to the number of proteins deposited in the GenBank. This may reflect that the 43 unigenes are PACC clade or C4 grass specific genes and may be valuable for the evolutionary research of Chloridoideae together with 2,882 unigenes that could not be annotated both by InterProScan and BLASTX.

The GO results of *C. dactylon* are not markedly different from those of other grasses, which indicate ‘generality’ of the Bermudagrass cDNA library constructed in this study. Initially, the normalized cDNA library was constructed in order to gain insight into basic genetic or

molecular information from *C. dactylon*, which, in turn, would be used to profile gene expression under various environmental conditions and to analyze the evolution of the Chloridoideae. Considering the purpose of the cDNA library, the results described in this chapter are very promising in that the redundancy is relatively low and the profile of annotated unigenes is in accordance with previous studies; however, the data and conclusions presented in this study must be considered in light of the current state of grass genome annotations, especially *O. sativa*. At present, only about one-third of the rice genes have been assigned to functional categories. More generally, there are many genes, especially those unique to plants or eukaryotes in general, the function of which has never been investigated in any organism. Even for relatively well-studied genes, GO categories can be difficult to assign accurately. In addition, there is no experimental proof of the existence of many computer-predicted genes, and even if such inferred genes are functional, their intron/exon boundaries have not been firmly established. Many nonfunctional sequences may have been misannotated as hypothetical genes based on unusual GC content, small size, and lack of EST data (Cruveiler et al. 2004).

Candidate DNA markers

Simple sequence repeats, or microsatellites, have been widely used as molecular markers because of their abundance and high level of polymorphism. Numerous examples of *in silico* mining of SSR markers out of EST data from diverse organisms have been published over the last several years (Rohrer et al. 2002; Jany et al. 2003; Bhat et al. 2005; Yu et al. 2006). This approach can reduce the need for the costly and time-consuming benchwork. In addition to requiring less time and money to develop, EST-derived simple sequence repeat markers (EST-SSRs) have a number of advantages over genomic-derived SSRs. A total of 143 unigenes of the

9,414 *Cynodon* were identified to contain candidate EST-SSRs and primers could be designed in 95 unigenes. Some of the 143 unigenes have significant similarities to ESTs from other grasses. An example of multiple sequence alignment is shown in Figure 3.6, which shows that primers of unigene 0263 are located within conserved regions (from switchgrass, sorghum, sugarcane, and rice) flanking the repeat motif. Thus, some candidate EST-SSRs from Bermudagrass are expected to function across taxa. EST-SSRs tend to be more widely transferable between species and even genera (Chagne et al. 2004; Fraser et al. 2005). Areshchenkova and Ganal (2002) suggested that this may be because EST-SSRs are more likely to be in gene-rich euchromatic regions of chromosomes than those developed by screening of genomic libraries. The high intertaxon transferability of EST-SSRs means that even if a particular organism has no EST sequence resources available, sequences from a related species can often be used for SSR development (Barkley et al. 2005; Varshney et al. 2005). Therefore, candidate EST-SSRs identified in this study could be useful in other grasses that still lack molecular information, especially grasses in the PACC clade. In addition, trinucleotide was the most frequent motif of the 143 candidate EST-SSRs, which is another advantage of EST-SSRs. Li et al. (2004) reported that EST-SSRs are typically composed of trinucloetide repeats, which are easier to score than dinucleotide repeats.

The *Cynodon* unigenes were also subjected to a search for conserved-intron scanning primers (CISPs; Feltus et al. 2006). A total of 6,014 candidate CISP sets were designed in the *Cynodon* unigenes based on BLASTN results against rice chromosomes 1 to 12. A strength of this approach is cross-taxon utilization like that of EST-SSRs. In other words, CISPs can also provide tools suitable for linking genomics research in many grasses that lack sequence information. Since only perfectly conserved regions (no mismatch) between Bermudagrass and

rice sequences were considered to design CISP sets, the CISPs in this study are assumed to be highly conserved in other grasses. Feltus et al. (2006) reported that about one-half of CISPs from their study worked in individual taxa for which DNA sequence information was either lacking (Chloridoids) or was not considered in primer design (Pooids), and about one-third work in all Panicoid, Chloridoid, and Oryzoid grasses tested. Polymorphisms such as SNPs and INDELs could be found in introns by amplifying and sequencing CISP sets. Since introns have less evolutionary constraint than exons and should therefore be more likely to identify polymorphism. Additionally, utilization of introns can increase the scannable genome space.

Although both EST-SSRs and CISPs found in this study need to be verified by PCR and sequencing procedures, they may be useful in Bermudagrass genetics as well as in comparative genomic study of the grass family due to their high levels of similarities across taxa. Particularly, they may be highly valuable to study grasses that still lack sequence information.

Since *C. dactylon* has been inadequately studied at the level of molecular biology despite its agricultural significance, the *C. dactylon* normalized cDNA library represents an initial step towards looking at gene-specific expression in this species, and will pave the way for creation of other resources such as microarray chips that can help provide a view of global gene expression in various environmental conditions and growth stages. In addition, the results described herein can likely be extrapolated to look at other related Chloridoids that have not been studied at the molecular level. For example, candidate EST-SSRs and CISPs from *C. dactylon* may be useful in other Chloridoids because the conserved nature of coding sequences may foster cross-taxon amplification.

The results will also be useful to help understand differences in gene expression between diploid and tetraploid *Cynodon* populations as well as *Cynodon* and other well-studied grasses,

and give insight into the effects of chromosome doubling and polyploidization events in the Chloridoideae subfamily, one of the most understudied grass subfamilies.

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang A, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402

Arenshchenkova T, Ganal MW (2002) Comparative analysis of polymorphism and chromosomal location ofo tomato microsatellite markers isolated from different sources. Theor Appl Genet 104:229-235

Attwood TK, Beck ME, Flower DR, Scordis P, Selley JN (1998) The PRINTS protein fingerprint database in its fifth year. Nucleic Acids Res 26:304-308

Austin R, Provart NJ, Sacadura NR, Nugent KG, Babu M, Saville BJ (2004) A comparative genomic analysis of ESTs from *Ustilago maydis*. Funct Integr Genomics 4:207-218

Baltimore D (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. Nature. 226:1209-1211

Barkley NA, Newman ML, Wang ML, Hotchkiss MW, Pederson GA (2005) Assessment of the genetic diversity and phylogenetic relationships of a temperate bamboo collection by using transferred EST-SSR markers. Genome 48:731-737

Beard JB (1973) Turfgrass: Science and culture, Prentice-Hall Inc., Englewood Cliffs New Jersey, pp 132-142

Bhat PR, Krishnakumar V, Hendre PS, Rajendrakumar P, Varshney RK, Aggarwal RK (2005) Identification and characterization of expressed sequence tags-derived simple sequence

repeats, markers from robusta coffee variety 'C × R' (an interspecific hybrid of *Coffea canephora* × *Coffea congensis*). Mol Ecol Notes 5:80-83

Bonaldo MF, Lennon G, Soares MB (1996) Normalization and subtraction: two approaches to facilitate gene discovery. Genome Res 6:791–806

Brautigam M, Lindlof A, Zakhrabekova S, Gharti-Chhetri G, Olsson B, Olsson O (2005) Generation and analysis of 9792 EST sequences from cold acclimated oat, *Avena sativa*. BMC Plant Biol 5:18

Brendel V, Kurtz S, Walbot V (2002) Comparative genomics of Arabidopsis and maize: prospects and limitations. Genome Biol 3:1005.1-1005.6

Bru C, Courcelle E, Carrere S, Beausse Y, Dalmar S, Kahn D (2005) The ProDom database of protein domain families: more emphasis on 3D. Nucleic Acids Res 33:D212-D215

Caetano-Anolles G (1998) DNA analysis of turfgrass genetic diversity. Crop Sci 38:1415-1424

Carninci P, Shibata Y, Hayatsu N, Sugahara Y, Shibata K, Itoh M, Konno H, Okazaki Y, Muramatsu M, Hayashizaki Y (2000) Normalization and Subtraction of Cap-Trapper-Selected cDNAs to Prepare Full-Length cDNA Libraries for Rapid Discovery of New Genes. Genome Res 10:1617-1630

Chai B, Sticklen M (1998) Applications of biotechnology in turfgrass genetic improvement. Crop Sci 38:1320-8

Chagne D, Chaumeil P, Ramoer A, Collada C, Guevara A, Cervera MTG, Vendramin GG, Garcia V, Frigerio J-M, Echt C, Richardson T, Plomion C (2004) Cross-species transferability and mapping of genomic and cDNA SSRs in pines. Theor Appl Genet 109:1204-1214

Cho SK, OK SH, Jeung JU, Shim KS, Jung KW, You MK, Kang KH, Chung YS, Choi HC, Moon HP, Shin JS (2004) Comparative analysis of 5,211 leaf ESTs of wild rice (*Oryza minuta*). Plant Cell Rep 22:839-847

Cruveiller S, Jabbari K, Clay O, Bernardi G (2004) Incorrectly predicted genes in rice? Gene 333:187-188

Eriksson EM, Bovy A, Manning K, Harrison L, Andrews J, Silva JD, Tucker GA, Seymour GB (2004) Effect of the Colorless non-ripening Mutation on Cell Wall Biochemistry and Gene Expression during Tomato Fruit Development and Ripening. Plant Physiol 136:4184-4197

Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred II Error probabilities. Genome Res 8:186-194

Ewing RM, Kahla AB, Poirot O, Lopez F, Audic S, Claverie J-M (1998) Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression. Genome Res 9:950-959

Feltus FA, Singh HP, Lohithaswa HC, Schulze SR, Silva TD, Paterson AH (2006) A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved-noncoding sequences in orphan crops. Plant Physiol 140:1183-1191

Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A (2006) Pfam: clans, web tools and services. Nucleic Acids Res 34: D247-D251

Fraser LG, McNeilage MA, Tsang GK, Harvey CF, De Silva HN (2005) Cross-species amplification of microsatellite loci within dioecious, polyploidy genus *Actinidia* (Actinidiaceae). Theor Appl Genet 112:149-157

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun W, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, Wing R, Dean R, Yu Y, Zharkikh A, Shen R, Sahasrabudhe S, Thomas A, Cannings R, Gutin A, Pruss D, Reid J, Tavtigian S, Mitchell J, Eldredge G, Scholl T, Miller RM, Bhatnagar S, Adey N, Rubano T, Tusneem N, Robinson R, Feldhaus J, Macalma T, Oliphant A, Briggs S (2002) A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92-100

Groth D, Lehrach H, Hennig S (2004) GOblet: a platform for Gene Ontology annotation of anonymous sequence data. Nucleic Acids Res 32:W313-W317

Haft DH, Selengut JD, White O (2003) The TIGRFAMs database of protein families. Nucleic Acids Res 31:371-373

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M, Balakrishnan R, Cherry JM, Christie KR, Costanzo MC, Dwight SS, Engel S, Fisk DG, Hirschman JE, Hong EL, Nash RS, Sethuraman A, Theesfeld CL, Botstein D, Dolinski K, Feierbach B, Berardini T, Mundodi S, Rhee SY, Apweiler R, Barrell D, Camon E, Dimmer E, Lee V, Chisholm R, Gaudet P, Kibbe W, Kishore R, Schwarz EM, Sternberg P, Gwinn M, Hannick L, Wortman J, Berriman M,

Wood V, de la Cruz N, Tonellato P, Jaiswal P, Seigfried T, White R (2004) The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32: D258-D261

Hulo N, Bairoch A, Bulliard V, Cerutti L, De Castro E, Langendijk-Genevaux PS, Pagni M, Sigrist CJ (2006) The PROSITE database. Nucleic Acids Res 34:D227-D230

Jany JL, Bousquet J, Khasa DP (2003) Microsatellite markers for Hebeloma species developed from expressed sequence tags in the ectomycorrhizal fungus *Hebeloma cylindrosporum*. Mol Ecol Notes 3:659-661

Krakowski K, Bunville J, Seto J, Baskin D, Seto D (1995) Rapid purification of fluorescent dyelabeled products in a 96-well format for high-throughput automated DNA sequencing. Nucleic Acids Res 23:4930-4931

Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P (2006) SMART 5: domains in the context of genomes and networks. Nucleic Acids Res 34:D257-D260

Ligon PC (1993) Seeds of change. Dealer Progress Magazine Nov-Dec:29-30

Li YC, Korol AB, Fahima T, Nevo E (2004) Microsatellites within genes: structure, function, and evolution. Mol Biol Evol 21:991-1007

Patanjali SR, Parimoo S, Weissman SM (1991) Construction of a uniform-abundance (normalized) cDNA library. Proc Natl Acad Sci USA 88:1943-1947

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651-652

Porubleva L, Velden KV, Kothari S, Qliver DJ, Chitnis PR (2001) The proteome of maize leaves: Use of gene sequences and expressed sequence tag data for identification of proteins with peptide mass fingerprints. Electrophoresis 22:1724-1738

Rohrer GA, Fahrenkrug SC, Nonneman D, Tao N, Warren WC (2002) Mapping microsatellite markers identified in porcine EST sequences. Anim Genet 33:372-376

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365-386

Soares M, Bonaldo M, Jelene P, Su L, Lawton L, Efstratiadis A (1994) Construction and characterization of a normalized cDNA library. Proc Natl Acad Sci USA 91:9228-9232

Temin HM, Mizutani S (1970) RNA-dependent DNA polymerase in virions of *Rous sarcoma* virus. Nature 226:1211-1213

Varshney RK, Sigmund R, Borner A, Korzun V, Stein N, Sorrells ME, Langridge P, Graner A (2005) Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. Plant Sci 168:195-202

Vettore AL, da Silva FR, Kemper EL, Souza GM, da Silva AM, Ferro MT, Henrique-Silva F, Giglioti EA, Lemos MVF, Coutinho LL, Nobrega MP, Carrer H, Franca SC, Bacci Jr M, Goldman MHS, Gomes SL, Nunes LR, Camargo LEA, Siqueira WJ, Sluys MV, Thiemann OH, Kuramae EE, Santelli RV, Marino CL, Targon MLPN, Ferro JA, Silveira HCS, Marini DC, Lemos EGM, Monteiro-Vitorello CB, Tambor JHM, Carraro DM, Roberto PG, Martins VG, Goldman GH, de Oliveira RC, Truffi D, Colombo CA, Rossi M, de Araujo PG, Sculaccio SA, Angella A, Lima MMA, de Rosa Jr VE, Siviero F, Coscrato VE, Machado MA, Grivet L, Di Mauro SMZ, Nobrega FG, Menck CFM, Braga MDV, Telles GP, Cara FAA, Pedrosa G, Meidanis J, Arruda P (2007) Analysis and functional annotation of an expressed sequence tag collection for tropical crop sugarcane. Genome Res 13:2725-2735

Vogel JP, Gu YQ, Twigg P, Lazo GR, Laudencia-Chingcuanco D, Hayden DM, Donze TJ, Vivian LA, Stamova B, Coleman-Derr D (2006) EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*. Theor Appl Genet 113:186-195

Yu J, Sun Q, La Rota M, Edwards H, Tefera H, Sorrells ME (2006) Expressed sequence tag analysis in tef (*Eragrostis tef* (Zucc) Trotter). Genome 49:365-372

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *indica*). Science 296:79-92

Zdobnov EM, Apwiler R (2001) InterProScan: an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17:847-848

Zhang D, Choi DW, Wanamaker S, Fenton RD, Chin A, Malatrasi M, Turuspekov Y, Walia H, Akhunov ED, Kianian P, Otto C, Simons K, Deal KR, Echenique V, Stamova B, Ross K, Butler GE, Strader L, Verhey SD, Johnson R, Altenbach S, Kothari K, Tanaka C, Shah MM, Laudencia-Chingcuanco D, Han P, Miller RE, Crossman CC, Chao S, Lazo GR, Klueva N, Gustafson JP, Kianian SF, Dubcovsky J, Walker-Simmons MK, Gill KS, Dvorak J, Anderson OD, Sorrells ME, McGuire PE, Qualset CO, Nguyen HT, Close TJ

(2004) Construction and evaluation of cDNA libraries for large-scale expressed sequence tag sequencing in wheat (*Triticum aestivum* L.). Genetics 168:595-608

Zhang J, Kirkham MB (1996) Antioxidant response to drought in sunflower and sorghum seedlings. New Phytol 132:361-373

Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, Matz MV, Meleshkevitch E, Moroz LL, Lukyanov SA, Shagin DA (2004) Simple cDNA normalization using kamchatka crab duplex-specific nuclease. Nucleic Acid Res 32(3):e37

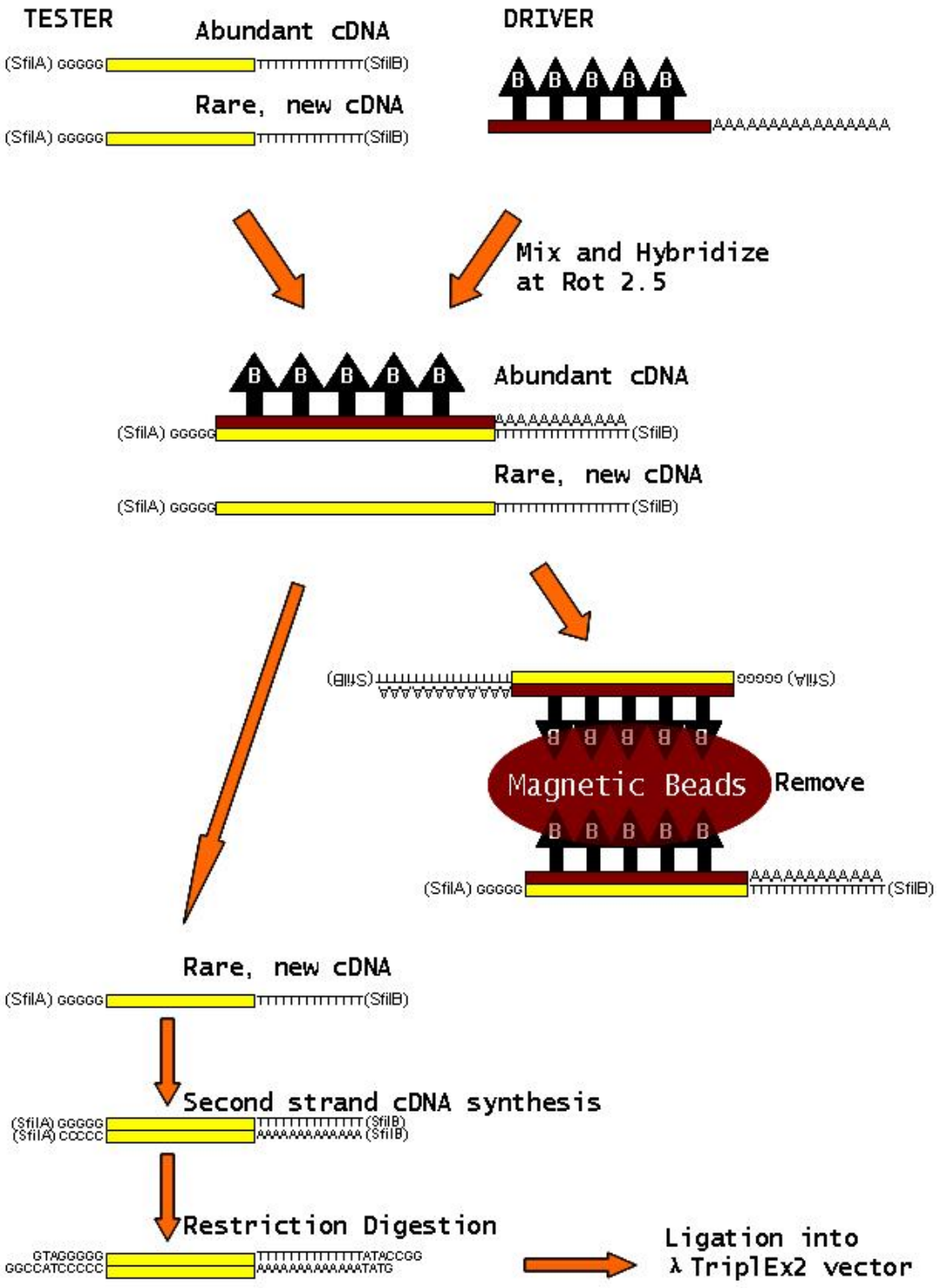


Figure 3.1. Normalization protocol applied to this study (Carninci et al. 2000). Rot value is a RNA-DNA reassociation constant. R_0t 2.5 is optimized for plant DNA structure in *Arabidopsis* and *Oryza*.

Table 3.1. Summary of the *Cynodon dactylon* normalized cDNA library.

Primary titer (pfu/mL)	**2.2×10^6**
Amplified titer (pfu/mL)	**1.1×10^{10}**
Average cDNA insert size	**1200 bp**
Total sequences	**21,504**
Sequences passed quality check	15,588 (73%)
Total good sequence length	9,303,554 bp
Average good sequence length	597 bp
Number of unigenes assembled by TGICL	**9,414**
Number of Clusters	2,306
Number of Contigs	2,467
Number of Singletons	6,947
Total unigene length	6,279,438 bp
Average unigene length	667 bp
Redundancy[a]	**39.6%**

[a]Redundancy (%) = [1-(number of unigenes/number of sequences passed quality check)] ×100

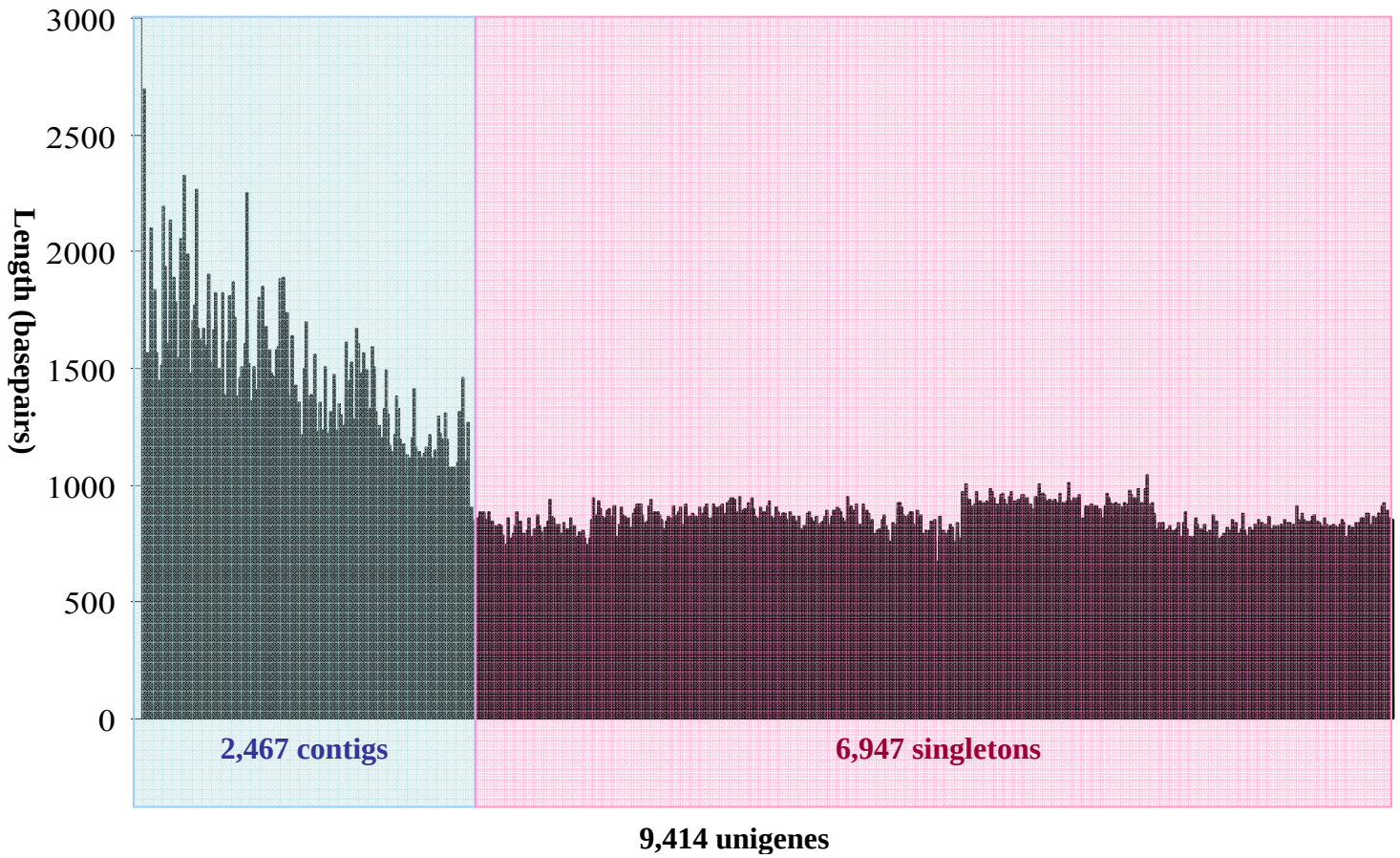


Figure 3.2. Length distribution of 9,414 unigenes.

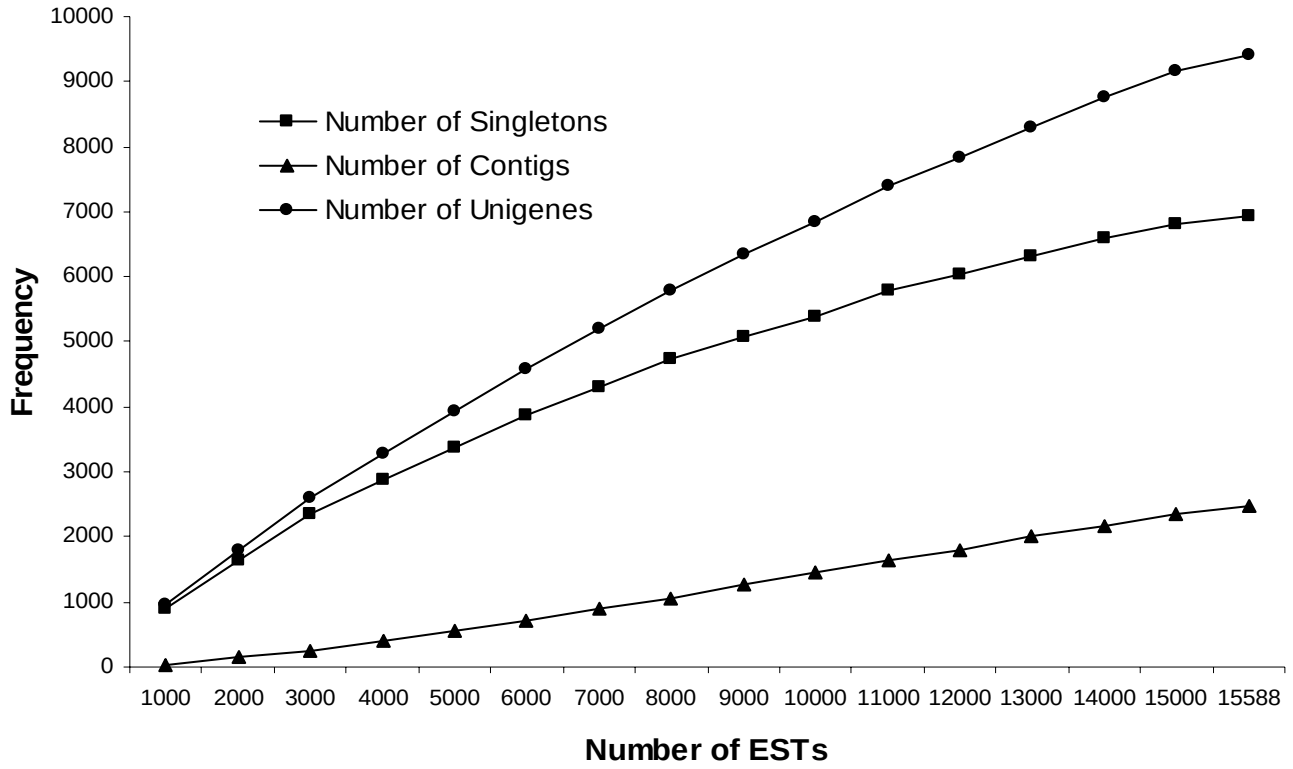


Figure 3.3. Number of contigs, singletons, and unigenes in 15,588 ESTs from *C. dactylon* normalized cDNA library.

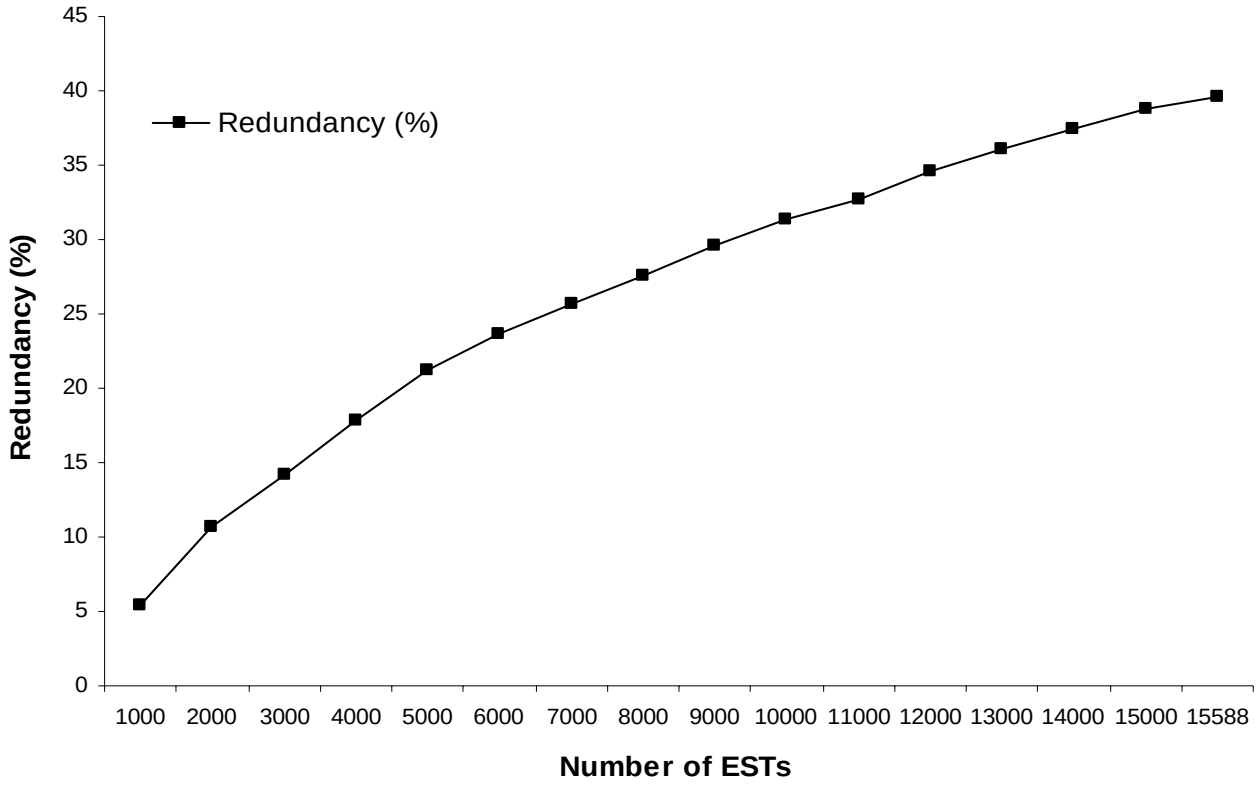
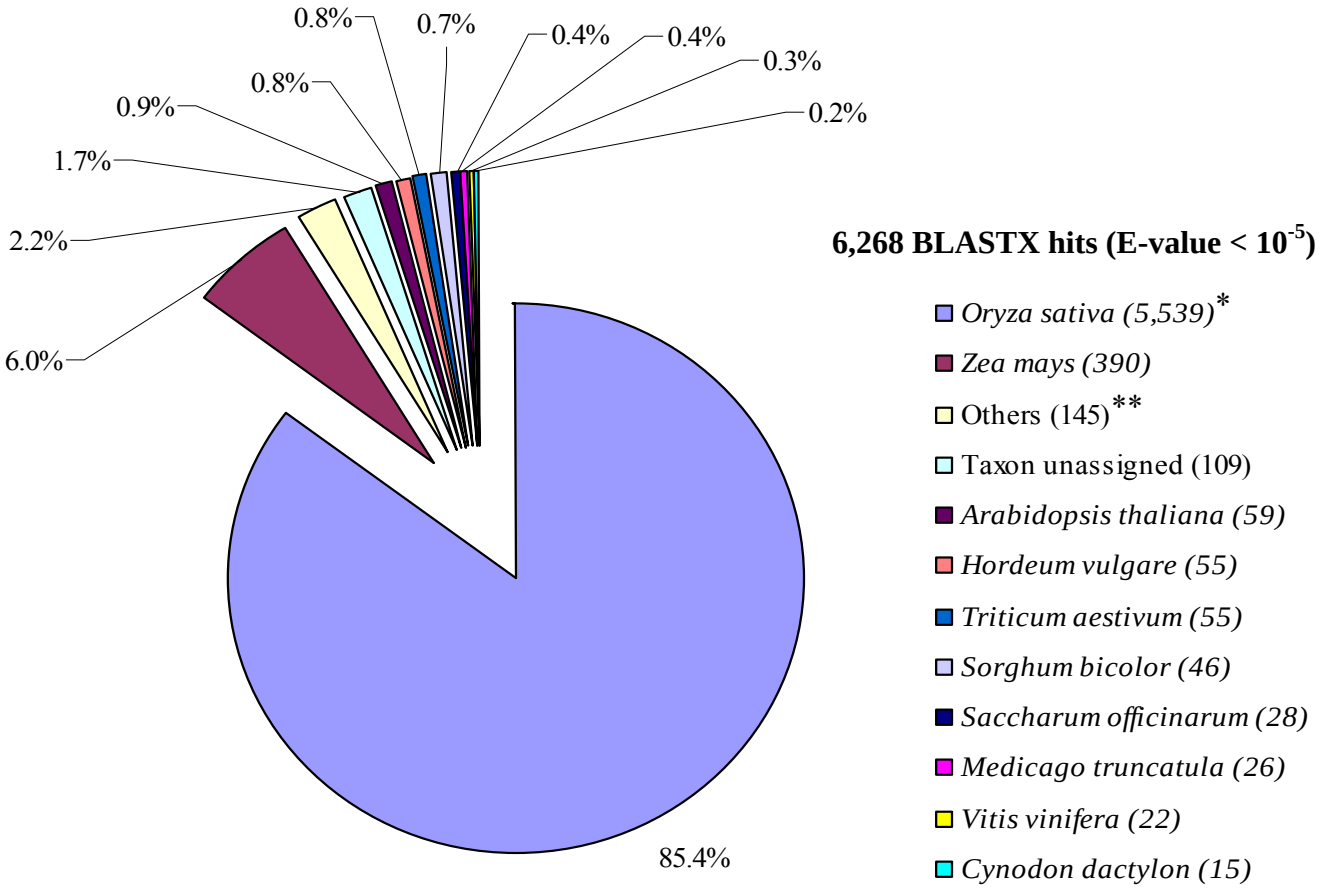


Figure 3.4. Redundancy of 15,588 ESTs from the *C. dactylon* normalized cDNA library. Redundancy (%) = [1-(number of unigenes/number of sequences passed quality check)] ×100.

Table 3.2. Summary of annotation of 9,414 unigenes.

Total number of unigenes	**9,414**	
BLASTX Search (NCBI nr database)		
Number of significant matches (E-value < 10^{-5})	6,489	
Number of no hits	2,925	
InterProScan Search		
Number of unigenes annotated by InterProScan	3,673	
Number of motifs assigned to 3,673 unigenes		9,534
Number of motifs identified by InterProScan (unique motifs only)		1,657
Number of BLASTX no hits annotated by InterProScan (E-value < 10^{-5})		43
Gene Ontology Search against GOblet's Plants Database (271,009 proteins)		
Number of unigenes with GO functional assignments (E-value < 10^{-10})	4,608	



* Numbers in parentheses indicate the number of BLASTX best hits for each species.
**'Others' includes 78 different species.

Figure 3.5. Taxonomic distribution of 6,268 unigenes showing significant BLASTX matches.

Table 3.3. The 100 most frequent protein signatures identified by the InterProScan application using 9,414 *C. dactylon* unigenes. Total 3,673 unigenes have more than one known protein domains below an E-value of 10^{-5}.

IPR ID[a]	Motifs[b]	Functional Description	IPR ID[a]	Motifs[b]	Functional Description
IPR000719	249	Protein kinase	IPR009072	35	Histone-fold
IPR001806	163	Ras GTPase	IPR001564	35	Nucleoside diphosphate kinase
IPR000504	117	RNA-binding region RNP-1	IPR003579	35	Ras small GTPase, Rab type
IPR001245	109	Tyrosine protein kinase	IPR002110	35	Ankyrin
IPR001680	105	G-protein beta WD-40 repeat	IPR012287	34	Homeodomain-related
IPR000608	100	Ubiquitin-conjugating enzyme, E2	IPR002198	34	Short-chain dehydrogenase/reductase SDR
IPR002016	81	Haem peroxidase, plant/fungal/bacterial	IPR002067	33	Mitochondrial carrier protein
IPR000425	80	Major intrinsic protein	IPR000558	33	Histone H2B
IPR001128	79	Cytochrome P450	IPR002452	31	Alpha tubulin
IPR002048	77	Calcium-binding EF-hand	IPR001509	31	NAD-dependent epimerase/dehydratase
IPR000823	76	Plant peroxidase	IPR001464	31	Annexin
IPR001951	72	Histone H4	IPR002100	30	Transcription factor, MADS-box
IPR000626	71	Ubiquitin	IPR011990	30	Tetratricopeptide-like helical
IPR012677	71	Nucleotide-binding, alpha-beta plait	IPR002453	30	Beta tubulin
IPR002130	70	Peptidyl-prolyl cis-trans isomerase	IPR000173	29	Glyceraldehyde 3-phosphate dehydrogenase
IPR012335	67	Thioredoxin fold	IPR002097	29	Profilin/allergen
IPR000894	66	RuBP carboxylase, small chain	IPR007125	28	Histone core
IPR000864	64	Proteinase inhibitor I13, potato inhibitor I	IPR001697	28	Pyruvate kinase
IPR002290	63	Serine/threonine protein kinase	IPR001650	28	Helicase, C-terminal
IPR001005	60	Myb, DNA-binding	IPR001789	28	Response regulator receiver
IPR006689	59	ARF/SAR superfamily	IPR000850	26	Adenylate kinase
IPR002347	56	Glucose/ribitol dehydrogenase	IPR012340	26	Nucleic acid-binding, OB-fold
IPR001841	54	Zinc finger, RING-type	IPR006649	26	ribonucleoprotein, eukaryotic/archaea-type
IPR000528	52	Plant lipid transfer protein/Par allergen	IPR001424	26	Superoxide dismutase, copper/zinc binding
IPR002401	52	Cytochrome P450, E-class, group I	IPR000007	26	Tubby, C-terminal
IPR001360	52	Glycoside hydrolase, family 1	IPR000795	25	Protein synthesis factor, GTP-binding
IPR001476	51	Chaperonin Cpn10	IPR003577	25	Ras small GTPase, Ras type
IPR004000	51	Actin/actin-like	IPR003578	25	Ras small GTPase, Rho type
IPR001395	50	Aldo/keto reductase	IPR000146	25	Inositol phosphatase
IPR002119	49	Histone H2A	IPR000702	25	Ribosomal protein L6
IPR001993	49	Mitochondrial substrate carrier	IPR013785	24	Aldolase-type TIM barrel
IPR001471	49	Pathogenesis-related transcriptional factor	IPR000582	24	Acyl-coA-binding protein, ACBP
IPR000164	47	Histone H3	IPR005455	24	Plant profiling
IPR000217	42	Tubulin	IPR000668	24	Peptidase C1A, papain C-terminal
IPR000308	41	14-3-3 protein	IPR003959	23	AAA ATPase, core
IPR002885	41	Pentatricopeptide repeat	IPR011989	23	Armadillo-like helical
IPR001563	40	Peptidase S10, serine carboxypeptidase	IPR001289	23	CCAAT-binding transcription factor
IPR001344	39	Chlorophyll A-B binding protein	IPR002207	22	Plant ascorbate peroxidase
IPR013753	39	Ras	IPR003245	22	Plastocyanin-like
IPR000008	39	C2 calcium-dependent membrane targeting	IPR000266	22	Ribosomal protein S17
IPR007118	39	Expansin/Lol pI	IPR001709	21	Flavoprotein cytochrome reductase
IPR006186	39	S/T-specific protein phosphatase	IPR006662	21	Thioredoxin-related
IPR001199	39	Cytochrome b5	IPR001781	21	LIM, zinc-binding
IPR001752	38	Kinesin, motor region	IPR001623	21	Heat shock protein DnaJ, N-terminal
IPR003612	37	Plant lipid transfer protein	IPR001164	21	Arf GTPase activating protein
IPR005795	37	Major pollen allergen Lol pI	IPR008162	20	Inorganic pyrophosphatase
IPR005225	36	Small GTP-binding protein domain	IPR001087	20	Lipolytic enzyme, G-D-S-L
IPR000941	36	Enolase	IPR003439	20	ABC transporter related
IPR013781	36	Glycoside hydrolase, catalytic core	IPR009072	35	Histone-fold
IPR005819	35	Histone H5	IPR001564	35	Nucleoside diphosphate kinase

[a]InterProScan Identifiers ; [b]Number of protein motifs assigned

Table 3.4. 25 different protein motifs further annotated by InterProScan for 43 unigenes that were not significantly matched (No hits) in a BLASTX search (E-value < 10^{-5}).

IPR ID[a]	Motifs[b]	Functional Description
IPR005819	9	Histone H5
IPR000347	4	Plant metallothionein, family 15
IPR007834	4	DSS1/SEM1
IPR007836	2	Ribosomal protein L41
IPR001813	2	Ribosomal protein 60S
IPR004184	2	Pyruvate formate-lyase, PFL
IPR010800	2	Glycine rich
IPR001878	1	Zinc finger, CCHC-type
IPR005124	1	Vacuolar (H+)-ATPase G subunit
IPR001437	1	Transcription elongation factor, GreA/GreB region, prokaryotic
IPR005516	1	Remorin, C-terminal region
IPR007608	1	Protein of unknown function DUF584
IPR005174	1	Protein of unknown function DUF295
IPR012866	1	Protein of unknown function DUF1644
IPR009806	1	Photosystem II protein PsbW, class 2
IPR005770	1	Phosphonate-binding periplasmic protein
IPR002130	1	Peptidyl-prolyl cis-trans isomerase, cyclophilin-type
IPR012906	1	PaaX-like, N-terminal
IPR007513	1	Four F5 protein
IPR000259	1	Fimbrial protein
IPR000167	1	Dehydrin
IPR008276	1	Concentrative nucleoside transporter
IPR002543	1	Cell divisionFtsK/SpoIIIE
IPR000515	1	Binding-protein-dependent transport systems inner membrane component
IPR001188	1	Bacterial periplasmic spermidine/putrescine-binding protein
Total	**43**	

[a]InterProScan Identifiers; [b]Number of protein motifs assigned

Table 3.5. Gene ontology (GO) mappings of 9,414 unigenes using GOblet's plants database. Note that individual GO categories can have multiple mappings.

Categories and subcategories	GO ID	Representation	Percentage of total
(a) Molecular function	**GO:0003674**	**4092**	**100**
motor activity	GO:0003774	31	0.8
microtubule motor activity	GO:0003777	30	0.7
catalytic activity	GO:0003824	2300	56.2
glycogen debranching enzyme activity	GO:0004133	4	0.1
helicase activity	GO:0004386	75	1.8
transposase activity	GO:0004803	6	0.1
3,4-dihydroxy-2-butanone-4-phosphate synthase activity	GO:0008686	2	0.0
integrase activity	GO:0008907	2	0.0
oxidoreductase activity	GO:0016491	427	10.4
transferase activity	GO:0016740	861	21.0
hydrolase activity	GO:0016787	771	18.8
lyase activity	GO:0016829	133	3.3
isomerase activity	GO:0016853	105	2.6
ligase activity	GO:0016874	123	3.0
signal transducer activity	GO:0004871	135	3.3
two-component sensor molecule activity	GO:0000155	38	0.9
two-component response regulator activity	GO:0000156	19	0.5
receptor activity	GO:0004872	99	2.4
receptor signaling protein activity	GO:0005057	7	0.2
receptor binding	GO:0005102	5	0.1
structural molecule activity	GO:0005198	243	5.9
structural constituent of ribosome	GO:0003735	200	4.9
structural constituent of cell wall	GO:0005199	19	0.5
structural constituent of cytoskeleton	GO:0005200	21	0.5
transporter activity	GO:0005215	380	9.3
amine transporter activity	GO:0005275	8	0.2
lipid transporter activity	GO:0005319	2	0.0
organic acid transporter activity	GO:0005342	14	0.3
oxygen transporter activity	GO:0005344	3	0.1
carrier activity	GO:0005386	250	6.1
intracellular transporter activity	GO:0005478	5	0.1
electron transporter activity	GO:0005489	102	2.5
protein transporter activity	GO:0008565	30	0.7
ion transporter activity	GO:0015075	157	3.8
carbohydrate transporter activity	GO:0015144	21	0.5
drug transporter activity	GO:0015238	9	0.2
channel or pore class transporter activity	GO:0015267	24	0.6
ATPase activity, coupled to movement of substances	GO:0043492	58	1.4

Table 3.5. Continued.

Categories and subcategories	GO ID	Representation	Percentage of total
(a) Molecular function	**GO:0003674**	**4092**	**100**
binding	GO:0005488	2447	59.8
nucleotide binding	GO:0000166	886	21.7
pattern binding	GO:0001871	5	0.1
nucleic acid binding	GO:0003676	777	19.0
chromatin binding	GO:0003682	8	0.2
steroid binding	GO:0005496	1	<0.1
protein binding	GO:0005515	483	11.8
lipid binding	GO:0008289	39	1.0
selenium binding	GO:0008430	5	0.1
vitamin binding	GO:0019842	26	0.6
carbohydrate binding	GO:0030246	54	1.3
carboxylic acid binding	GO:0031406	9	0.2
peptide binding	GO:0042277	1	<0.1
ribonucleoprotein binding	GO:0043021	2	0.1
ion binding	GO:0043167	721	17.6
amine binding	GO:0043176	9	0.2
tetrapyrrole binding	GO:0046906	58	1.4
cofactor binding	GO:0048037	126	3.1
metal cluster binding	GO:0051540	7	0.2
antioxidant activity	GO:0016209	54	1.3
glutathione-disulfide reductase activity	GO:0004362	1	<0.1
peroxidase activity	GO:0004601	52	1.3
thioredoxin-disulfide reductase activity	GO:0004791	1	<0.1
enzyme regulator activity	GO:0030234	36	0.9
enzyme inhibitor activity	GO:0004857	16	0.4
enzyme activator activity	GO:0008047	4	0.1
kinase regulator activity	GO:0019207	5	0.1
phosphatase regulator activity	GO:0019208	7	0.2
GTPase regulator activity	GO:0030695	8	0.2
transcription regulator activity	GO:0030528	184	4.5
two-component response regulator activity	GO:0000156	19	0.5
transcription factor activity	GO:0003700	281	6.9
RNA polymerase II transcription factor activity	GO:0003702	4	0.1
transcriptional elongation regulator activity	GO:0003711	6	0.1
transcription cofactor activity	GO:0003712	31	0.8
transcription termination factor activity	GO:0003715	1	0.0
transcriptional activator activity	GO:0016563	8	0.2
transcriptional repressor activity	GO:0016564	14	0.3
transcription initiation factor activity	GO:0016986	7	0.2

Table 3.5. Continued.

Categories and subcategories	GO ID	Representation	Percentage of total
(a) Molecular function	**GO:0003674**	**4092**	**100**
translation regulator activity	GO:0045182	60	1.5
translation factor activity, nucleic acid binding	GO:0008135	108	2.6
nutrient reservoir activity	GO:0045735	11	0.3
(b) Cellular component	**GO:0005575**	**1882**	**100**
extracellular region	GO:0005576	34	1.8
apoplast	GO:0048046	20	1.1
cell	GO:0005623	1832	97.3
cell fraction	GO:0000267	1	0.1
intracellular	GO:0005622	1510	80.2
cell surface	GO:0009986	3	0.2
membrane	GO:0016020	668	35.5
external encapsulating structure	GO:0030312	32	1.7
periplasmic space	GO:0042597	5	0.3
cell projection	GO:0042995	7	0.4
virion	GO:0019012	9	0.5
viral capsid	GO:0019028	7	0.4
viral envelope	GO:0019031	2	0.1
membrane-enclosed lumen	GO:0031974	15	0.8
organelle lumen	GO:0043233	15	0.8
envelope	GO:0031975	177	9.4
cell envelope	GO:0030313	2	0.1
organelle envelope	GO:0031967	175	9.3
organelle	GO:0043226	1181	62.8
vesicle	GO:0031982	36	1.9
membrane-bound organelle	GO:0043227	928	49.3
non-membrane-bound organelle	GO:0043228	333	17.7
intracellular organelle	GO:0043229	1167	62.0
organelle lumen	GO:0043233	15	0.8
protein complex	GO:0043234	548	29.1
phosphopyruvate hydratase complex	GO:0000015	31	1.6
exocyst	GO:0000145	17	0.9
1,3-beta-glucan synthase complex	GO:0000148	17	0.9
ubiquitin ligase complex	GO:0000151	11	0.6
exosome (RNase complex)	GO:0000178	1	0.1
proteasome complex (sensu Eukaryota)	GO:0000502	32	1.7
nucleosome	GO:0000786	44	2.3
nuclear pore	GO:0005643	95	5.0

Table 3.5. Continued.

Categories and subcategories	GO ID	Representation	Percentage of total
(b) Cellular component	**GO:0005575**	**1882**	**100**
protein complex (Continued)	GO:0043234	548	29.1
transcription factor complex	GO:0005667	44	2.3
signal peptidase complex	GO:0005787	8	0.4
heterotrimeric G-protein complex	GO:0005834	1	0.1
fatty acid synthase complex	GO:0005835	3	0.2
proteasome regulatory particle (sensu Eukaryota)	GO:0005838	8	0.4
eukaryotic translation initiation factor 3 complex	GO:0005852	2	0.1
eukaryotic translation elongation factor 1 complex	GO:0005853	3	0.2
microtubule associated complex	GO:0005875	35	1.9
unlocalized protein complex	GO:0005941	26	1.4
6-phosphofructokinase complex	GO:0005945	20	1.1
protein kinase CK2 complex	GO:0005956	5	0.3
glycine cleavage complex	GO:0005960	8	0.4
glycine dehydrogenase complex (decarboxylating)	GO:0005961	3	0.2
voltage-gated potassium channel complex	GO:0008076	9	0.5
signalosome complex	GO:0008180	19	1.0
F-actin capping protein complex	GO:0008290	2	0.1
cytochrome b6f complex	GO:0009512	1	0.1
photosystem I	GO:0009522	26	1.4
photosystem II	GO:0009523	20	1.1
oxygen evolving complex	GO:0009654	20	1.1
prefoldin complex	GO:0016272	8	0.4
eukaryotic 43S preinitiation complex	GO:0016282	2	0.1
myosin	GO:0016459	17	0.9
proton-transporting two-sector ATPase complex	GO:0016469	107	5.7
hydrogen-translocating V-type ATPase complex	GO:0016471	2	0.1
chromatin remodeling complex	GO:0016585	17	0.9
DNA-directed RNA polymerase II, holoenzyme	GO:0016591	3	0.2
light-harvesting complex	GO:0030076	1	0.1
ribonucleoprotein complex	GO:0030529	252	13.4
Mre11 complex	GO:0030870	12	0.6
RNA polymerase complex	GO:0030880	2	0.1
NADH dehydrogenase complex (quinone)	GO:0030964	6	0.3
mitochondrial intermembrane space protein transporter complex	GO:0042719	6	0.3
GPI-anchor transamidase complex	GO:0042765	9	0.5
receptor complex	GO:0043235	1	0.1
oxoglutarate dehydrogenase complex	GO:0045252	4	0.2

Table 3.5. Continued.

Categories and subcategories	GO ID	Representation	Percentage of total
(b) Cellular component	**GO:0005575**	**1882**	**100**
protein complex (Continued)	GO:0043234	548	29.1
pyruvate dehydrogenase complex	GO:0045254	6	0.3
proton-transporting ATP synthase complex	GO:0045259	9	0.5
proton-transporting ATP synthase complex, catalytic core	GO:0045261	9	0.5
proton-transporting ATP synthase complex, coupling factor	GO:0045263	1	0.1
respiratory chain complex I	GO:0045271	4	0.2
respiratory chain complex III	GO:0045275	5	0.3
ubiquinol-cytochrome-c reductase complex	GO:0045285	5	0.3
ribulose bisphosphate carboxylase complex	GO:0048492	11	0.6
(c) Biological process	**GO:0008150**	**3195**	**100**
reproduction	GO:0000003	8	0.3
sexual reproduction	GO:0019953	8	0.3
development	GO:0007275	40	1.3
pattern specification	GO:0007389	1	<0.1
aging	GO:0007568	2	0.1
morphogenesis	GO:0009653	1	<0.1
post-embryonic development	GO:0009791	5	0.2
cell differentiation	GO:0030154	9	0.3
regulation of gene expression, epigenetic	GO:0040029	14	0.4
root development	GO:0048364	2	0.1
meristem development	GO:0048507	1	<0.1
organ development	GO:0048513	2	0.1
regulation of development	GO:0050793	1	<0.1
physiological process	GO:0007582	3073	96.2
metabolism	GO:0008152	2603	81.5
photosynthesis	GO:0015979	63	2.0
death	GO:0016265	57	1.8
homeostasis	GO:0042592	18	0.6
regulation of physiological process	GO:0050791	359	11.2
coagulation	GO:0050817	1	0.03
organismal physiological process	GO:0050874	6	0.2
cellular physiological process	GO:0050875	2801	87.7
localization	GO:0051179	626	19.6
cellular process	GO:0009987	2853	89.3
cell communication	GO:0007154	137	4.3
cell adhesion	GO:0007155	7	0.2
cell differentiation	GO:0030154	9	0.3

Table 3.5. Continued.

Categories and subcategories	GO ID	Representation	Percentage of total
(c) Biological process	**GO:0008150**	**3195**	**100**
cellular process (Continued)	GO:0009987	2853	89.3
regulation of cellular process	GO:0050794	353	11.0
cellular physiological process	GO:0050875	2801	87.7
growth	GO:0040007	1	0.03
cell growth	GO:0016049	1	0.03
regulation of biological process	GO:0050789	388	12.1
regulation of gene expression, epigenetic	GO:0040029	14	0.4
positive regulation of biological process	GO:0048518	8	0.3
negative regulation of biological process	GO:0048519	26	0.8
regulation of enzyme activity	GO:0050790	23	0.7
regulation of physiological process	GO:0050791	359	11.2
regulation of development	GO:0050793	1	0.03
regulation of cellular process	GO:0050794	353	11.0
response to stimulus	GO:0050896	267	8.4
response to stress	GO:0006950	220	6.9
response to external stimulus	GO:0009605	32	1.0
response to biotic stimulus	GO:0009607	96	3.0
response to abiotic stimulus	GO:0009628	119	3.7
response to endogenous stimulus	GO:0009719	78	2.4
detection of stimulus	GO:0051606	3	0.1
interaction between organisms	GO:0051704	4	0.1
interspecies interaction between organisms	GO:0044419	4	0.1
physiological interaction between organisms	GO:0051706	1	<0.1

(a) 4,092 clusters generated 13,301 multiple mappings. Percentage representation is based on 4,092. (b) 1,882 clusters generated 9,732 multiple mappings. Percentage representation is based on 1,882. (c) 3,195 clusters generated 17,859 multiple mappings. Percentage representation is based on 3,195.

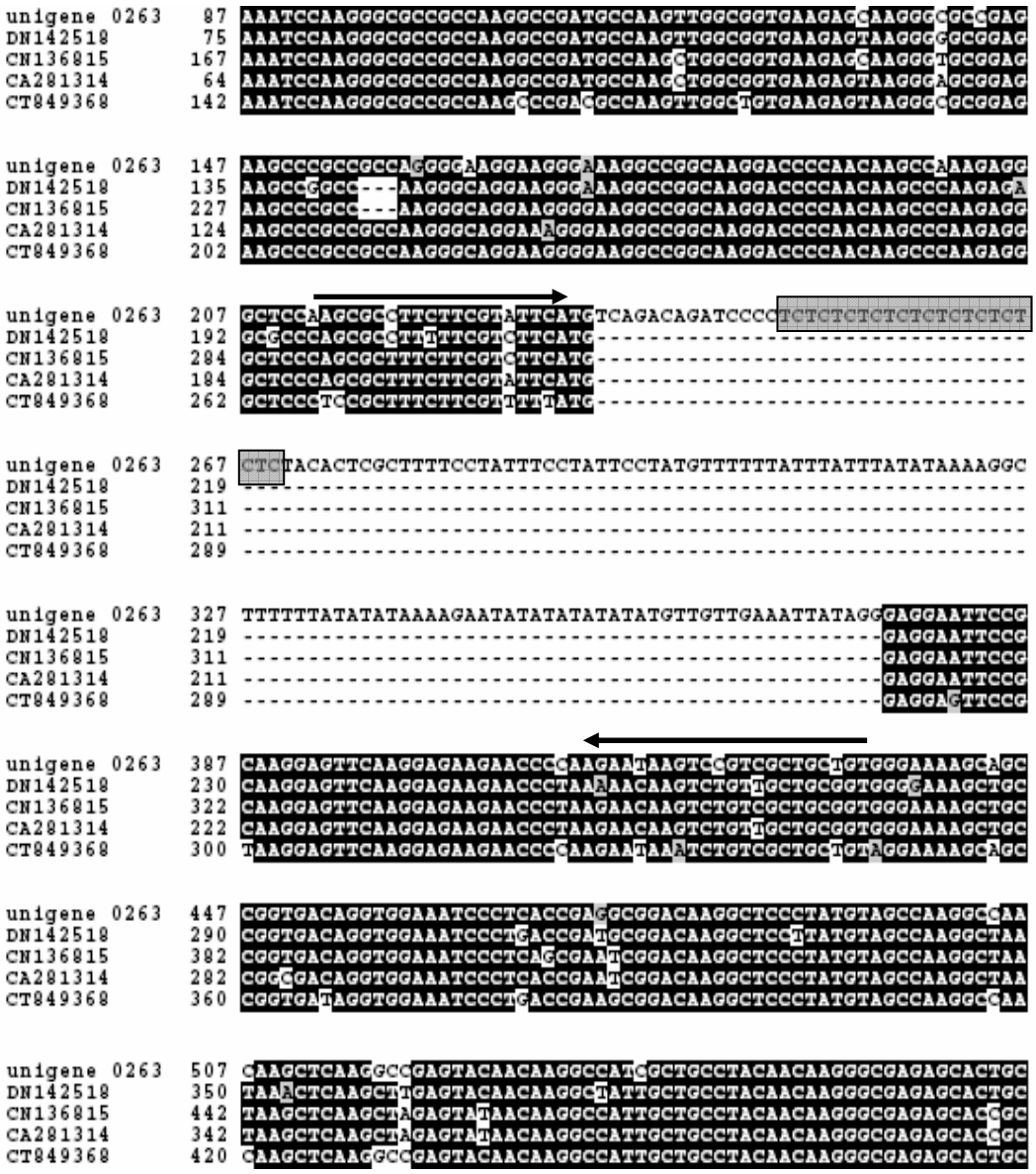


Figure 3.6. Multiple sequence alignment of unigene 0263 with nucleotide sequences from *Panicum virgatum* (DN142518), *Sorghum bicolor* (CN136815), *Saccharum officinarum* (CA291314), and *Oryza sativa* (CT849368). Arrows above nucleotide sequence indicate forward and reverse primers, and shaded box indicates the repeat motif. ClustalW (http://www.ebi.ac.uk/clustalw/index.html) was used for alignment of sequence with "gap open penalty: 10" and "gap extention penaly: 1". BOXSHADE 3.21 (http://www.ch.embnet.org/software/BOX_form.html) software was used for displaying multiple sequence alignment.

CHAPTER 4

PROFILE OF DROUGHT STRESS-RESPONSIVE GENES IN *Cynodon dactylon* L.[1]

[1]Kim C, Lemke C, and Paterson AH To be submitted to *Plant, Cell & Environment.*

ABSTRACT

Water deficit is one of the main abiotic factors that affect warm-season turfgrasses in subtropical regions. Common Bermudagrass (*Cynodon dactylon*), one of the most prevalent turfgrasses used in the southern United States, is a moderately drought-tolerant species. To identify genes induced during the water stress response in Bermudagrass, cDNA macroarrays including 4,608 clones from a normalized cDNA library were used for expression profiling. Bermudagrass was grown on desalted sand media with a nutrient solution, and drought stress was imposed by PEG 8,000 (polyethyleneglycol molecular weight 8,000) to minimize extraneous variables such as soil physical properties and soil nutritional status. The experiment included three different treatments and four biological replications. At three different time points, mRNA was extracted from each plant sample and used as probe for macroarray hybridization. The macroarray analysis identified 189 drought-responsive candidate genes from *C. dactylon*, of which 120 were up-regulated and 69 were down-regulated. BLASTX annotation suggested that up-regulated genes may be involved in osmotic adjustment, signal transduction pathways, protein repair systems, and removal of toxins, while down-regulated genes were mostly related to basic plant metabolism such as photosynthesis and glycolysis. Several GO functional categories such as peroxidase activity, two-component sensor molecule activity, response to abiotic stress, and response to external stimulus were specific to up-regulated genes, whereas photosystem I and some transporter activities were specific to down-regulated genes. The experimental procedures and analytical methods are effective tools to profile transcripts responsive to a variety of plant growth conditions. Identifying *Cynodon* genes that are turned on or off in response to drought stress is an important step toward enhanced stress tolerance by identifying genes that could be deployed for testing by many biotechnology-based approaches.

INTRODUCTION

Abiotic stress can severely impair plant growth and performance. Environmental factors such as drought, extreme temperatures, or high and fluctuating salinity are responsible for significant yield reductions in cultivated areas worldwide (Boyer 1982). Leading to a series of morphological, physiological, biochemical, and molecular changes, abiotic stress adversely affects plant growth and productivity (Bray 1997). The complexity of these responses is not surprising because plants must be able to tolerate significant variations in soil composition, temperature, and water potential during the life cycle, by changes in gene expression. Many genes have been described that respond to environmental stresses such as drought, high salinity, and low temperatures in plants (Ingram & Bartels 1996; Shinozaki & Yamaguchi-Shinozaki 1997; Zhu 2002; Shinozaki, Yanaguchi-Shinozaki & Seki 2003).

Drought stress has been a central topic of plant physiology because it significantly limits plant productivity. For example, loss to drought in the tropics alone is thought to exceed 20 million tons of grain per year, or approximately 17% of well-watered production, reaching up to 60% in severely affected regions such as southern Africa from 1991 to 1992 (Ribaut, Banziger & Hoisington 2002). To overcome these limitations and improve production efficiency to feed an increasing world population, drought tolerant crops must be developed; however, drought tolerance is a particularly challenging trait, due in large part to its unpredictable nature. Traditional breeding strategies that have attempted to improve drought tolerance utilizing genetic variation arising from intraspecific or interspecific hybridization and induced mutation have met with only limited success. Traditional approaches are limited by the complexity of stress tolerance traits, low genetic variance of yield components under stress conditions, and the lack of efficient selection techniques (Frova *et al*. 1999). Understanding physiological responses to

water stress at the molecular level may help to resolve these problems. For the last decade, a large and still-increasing number of genes, transcripts, and proteins have been implicated in drought stress pathways in major crops by virtue of many tools for gene discovery.

Like food crops, many turfgrasses require appreciable water to maintain high quality and growth (Huang, Duncan & Carrow 1997). One strategy to reduce irrigation requirements and water stress is to use drought resistant species and cultivars; however, the genetics and physiology of turfgrass are not yet as well developed as those of major crops such as corn, sorghum, barley, wheat, and rice. Advances in turfgrass relatives such as the major cereals may, on the other hand, be leveraged to accelerate understanding of turfgrass molecular and physiological biology.

Bermudagrass (*Cynodon dactylon*) is widely used as turfgrass in tropical and subtropical regions. Although the need for new Bermudagrass cultivars is increasing, breeding programs still depend on traditional methods such as inter-/intraspecific hybridization and selection. In order to expand the tools and approaches available to improve Bermudagrass, it is crucial to have more information about changes in gene expression during drought stress. Such information should not only incorporate knowledge of previous studies from other plants but also provide valuable resources for the development of drought tolerant Bermudagrass.

The objective of this study is to identify genes in *C. dactylon* for which expression patterns are correlated with physiological responses to drought conditions. Identifying genes turned on or off in response to water stress will fulfill an important step toward enhancing drought tolerance and provide genes that can be deployed for testing by many biotechnology-based approaches.

MATERIALS AND METHODS

Plant materials and growth conditions

C. dactylon genotype T89 (PI 290869) was grown with half-strength Hoagland's solution in sand culture in the greenhouse at temperatures of 25°C/30°C (Night/Day). Drought stress was imposed with PEG 8,000 (polyethylene glycol, MW 8,000, Sigma-Aldrich, St. Louis, MO) by adjusting the water potential of roots as described by Verslues & Sharp (1999). Three ψ_ws were compared in this experiment: high ψ_w (-0.02 MPa, no PEG added), intermediate ψ_w (-0.7 Mpa, 10% PEG) and low ψ_w (-1.6 MPa, 20% PEG), representing normal conditions, moderate drought, and severe drought, respectively. Each treatment was replicated four times in a randomized complete block design. The treatment continued until severe leaf rolling occurred (about twenty days) and samples were obtained at 3 DAT (Days After Treatment), 6 DAT, and 9 DAT, which represent the early, intermediate, and late stages of water stress, respectively. Consequently, the total number of plant samples was 36 (3 different treatments × 3 different time points × 4 biological replications). The plant samples were frozen in liquid nitrogen for subsequent laboratory analysis.

To assess the condition of the plants during drought stress treatment, percent coverage levels were obtained on each sampling day. Color ratings obtained on the same days as percent coverage levels were based on Carrow & Duncan (2003): 9.0 = dark green color, 1.0 = no green color.

Preparation of high density filters

The first 4,608 clones (12 384-well plates) from a normalized cDNA library were double-spotted using a 2 × 2 grid pattern by a Qbot (Genetix, New Milton, Hampshire, UK) onto

12 nylon Hybond N+ membranes (200 cm^2; Amersham, Piscataway, NJ). The gridding pattern was designed so that each membrane completely contains the 4,608 clones with each clone being inoculated twice. Membranes were placed on Q-Trays (Genetix) containing LB with 50 μg/mL of ampicillin, and grown for 16-21 h at 37°C. After air drying, membranes (with nucleic acid spots facing up) were denatured in 0.6 M NaOH for 3 min., neutralized in 0.5 M Tris-HCl (pH 7.5) for 3 min., rinsed in distilled water for 30 seconds, and stored at -20°C until use.

RNA extraction and probe preparation

Total RNAs were extracted from all 36 samples using TRIZOL (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. mRNA was purified from total RNA using PolyATtract® mRNA Isolation Systems (Promega, Madison, WI) and quantified by spectrophotometry. 5 μL of polyadenylated RNA (2 μg/μL) was combined with 2 μL of oligo$(dT)_{12\text{-}18}$ (0.5 μg/μL) and incubated at 70°C for 10 minutes, then chilled on ice. This mixture was then combined with 4 μL of first-strand buffer [250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM $MgCl_2$], 1 μL of dATP/dGTP/dTTP (10 mM), 2 μL of 0.1 M DTT, and 1 μL of RNaseOUT Recombinant Ribonuclease Inhibitor (Invitrogen), 4 μL of ^{32}P-dCTP at 6000 Ci/mmol (Amersham), and 1 μL of SuperScript II Reverse Transcriptase (Invitrogen). The labeling reaction was then allowed to proceed at 37°C for two hours, after which 2 μL of stop buffer (2 M NaOH, 2 mM EDTA) was added. The labeled first-strand cDNA probe was then passed through a purification column prepared with Sephadex G-50 beads (Sigma-Aldrich) and assayed for specific activity using a scintillation counter before hybridization.

Hybridization

The membranes were prehybridized for at least two hours in 100 mL of hybridization buffer [0.5 M sodium phosphate (pH 7.2), 7% sodium dodecyl sulfate (SDS), 1 mM EDTA [pH 8.0], 1% bovine serum albumin] at 65°C in a rotating incubator. After prehybridization, the buffer was discarded, and 35 mL of fresh buffer was added along with the labeled probe. The hybridization took place at 65°C for 20 hours while the tube was rotated at 5-6 rpm. The buffer was discarded again, and each filter was washed twice with 100 mL of washing buffer (0.25× SSPE, 0.25% SDS) at 65°C at 12 rpm. The filters were then transferred into a solution of 2× SSC (0.3 M NaCl, 30 mM sodium citrate dehydrate, pH 7.0) briefly before being blotted dry and wrapped in clear plastic for further processing. After acquisition of data, the filters were stripped by adding a solution of boiling 0.5% SDS over them and shaking vigorously for five minutes.

Image analysis

The hybridized filters wrapped in plastic were placed into phosphorimager cassettes and exposed for 20 hours. Exposed screens were scanned and signal intensity for each clone recorded by a STORM 820 PhosphorImager (Molecular Dynamics, Sunnyvale, CA). Individual signal intensities were quantified by ImageQuant software (Molecular Dynamics) by aligning a 96 × 96 grid on the imaged filter such that each spot occurs within one square. Background correction was performed using a 'local average' function in ImageQuant. Negative values from the local average function were converted to 1 for further analysis. Each signal resulting from background correction was averaged with its duplicate on the same filter and transferred into Microsoft Excel spreadsheets, pre-designed to associate the ImageQuant data format to the correct gene identities.

Profiling drought-responsive genes

Raw intensity data for each sample were $\log_{10}$ transformed and then used for the calculation of Z-scores as described by Cheadle *et al.* (2003). Z-scores were calculated by subtracting the overall average gene intensity within a single membrane from the raw intensity data for each gene, and dividing that result by the standard deviation of all of the measured intensities according to the formula:

$$\text{Z-score} = (\text{intensity}_{G} - \text{mean intensity}_{G1...Gn}) / \text{SD}_{G1...Gn}$$

where G is any gene on the macroarray and G1...Gn represents the aggregate measure of all of the genes. In order to profile gene expression changes during water stress, Z-ratios were calculated by taking the difference between the averages of the observed gene Z-scores and dividing by the SD of all of the differences for that particular comparison (Cheadle *et al.* 2003):

$$\text{Z-ratio} = [(\text{Z-score}_{G1ave})_{Drought} - (\text{Z-score}_{G1ave})_{Control}] / \text{SD of Z-score differences}_{G1...Gn}$$

where G1 represents the average Z-score for any particular gene being tested under multiple drought stress conditions and G1...Gn represent the aggregate measure of all of the genes. The calculated Z-ratios have the advantage that they can be used in multiple comparisons without further reference to the individual conditional standard deviations by which they were derived. A Z-ratio of ± 1.96 is inferred as significant ($P < 0.05$). For example, if Z-ratios of two different genes in any particular condition are -2.00 and 2.00, they are considered significantly down-regulated and up-regulated, respectively.

Annotation of drought candidate genes

Drought candidate genes showing Z-ratios of ± 1.96 were annotated using two comparative genomic tools, BLASTX and Gene Ontology. The drought candidate genes were

queried against the NCBI non-redundant (nr) protein reference library, using the NCBI standalone blastall program and the BLASTX algorithm at default settings (Altschul *et al.* 1997). Within each candidate gene, the existence of a top high-scoring pair (HSP) with an E-value below 10^{-5} was taken as indicative of significant similarity. The drought candidate genes were also assigned functions according to Gene Ontology terms (Harris *et al.* 2004) based on BLAST definitions with Goblet's plant database (271,009 plant proteins with 154,140 GO Identifiers, http://goblet.molgen.mpg.de; Groth, Lehrach & Hennig 2004) using an E-value threshold of 10^{-10}.

Cluster analysis

Hierarchical clustering of experimental variation in gene expression was determined using software programs developed at Stanford University (Eisen *et al*. 1998). The cluster algorithm was set to complete linkage clustering using the uncentered Pearson correlation. In this approach, all gene expression vectors are compared with each other, such that a matrix of correlations is generated.

Promoter analysis

Drought-responsive genes significantly up- or down-regulated during water stress were subjected to promoter analyses. Since little genomic sequence from *C. dactylon* exists, the analysis was dependent upon rice 1kb-Upstream sequences from the ATG translation start site containing putative promoter sequences downloaded from TIGR (http://www.tigr.org.tdb/e2kq/osa1/data_download.shtml). In order to find rice locus identifiers (e.g. LOC_Os02g52650.1), drought candidate genes were searched against the TIGR rice protein database using BLASTX with the cut-off E value of 10^{-10}. With those locus identifiers of drought candidate genes, a rice

1kb-Upstream fasta file was generated by an in-house Perl script and submitted to the PLACE (http://www.dna.affrc.go.jp/PLACE/index.html) database (Higo *et al*. 1999) to explore for differences between up- and down-regulation in the frequencies of *cis*-acting regulatory elements. The confidence limit for a binomial proportion (P = 0.01) was used to evaluate the differences.

RESULTS

Evaluation of treatment effect by turf density and leaf color

In order to evaluate the effect of PEG treatment, percent coverage and leaf color were visually estimated until 12 DAT (Days After Treatment). Figure 4.1 and 4.2 show percent coverage levels and leaf color ratings, respectively. While percent coverage levels tend to decrease as PEG concentration increase, they are not significantly different among various time points in each treatment. Leaf color has a slightly different tendency from percent coverage. In 10 % and 20 % PEG treatments, leaf color ratings decrease until 9 DAT but significantly increase at 12 DAT. On the other hand, leaf color ratings significantly decrease at all time points as PEG concentration increases.

Drought stress candidate genes in C. dactylon

For data normalization, raw data from ImageQuant software were subjected to calculation of Z-scores for every single gene on each membrane and Z-ratios were calculated by comparing the Z-scores of treatments with those of controls. Since Z-ratios follow a normal distribution, a Z-ratio of ± 1.96 was inferred as significant at the 95 % confidence level. As a result, 189 drought candidate genes (120 up- and 69 down-regulated genes) were identified from two different treatments (10 % and 20% PEG) and three different time points (3, 6, and 9 DAT).

The candidate genes were searched against the NCBI nr database using BLASTX algorithm (E-value < 10^{-5}) in order to find sequence similarities to known plant proteins.

The Venn diagram in Figure 4.3 indicates the number of genes involved in each treatment and time point. During water stress treatment, 12 genes and 2 genes were always up- and down-regulated, respectively. Table 4.1 shows the list of those 14 genes. Of the 12 up-regulated genes, a putative delta-pyrroline-5-carboxylate synthetase (ES292694) is an essential enzyme to synthesize proline, which is a stress-responsive amino acid, and a putative MYB17 protein (ES295217) is known as a transcription factor able to induce a variety of stress-responsive genes. Tables 4.2 and 4.3 indicate the list of up- and down-regulated genes in any treatments or time points. Obvious differences in gene expression profile were found between up- and down-regulated genes. For example, some drought-responsive genes known in other plants such as the transcripts of a peroxiredoxin (ES292613), a drought inducible 22kD protein (ES294446), and an early-responsive to dehydration protein (ES293944) were up-regulated whereas genes related to basic plant metabolism, such as photosynthesis and glycolysis, were down-regulated. On the other hand, no significant pattern of gene expression was identified between treatments or among time points; however, genes related to detoxification, such as a putative peroxiredoxin (ES292613), were specifically up-regulated at 3 DAT and gradually decreased at 6 and 9 DAT.

Functional classification of 189 drought candidate genes

For functional classification, 189 drought candidate genes were subjected to a Gene Ontology (GO) search. Table 4.4 lists the results of 73 of 120 up-regulated genes and 46 of 69 down-regulated genes classified by GO terms. For the genes without GO terms, 17 of up-

regulated and 10 of down-regulated genes have not been classified by the GO, and the rest of them do not have significant similarity in BLASTX search. A total of 704 and 371 multiple GO terms from the molecular function, the cellular component, and the biological process were assigned to the 73 up-regulated and 46 down-regulated genes, respectively. GO search results suggested detailed differences between up- and down-regulations based on the functional groups of the drought candidate genes. Comparisons of GO terms (level 3) from the molecular function category are shown in Figure 4.4. Translation factor activity (GO:0008135), transcription factor activity (GO:0003700), peroxidase activity (GO:0004601), receptor activity (GO:0004872), two-component sensor molecule activity (GO:0000155), transposase activity (GO:0016740), and helicase activity (GO:0004386) were specifically found in up-regulated genes. In contrast, ion transporter activity (GO:0015075), protein transporter activity (GO:0008565), electron transporter activity (GO:0005489), carrier activity (GO:0005386), and lyase activity (GO:0016829) were specifically represented in down-regulated genes. Comparisons of GO terms from the cellular component category revealed that mitochondria protein transporter complex (GO: 0042719), RNA polymerase complex (GO: 0030880), and cell surface (GO: 0009986) were specifically represented in up-regulated genes whereas photosystem I (GO: 0009522), phosphopyruvate hydratase complex (GO: 0000015), and organelle envelope (GO: 0031967) were only found in down-regulated genes. Membrane (GO: 0016020) was found both in up- and down-regulated genes but showed significantly higher representation in down-regulated genes than in up-regulated genes (Figure 4.5). Comparisons of GO terms from the biological process category provided the most obvious differences between up- and down-regulated genes (Figure 4.6). In this category, metabolism (GO: 0008152) and the cellular physiological process (GO: 0050875) were excluded because those groups were so general that they acted like outliers.

Response to endogenous stimulus (GO:0009717), response to abiotic stimulus (GO:0009628), response to external stimulus (GO:0009605), regulation of cellular process (GO:0050875), and regulation of the physiological process (GO:0050791) were peculiar to up-regulated genes. In addition, response to biotic stimulus (GO:0009607) and response to stress (GO:0006950) showed higher proportion in up-regulated genes than in down-regulated genes. In contrast, the proportion of photosynthesis (GO:0015979) and sexual reproduction (GO:0019953) genes in the down-regulated group were significantly higher than in up-regulated genes.

Comparison of putative cis-acting regulatory elements

Since the genomic sequence of *C. dactylon* is still unknown, 1-kb upstream regions from the ATG translation start site were retrieved for corresponding genes from the rice pseudomolecules to compare putative cis-acting regulatory elements between up- and down-regulated genes. Putative upstream promoter regions could be retrieved for 144 loci, including 87 up-regulated genes and 57 down-regulated genes. Using the PLACE database, about 56,000 candidate cis-acting regulatory elements were identified from 144 kb of upstream sequences. In order for reliable interpretation of this comparative analysis, a strict statistical threshold ($P < 0.01$) was applied and any cis-elements less than 6 bp in length were excluded from this analysis. Also, cis-elements that have a higher proportion in down-regulated genes were not considered in the analysis because these elements were too common to provide proper inference about drought-responsive cis-acting regulatory elements.

The proportions of cis-acting regulatory elements significantly represented in up-regulated genes are shown in Table 4.5. Apparent correlations between treatments or time points could not be found, but five cis-acting regulatory elements, the GAGA [$(GA)_9$], the LTRE-1

(CCGAAA), the ACTCAT element (ACTCAT), the S-box (CACCTCCA), and the Up2 (AAACCCTA), showed consistently higher proportions in up-regulated genes than in down-regulated genes. Of the five elements, the LTRE-1 and the ACTCAT elements showed significant abundance both in 10% and 20% PEG treatment. In addition to the LTRE-1 and the ACTCAT elements, in 10 % PEG treatment, the Dc3 (ACACNNG) also showed significant abundance whereas the S-box and the TATCCAC box (TATCCAC) were significantly over-represented in the 20 % PEG treatment. The ACTCAT element, the PRE [SCGAYNR$(N)_{15}$HD], and the S-box had significantly high proportions at 3 DAT. None of the cis-acting regulatory elements showed significant abundance at 6 DAT. However, the LTRE-1, the Pyrimidine box (TTTTTTCC), and the Up2 were significantly over-represented at 9 DAT. Details of each element will be discussed below, but most cis-acting regulatory elements described in Table 4.5 are directly or indirectly related to drought stress response in other plants.

Cluster analysis

Clustering is a simple but proven method for analyzing gene expression data. With this method, the gene expression vectors calculated in length units based on intensities of hybridization data, which make up the hybridization data matrix, are reordered to place similar vectors closer to each other within the matrix. In this study, the Z-scores of 189 drought candidate genes were subjected to vector calculations using the Cluster program (Eisen *et al*. 1998). All the gene expression vectors from the drought candidate genes were compared with each other, such that a matrix of correlations was generated. The largest correlation in the matrix defined the two most similar vectors, and these were then joined to form a node. The node had a compound vector associated with it, which was calculated as the average of the vectors that

contributed to it. This compound vector was then compared with all existing unjoined gene expression and compound vectors, and the process was repeated. Thus, single expression profiles were successively joined to form nodes, which in turn were joined further. The process continued until all individual profiles and nodes were joined to form a single hierarchical tree as shown in Figure 4.7. The Cluster program generated seven different clusters according to the gene expression patterns of the 189 drought candidate genes. Clusters I-III include the 69 down-regulated genes whereas clusters IV-VII include the 120 up-regulated genes. Of the four clusters including the up-regulated genes, cluster IV indicates slight differences whereas cluster V indicates extreme differences between the control and the PEG treatments.

The drought candidate genes involved in each cluster were compared with the GO terms shown in Table 4.4, with the results summarized in Table 4.6. The cluster-specific GO terms enable us to deduce which functions or processes are specifically expressed in similar patterns regardless of PEG concentrations or sampling time points.

Of the three clusters including down-regulated genes, cluster I has only one specific GO term, photosystem I (GO:0009522), assigned to ES294943 putatively encoding photosystem I reaction center subunit. Cluster II includes two specific GO terms, phosphopyruvate hydratase complex (GO:0000015) and sexual reproduction (GO:0019953). The phosphopyruvate hydratase complex is directly connected to glycolysis (GO:0006096, data not shown). Cluster III mostly includes organic and electron transporter activities (GO:0005342 and GO:0005489, respectively), which are further annotated as sodium symporter activity (GO:0008508, data not shown).

Of the four clusters including up-regulated genes, cluster IV includes some drought-responsive GO terms, such as two-component sensor molecule activity (GO:0000155) and regulation of gene expression, epigenetic (GO:0040029), which are related to signal transducer

activity (GO:0004871, data not shown) and DNA methylation (GO:0006306, data not shown). Cluster V specifically includes response to external stimulus (GO:0009605) in the same subcategory with response to abscisic acid stimulus (GO:0009737, data not shown). Cluster VI includes pattern binding (GO:0001871), which is connected to response to pest, pathogen, or parasite (GO:0009613, data not shown) under response to stress (GO:0006950, data not shown). Cluster VII includes transposase activity (GO:0004803) and tetrapyrrole binding (GO:0046906) which correspond to ES293251 and ES295052, respectively. These two genes have significant similarity to abscisic acid-induced-like protein (ABA93823) and ABA 8-hydroxylase 2 (ABB71586) in the NCBI nr database, respectively.

DISCUSSION

Visual estimation of turf density and leaf color under water stress

To evaluate the PEG treatment effect, percent coverage and leaf color were visually estimated as shown in Figure 4.1 and 4.2. Percent coverage levels were significantly different between 10 % PEG and 20 % PEG while it was not significantly different among different time points. The leaf color was significantly different between the different PEG concentrations and among different time points. Also, the leaf color recovered in both PEG concentrations at 12 DAT. Common Bermudagrass is known to have a certain intensity of drought resistance by virtue of its C4 photosynthetic pathway (Casler & Duncan, 2003). No evidence of drought-stress acclimation in Bermudagrass has been found; however, Tabaei-aghadei, Harrison & Pearce (2000) compared drought-stress acclimation of *Lophopyrum elongatum* (drought-susceptible) with that of *Agropyron desertorum* (drought-resistant), which revealed that the drought acclimation of *A. desertorum* was initiated at 9-10 days after drought stress treatment. Thus,

recovery of leaf color at 12 DAT may be explained by drought-stress acclimation. Percent coverage levels might be less affected by drought stress treatment than leaf color because the greenhouse experiment was performed in a short period of time.

Since drought stress was imposed by adding PEG in the nutrient solution, symptoms may be somewhat different from field conditions. The purpose of this treatment is to obtain uniform plant samples under different PEG concentrations and time points without various environmental factors. Considering the results of the visual estimations, the PEG treatment had a significant effect on turf growth although obvious chlorosis or leaf rolling was not observed during drought treatment.

Profile of drought candidate genes

The 189 drought candidate genes (up- and down-regulated genes during drought-stressed conditions), their BLASTX results against the NCBI nr database (E-value < 10^{-5}), and their Z-ratios were listed in Tables 4.1, 4.2, and 4.3. The transcripts of a delta1-pyrroline-5-carboxylate synthase (P5CS, similar to BAA19916) and MYB17 protein (similar to CAD44611) were up-regulated during the entire drought-stress treatment (Table 4.1). Ingram & Bartels (1996) reviewed that the P5CS is an essential enzyme for proline biosynthesis which has long been accepted as a stress-responsive amino acid in many plant species. The role of proline under various stress conditions is still obscure but several possible roles have been proposed, including stabilization of macromolecules, a sink for excess reductant, and a store of carbon and nitrogen for use after relief of water deficit (Raymond and Smirnoff, 2002). Abe *et al*. (2003) reported that AtMYC2 protein functions as a transcriptional activator in the ABA-responsive gene expression of the rd22 gene in *Arabidopsis thanliana*. Recently, Suprunova *et al*. (2007) also

reported the MYC protein is induced in wild barley (*Hordeum spontaneoum*) under water-stressed conditions. Thus, it can be assumed that the MYC protein also acts as a drought-induced protein in *C. dactylon*.

Table 4.2 indicates the up-regulated genes in at least one treatment or time point. Up-regulated genes in the early stage of water stress (3 DAT) constitute putative peroxiredoxin (ES292613) and 10 kDa chaperonin (ES292573), especially the latter of which was always up-regulated in 20 % PEG treatment representing severe drought-stressed conditions. Ingram & Bartels (1996) reported that the peroxiredoxin increases in response to drought stress because of elevated photorespiratory activity during drought, which acts as a scavenger of active oxygen species. They also reported that chaperonins are involved in protein repair by helping other proteins to recover their native conformation after denaturation or misfolding during water stress. At 6 DAT and 9 DAT, genes similar to ABA-induced or stress-induced proteins, transcription factors involved in water stress response, protein kinases related to stress signal transduction pathway, and low-molecular-weight heat-shock proteins were up-regulated. These genes were analogous to what has been reported for drought stress-dependent transcripts in barley (Ozturk *et al.* 2002), rice (Gorantla *et al*. 2005), and maize (Zheng *et al*. 2004), which reflects that *C. dactylon* shares many drought-responsive mechanisms with other plant species.

Ozturk *et al*. (2002) also claimed that transcripts related to basic metabolism were down-regulated upon drought shock in barley leaves and roots. From the hybridizations with mRNA from water-stressed leaves of *C. dactylon*, down-regulation was obvious for a number of basic biosynthetic functions, including photosynthesis and photorespiration, and amino acid and carbohydrate metabolism (Table 4.3). Although Tables 4.1, 4.2, and 4.3 provide a glimpse of the many different functions regulated under water-stressed conditions, their role and importance in

tolerance or sensitivity is impossible to judge from the limited selection of ESTs that are contained on the macroarray membrane. With the 4,608 ESTs, the hybridization covered approximately 20% or less of all transcripts expressed in the *C. dactylon* leaf tissues, and is a merely a beginning; however, the drought candidate genes will be an important step toward improving drought tolerance of *C. dactylon* together with other warm season turfgrasses by providing basic molecular information about plant response under drought-stressed conditions.

Functional classification of the drought candidate genes using GO terms

GO terms were assigned to 189 up- or down-regulated genes for further classification based on their functional categories. This analysis provided for straightforward comparison between up- and down-regulations. Table 4.4 listed the GO terms for the 189 drought candidate genes and Figures 4.4, 4.5, and 4.6 indicate schematized summaries of the candidate genes according to the GO's third level annotations involved in three basal level categories including molecular function, cellular component, and biological process, respectively. In the molecular function category (Figure 4.4), translation factor activity (GO:0008135), transcription factor activity (GO:0003700), peroxidase activity (GO:0004601), two-component sensor molecule activity (GO:0000155), and receptor activity (GO:0004872) were specifically found in up-regulated genes.

The translation factor activity includes two clones, ES292999 (similar to geranylgeranyl reductase) and ES294550 (similar to ankyrin protein). Tanaka *et al*. (1999) reported that loss of geranylgeranyl reductase activity leads to loss of chlorophyll and tocopherol in transgenic tobacco, which resulted in susceptibility to oxidative stress under various abiotic stress conditions. Lu *et al*. (2003) reported that ACD6, an ankyrin protein, is a regulator of salicylic

acid signaling in *Arabidopsis* defense response. Although the ankyrin protein is a member of one of the largest uncharacterized gene families in higher plants, its relationship with biotic and abiotic stresses is gaining significance in *Arabidopsis* (AbuQamar *et al.* 2006; Stone *et al*. 2006).

The transcription factor activity includes five clones, ES293896 (similar to an unknown protein from rice), ES293910 [similar to ATP-dependent Clp (chloroplast) protease ATP-binding subunit], ES295217 (similar to MYB 17 protein), ES294166 (similar to ABA-responsive element-binding protein 3), and ES292269 (similar to MADS box-like protein). Zheng *et al*. (2002) showed that the ATP-dependent Clp protease incorporates the activity of molecular chaperones to target specific polypeptide substrates and avoid inadvertent degradation of others. Two clones similar to 10 kDa chaperonin (ES292573) and heat shock protein 82 (ES292856), identified as up-regulated in this study, represent a ubiquitous group of regulatory proteins well known to directly influence the structure and function of many polypeptides. They assist in the folding, assembly, and translocation of numerous cellular proteins during both normal and adverse growth conditions. Thus, it can be assumed that *C. dactylon* has a protein repair system identified in other plant species under abiotic stress conditions. Transcription factors involved in various abiotic stresses have been intensively studied by many researchers. According to Yamaguchi-Shinozaki & Shinozaki (2006), the MYB 17 and the ABA-responsive element-binding proteins up-regulated in their study mediate ABA-dependent gene expression under drought or cold stress. In addition, ES293944 (similar to early-responsive to dehydration protein), involved in ABA-independent gene expression under abiotic stress, is also up-regulated in *C. dactylon*. Thus, it can also be assumed that both ABA-dependent and independent pathways are activated in *C. dactylon* in response to drought stress.

The peroxidase activity includes two clones, ES294956 (similar to catalase) and ES292613 (similar to peroxiredoxin), which are able to detoxify active oxygen species such as O_2^-, H_2O_2, and OH^- which damage membranes and macromolecules. Plants have developed several antioxidation strategies to scavenge these toxic compounds. Therefore, enhancement of antioxidant defense in *C. dactylon* may increase tolerance to drought stress.

The two-component sensor molecule activity includes four clones, ES294741 (similar to OSJNBb0003B01.20), ES293931 (similar to DNA methyltransferase ZMET 4), ES292169 (unknown protein from rice), and ES292751 (similar to glycosyltransferase). The receptor activity includes three clones, ES292933 (similar to protein kinase domain), ES293117 (similar to kinase), and ES292169 (similar to an unknown protein from rice). Those two categories are subcategories of the signal transducer activity (GO:0004871). The two-component system, consisting of a membrane-localized histidine (His) protein kinase that senses a signal input and a response regulator that mediates the output, is an ancient and evolutionarily conserved signaling mechanism in prokaryotes and eukaryotes (Hwang, Chen & Sheen 2002). Based on the GO annotation, the four clones involved in the two-component sensor molecule activity have a protein histidine kinase activity (GO:0004673, data not shown), and the three clones involved in the receptor activity have a protein kinase activity (GO:0004672, data not shown). It has been known that many protein kinases such as mitogen-activated protein kinases (MAPKs), calcium-dependent protein kinase (CDPK), and Suc nonfermenting-related kinase (SnRK) are involved in the stress signal transduction pathway (Boudsocq & Lauriere, 2005). All up-regulated protein kinases during drought stress may provide clues to protein kinase-mediated signal transduction pathways in *C. dactylon*.

No significant differences exist between up- and down-regulations in the cellular component category (Figure 4.5) except membrane (GO:0016020). A higher proportion of the membrane subcategory was found in down-regulations, suggesting that targets of water stress were often localized in membranes. Membranes are a primary target for active oxygen species and lose integrity under conditions of water stress, which may explain why the membrane category has a higher proportion in down-regulations than in up-regulations.

In the biological process category (Figure 4.6), there is a response to the stimulus category (GO:0050896) including several subcategories such as response to stress (GO:0006950), response to external stimulus (GO:0009605), response to biotic stimulus (GO:0009607), response to abiotic stimulus (GO:0009628), and response to endogenous stimulus (GO:0009717). The response to the stimulus category includes 10 clones – ES294550 (similar to ankyrin protein), ES293910 (similar to ATP-dependent Clp protease ATP-binding subunit), ES295217 (similar to MYB17 protein), ES294956 (similar to catalase), ES291977 (similar to abscisic acid- and stress-inducible protein), ES293626 (similar to *Magnaporthe grisea* pathogenicity protein), ES294839 (similar to MLA1), ES292582 (similar to endochitinase), ES294446 (similar to drought inducible 22 kD protein), and ES293100 (similar to Clp amino terminal domain) – which are also included in regulation of the cellular process (GO:0050794) and regulation of the physiological process (GO:0050791) because the GO terms can have multiple annotations. It is quite interesting that four genes – ES295217, ES293626, ES294839, and ES292582 – included in the response to the biotic stimulus category are induced by drought stress in *C. dactylon*. The MYB 17 protein encoded by the ES295217 is also involved in the response to the abiotic stimulus category, while the MLA1 protein encoded by the ES294839 is known to confer powdery mildew resistance in barley (Mohler *et al*. 2002). Fujita *et al*. (2006) reported that biotic

and abiotic stress share some parts of their signal transduction pathways resulting in crosstalk between abiotic and biotic stress responses. According to the report, MYB-related gene expressions in response to drought stress are mediated by ABA, which can also regulate gene expressions under biotic stress conditions. Abundance of the barley R proteins, MLA1 and MLA6, decreased dramatically in response to heat stress although the level of *MLA1* and *MLA6* transcripts was elevated (Bieri *et al*. 2004) , suggesting that R protein instability could be imposed by environmental stresses regardless of its gene expression. Thus the four genes included in the response to the biotic stimulus category might be evidence that some gene expression pathways in *C. dactylon* are shared between abiotic and biotic stress.

On the other hand, sexual reproduction category (GO:0019953) including ES295002 (putative beta-expansin 5) is exclusively found in down-regulation. Munns *et al*. (2000) suggested that drought-induced inhibition of leaf growth in plants generally results from decreases in cell wall extensibility. McQueen-Mason, Durachko & Cosgrove (1992) proposed that cell wall extensibility in plants is determined both by the underlying structure of the wall and the activity of wall-modifying proteins such as expansins. The expansins are also essential for growth of silk in maize (Wu, Meeley & Cosgrove 2001), which, in turn, may affect yields of corn in that Anthesis-Silking Interval (ASI) might be negatively correlated with the activity of expansin. A large ASI typically affects ovule pollination and therefore jeopardizes grain production in maize (Ribaut & Ragot, 2007). As a result, drought stress at flowering time most greatly reduces yields in maize. Several extensive breeding programs of maize have selected against the increase in ASI with water deficit (Welcker *et al*. 2007). Thus, to increase the activity of expansin during drought stress may be a potential breeding target.

Putative promoter analysis

Table 4.5 shows proportions of cis-acting regulatory elements significantly represented in up-regulated genes. The results of this procedure only permit indirect inference about promoters in the drought candidate genes from *C. dactylon*, and may also be affected by various factors such as length and complexity of cis-acting regulatory elements. Thus, data must be carefully interpreted in order to have reliable results from this comparative analysis. Evidence that some cis-acting regulatory elements are responsive to drought stress could be found in other plants.

GAGA elements comprising the dinucleotide repeat sequence $(GA)_n$ regulate transcription of the *hsp70.1* gene in animals in response to various stresses (Bevilacqua, Fiorenza & Mangia 2000). In plants, the $(GA)_n$ sequences in regulatory regions of some genes from soybean and barley have protein-binding affinity (Sangwan & O'Brian 2002; Santi *et al*. 2003). In addition, Gromoff *et al*. (2006) reported that PRE (plastid-responsive element) acts as an enhancer within the *Chlamydomonas HSP70A* promoter. However, no direct evidence has been found that GAGA elements and PREs are related to expression of heat shock proteins in plants. Thus, GAGA elements and PREs in plants need to be further investigated based on their relationship with expression of heat shock proteins because some heat shock proteins were identified as drought-responsive genes in *C. dactylon*.

Dunn *et al*. (1998) suggested that low-temperature-responsive elements (LTRE) induce transcription of stress-responsive genes in response to cold stress as well as drought stress in barley. This may explain why the LTRE-1 is highly represented in up-regulated genes in *C. dactylon*. The ACTCAT is recognized by bZIP transcription factors in order to induce *ProDH* gene encoding proline dehydrogenase (Satoh *et al*. 2002). Proline dehydrogenase (ProDH) is

catalyzing the first step in proline degradation. The toxicity of proline to plant growth has raised questions despite its protective functions in response to environmental stresses. Nanjo *et al*. (2003) suggested that an excess of proline might lead to feedback inhibition of cell wall-associated proteins responsible for normal morphogenesis at a transcriptional level. Thus, it can be speculated that the feedback inhibition regulates the level of proline in *C. dactylon*.

The Dc3 and the S-box elements known to be involved in ABA response were highly represented in up-regulations. *Dc3* is a carrot LEA class gene that is abundantly expressed during somatic and zygotic embryogenesis. Its expression is normally embryo-specific and can also be induced by ABA. Kim, Chung & Thomas (1997) found a regulatory element in the *Dc3* gene, which is essential for ABA-induced expression, and named it the Dc3 element. The Dc3 element is a binding site of two basic leucine zipper proteins, named DPBF-1 and 2 (Dc3 Promoter-Binding Factor-1 and 2). Finkelstein & Lynch (2000) confirmed that the DPBF-1 has significant similarity to the predicted amino acid sequence of *ABI5* (ABA-Insensitive-5) gene in *Arabidopsis*. CMA5 (Conserved Modular Arrangement 5) is characterized as a minimal light-responsive unit (Martinez-Hernandez *et al*. 2002). Acevedo-Hernandez, Leon & Herrera-Estrella. (2005) showed that the CMA5 is able to respond not only to light and chloroplast signals, but also to sugar signals in a pathway involving ABA. The CMA5 has a cis-regulatory element named S-box, which is a binding site of ABI4 (ABA-Insensitive-4) protein. Based on the two promoters together with several ABA-induced genes identified as drought candidate genes in *C. dactylon*, it can be inferred that ABA in *C. dactylon* plays an important role in response to drought stress.

Cluster analysis

No obvious pattern was observed in the cluster-specific GO terms (Table 4.6). However, cluster III specifically includes two genes, ES292409 (similar to bile acid transporter family protein) and ES294456 (similar to glutaredoxin), which share a common GO term 'transporter activity (GO:0005215)'. Rzewuski & Sauter (2002) reported the novel rice gene *OsSbf1* encoding a putative member of the Na^+/bile acid symporter family. They suggested that a possible function of SBF proteins might be in the transport of sulphonated brassinosteroids which are structurally related to bile acids, since bile acids had not been found in plants. On the other hand, Rouhier, Gelhaye & Jacquot (2004) reported that knowledge of plant glutaredoxin is still scarce, and proposed that glutaredoxins in plants are involved in the oxidative stress response based on the fact that glutaredoxins need to be reduced in order to function, the reducing system being composed of an NADPH-dependent pyridine nucleotide oxidoreductase called glutathione reductase and the small tripeptide, glutathione.

Interestingly, the two clones, ES292409 and ES294456, are also found to have the sodium symporter activity (GO:0008508, data not shown) by further annotation using the GO terms. Initially, some transporter activities were expected to be represented in up-regulations based on Gorantla *et al*. (2005), but they are specifically down-regulated as shown in Table 4.4 and Figure 4.4. Particularly, two clones, sharing 'the sodium symporter activity' GO term, are only found in cluster III, showing the most extreme difference between the highest and the lowest Z-scores (data not shown); however, it is still unclear why transporter activities are only found in down-regulations and several transporter activities are specifically found in cluster III.

Evidence about levels and patterns of gene expression in *Cynodon* in response to water deficit is an important first step toward dissecting the complex underlying pathways. Irrespective

of the many experimental conditions, the analysis identified many genes that had previously been implicated in plant stress responses in detailed studies that focused on individual genes. Comparative analysis of cis-acting elements also largely corresponded to the results of other studies, while suggesting the existence of additional putative cis-elements, which await detailed analysis. GO functional classification clearly showed differences between up- and down-regulated genes. In cluster analysis, the drought candidate genes were scattered in virtually all the clusters, which appears to indicate the fundamental roles that these proteins play in diverse response to drought stress such as detoxification, osmotic adjustment, and signaling pathways.

While not as sensitive as microarray, these low-cost macroarray analyses provided for the systematic analysis of quantitative gene expression and also an effective tool with which to find novel genes that are expressed in certain conditions or in certain tissues. The experimental procedures used in this chapter can also be applied to various abiotic or biotic stress conditions other than drought stress. By accumulating data on gene expression by tissue type, developmental stage, hormone and herbicide treatment, genetic background, and environmental conditions, it should be possible to identify genes involved in many important processes of development and responses to environmental conditions in *Cynodon*.

REFERENCES

Abe H., Urao T., Ito T., Seki M., Shinozaki K. & Yamaguchi-Shinozaki K. (2003) Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling. *Plant Cell* **15**, 63-78.

AbuQamar S., Chen X., Dhawan R., Bluhm B., Salmeron J., Lam S., Dietrich R.A. & Mengiste T. (2006) Expression profiling and mutant analysis reveals complex regulatory networks involved in Arabidopsis response to *Botrytis* infection. *Plant Journal* **48**, 28-44.

Acevedo-Hernandez G.J., Leon P. & Herrera-Estrella L.R. (2005) Sugar and ABA responsiveness of a minimal *RBCS* light-responsive unit is mediated by direct binding of ABI4. *Plant Journal* **43**, 506-519.

Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang A., Miller W. & Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.

Bevilacqua A., Fiorenza M.T. & Mangia F. (2000) A developmentally regulated GAGA box-binding factor and Sp1 are required for transcription of the hsp70.1 gene at the onset of mouse zygotic genome activation. *Development* **127**, 1541-1551.

Bieri S., Mauch S., Shen Q.H., Peart J., Devoto A., Casais C., Ceron F., Schulze S., Steinbiss H.H., Shirasu K. & Schultz-Lefert P. (2004) Rar1 positively controls steady state levels of barley MLA resistance proteins and enables sufficient MLA6 accumulation for effective resistance. *Plant Cell* **16**, 3480-3495.

Boudsocq M. & Lauriere C. (2005) Osmotic signaling in Plants. Multiplel pathways mediated by emerging kinase families. *Plant Physiology* **138**, 1185-1194.

Boyer J.S. (1982) Plant productivity and environment. *Science* **218**, 443-448.

Bray E.A. (1997) Molecular response to water deficit. *Trends in Plant Science* **2**, 48-54.

Carrow R.N. & Duncan R.R. (2003) Improving Drought Resistance and Persistence in Turf-Type Tall Fescue. *Crop Science* **43**, 978-984.

Casler, M.D. & Duncan R.R. (2003) Bermudagrass In *Turfgrass Biology, Genetics, and Breeding*, p 236. John Wiley & Sons Inc., Hoboken, New Jersey, USA.

Cheadle C., Vawter M.P., Freed W.J. & Becker K.G. (2003) Analysis of Microarray Data Using Z-score Transformation. *Journal of Molecular Diagnostics* **4**, 73-81.

Dunn M.A., White A.J., Vural S. & Hughes M.A. (1998) Identification of promoter elements in a low-temperature-responsive gene (blt4.9) from barley (*Hordeum vulgare* L.). *Plant Molecular Biology* **38**, 551-564.

Eisen M.B., Spellman P.T., Brown P.O. & Botstein D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences, USA* **95**, 14863-14868.

Finkelstein R.R. & Lynch T.J. (2000) The Arabidopsis abscisic acid response gene *ABI5* encodes a basic leucine zipper transcription factor. *Plant Cell* **12**, 599-609.

Frova C., Krajewski P., di Fonzo N., Villa M. & Sari-Gorla M. (1999) Genetic analysis of drought tolerance in maize by molecular markers. 1. Yield components. *Theoretical and Applied Genetics* **99**, 280-288.

Fujita M., Fujita Y., Noutoshi Y., Takahashi F., Narusaka Y., Yamaguchi-Shinozaki K. & Shinozaki K. (2006) Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Current Opinion in Plant Biology* **9**, 436-442.

Gorantla M., Babu P.R., Reddy Lachagari V.B., Feltus F.A., Paterson A.H. & Reddy A.R. (2005) Functional genomics of drought stress response in rice: Transcript mapping of annotated unigenes of an indica rice (*Oryza sativa* L. cv. Nagina 22). *Current Science* **89**, 496-514.

Gromoff E.D., Schroda M., Oster U. & Beck C.F. (2006) Identification of a plastid response element that acts as an enhancer within the Chlamydomonas HSP70A promoter. *Nucleic Acid Research* **34**, 4767-4779.

Groth D., Lehrach H. & Hennig S. (2004) GOblet: a platform for Gene Ontology annotation of anonymous sequence data. *Nucleic Acids Research* **32**, W313-W317.

Harris M.A., Clark J., Ireland A., Lomax J., Ashburner M., Foulger R., Eilbeck K., Lewis S., Marshall B., Mungall C., Richter J., Rubin G.M., Blake J.A., Bult C., Dolan M., Drabkin H., Eppig J.T., Hill D.P., Ni L., Ringwald M., Balakrishnan R., Cherry J.M., Christie K.R., Costanzo M.C., Dwight S.S., Engel S., Fisk D.G., Hirschman J.E., Hong E.L., Nash R.S., Sethuraman A, Theesfeld C.L., Botstein D., Dolinski K., Feierbach B., Berardini T., Mundodi S., Rhee S.Y., Apweiler R., Barrell D., Camon E., Dimmer E., Lee V., Chisholm R., Gaudet P., Kibbe W., Kishore R., Schwarz E.M., Sternberg P., Gwinn M., Hannick L., Wortman J., Berriman M., Wood V., de la Cruz N., Tonellato P., Jaiswal P., Seigfried T. & White R. (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* **32**, D258-D261.

Higo K., Ugawa Y., Iwamoto M. & Korenaga T. (1999) Plant cis-acting regulatory DNA elements (PLACE) database. *Nucleic Acid Research* **27**, 279-300.

Huang B., Duncan R.R. & Carrow R.N. (1997) Drought-resistance mechanisms of seven warm-season turfgrasses under surface soil drying: I. Shoot response. *Crop Science* **37**, 1858-1863.

Hwang I., Chen H. & Sheen J. (2002) Two-component signal transduction pathways in Arabidopsis. *Plant Physiology* **129**, 500-515.

Ingram J. & Bartels D. (1996) The molecular basis of dehydration tolerance in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **44**, 377-403.

Kim S.Y., Chung H. & Thomas T.L. (1997) Isolation of a novel class of bZIP transcription factors that interact with ABA-responsive and embryo-specification elements in the *Dc3* promoter using a modified yeast one-hybrid system. *Plant Journal* **11**, 1237-1251.

Lu H., Rate D.N., Song J.T. & Greenberg J.T. (2003) ACD6, a Novel Ankyrin Protein, Is a Regulator and an Effector of Salicylic Acid Signaling in the Arabidopsis Defense Response. *Plant Cell* **15**, 2408-2420.

Martinez-Hernandez A., Lopez-Ochoa L., Arguello-Astorga G. & Herrera-Estrella L. (2002) Functional properties and regulatory complexity of a minimal *RBCS* light-responsive unit activated by phytochrome, cryptochrome, and plastid signals. *Plant Physiology* **128**, 1223-1233.

McQueen-Mason S., Durachko D.M. & Cosgrove D.J. (1992) Two endogenous proteins that induce cell wall extension in plants. *Plant Cell* **4**, 1425-1433.

Mohler V., Klahr A., Wenzel G. & Schwarz G. (2002) A resistance gene analog useful for targeting disease resistance genes against different pathogens on group 1S chromosomes of barley, wheat and rye. *Theoretical and Applied Genetics* **105**, 364-368.

Munns R., Passioura J.B., Guo J., Chazen O. & Cramer G.R. (2000) Water relations and leaf expansion: importance of time scale. *Journal of Experimental Botany* **51**, 1495-1504.

Nanjo T., Fujita M., Seki M., Kato T., Tabata S. & Shinozaki K. (2003) Toxicity of free proline revealed in an Arabidopsis T-DNA-tagged mutant deficient in proline dehydrogenase. *Plant and Cell Physiology* **44**, 541-548.

Ozturk Z.N., Talame V., Deyholos M., Michalowski C.B., Galbraith D.W., Gozukirmizi N., Tuberosa R. & Bohnert H.J. (2002) Monitoring large-scale changes in transcript abundance in drought-and salt-stressed barley. *Plant Molecular Biology* **48**, 551-573.

Raymond M.J. & Smirnoff N. (2002) Proline metabolism and transport in maize seedlings at low water potential. *Annals of Botany* **89**, 813-823.

Ribaut J-.M., Banziger M. & Hoisington D. (2002) Genetic dissection and plant improvement under abiotic stress conditions: drought tolerance in maze as an example. *JIRCAS Working Report* **23**, 85-92.

Ribaut J-.M. & Ragot M. (2007) Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. *Journal of Experimental Botany* **58**, 351-360.

Rouhier N., Gelhaye E. & Jacquot J.P. (2004) Plant glutaredoxins: still mysterious reducing systems. *Cellular and Molecular Life Sciences* **61**, 1266-1277.

Rzewuski, G. & Sauter M. (2002) The novel rice (*Oryza sativa* L.) gene *OsSbf1* encodes a putative member of the Na^+/bile acid symporter family. *Journal of Experimental Botany* **53**, 1991-1993.

Sangwan I. & O'Brian M.R. (2002) Identification of a soybean protein that interacts with GAGA element dinucleotide repeat DNA. *Plant Physiology* **129**,1788-1794.

Santi L., Wang Y., Stile M.R., Berendzen K., Wanke D., Roig C., Pozzi C., Muller K., Muller J., Rohde W. & Salamini F. (2003) The GA octodinucleotide repeat binding factor BBR participates in the transcriptional regulation of the homeobox gene Bkn3. *Plant Journal* **34**, 813-826.

Satoh R., Nakashima K., Seki M., Shinozaki K. & Yamaguchi-Shinozaki K. (2002) ACTCAT, a novel cis-acting element for proline- and hypoosmolarity-responsive expression of the ProDH gene encoding proline dehydrogenase in Arabidopsis. *Plant Physiology* **130**, 709-719.

Shinozaki K., Yamaguchi-Shinozaki K. & Seki M. (2003) Regulatory network of gene expression in the drought and cold stress responses. *Current Opinion in Plant Biology* **6**, 410-417.

Shinozaki K. & Yamaguchi-Shinozaki K. (1997) Gene expression and signal transduction in water-stress response. *Plant Physiology* **115**, 327-334.

Stone S.L., Williams L.A., Farmer L.M., Vierstra R.D. & Callis J. (2006) KEEP ON GOING, a RING E3 ligase essential for Arabidopsis growth and development, is involved in abscisic acid signaling. *Plant Cell* **18**, 3415-3428.

Suprunova T., Krugman T., Distelfeld A., Fahima T., Nevo E. & Korol A. (2007) Identification of a novel gene (Hsdr4) involved in water-stress tolerance in wild barley. *Plant Molecular Biology* **64**, 17-34.

Tabaei-Aghadaei S.R., Harrison P. & Pearce R.S. (2000) Expression of dehydration-related-genes in the crowns of wheatgrass species [*Lophopyrum elongatum* (Host) A. Love and *Agropyron desertorum* (Fisch. ex Link.) Schult.] having contrast acclimation to salt, cold, and drought. *Plant, Cell & Environment* **23**, 561-571.

Tanaka R., Oster U., Kruse E., Rudiger W. & Grimm B. (1999) Reduced Activity of Geranylgeranyl Reductase Leads to Loss of Chlorophyll and Tocopherol and to Partially Geranylgeranylated Chlorophyll in Transgenic Tobacco Plants Expressing Antisense RNA for Geranylgeranyl Reductase. *Plant Physiology* **120**, 695-704.

Verslues P.E. & Sharp R.E. (1999) Proline Accumulation in Maize (*Zea mays* L.) Primary Roots at Low Water Potentials. II. Metabolic Source of Increased Proline Deposition in the Elongation Zone. *Plant Physiology* **119**, 1349-1360.

Welcker C., Boussuge B., Bencivenni C., Ribaut J-.M. & Tardieu F. (2007) Are source and sink strengths genetically linked in maize plants subjected to water deficit? A QTL study of the responses of leaf growth and of Anthesis-Silking Interval to water deficit. *Journal of Experimental Botany* **58**, 339-349.

Wu Y., Meeley R.B. & Cosgrove D.J. (2001) Analysis and Expression of the α-Expansin and β-Expansin Gene Families in Maize. *Plant Physiology* **126**, 222–232.

Yamaguchi-Shinozaki K & Shinozaki K. (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annual Review of Plant Biology* **57**, 781-803.

Zheng B., Halperin T., Hruskova-Heidingsfeldova O., Adam Z. & Clarke A.K. (2002) Characterization of chloroplast Clp proteins in Arabidopsis: Localization, tissue specificity and stress responses. *Physiologia Plantarum* **114**, 92-101.

Zheng J., Zhao J., Tao Y., Wang J., Liu Y., Fu J., Jin Y., Gao P., Zhang J., Bai Y. & Wang G. (2004) Isolation and analysis of water stress induced genes in maize seedlings by subtractive PCR and cDNA macroarray. *Plant Molecular Biology* **55**, 807-823.

Zhu J.-K. (2002) Salt and drought stress signal transduction in plants. *Annual Review of Plant Biology* **53**, 247-273.

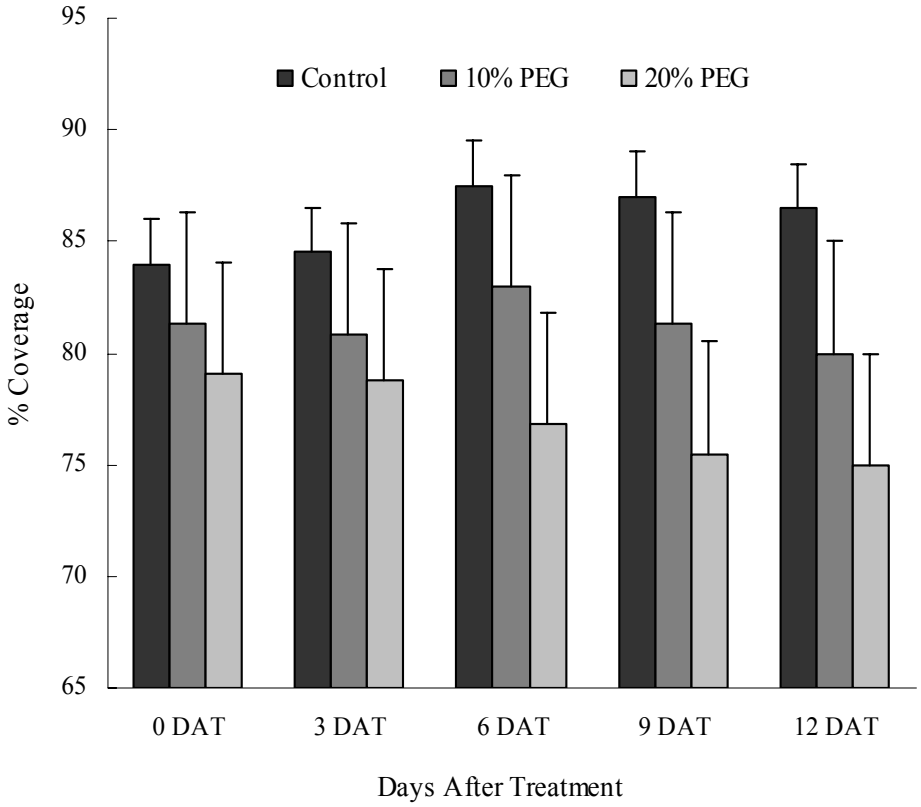
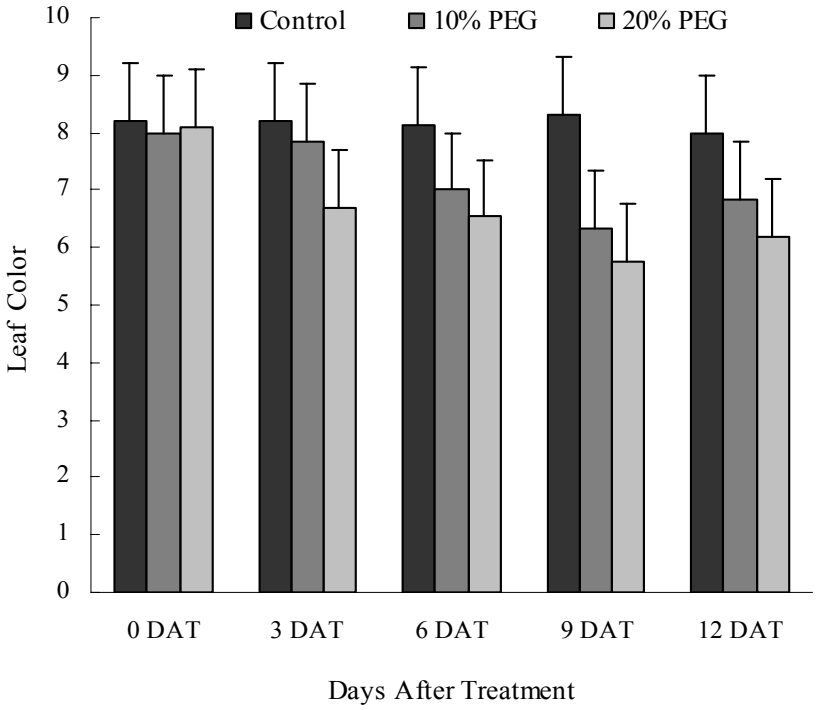


Figure 4.1. Turfgrass density represented by percent coverage in each drought stress treatment (Each standard error bar is shown on top of column).



9 = dark green color; 1 = no green color; 6.5 = minimum acceptable color for a good turf

Figure 4.2. Turfgrass color as affected by three different intensities of drought stress treatment (Each standard error bar is shown on top of column).

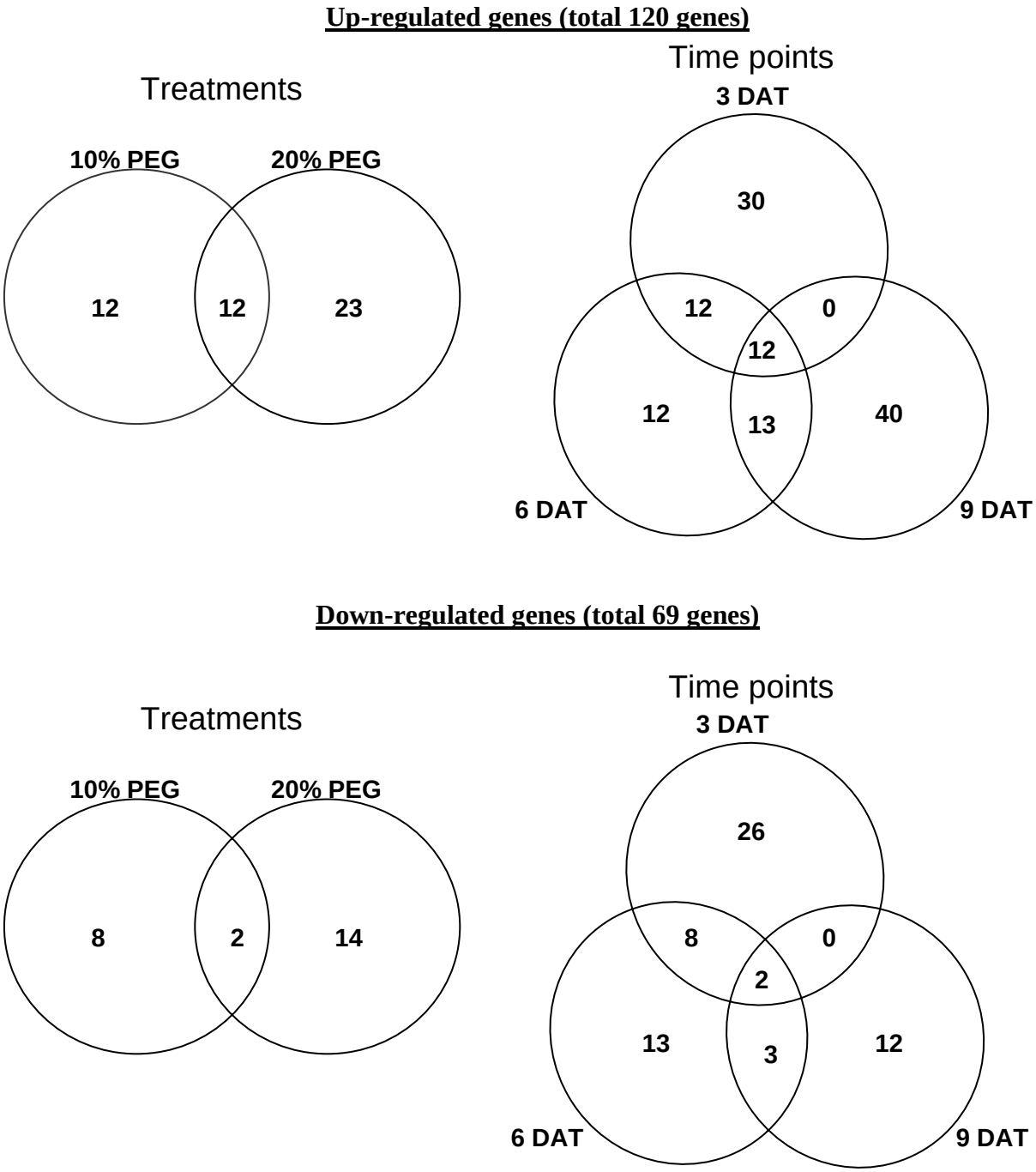


Figure 4.3. Venn diagrams showing the classification of genes inducible in different PEG concentrations or at different sampling time points. Note that a gene is represented in multiple groups.

Table 4.1. Significantly up- or down-regulated genes during entire drought-stress treatment.

Accession No.	GenBank Match	Putative Function	Z-ratios[a]					
			10% PEG			20% PEG		
			3 DAT	6 DAT	9 DAT	3 DAT	6 DAT	9 DAT
	Significantly up-regulated transcripts							
ES295333	CAA70175	stress responsive protein putative expressed [*Oryza sativa*]	**2.00**	**1.96**	**2.51**	**2.32**	**2.46**	**2.74**
ES292386	AAL86300	unknown protein [*Arabidopsis thaliana*]	**2.01**	**3.49**	**3.37**	**3.48**	**3.72**	**2.47**
ES292694	BAA19916	deltal-pyrroline-5-carboxylate synthetase [*Oryza sativa*]	**2.26**	**2.24**	**2.33**	**2.01**	**2.00**	**2.76**
ES292628	BAD25454	unknown protein [*Oryza sativa*]	**3.64**	**3.57**	**2.72**	**3.65**	**2.90**	**2.29**
ES293523	XP_467946	putative light-harvesting chlorophyll-a/b protein of photosystem I [*Oryza sativa*]	**2.54**	**2.68**	**2.39**	**2.61**	**2.10**	**2.17**
ES295217	CAD44611	MYB17 protein [*Oryza sativa*]	**2.69**	**2.26**	**2.57**	**2.84**	**2.36**	**2.52**
ES294101	BAA90810	BAG domain containing protein-like [*Oryza sativa*]	**2.88**	**3.14**	**2.79**	**3.40**	**3.24**	**2.71**
ES295143	ABA96309	Clp amino terminal domain putative [*Oryza sativa*]	**2.47**	**4.23**	**5.33**	**2.73**	**3.73**	**4.77**
ES293349	N.A.[b]	No Hits Found	**3.79**	**3.74**	**2.87**	**4.04**	**3.47**	**2.83**
ES293785	N.A.	No Hits Found	**3.67**	**2.95**	**2.02**	**3.38**	**3.06**	**2.07**
ES294152	N.A.	No Hits Found	**2.10**	**3.21**	**2.33**	**2.53**	**2.91**	**2.09**
ES294316	N.A.	No Hits Found	**2.84**	**2.74**	**2.80**	**3.32**	**2.82**	**2.51**
	Significantly down-regulated transcripts							
ES295675	XP_478841	putative photosystem I antenna protein [*Oryza sativa*]	**-2.33**	**-2.43**	**-1.98**	**-2.96**	**-2.47**	**-1.97**
ES291925	N.A.	No Hits Found	**-2.43**	**-2.73**	**-2.13**	**-3.08**	**-3.34**	**-2.41**

[a]Z-ratios were calculated based on the comparison of Z-score between treatments and control. Since Z-ratios show a normal distribution, Z-ratios of ± 1.96 are considered significantly up- or down-regulated (numbers in bold).
[b]Not Available

Table 4.2. Significantly up-regulated genes in at least one treatment or time point.

Accession No.	GenBank Match	Putative Function	Z-ratios[a]					
			10% PEG			20% PEG		
			3 DAT	6 DAT	9 DAT	3 DAT	6 DAT	9 DAT
ES292613	AAG40130	peroxiredoxin [*Oryza sativa*]	**2.19**	1.96	1.64	**2.07**	1.77	0.75
ES293602	AAK13157	dehydration-responsive protein putative expressed [*Oryza sativa*]	1.94	**2.95**	**3.29**	**2.09**	**2.54**	**3.22**
ES292985	AAK28970	YDG/SRA domain containing protein expressed [*Oryza sativa*]	**2.19**	**2.40**	1.45	**2.59**	**2.47**	1.25
ES293931	AAK40306	DNA methyltransferase ZMET4 [*Zea mays*]	1.81	**2.34**	**2.33**	**2.90**	**2.74**	**2.66**
ES293808	AAB07458	14-3-3 putative express [*Oryza sativa*]	1.40	**2.09**	1.09	1.62	**2.18**	1.20
ES294446	BAB68268	drought inducible 22 kD protein [*Saccharum officinarum*]	1.60	**2.49**	**3.20**	**2.14**	**2.20**	**2.60**
ES295401	CAD10342	1, 2-dihydroxy-3-keto-5-methylthiopentene dioxygenase [*Oryza sativa*]	**2.35**	**2.20**	**2.01**	**2.15**	**2.11**	1.66
ES294174	AAL01113	Su(VAR)3-9-related protein 4 [*Arabidopsis thaliana*]	**2.45**	**2.45**	1.72	**2.56**	**2.04**	0.60
ES293117	AAL58279	putative kinase [*Oryza sativa*]	-0.68	1.03	**2.14**	1.26	**2.88**	**2.56**
ES293626	AAL77134	Putative Magnaporthe grisea pathogenicity protein [*Oryza sativa*]	1.19	1.74	**2.74**	1.38	1.37	**2.29**
ES293135	AAL92029	cytosolic 6-phosphogluconate dehydrogenase [*Oryza sativa*]	0.37	1.66	**2.27**	1.89	1.94	**2.81**
ES293454	BAB89662	putative ubiquitin conjugating enzyme [*Oryza sativa*]	**2.15**	1.94	0.80	**2.28**	1.30	0.90
ES292684	AAB53748	Isolation and Characterization of a cDNA Encoding 3-Hydroxy-3-Methylglutaryl-CoA Reductase [*Oryza sativa*]	1.79	**2.14**	**2.07**	**3.29**	**2.67**	**2.44**
ES294758	AAM83219	kelch repeat-containing serine/threonine phosphoesterase [*Oryza sativa*]	**2.10**	**2.39**	**2.70**	**2.24**	1.84	1.33
ES293057	BAC07023	putative DNA-directed RNA polymerase Iia [*Oryza sativa*]	-0.56	1.21	**2.53**	1.01	1.74	**2.21**
ES294388	NP_199758	ATP binding / kinase/ protein kinase [*Arabidopsis thaliana*]	1.05	**2.09**	**2.59**	**2.02**	**1.98**	**2.21**
ES294960	AAM97107	ATP-dependent Clp protease proteolytic subunit (ClpR3) putative [*Arabidopsis thaliana*]	**1.98**	**2.01**	**2.03**	1.22	1.12	1.79
ES292573	AAB63591	10 kDa chaperonin [*Oryza sativa*]	**2.32**	1.85	1.36	**3.14**	**2.99**	**2.24**
ES292896	CAD57072	unnamed protein product [*Zea mays*]	1.08	**2.05**	**2.05**	1.83	**2.13**	1.85
ES291977	AAB96681	abscisic acid- and stress-inducible protein [*Oryza sativa*]	0.16	**1.97**	**2.69**	**2.84**	**2.87**	1.42
ES293877	AAO66563	putative reverse transcriptase [*Oryza sativa*]	0.00	0.60	**2.36**	1.52	2.21	**3.62**
ES293618	AAP06842	unknown protein [*Oryza sativa*]	1.54	0.80	1.21	2.07	**2.66**	**3.68**
ES294783	AAP06907	hypothetical protein [*Oryza sativa*]	**2.34**	1.37	1.20	**2.15**	0.45	0.27
ES294092	AAP50991	putative 3-isopropylmalate dehydrogenase [*Oryza sativa*]	1.20	1.63	**3.01**	**2.72**	**2.26**	**2.99**
ES292582	CAB01591	endochitinase [*Persea americana*]	-1.83	0.48	**2.78**	1.20	1.61	**2.63**
ES295499	BAC79184	putative water stress induced tonoplast intrinsic protein [*Oryza sativa*]	-0.48	1.06	**2.28**	-0.12	1.35	**2.13**
ES293896	BAC79975	unknown protein [*Oryza sativa*]	**2.13**	1.83	1.24	**2.00**	1.45	1.17
ES295606	BAC79847	putative TGF(transfoming growth factor) beta inducible nuclear protein TINP1 [*Oryza sativa*]	**2.43**	1.25	0.59	**2.60**	1.33	0.47

Table 4.2. Continued.

Accession No.	GenBank Match	Putative Function	Z-ratios[a]					
			10% PEG			20% PEG		
			3 DAT	6 DAT	9 DAT	3 DAT	6 DAT	9 DAT
ES295067	AAQ62066	single myb histone 3 [Zea mays]	**2.17**	**2.07**	**2.72**	1.69	1.62	**2.30**
ES292967	BAC84637	putative small nuclear ribonucleoprotein E [*Oryza sativa*]	1.93	**2.09**	1.67	**2.25**	**2.02**	0.41
ES294337	NP_916313	putative ubiquitin carboxyl-terminal hydrolase [*Oryza sativa*]	1.04	**2.00**	**2.67**	**2.45**	**2.63**	**2.91**
ES294493	NP_917327	phosphoenolpyruvate carboxylase kinase 4 putative express [*Oryza sativa*]	-1.56	-0.06	**1.98**	0.85	1.57	**2.71**
ES294839	NP_917546	putative MLA1 [*Oryza sativa*]	**2.71**	**2.01**	1.94	**2.07**	1.46	1.46
ES295295	NP_921728	putative pol polyprotein [*Oryza sativa*]	1.49	**2.20**	**2.31**	**2.04**	**2.56**	**2.82**
ES292285	AAO83391	GCK-like kinase MIK [Zea mays]	1.03	1.36	**1.98**	1.57	1.52	**1.97**
ES295206	BAC79924	small zinc finger-like protein [*Oryza sativa*]	0.47	1.51	**2.07**	1.28	**2.27**	**2.42**
ES294956	BAA34205	catalase [*Oryza sativa*]	1.69	1.56	**2.40**	1.57	**2.04**	**2.48**
ES293946	BAD07676	unknown protein [*Oryza sativa*]	1.27	**2.63**	**3.63**	**2.90**	**2.61**	**2.51**
ES294268	BAD09604	putative auxin-regulated protein [*Oryza sativa*]	-0.26	1.59	**3.05**	1.88	**3.12**	**3.77**
ES295683	BAD09990	ternary complex factor-like [*Oryza sativa*]	2.88	**3.13**	**3.04**	**3.03**	**2.37**	1.70
ES294989	AAS48644	putative ABA-induced protein [Cynodon dactylon]	1.50	1.69	**2.98**	1.15	**1.96**	**2.81**
ES293850	AAA16225	alpha-tubulin [*Oryza sativa*]	**2.66**	1.89	0.91	**2.11**	1.46	0.84
ES295273	AAS76269	At2g25605 [Arabidopsis thaliana]	**2.31**	**2.89**	**2.71**	**2.51**	**2.06**	1.71
ES292970	AAT01314	putative GTP-binding protein Rab7a [*Oryza sativa*]	-0.01	0.70	**2.49**	1.00	1.19	**2.29**
ES294095	AAT38062	unknown protein [*Oryza sativa*]	**1.96**	**3.16**	1.69	**2.71**	**2.80**	0.41
ES292169	AAT44132	unknown protein [*Oryza sativa*]	**2.15**	1.79	1.30	**2.25**	**2.36**	**2.84**
ES293479	BAD30961	putative phospholipase A2 activating protein [*Oryza sativa*]	**2.72**	**2.67**	1.60	**2.33**	1.69	1.73
ES294166	AAT77290	putative ABA-responsive element-binding protein 3 [*Oryza sativa*]	0.81	1.41	**1.97**	**2.06**	**2.24**	**2.89**
ES292737	AAT77407	unknown protein [*Oryza sativa*]	0.10	1.35	**2.10**	1.00	1.74	**2.21**
ES292780	XP_466302	unknown protein [*Oryza sativa*]	**3.36**	**2.73**	1.31	**3.17**	1.93	0.87
ES292999	XP_467759	putative geranylgeranyl reductase [*Oryza sativa*]	1.08	1.55	1.63	**2.10**	**2.58**	**2.26**
ES293910	XP_472335	ATP-dependent Clp protease ATP-binding subunit [*Oryza sativa*]	1.72	**2.76**	**2.97**	1.83	**2.66**	**3.25**
ES293564	XP_474031	N-terminal acetyltransferase complex ARD1 subunit [*Oryza sativa*]	**2.17**	**2.44**	1.26	**2.35**	1.45	0.33
ES295542	XP_474314	OSJNBb0004A17.12 [*Oryza sativa*]	0.00	1.40	**2.64**	1.80	**2.47**	**3.04**
ES294550	XP_475059	putative ankyrin protein [*Oryza sativa*]	**2.36**	1.72	1.17	**2.29**	1.46	0.57
ES292232	XP_478169	putative FKBP-type peptidyl-prolyl cis-trans isomerase; protein [*Oryza sativa*]	1.75	**2.30**	**2.10**	1.72	1.85	**2.04**

Table 4.2. Continued.

Accession No.	GenBank Match	Putative Function	Z-ratios[a]					
			10% PEG			20% PEG		
			3 DAT	6 DAT	9 DAT	3 DAT	6 DAT	9 DAT
ES292856	XP_483191	heat shock protein 82 [*Oryza sativa*]	**3.92**	**3.83**	**2.48**	**3.97**	**3.35**	1.72
ES294018	BAD36432	unknown protein [*Oryza sativa*]	0.65	**2.71**	**3.52**	**2.33**	**2.96**	**2.30**
ES295214	BAD37864	CREG2-protein-like [*Oryza sativa*]	**2.10**	**2.24**	1.71	**2.74**	1.86	0.36
ES293215	AAU44135	putative protein kinase [*Oryza sativa*]	-1.52	0.18	**2.56**	1.09	1.34	**2.85**
ES292269	BAA81884	MADS box-like protein [*Oryza sativa*]	-0.51	1.55	**2.79**	1.26	**2.46**	**2.40**
ES293018	AAU90215	putative uracil phosphoribosyltransferase [*Oryza sativa*]	1.56	1.78	**2.40**	**2.41**	**2.17**	**1.98**
ES293955	BAD53895	unknown protein [*Oryza sativa*]	0.88	1.71	**2.11**	1.51	**2.25**	**2.17**
ES294050	BAD54213	unknown protein [*Oryza sativa*]	**2.63**	**2.03**	**1.96**	**2.27**	1.75	1.82
ES291937	BAD62412	MYB transcription factor-like [*Oryza sativa*]	1.21	**2.19**	0.81	1.81	**1.97**	-0.06
ES294702	AAV32218	cytoplasmic ribosomal protein L18 [*Oryza sativa*]	**2.46**	**2.28**	1.30	**2.63**	1.86	0.51
ES292751	CAI30072	glycosyltransferase [*Sorghum bicolor*]	1.73	**2.56**	**2.03**	**2.08**	**2.23**	1.86
ES294741	CAE03629	OSJNBb0003B01.20 [*Oryza sativa*]	1.57	**3.37**	**3.14**	**2.16**	**2.10**	1.01
ES294368	CAE05469	OSJNBa0006A01.5 [*Oryza sativa*]	**2.49**	1.13	0.35	**2.14**	1.06	0.58
ES294834	CAB59202	serine carboxylase II-2 [*Hordeum vulgare*]	-0.53	0.81	**2.51**	1.14	1.62	**2.36**
ES292088	AAX96864	Transposon protein putative mutator sub-class [*Oryza sativa*]	0.13	0.63	1.82	**2.11**	**2.60**	**3.47**
ES291948	ABA91996	expressed protein [*Oryza sativa*]	**2.11**	1.86	1.41	**2.72**	**2.70**	1.63
ES293251	ABA93823	abscisic acid-induced-like protein [*Oryza sativa*]	**1.98**	1.75	**2.00**	1.64	1.53	**2.10**
ES295566	ABA95188	salt-inducible protein putative [*Oryza sativa*]	1.13	1.73	**2.77**	1.30	1.74	**2.33**
ES295487	ABA95281	U2 snRNP auxilliary factor large subunit splicing factor putative [*Oryza sativa*]	1.62	1.68	**2.34**	**2.47**	**2.68**	**4.02**
ES293100	ABA96309	Clp amino terminal domain putative [*Oryza sativa*]	1.81	**3.23**	**3.52**	1.00	1.75	**3.02**
ES293011	ABA98638	RNA-binding protein precursor [*Oryza sativa*]	0.85	**2.11**	**3.10**	**2.60**	**2.20**	**2.48**
ES292933	ABA99275	Protein kinase domain putative [*Oryza sativa*]	0.70	**2.67**	**3.31**	**2.35**	**2.49**	**2.45**
ES293944	ABA99695	early-responsive to dehydration protein [*Oryza sativa*]	**2.03**	**2.57**	1.04	1.59	**1.97**	0.82
ES295052	ABB71586	ABA 8'-hydroxylase 2 [*Hordeum vulgare*]	1.12	**1.89**	**2.90**	1.33	1.65	**2.75**
ES294426	ABD32900	Conserved hypothetical protein [*Medicago truncatula*]	0.64	1.50	**2.24**	1.39	1.67	**2.06**
ES294506	BAB03388	unknown protein [*Oryza sativa*]	**2.28**	1.56	0.75	**2.84**	**2.13**	1.59
ES292188	N.A.[b]	No Hits Found	**2.28**	1.88	1.29	**2.32**	1.79	1.26
ES292235	N.A.	No Hits Found	**2.63**	**2.26**	1.24	**2.61**	1.37	0.01

Table 4.2. Continued.

Accession No.	GenBank Match	Putative Function	Z-ratios[a]					
			10% PEG			20% PEG		
			3 DAT	6 DAT	9 DAT	3 DAT	6 DAT	9 DAT
ES292459	N.A.	No Hits Found	-1.01	1.04	**2.42**	1.26	1.77	**2.41**
ES292580	N.A.	No Hits Found	**2.92**	**2.56**	1.51	**2.56**	0.93	-0.09
ES292583	N.A.	No Hits Found	0.40	1.14	**2.22**	1.46	**2.01**	**2.24**
ES292663	N.A.	No Hits Found	1.60	**2.34**	1.53	**2.29**	**2.87**	1.26
ES292811	N.A.	No Hits Found	1.67	**2.38**	**2.14**	**2.64**	**2.46**	1.48
ES292845	N.A.	No Hits Found	-0.46	0.77	**2.62**	**2.29**	**2.23**	**2.53**
ES292860	N.A.	No Hits Found	**2.39**	**2.90**	**2.01**	**2.65**	1.94	1.27
ES292940	N.A.	No Hits Found	-0.99	0.69	**2.04**	1.18	1.63	**2.47**
ES292971	N.A.	No Hits Found	0.27	1.40	**2.45**	1.62	1.74	**2.25**
ES293005	N.A.	No Hits Found	1.79	**3.44**	**3.34**	**2.80**	**3.53**	**2.81**
ES294303	N.A.	No Hits Found	**2.53**	**2.72**	1.95	**2.19**	**2.23**	1.62
ES294367	N.A.	No Hits Found	**2.90**	**2.51**	**2.39**	**3.37**	**2.74**	1.60
ES294397	N.A.	No Hits Found	**3.00**	**2.61**	1.50	**2.17**	**2.36**	1.83
ES294572	N.A.	No Hits Found	0.47	1.24	**2.27**	1.83	1.26	**1.99**
ES294889	N.A.	No Hits Found	**2.38**	**2.09**	**2.03**	**2.20**	0.98	0.80
ES294935	N.A.	No Hits Found	0.60	**2.82**	**3.14**	**2.14**	**2.59**	1.75
ES294948	N.A.	No Hits Found	0.67	1.86	**3.20**	1.83	**2.22**	**2.38**
ES295029	N.A.	No Hits Found	**4.05**	**3.12**	1.15	**2.95**	**2.00**	1.25
ES295266	N.A.	No Hits Found	**3.00**	1.77	1.04	**2.74**	1.59	1.05
ES295328	N.A.	No Hits Found	0.12	1.11	**2.07**	1.90	**2.89**	**3.04**
ES295357	N.A.	No Hits Found	1.66	**2.21**	1.73	**2.04**	**2.17**	1.87
ES295441	N.A.	No Hits Found	**2.08**	1.90	1.48	**2.03**	1.42	1.22
ES295453	N.A.	No Hits Found	**2.39**	**3.02**	1.69	**2.63**	**2.01**	0.87
ES295475	N.A.	No Hits Found	**2.73**	**2.18**	**2.02**	**2.82**	1.31	1.72

[a]Z-ratios were calculated based on the comparison of Z-score between treatments and control. Since Z-ratios show a normal distribution, Z-ratios of $\pm$ 1.96 are considered significantly up- or down-regulated (numbers in bold).
[b]Not Available

Table 4.3. Significantly down-regulated genes in at least one treatment or time point.

Accession No.	GenBank Match	Putative Function	Z-ratios[a]					
			10% PEG			20% PEG		
			3 DAT	6 DAT	9 DAT	3 DAT	6 DAT	9 DAT
ES294965	AAG13621	putative LIM domain protein [*Oryza sativa*]	-1.68	**-2.11**	-1.69	**-2.30**	**-2.41**	**-2.29**
ES293171	AAK40310	putative methyl-binding domain protein MBD111 [*Zea mays*]	-1.14	**-2.19**	-1.92	-1.44	**-2.02**	-1.66
ES294955	BAB55686	putative malate dehydrogenase [*Oryza sativa*]	**-2.05**	-1.63	-1.27	**-1.96**	-1.15	-1.30
ES295002	AAK56128	beta-expansin 5 [*Zea mays*]	-1.54	-1.43	**-2.23**	**-2.59**	-1.31	**-2.12**
ES295437	NP_201222	unknown protein [*Arabidopsis thaliana*]	-0.73	**-2.22**	-1.77	-1.25	**-2.19**	-1.43
ES295640	CAA69075	S-adenosylmethionine decarboxylase [*Zea mays*]	**-2.49**	**-2.28**	-0.27	**-2.32**	-1.91	-0.69
ES294405	AAB36545	ubiquitin-like protein [*Phaseolus vulgaris*]	**-2.06**	**-1.99**	**-2.13**	**-2.76**	**-2.29**	-1.74
ES293541	AAL32838	putative ubiquitin-conjugating enzyme [*Arabidopsis thaliana*]	**-1.98**	-0.95	-0.79	**-2.08**	-1.20	-0.51
ES293709	AAL35606	peroxisomal multifunctional protein [*Oryza sativa*]	**-2.72**	**-2.02**	**-1.54**	**-2.49**	-1.19	-1.01
ES294603	BAA00009	triosephosphate isomerase [*Zea mays*]	**-2.72**	**-2.88**	**-2.65**	**-2.71**	-1.62	-0.45
ES293034	CAA39454	enolase [*Zea mays*]	**-2.32**	-1.70	-1.30	**-2.53**	-1.66	-1.42
ES295470	AAN31845	putative polyubiquitin [*Arabidopsis thaliana*]	**-2.82**	**-2.51**	-1.64	**-2.66**	**-2.06**	**-2.00**
ES292767	AAC05717	alpha tubulin 1 [*Eleusine indica*]	**-2.34**	**-2.31**	**-2.09**	**-2.30**	-1.38	-0.52
ES293825	AAB86939	NOI protein [*Oryza sativa*]	-1.23	-0.82	-1.10	**-2.07**	**-2.23**	**-2.90**
ES292268	AAO41140	cellulose synthase [*Oryza sativa*]	-0.70	-1.25	**-2.16**	-1.59	-1.72	**-2.68**
ES295172	BAC57826	putative Acyl-CoA binding protein [*Oryza sativa*]	-1.21	-1.63	**-2.03**	-1.23	-1.30	**-2.33**
ES294198	AAC03055	methylmalonate semi-aldehyde dehydrogenase [*Oryza sativa*]	-1.07	-0.79	-1.31	**-2.12**	**-2.02**	**-2.45**
ES292734	AAO72664	wheat adenosylhomocysteinase-like protein [*Oryza sativa*]	**-2.31**	**-2.80**	**-2.48**	**-2.92**	**-2.44**	-1.02
ES294456	AAP80853	glutaredoxin [*Triticum aestivum*]	-0.80	**-2.37**	-1.62	**-1.98**	**-2.13**	-0.98
ES295580	AAC26045	aconitase-iron regulated protein 1 [*Citrus limon*]	-0.02	-1.80	**-2.93**	-1.91	**-2.83**	**-2.17**
ES293751	NP_915603	hydrolase alpha/beta fold family protein [*Oryza sativa*]	**-2.37**	-1.65	-1.30	**-2.44**	-1.27	-1.11
ES292376	NP_915626	putative uricase [*Oryza sativa*]	-0.96	**-2.28**	**-3.14**	**-2.66**	**-3.08**	**-2.91**
ES292409	ABA96556	bile acid transporter family protein [*Oryza sativa*]	**-2.28**	**-2.41**	**-2.15**	-1.51	-1.44	-1.63
ES294475	NP_921725	putative acetohydroxyacid isomeroreductase [*Oryza sativa*]	-1.07	-1.57	-1.71	**-2.27**	**-2.27**	**-2.06**
ES294066	AAR89836	putative methylenetetrahydrofolate reductase [*Oryza sativa*]	**-2.44**	-1.53	-0.41	**-2.33**	-1.25	-1.01
ES291898	AAR99579	60S ribosomal protein L44 [*Phalaenopsis* hybrid cultivar]	**-2.33**	**-2.11**	-1.09	**-2.25**	-1.87	-1.66
ES294042	CAB41117	putative protein [*Arabidopsis thaliana*]	0.02	-0.83	-1.87	**-2.18**	**-2.75**	**-2.04**
ES293960	AAT11797	MT-like protein [*Cynodon dactylon*]	**-2.37**	-1.52	-0.82	**-2.22**	-1.31	-1.29

Table 4.3. Continued.

Accession No.	GenBank Match	Putative Function	Z-ratios[a]					
			10% PEG			20% PEG		
			3 DAT	6 DAT	9 DAT	3 DAT	6 DAT	9 DAT
ES294701	BAD21122	ATP-dependent RNA helicase [*Hordeum vulgare*]	**-2.49**	-1.87	-1.52	**-2.39**	-1.02	-1.85
ES294387	BAD22517	putative hydroxyproline-rich glycoprotein 1 [*Oryza sativa*]	-0.59	**-2.21**	**-2.04**	-1.63	**-2.49**	-1.32
ES295625	BAD22518	glycolipid transfer protein-like [*Oryza sativa*]	-0.12	-1.08	**-2.60**	-1.53	-1.67	**-2.30**
ES294567	BAD21528	putative brown planthopper-induced resistance protein 1 [*Oryza sativa*]	**-2.35**	**-2.25**	**-2.18**	**-3.04**	**-2.53**	-1.57
ES294034	BAD21676	alcohol dehydrogenase class III [*Oryza sativa*]	-0.33	-1.03	**-1.97**	-1.51	-1.74	**-2.26**
ES294698	AAT39155	Peroxisomal membrane anchor protein conserved region containing [*Oryza sativa*]	-1.66	**-2.78**	-1.55	**-2.63**	**-1.99**	-0.65
ES292479	AAT77005	FAD dependent oxidoreductase family protein expressed [*Oryza sativa*]	**-2.22**	-1.82	-1.38	**-2.41**	-1.22	-1.59
ES292325	XP_467481	mitochondrial import inner membrane translocase subunit T [*Oryza sativa*]	0.30	-1.41	**-2.40**	-1.38	**-2.70**	**-2.50**
ES295326	XP_473159	transposon protein putative unclassified [*Oryza sativa*]	0.88	-0.50	**-2.24**	-0.74	-1.63	**-3.01**
ES295092	XP_473344	OSJNBa0091D06.22 [*Oryza sativa*]	-1.74	**-2.93**	**-2.70**	**-2.54**	**-2.23**	-1.67
ES294535	XP_474157	ZIM motif family protein [*Oryza sativa*]	**-2.14**	**-2.10**	-1.75	**-2.23**	-1.82	-0.37
ES295447	XP_474434	Soluble inorganic pyrophosphatase putative expressed [*Oryza sativa*]	-1.18	**-2.25**	-1.87	**-2.54**	**-3.19**	-1.74
ES293583	BAD53633	putative CBF1 interacting corepressor CIR [*Oryza sativa*]	**-2.09**	**-2.41**	**-1.98**	-1.92	-1.85	-1.64
ES294555	AAV28626	Bet v I allergen [*Zea mays*]	-0.28	-1.72	**-2.10**	-1.94	**-2.16**	**-2.11**
ES294786	AAV44072	unknown protein [*Oryza sativa*]	-0.95	**-1.99**	-1.26	-1.56	**-2.00**	-0.66
ES293441	BAD68410	putative anthocyanin 5-aromatic acyltransferase [*Oryza sativa*]	**-2.13**	**-2.61**	**-2.74**	**-2.73**	-1.41	-1.33
ES295502	AAA86977	jasmonate-induced protein homolog; similar to barley jasmonate-induced protein Swiss-Prot Accession Number P32024	**-2.97**	**-2.46**	-0.63	**-2.86**	**-2.66**	-1.16
ES294976	AAX96587	expressed protein [*Oryza sativa*]	-0.77	**-2.20**	**-2.64**	-1.89	**-2.07**	**-1.96**
ES292853	CAB65537	Toc34-1 protein [*Zea mays*]	-0.38	**-2.05**	-1.41	-1.26	**-2.17**	-1.67
ES293025	AAA64414	chlorophyll a/b-binding apoprotein CP26 precursor [*Oryza sativa*]	-0.35	-0.93	-1.42	**-2.09**	**-2.74**	**-2.01**
ES295657	AAF65512	ADP-ribosylation factor [*Capsicum annuum*]	**-1.99**	**-2.27**	-1.42	**-2.12**	-1.01	-0.44
ES294351	ABA93776	ubiquitin activating enzyme - like protein [*Oryza sativa*]	-1.31	**-2.33**	-1.85	-1.93	**-2.45**	-1.25
ES293286	ABA94940	vacuolar sorting protein-like; embryogenesis protein H beta 58-like protein [*Oryza sativa*]	**-2.65**	-1.24	-0.19	**-1.99**	-0.45	-0.21
ES294943	ABA96024	photosystem i reaction centre subunit n chloroplast precursor [*Oryza sativa*]	**-2.35**	**-1.99**	-1.31	**-2.54**	-1.79	-0.65
ES292926	AAA69028	carbonic anhydrase 1 [*Oryza sativa*]	**-2.68**	**-2.39**	-1.41	**-2.89**	**-2.41**	**-2.42**
ES293487	CAA58474	methionine synthase [*Catharanthus roseus*]	**-2.16**	-1.84	-1.79	**-2.14**	-1.61	-0.45
ES293074	AAA82047	glyceraldehyde-3-phosphate dehydrogenase [*Oryza sativa*]	**-2.16**	**-1.89**	-1.32	**-2.09**	-1.40	-0.55
ES291890	N.A.[b]	No Hits Found	-1.61	-1.85	**-2.71**	-1.91	**-2.10**	**-1.96**

Table 4.3. Continued.

Accession No.	GenBank Match	Putative Function	Z-ratios[a]					
			10% PEG			20% PEG		
			3 DAT	6 DAT	9 DAT	3 DAT	6 DAT	9 DAT
ES292782	N.A.	No Hits Found	-0.23	**-2.21**	**-2.25**	-1.80	-1.89	**-2.01**
ES292880	N.A.	No Hits Found	-1.19	-1.28	-1.42	**-2.43**	**-3.43**	**-2.97**
ES292994	N.A.	No Hits Found	**-2.02**	-1.49	-1.12	**-2.05**	-0.62	-0.96
ES293232	N.A.	No Hits Found	-0.72	-1.11	-1.66	**-2.26**	**-2.74**	**-2.07**
ES293296	N.A.	No Hits Found	-1.24	-1.66	-1.86	**-2.47**	**-2.27**	**-2.30**
ES293431	N.A.	No Hits Found	-1.88	**-2.85**	**-2.94**	**-3.35**	**-3.17**	-1.86
ES293729	N.A.	No Hits Found	**-2.15**	**-2.05**	-1.14	**-2.57**	-1.44	-1.77
ES293801	N.A.	No Hits Found	1.21	-0.39	**-1.97**	-1.20	**-1.97**	**-2.27**
ES293969	N.A.	No Hits Found	-0.45	**-2.82**	**-2.99**	**-2.31**	**-2.97**	**-2.52**
ES294447	N.A.	No Hits Found	**-2.33**	**-2.21**	-0.84	**-2.21**	-1.80	0.03
ES295647	N.A.	No Hits Found	-0.96	**-2.91**	**-2.69**	-1.47	**-1.99**	-0.05

[a]Z-ratios were calculated based on the comparison of Z-score between treatments and control. Since Z-ratios show a normal distribution, Z-ratios of $\pm$ 1.96 are considered significantly up- or down-regulated (numbers in bold).
[b]Not Available

Table 4.4. Gene ontology (GO) mappings of the 189 drought candidate genes using Goblet's plant database. Note that individual GO categories can have multiple mappings and percentage representations are based on the number of first-level GO annotations (the molecular function, the cellular component, and the biological process).

		Up-regulated genes		Down-regulated genes	
Categories and subcategories	GO ID	Rep.	%	Rep.	%
Molecular function	**GO:0003674**	**65**	**100**	**41**	**100**
Catalytic activity	GO:0003824	39	60	26	63
Helicase activity	GO:0004386	2	3	0	0
Transposase activity	GO:0004803	1	2	0	0
Oxidoreductase activity	GO:0016491	8	12	9	22
Transferase activity	GO:0016740	18	28	5	12
Hydrolase activity	GO:0016787	11	17	5	12
Lyase activity	GO:0016829	0	0	5	12
Isomerase activity	GO:0016853	1	2	2	5
Ligase activity	GO:0016874	1	2	2	5
Signal transducer activity	GO:0004871	4	6	0	0
two-component sensor molecule activity	GO:0000155	4	6	0	0
receptor activity	GO:0004872	3	5	0	0
Structural molecule activity	GO:0005198	5	8	3	7
Structural constituent of ribosome	GO:0003735	4	6	3	7
Structural constituent of cell wall	GO:0005199	1	2	0	0
Transporter activity	GO:0005215	1	2	4	10
Organic acid transporter activity	GO:0005342	0	0	1	2
Carrier activity	GO:0005386	0	0	3	7
Electron transporter activity	GO:0005489	0	0	1	2
Protein transporter activity	GO:0008565	0	0	2	5
Ion transporter activity	GO:0015075	0	0	1	2
Binding	GO:0005488	42	65	22	54
Nucleotide binding	GO:0000166	17	26	7	17
Pattern binding	GO:0001871	1	2	0	0
Nucleic acid binding	GO:0003676	16	25	2	5
Protein binding	GO:0005515	8	12	4	10
Lipid binding	GO:0008289	1	2	1	2
Carbohydrate binding	GO:0030246	2	3	0	0
Ion binding	GO:0043167	9	14	9	22
Tetrapyrrole binding	GO:0046906	1	2	0	0
Cofactor binding	GO:0048037	1	2	3	7
Antioxidant activity	GO:0016209	2	3	0	0
Peroxidase activity	GO:0004601	2	3	0	0

Table 4.4. Continued.

Categories and subcategories	GO ID	Up-regulated genes		Down-regulated genes	
		Rep.	%	Rep.	%
Molecular function (Continued)	**GO:0003674**	**65**	**100**	**41**	**100**
Transcription regulator activity	GO:0030528	3	5	0	0
Transcription factor activity	GO:0003700	5	8	0	0
Translation regulator activity	GO:0045182	1	2	0	0
Translation factor activity	GO:0008135	2	3	0	0
Nutrient reservoir activity	GO:0045735	1	2	0	0
Cellular component	**GO:0005575**	**24**	**100**	**16**	**100**
Extra cellular region	GO:0005576	0	0	1	6
Cell	GO:0005623	24	100	15	94
Intracellular	GO:0005622	20	83	11	69
Cell surface	GO:0009986	1	4	0	0
Membrane	GO:0016020	4	17	9	56
Envelope	GO:0031975	0	0	2	13
Organelle envelope	GO:0031967	0	0	2	13
Organelle	GO:0043226	17	71	8	50
Membrane-bound organelle	GO:0043227	13	54	5	31
Non-membrane-bound organelle	GO:0043228	4	17	3	19
Intracellular organelle	GO:0043229	17	71	8	50
Protein complex	GO:0043234	8	33	5	31
Phosphopyruvate hydratase complex	GO:0000015	0	0	2	13
Photosystem I	GO:0009522	0	0	1	6
Ribonucleoprotein complex	GO:0030529	4	17	3	19
RNA polymerase complex	GO:0030880	2	8	0	0
Mitochondrial intermembrane space protein transporter complex	GO:0042719	2	8	0	0
Biological process	**GO:0008150**	**57**	**100**	**35**	**100**
Reproduction	GO:0000003	0	0	1	3
Sexual reproduction	GO:0019953	0	0	1	3
Development	GO:0007275	1	2	0	0
Regulation of gene expression, epigenetic	GO:0040029	1	2	0	0

Table 4.4. Continued.

Categories and subcategories	GO ID	Up-regulated genes		Down-regulated genes	
		Rep.	%	Rep.	%
Biological process (Continued)	**GO:0008150**	**57**	**100**	**35**	**100**
Physiological process	GO:0007582	56	98	35	100
Metabolism	GO:0008152	51	89	29	83
Photosynthesis	GO:0015979	1	2	3	9
Death	GO:0016265	2	4	1	3
Homeostasis	GO:0042592	0	0	1	3
Regulation of physiological process	GO:0050791	11	19	0	0
Coagulation	GO:0050817	1	2	0	0
Cellular physiological process	GO:0050875	52	91	31	89
Localization	GO:0051179	6	11	8	23
Cellular process	GO:0009987	52	91	31	89
cell communication	GO:0007154	2	4	1	3
regulation of cellular process	GO:0050794	10	18	0	0
cellular physiological process	GO:0050875	52	91	31	89
Regulation of biological process	GO:0050789	12	91	0	0
Regulation of gene expression, epigenetic	GO:0040029	1	4	0	0
Negative regulation of biological process	GO:0048519	2	18	0	0
Regulation of physiological process	GO:0050791	11	91	0	0
Regulation of cellular process	GO:0050794	10	21	0	0
Response to stimulus	GO:0050896	10	18	1	3
Response to stress	GO:0006950	10	18	1	3
Response to external stimulus	GO:0009605	4	7	0	0
Response to biotic stimulus	GO:0009607	4	7	1	3
Response to abiotic stimulus	GO:0009628	4	7	0	0
Response to endogenous stimulus	GO:0009719	5	9	0	0

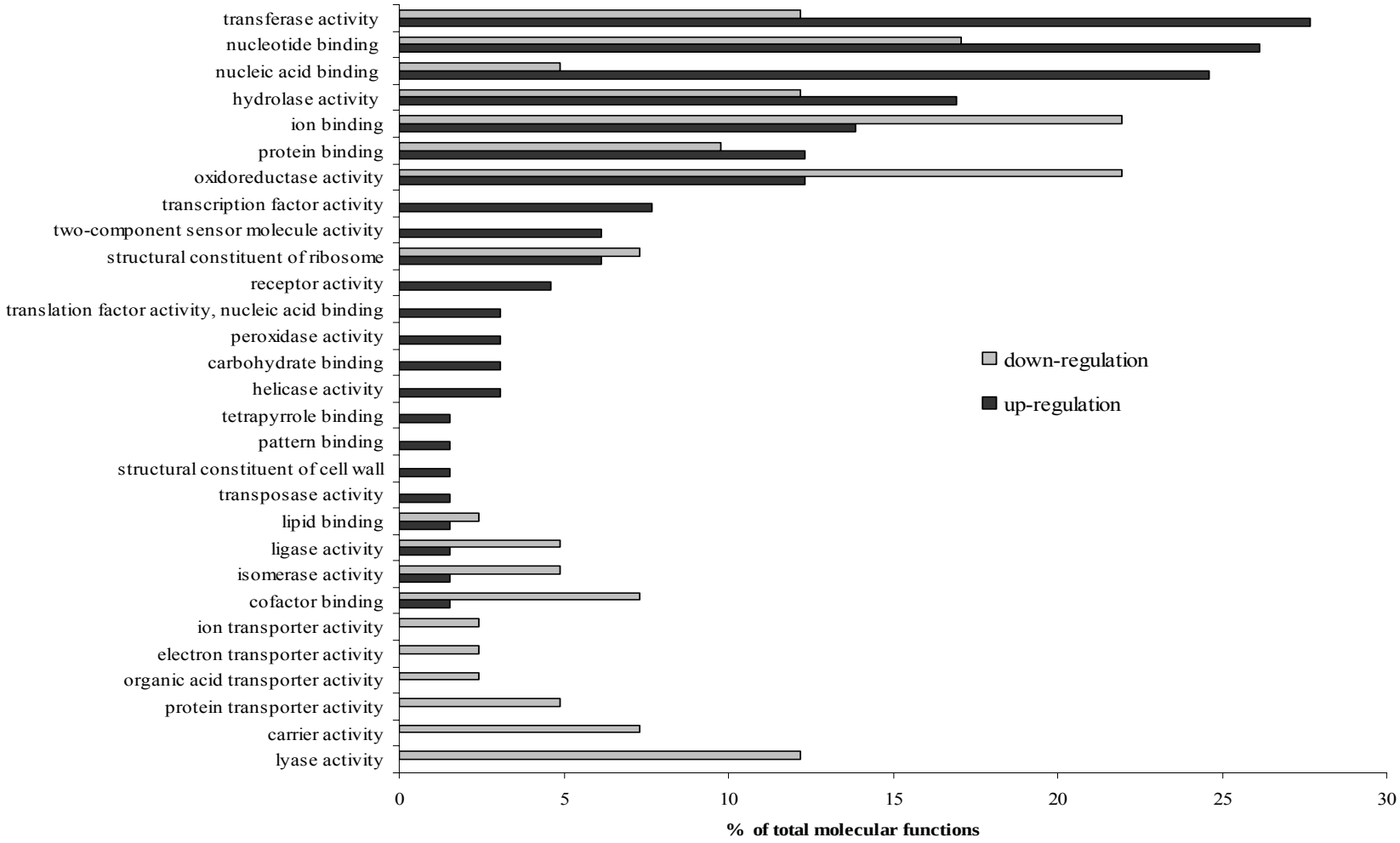


Figure 4.4. Percentage representations of third-level annotations from GO molecular function category between up- and down-regulated genes.

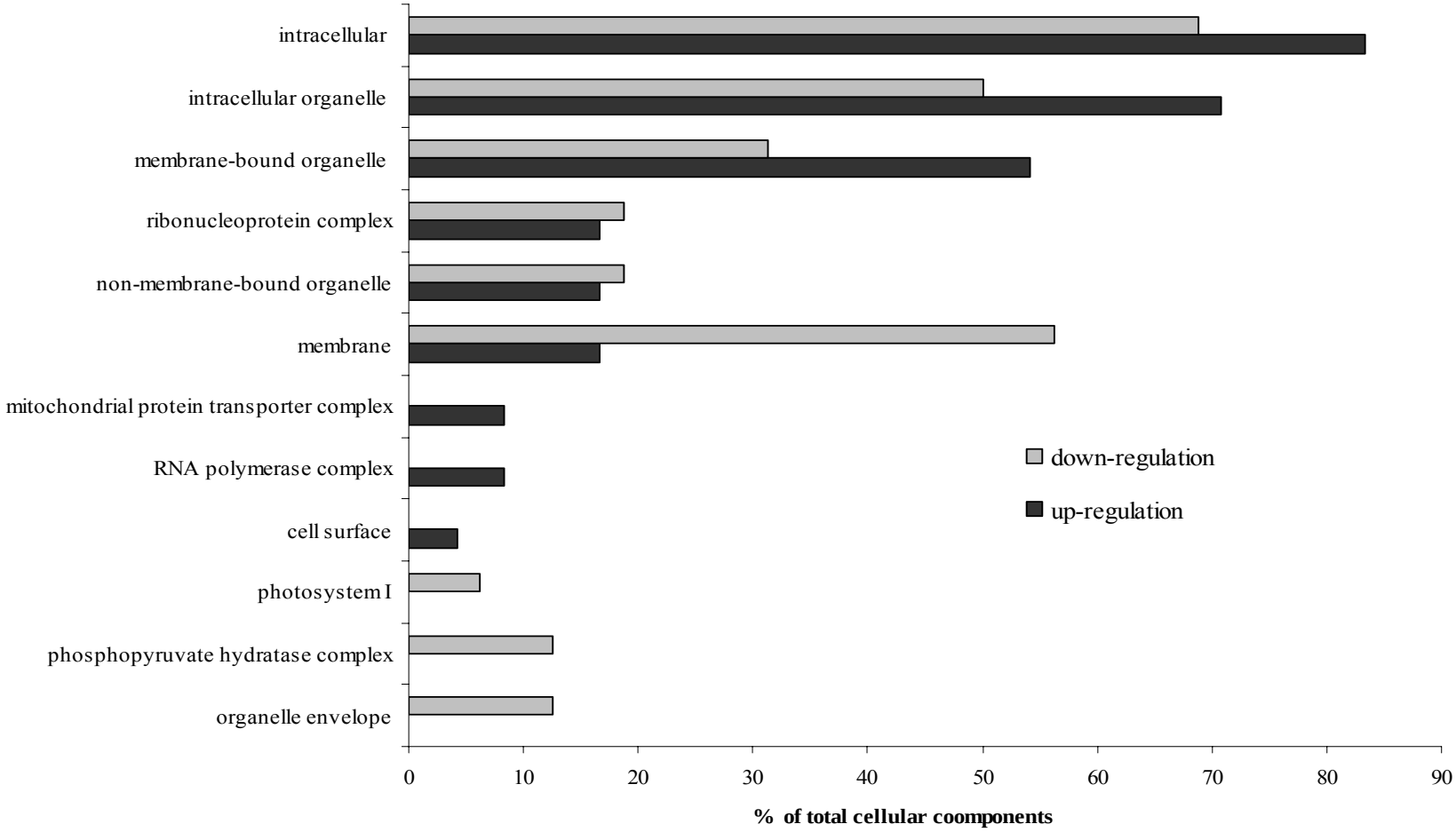


Figure 4.5. Percentage representations of third-level annotations from GO cellular component category between up- and down-regulated genes.

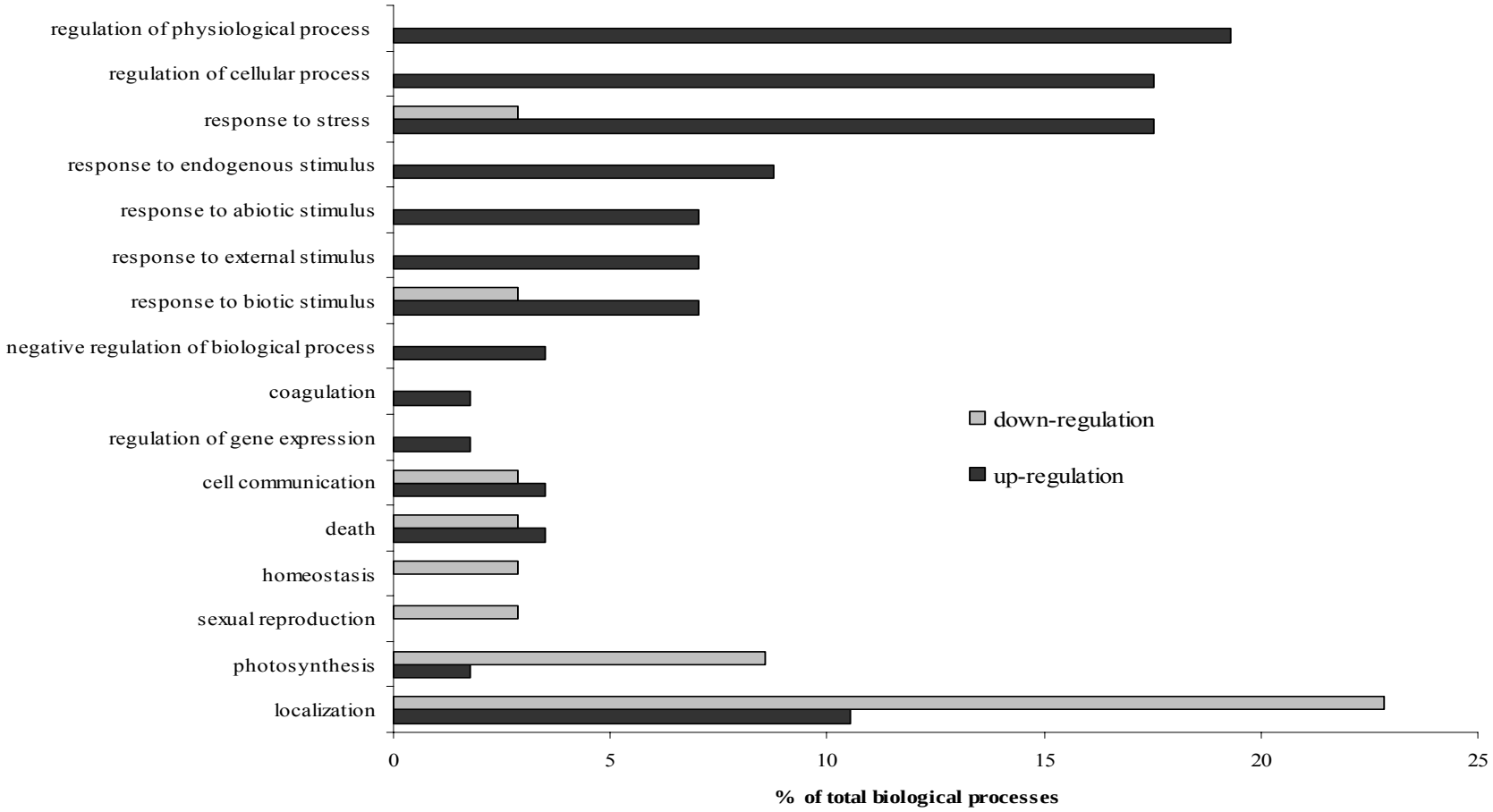


Figure 4.6. Percentage representations of third-level annotations from GO biological process category between up- and down-regulated genes. Note that the metabolism (GO:0008152) and the cellular physiological process (GO:0050875) groups were excluded in this chart because most candidate genes were involved in those two groups.

Table 4.5. Summary of selected cis-acting regulatory elements highly represented in rice homologs of up-regulated *C. dactylon* genes.

Profiling Group[a]			SU3D	SD3D	SU6D	SD6D	SU9D	SU9D	SU10P	SD10P	SU20P	SD20P	TU	TD
Gene number			37	29	34	20	52	14	17	9	29	11	87	57
cis-element	Sequences	Site #[b]	%	%	%	%	%	%	%	%	%	%	%	%
GAGA	$(GA)_9$	S000405	10.8	3.4	11.8	0.0	9.6	0.0	17.6	0.0	13.8	0.0	**9.2**[c]	**0.0**
Dc3	ACACNNG	S000292	83.8	69.0	76.5	70.0	80.8	85.7	**82.4**	**44.4**	75.9	90.9	78.2	75.4
LTRE-1	CCGAAA	S000250	32.4	24.1	29.4	15.0	**26.9**	**7.1**	**35.3**	**11.1**	**24.1**	**9.1**	**31.0**	**17.5**
ACTCAT	ACTCAT	S000450	**37.8**	**6.9**	35.3	35.0	46.2	21.4	**29.4**	**0.0**	**41.4**	**27.3**	**40.2**	**22.8**
PRE	$SCGAYNR(N)_{15}HD$	S000506	**70.3**	**34.5**	52.9	45.0	61.5	64.3	58.8	44.4	65.5	63.6	**69.0**	**47.4**
Pyrimidine box	TTTTTTCC	S000298	16.2	27.6	26.5	10.0	**23.1**	**0.0**	23.5	44.4	20.7	9.1	25.3	15.8
S-box	CACCTCCA	S000500	**10.8**	**0.0**	5.9	4.7	9.6	0.0	5.9	0.0	**10.3**	**0.0**	**10.3**	**1.8**
TATCCAC box	TATCCAC	S000416	16.2	6.9	11.8	15.0	17.3	7.1	29.4	11.1	**20.7**	**0.0**	16.1	8.8
Up2	AAACCCTA	S000472	18.9	10.3	14.7	5.0	**11.5**	**0.0**	23.5	11.1	17.2	9.1	**20.7**	**5.3**

[a]SU3D, upregulation at 3DAT; SD3D, downregulation at 3 DAT; SU6D, upregulation at 6 DAT; SD6D, downregulation at 6 DAT; SU9D, upregulation at 9 DAT; SD9D, downregulation at 9 DAT; SU10P, upregulation in 10 % PEG; SD10P, downregulation in 10 % PEG; SU20P, upregulation in 20 % PEG; SD20P, downregulation in 20 % PEG; TU, total up-regulations; TD, total down-regulations.
[b]PLACE database (www.dna.affrc.go.jp/PLACE/) accession numbers.
[c]Numbers in bold indicate that two proportions are significantly different at the 99 % confident level.

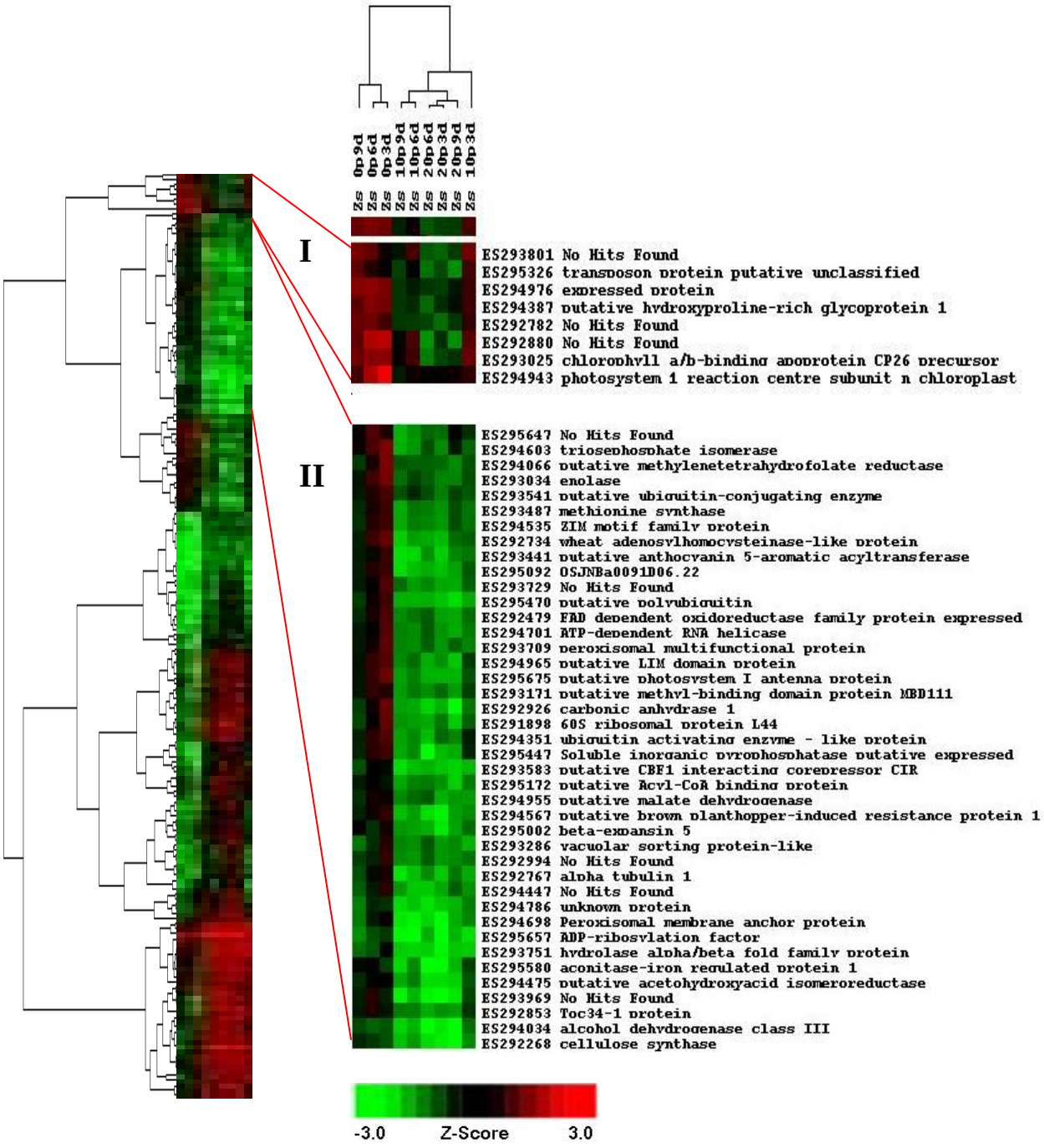


Figure 4.7. Clustering of the 189 drought candidate genes. Each row corresponds to the drought candidate gene and each column corresponds to Z-score for different treatments and time points (*Zs, Z-score; 0p9d, Control at 9 DAT; 0p6d, Control at 6 DAT; 0p3d, Control at 3 DAT; 10p9d, 10 % PEG at 9 DAT; 10p6d, 10 % PEG at 6 DAT; 20p6d, 20 % PEG at 6 DAT; 20p3d, 20 % PEG at 3 DAT; 20p9d, 20 % PEG at 9 DAT; 10p3d, 10 % PEG at 3 DAT).

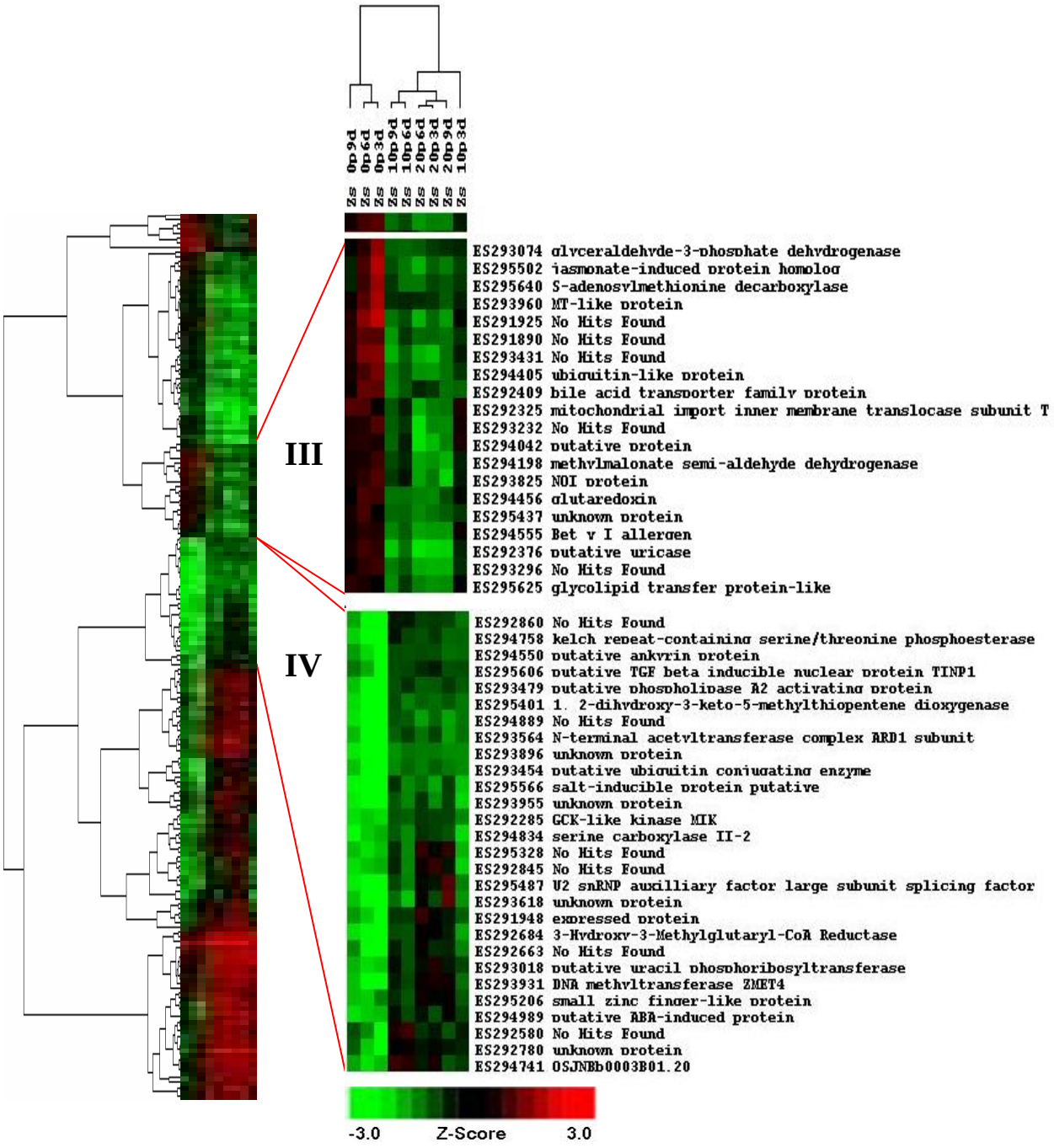


Figure 4.7. Continued.

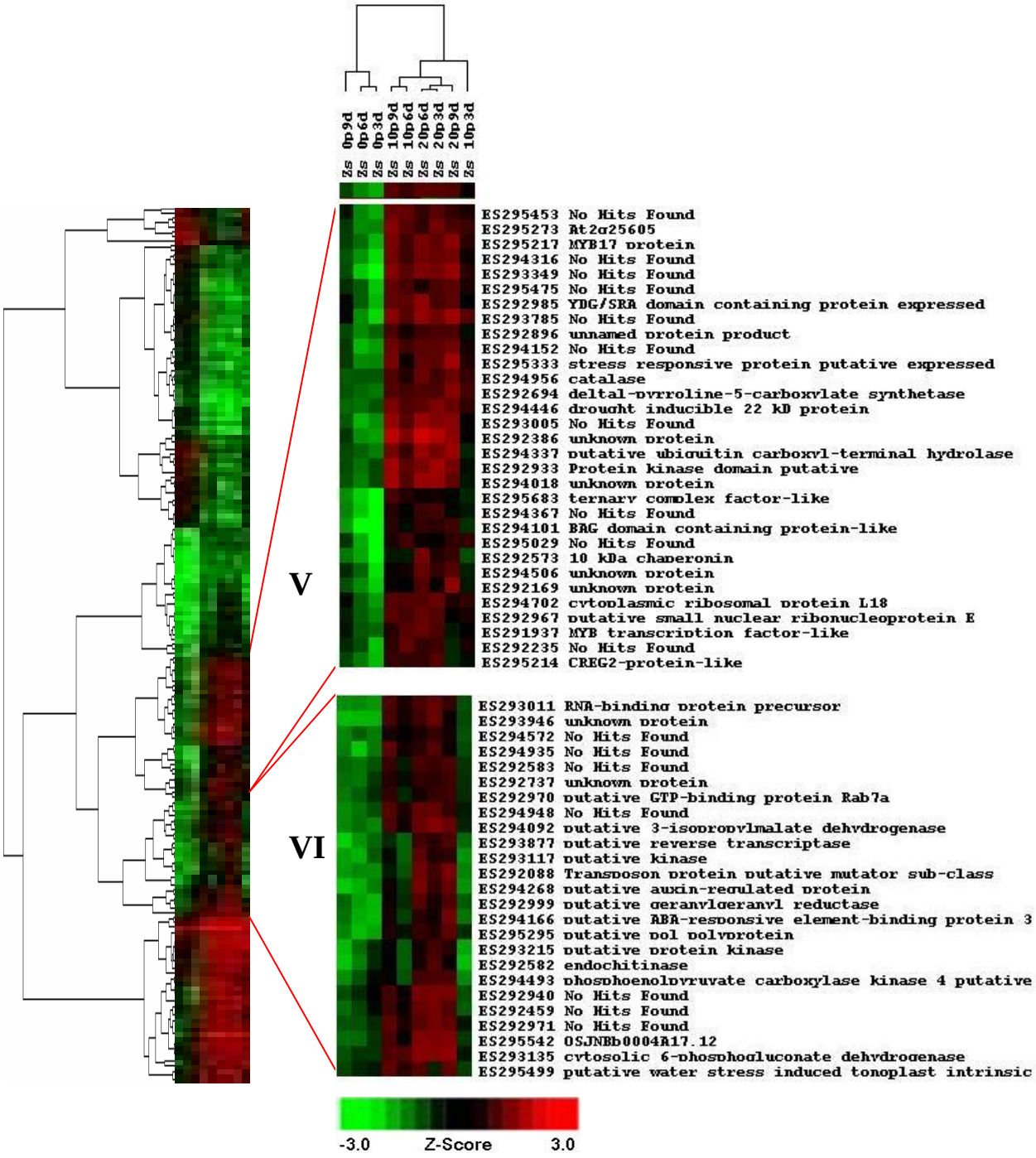


Figure 4.7. Continued.

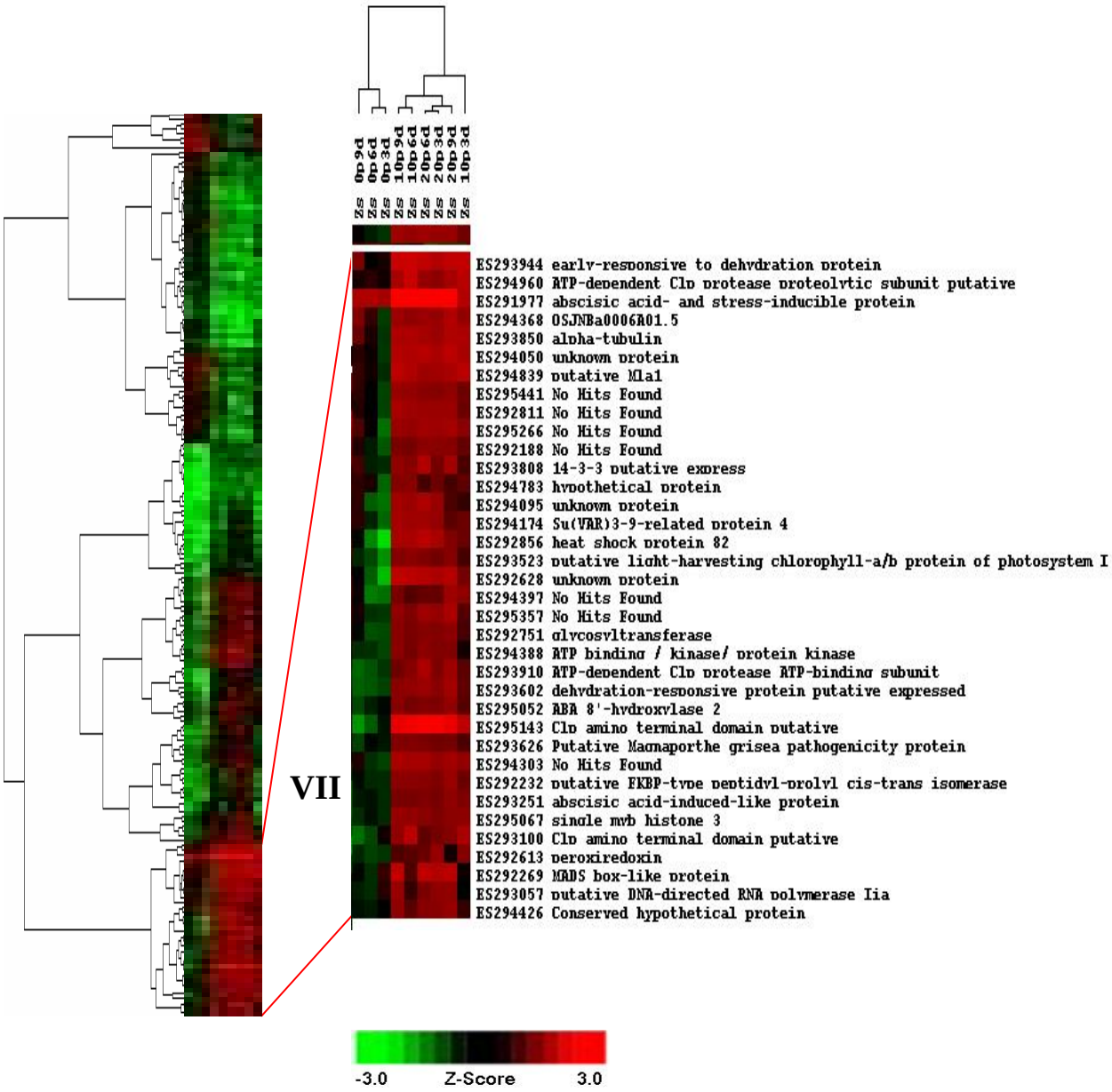


Figure 4.7. Continued.

Table.4.6. Summary of comparison between the genes from each cluster and the corresponding GO terms. Note that one gene can have multiple GO terms and vice versa.

	69 down-regulated genes			120 up-regulated genes			
Cluster	I	II	III	IV	V	VI	VII
Number of genes in each cluster	8	41	20	28	31	25	36
Number of genes with GO terms	4	31	11	17	15	15	26

Cluster-specific GO terms (level 3)

Cluster	Accession number	GO ID	Description
I	ES294943	GO:0009522	photosystem I
II	ES293034 ES295447	GO:0000015	phosphopyruvate hydratase complex
	ES295002	GO:0019953	sexual reproduction
III	ES292409	GO:0005342	organic acid transporter activity
		GO:0042592	homeostasis
	ES294456	GO:0005489	electron transporter activity
		GO:0015075	ion transporter activity
IV	ES294741	GO:0000155	two-component sensor molecule activity
	ES295206	GO:0042719	mitochondrial intermembrane space protein transporter complex
	ES293931	GO:0040029	regulation of gene expression, epigenetic
V	ES292896	GO:0004386	helicase activity
	ES295217	GO:0009605	response to external stimulus
VI	ES295295	GO:0005199	structural constituent of cell wall
		GO:0009986	cell surface
	ES292582	GO:0001871	pattern binding
	ES292999	GO:0008135	translation factor activity, nucleic acid binding
VII	ES293251	GO:0004803	transposase activity
	ES295052	GO:0046906	tetrapyrrole binding
	ES293057 ES293850	GO:0030880	RNA polymerase complex
	ES294095	GO:0050817	coagulation

CHAPTER 5

PHYLOGENETIC PREDATING OF GENE DUPLICATION EVENTS IN *Cynodon dactylon* L. AND MODEL GRASS SPECIES BY COMPARATIVE GENOMIC APPROACHES[1]

[1]Kim C, Tang H and Paterson AH. To be submitted to *New Phytologist*.

ABSTRACT

ESTs (Expressed Sequence Tags) from a variety of plant species have been useful for comparative genomics. Using EST collections, we generated unigene sets and analyzed them to further elucidate the evolutionary history of grass subfamilies. A total of eight grasses [*Cynodon dactylon* (Bermudagrass), *Sorghum bicolor* (sorghum), *Saccharum officinarum* (sugarcane), *Zea mays* (maize), *Oryza sativa* (rice), *Hordeum vulgare* (barley), *Festuca arundinacea* (tall fescue), and *Triticum aestivum* (wheat)] in four subfamilies and five tribes were analyzed using two different approaches. The evolution of the Chloridoideae subfamily, previously lacking sequence data, was clarified by virtue of Bermudagrass ESTs generated from a normalized cDNA library. Age distributions of duplicated genes based on synonymous substitution rate (Ks) suggested several duplication events in Bermudagrass, sorghum, barley, tall fescue, and wheat. Phylogenetic analysis with the unigene sets indicated that the analyzed grasses diverged from a common ancestor after a shared ancient polyploidization (ca. 50.0 ~ 67.8 million years ago). Additional duplication events were indicated after divergence of the PACC and BEP clade in sorghum, tall fescue, and rice. Both age distributions and phylogenetic analyses are attractive approaches for using partial sequences to find large-scale genomic changes in any species if a sufficient number of ESTs is available.

INTRODUCTION

Genome duplication or polyploidy is common in flowering plants (Stebbins, 1971). The once controversial proposal that evolution moves forward through whole genome duplication is gaining support from sequence analysis, which is more sensitive than methods such as chromosome counting, analysis of meiotic chromosome pairing, and marker-based mapping, which have previously been used to assess polyploidy. Accumulating evidence shows that genome duplications occurred in the lineages of all vertebrates (Wolfe, 2001). In plants, polyploidy was traditionally proposed to have occurred in the lineage of at least 70% of angiosperms (Masterson, 1994) and 95% of pteridophytes (Soltis & Soltis, 1999). The full genomic sequencing of *Arabidopsis*, rice, and sorghum facilitates the investigation of whole genome duplications. While these particular species are considered to be classical diploids, they have been revealed to be ancient polyploids (paleopolyploids) and indeed, the timing of these polyploidizations suggests that all angiosperms are paleopolyploids (Bowers *et al.*, 2003; Paterson *et al.*, 2004). Methods have also been developed to estimate whole genome duplication events in the evolutionary history of the organism for which only partial sequences are available, by comparative genomic approaches (Vandepoele *et al.*, 2003; Bowers *et al.*, 2003; Blanc & Wolfe, 2004; Paterson *et al.*, 2004).

The grass family (Poaceae) is the main source of food for humans and animals, and an important part of the urban and suburban landscape. Ecologically, the significant impact of the family is reflected by its dominance in nature. The Poaceae contains approximately 10,000 species and 700 genera, and covers approximately 20% of the earth's land surface (Gaut, 2002). Owing to its importance, the evolutionary history of the family has been intensively investigated, facilitated in the last two decades by molecular approaches.

Bermudagrass (*Cynodon* spp.) belongs to the subfamily Chloridoideae, which is a part of the clade 'PACC' containing subfamilies Panicoideae, Arundinoideae, Centothecoideae, and Chloridoideae. All C4 species fall within the PACC clade (Kellogg, 2000). The panicoideae has been actively studied because it includes crops such as maize, sorghum, sugarcane, pearl millet, and foxtail millet, whereas the other subfamilies have been less well studied. The subfamily Chloridoideae is especially economically important in that it contains most warm-season turfgrass species, such as *Cynodon* spp. (bermudagrass), *Buchloe* spp. (buffalograss), and *Zoysia* spp. (zoysiagrass). Nevertheless, the number of nucleotide sequences is only 13,297 for the 215 species in 68 genera of this subfamily, as of February 2007 (Taxonomy browser, the National Center for Biotechnology Information (NCBI)).

The objectives of this study are to clarify the evolutionary history of *Cynodon* and other major crop species using comparative genomic tools and large expressed sequence tag (EST) databases. Whole genome duplications have played a major role in determining the structure of eukaryotic genomes. The nearly completed sequences of Arabidopsis, rice, and sorghum enable researchers to reveal the history of angiosperm genome evolution and provide a foundation for advancing knowledge about many other flowering plants by comparative approaches. This study includes new evidence regarding the evolutionary history of *Cynodon* (Chloridoideae), using a large body of data generated by ESTs, and taking advantage of genome sequences for other species of angiosperms.

MATERIALS AND METHODS

This study was conducted by two different methods, described by Blanc & Wolfe (2004), and Chapman *et al.* (2004), respectively. However, the overall procedures were modified

and new python scripts were generated. The procedures and tools used in this study are summarized in Figure 5.1.

I. Age distributions of duplicate genes from *Cynodon* and other grass species based on the synonymous substitution rate

Collection of unigene sets

EST sequences for *Oryza sativa* (spp. *japonica*), *Festuca arundinacea*, *Sorghum bicolor*, and *Saccharum officinarum* were downloaded from the NCBI dbEST database. For *C. dactylon*, EST sequences from the cDNA library described in Chapter 3 and the NCBI dbEST were combined to generate unigenes. The collected EST and cDNA sequences were cleaned of potential vector contamination using the Institute for Genomic Research (TIGR) SeqClean tool (http://www.tigr.org/tdb/tgi/software) and the NCBI UniVec database (http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html). The SeqClean tool also screens for low-complexity regions, poly A and poly T tracts, and sequence ends rich in undetermined bases, as well as removes low quality sequences. EST sequences corresponding to the same transcript were then assembled into unigenes with the program TGICL (Pertea *et al.*, 2003) using default parameters. Unigene sets for *Hordeum vulgare, Zea mays, and Triticum aestivum* were downloaded from TIGR gene indices (http://compbio.dfci.harvard.edu/tgi/plant.html).

Identification of paralogs/orthologs and dataset cleaning

After constructing unigene sets, all unigene sequences were searched against plant repeated sequences downloaded from the TIGR Plant Repeat Database (Ouyang & Buell, 2004;

http://www.tigr.org/tdb/e2k1/plant.repeats/index.shtml) using BLASTN (Altschul *et al.*, 1997). Any unigene matching a repeated sequence over 150 bp or more and with E-value less than 10^{-15} was removed from the dataset.

In order to identify paralogous sequences, all-against-all nucleotide sequence similarity searches were performed among the unigene sequences for each species using the program BLASTN. Sequences aligned over 300 bp and showing at least 40% identity were defined as pairs of paralogs. To identify putative orthologs between two species, each sequence from one species was searched against all sequences from the other species, and the same procedure was done conversely. Two sequences were defined as orthologs if each was the best match to the other and if the sequences were aligned over 300 bp or more and with E-value less than 10^{-20}.

Each member of a pair of sequences was searched using BLASTX (Altschul *et al.*, 1997) against all plant protein sequences available in the NCBI nr protein database. The best match was considered significant if the alignment length was more than 100 amino acids and E-value was less than 10^{-15}. If no significant match was found, the pair of sequences was discarded.

Estimation of the level of synonymous substitution between two sequences

The cleaned pairs of sequences were translated using the Genewise program, which can infer frameshift sites (Birney *et al.*, 2004), with the corresponding best match protein from the BLASTX search as a guide. For each pair of paralogs, the two translation products were aligned using ClustalW (Thompson *et al.*, 1994), and the resulting alignment was used as a guide to align the nucleotide sequences. After removing gaps and N-containing codons, the level of synonymous substitution was estimated using the maximum likelihood approach implemented in the program CODEML (Goldman & Yang, 1994; http://bcr.musc.edu/manuals/paml.html) which

is a part of the PAML package (Yang, 1997). Batch jobs were performed using various in-house python scripts.

Eliminating redundant Ks values in gene families

A gene family comprised of n members may be as a result of n - 1 gene duplication events. However, the number of possible pairwise comparisons within a gene family [$n \times (n - 1) / 2$] can be substantially larger than the number of gene duplications, which causes multiple estimates of the ages of some duplications. To eliminate those redundant Ks values, pairs of duplicated sequences were first grouped into gene families using a single linkage clustering method. For example, if A/B is one pair of paralogous sequences and B/C is another pair, then A, B, and C were defined as members of the same family, even if the pair A/C was absent in the whole set of pairs. After grouping all paralogous sequences into corresponding gene families, a hierarchical clustering method was used to reconstruct a tentative phylogeny of each gene family. For each gene family, each sequence in the family was initially treated as a cluster in order to compare the Ks value of each cluster with all possible pairs of clusters. Then, the pair of clusters having the smallest Ks value was replaced by a single new cluster containing all their sequences. The median Ks value was taken to represent the duplication event that gave rise to the two merged clusters. The hierarchical clustering steps were repeated until all sequences were contained in a single cluster. For example, if A, B, and C are members of a gene family, and A/B and B/C are paralogous pairs of which Ks values are 1.0 and 1.2, respectively, a median Ks value (1.1) of two paralogous pairs will represent the Ks value of this gene family.

II. Age distributions of duplicate genes from *Cynodon* and other grass species based on the analysis of phylogenetic trees

Rice duplication dataset and organism data creation

A revised version of the rice genome duplication dataset published by Paterson *et al.* (2004) was downloaded from the Plant Genomic Duplication Database (PGDD, http://chibba.agtec.uga.edu/duplication/index/home). The sequence databases for a set of grasses were established using the unigene sets previously used for Ks calculations. The unigene set for *Arabidopsis thaliana* was downloaded from TIGR gene indices (http://compbio.dfci.harvard.edu/tgi/plant.html). The protein sequences of *Physcomitrella patens* and *A. thaliana* predicted by whole genome sequences were downloaded from COSMOSS (http://www.cosmoss.org/) and TAIR (The Arabidopsis Information Resource; http://www.arabidopsis.org/), respectively. The predicted protein sequences of *S. bicolor* were obtained from an internal database (unpublished data) in the Plant Genome Mapping Laboratory.

Detection of homologs

To find homologs and generate phylogenetic trees, duplicated gene pairs from rice were searched against the organism databases. Homologs were defined as the TBLASTN (Altschul *et al.*, 1997) hits with the longest match region above a configurable significance (*E*-value) threshold (the default is an *E*-value of 10^{-5}). Information about best hits was collected to construct phylogenetic trees.

Generation of phylogenetic trees

Phylogenetic trees were constructed using four different protein sequences: (a) the two protein sequences from a duplicate gene pair (rice); (b) the best homolog from a comparison organism representing a particular taxonomic node (e.g. *C. dactylon* representing Chloridoideae); and (c) the best homolog from *Physcomitrella* (moss) as an outgroup known to be a very distant relative and thus used as the root in the final phylogeny. Since the sequence (b) was a cDNA sequence, it was translated using the Genewise program (Birney *et al.*, 2004) with the corresponding best match protein from BLASTX search (of all plant protein sequences available in GenBank) as a guide. For the four protein sequences, multiple alignments were performed by ClustalW (Thompson *et al.*, 1994), and used as input into phylogenetic analyses to produce rooted trees comparing the duplicate pairs and best homolog. The PHYLIP set of programs (version 3.67, http://evolution.genetics.washington.edu/phylip.html) was used for bootstrapped maximum likelihood ('proml') and protein parsimony ('protpars') analyses.

Interpretation of individual trees

The results of phylogenetic analysis led to only two possible rooted tree topologies. In one topology (external tree), the members of the duplicated pair are more similar to one another than either is to the best homolog, suggesting that gene duplication is more recent than taxon divergence. In the other topology (internal tree), the homolog is more similar to one member of the duplicate pair than to the other member of the duplicate pair, suggesting that taxon divergence is more recent than gene duplication.

Inferences about the timing of genomic duplication are based on differences in the frequencies of 'internal trees'. The fraction of 'internal trees' associated with each duplication

block of rice were compared by using one-way ANOVA for corrected samples and Tukey's studentized range test for post-ANOVA comparisons among species. Individual duplicated pairs were considered treatments, and the indicated species were conditions, accounting for correlations that may result from comparing identical genes in different species as described by Bowers *et al.* (2003).

RESULTS

Age distributions of duplicated genes using synonymous substitution rates

The numbers of initial cDNA sequences, unigenes, paralogous proteins, and Ks values used for each species were summarized in Table 5.1. Since initial unigene sets were filtered using the TIGR Plant Repeat Database, the unigenes listed in Table 5.1 are largely repeat-free. A large number of putative paralogous sequences were identified for each species after self-BLASTN search. The putative paralogs could have originated from the same cDNA sequence as a result of nonoverlapping EST sequences. Thus, this procedure to find paralogous sequences may overestimate the real number of paralogous sequences. The protein sequence for each paralogous sequence was translated with GeneWise software with the corresponding best-match proteins from BLASTX search against all plant protein sequences available in GenBank. If no significant best match was found, the pair of paralogous sequences was discarded. This step filtered a number of ambiguous sequences, and the synonymous substitution rate between pairs of paralogous sequences was calculated using the translated proteins. Only Ks values less than 2.50 were plotted in the Ks distributions.

In theory, the number of synonymous substitutions per site increases with time. Thus, the synonymous substitution rate (Ks) between coding sequences of each paralogous pair may

provide a relative chronology of duplication events. Since gene duplication is an ongoing process (Blanc & Wolfe, 2004), usually the peak of a distribution of Ks values against the number of paralogous pairs is at a very small Ks value, reflecting very recent single-gene duplications. If gene duplications are assumed to be random and occur at relatively steady rates during evolution, the Ks distribution may exponentially decrease from the initial peak as age (Ks) increases (null hypothesis). Therefore, any remarkable secondary peak in the Ks distribution may be recognized as a large-scale duplication such as whole genome duplication, segmental duplication, aneuploidy, or polyploidy.

In our study, eight different grasses were analyzed. Bermudagrass, sorghum, sugarcane, and maize represent the PACC clade of the grass family; on the other hand, rice, barley, tall fescue, and wheat represent the BEP clade. The tested grasses fall into four different subfamilies and five different tribes. Any secondary peak less than Ks = 1.0 was only considered a large-scale duplication event because Ks values greater than 1.0 were considered to have reached a saturation point where multiple substitutions may have occurred at a synonymous site (Li, 1997). A certain degree of subjectivity to interpret Ks plots for paralogous pairs is inevitable as no appropriate statistical method was available to formally test the null hypothesis of steady exponential decrease of frequencies of paralogous pairs with increased Ks.

Table 5.2 indicates possible secondary peaks that deviate from the null hypothesis, and their corresponding ages. The ages of the modes of the secondary peaks were estimated using a molecular clock rate of 6.5×10^{-9} synonymous substitutions per synonymous site per year, as proposed by Gaut *et al.* (1996). For grasses included in the PACC clade, a possible secondary peak was found from Ks = 0.70 to 0.75 in Bermudagrass [53.8 to 57.7 million years ago (MYA)] and two possible secondary peaks were found at 0.30 and 0.85 in sorghum (23.1 and 65.4 MYA,

respectively). However, no traces were found in sugarcane and maize (Table 5.2; Figure 5.2). For the BEP clade, a possible secondary peak was identified in barley (Ks = 0.75; ca. 57.7 MYA) and wheat (Ks = 0.10; ca .7.8 MYA), respectively, and two possible secondary peaks were found in tall fescue at Ks = 0.25 (ca. 19.2 MYA) and from Ks = 0.60 to 0.65 (46.2 to 50.0 MYA). No remarkable traces were found in rice because secondary peaks may be masked by low Ks values (Table 5.2; Figure 5.3). The signals of the secondary peaks for wheat and tall fescue were relatively strong; in contrast, those for barley, Bermudagrass, and sorghum were relatively weak.

In order to assess the relative age of speciation, the synonymous substitution rates for orthologous pairs identified between grasses was estimated with the same method used for paralogous pairs. Table 5.3 indicates the numbers of orthologous pairs and modal Ks values (Ks peak formed by orthologous pairs) between grasses. For the PACC clade (Table 5.3; Figure 5.4), modal Ks values between Chloridoideae (Bermudagrass) and Panicoideae (sorghum, sugarcane, and maize) ranged from 0.45 to 0.50 (34.6 to 38.5 MYA), reflecting the speciation time of those two subfamilies. The modal Ks value between maize and sorghum/sugarcane was 0.15 (ca. 11.5 MYA) whereas the modal Ks value between sorghum and sugarcane was 0.10 (ca. 7.8 MYA), indicating that maize was separated before the speciation of sorghum and sugarcane. For the BEP clade (Table 5.3; Figure 5.5), Ehrhartoideae (rice) and Pooideae (barley, tall fescue, and wheat) subfamilies showed modal Ks values from 0.50 to 0.60 (38.5 to 46.2 MYA), which indicate the period of evolution corresponding to their common ancestor. Tall fescue is classified into the Poeae tribe, while barley and wheat are included in the Triticeae tribe of the Pooideae subfamily. The modal Ks value from 0.25 to 0.30 (tall fescue vs. barley/wheat) may represent the speciation event of the common ancestor between tribes Poeae and Triticeae, estimated at from

19.2 to 23.1 MYA. The orthologous pairs between barley and wheat showed a modal Ks value of 0.15 (ca. 11.5 MYA).

A comparison of orthologous pairs between the PACC (Bermudagrass and sorghum) and the BEP clade (rice and barley) was also performed in the same manner, which revealed that the two different clades were separated by from Ks = 0.55 to 0.65, corresponding to 42.3 to 50.0 MYA (Table 5.3; Figure 5.6). The divergence of those two clades may have provided major evolutionary changes in grasses because C4 grasses fall into the PACC clade and have independently evolved from C3 grasses included in the BEP clade (Kellogg, 2000).

Age distributions of duplicated genes using phylogenetic trees

Since *O. sativa* (rice) is wholly sequenced and well-annotated at present, rice duplicated proteins were used as the initial source for this study. The rice duplication dataset was downloaded from the Plant Genomic Duplication Database (PGDD, http://chibba.agtec.uga.edu/duplication/index/home). The dataset includes 3,506 rice duplicated protein pairs based on 66,710 gene models in a rice pseudomolecule, released by TIGR (TIGR version 5.0, January, 2007).

To make inference between the duplication event of rice and other model grasses, phylogenetic trees were constructed with four different protein sequences including a pair of rice duplicates, the best-matching sequence from the plant species of interest, and a homologous protein of *Physcomitrella* as an outgroup, as described by Chapman *et al.* (2004). The analysis was performed using multiple sequence alignments of the four proteins. One hundred bootstrap replicates were examined using protein parsimony, with PHYLIP's 'protpars' default implementation for scoring amino acid changes. Only trees with at least 70 (out of 100) percent

bootstrap confidence were used in the analysis. Table 5.4 indicates the frequency of internal trees for each tested plant corresponding to rice duplication blocks. The frequencies of internal trees in grasses were significantly higher than in *Arabidopsis* based on Tukey's studentized range test ($P < 0.01$), indicating that the rice duplication event occurred before speciation of the tested grasses and after the divergence of rice - *Arabidopsis* (monocot - dicot). The results are consistent with the results of Paterson *et al.* (2004) and Chapman *et al.* (2004); however, our results included much higher frequencies of internal trees than the two previous studies. The result seems to be affected by the number of unigenes used for the phylogenetic analysis. Practically, the frequencies of internal trees tend to be positively correlated with the number of unigenes within each block in our study.

The corresponding frequencies of internal trees for block 10 (chromosome 11 - 12) in tested grasses were remarkably different from those for other blocks, which may reflect that block 10 resulted from a segmental duplication event as previously described by Wang *et al.* (2005) and Yu *et al.* (2005). To support the result, average Ks values for each block were statistically tested and shown in Table 5.5. The average Ks value of block 10 was significantly different from that of other blocks based on the Student-Newman-Keuls (SNK) test ($P < 0.01$).

To see whether the types of sequence database could affect the frequency of internal trees, two different types of sequence database (unigene sets and proteins predicted by whole genome sequences) were applied to *Arabidopsis* and sorghum using the same procedure. Table 5.6 shows pairwise comparison of the frequency for each duplicated block between the two different sequence databases. For *Arabidopsis*, the frequencies were not significantly different between the database types within each block, whereas for sorghum the frequencies of all the blocks other than block 10 were significantly higher in predicted protein than in unigenes at the

95 % or more confident level. The result of *Arabidopsis* can be explained in two different ways. One explanation could be the exact translation of *Arabidopsis* unigenes. Since the unigenes were translated with the GeneWise program using the corresponding best match protein from the BLASTX search against all plant protein sequences available in the NCBI nr protein database and all the proteins of *Arabidopsis* are already included in the database, the translated unigenes may provide precise translation of *Arabidopsis* unigenes. This could be the reason why the internal tree ratios of predicted proteins were not significantly different from those of unigenes. The other explanation could be the loss of signal caused by the evolutionary distance between rice and *Arabidopsis*. In other words, a number of signals for internal trees have been lost because *Arabidopsis* was separated from the rice lineage long before the ancient duplication of rice. As a result, the signals are too weak to make differences between the two different sequence databases. In contrast, the sorghum unigene set tends to underestimate internal trees compared with predicted proteins. The sorghum genome is less complete and less well-annotated than *Arabidopsis*, and more prone to false inferred translation of sorghum unigenes. Additionally, a number of internal trees may be lost unless the unigene set for sorghum sufficiently covers all sorghum genes.

DISCUSSION

Duplication and speciation events across the evolution of grass lineage

The primary goal of the current study is to clarify the evolutionary history of the grass family, especially with regard to the Chloridoideae. Although several evolutionary studies of the grass family using large-scale sequence data have been attempted recently (Vandepoele *et al.*, 2003; Blanc & Wolfe, 2004; Paterson *et al.*, 2004; Schlueter *et al.*, 2004; Swigonova *et al.*, 2004;

Wang *et al.*, 2005; Yu *et al.*, 2005; Wei *et al.*, 2007), sequence data for some subfamilies such as Chloridoideae was previously lacking. By virtue of a normalized cDNA library from *C. dactylon*, the Chloridoideae subfamily could now be studied in much more detail.

The duplication and speciation events observed or expected to be detected in our study are summarized in Figure 5.7. Chaw *et al.* (2004) proposed that monocots and dicots diverged at ca. 140 ~ 150 MYA (late Jurassic - early Cretaceous) based on whole chloroplast genome analysis. Prasad *et al*. (2005) suggested that grasses originated roughly at 80 ~ 85 MYA based on the analysis of phytoliths – microscopic pieces of silica formed in plant cells – in coprolites of dinosaurs. Our study suggests (Figure 5.7) that an ancient duplication occurred before the divergence of the BEP and the PACC clade, shared by most if not all of the grass family. Paterson *et al.* (2004) reported that the event occurred at ca. 70 MYA, and our study suggests slightly more recent dates based on Ks values between rice duplicated proteins (50.0 to 67.8 MYA; Table 5.5). Traces of the ancient polyploidization, called *rho* by some, were observed in the Ks distribution of paralogous pairs in Bermudagrass, sorghum, barley, and tall fescue, with modal secondary Ks peaks ranging 0.60 to 0.85 and corresponding to 46.2 to 65.4 MYA (Figure 5.7). Curiously, but consistent with the findings of others (Blanc & Wolfe, 2004), no traces of the *rho* were found in the Ks distribution of rice (Figure 5.3). This might be caused by the number of recent paralogous pairs in rice. In other words, relatively small signals reflecting ancient polyploidization or segmental duplication might be masked by a large number of young duplicated genes. When low Ks values (< 0.15) were removed from the Ks distribution of rice (data not shown), small secondary peaks were found at Ks = 0.40 and Ks = 0.90. The latter (Ks = 0.90) peak is consistent with our earlier estimate of the ancient polyploidization, although slightly later than the new estimates herein. Additionally, rice duplication block 10 (chromosome

11 - 12; Table 5.5) provided evidence of a segmental duplication event suggested by Wang *et al.* (2005) and Yu *et al.* (2005). To test if the former (Ks = 0.40) peak corresponds to the segmental duplication, paralogs of which Ks values are close to 0.40 were extracted and searched against the rice duplicated protein database using BLASTX. The paralogs were significantly concentrated on chromosome 11 or 12, indicating that the peak represents the segmental duplication (based on a contingency test, $\chi^2 = 20.43$, df = 1, $P < 0.001$); however, an estimated time of the segmental duplication based on Ks = 0.40 (ca. 30.8 MYA) was much earlier than ca. 18.9 MYA, which was estimated by rice duplicated protein pairs (Table 5.5). This discrepancy seems to result from different sources of sequence data (EST and whole genome sequence).

The BEP and the PACC clade diverged from 42.3 to 50.0 MYA based on Ks distributions between orthologous pairs, with the ancient polyploidization event predating their divergence. Gaut (2002) suggested that the assumed rice – maize divergence time (ca. 50 MYA) by White & Doebley (1999) is based on unconvincing evidence and needs to be reexamined with more thorough data; however, the divergence time estimated in our study is quite consistent with the estimate of ca. 50 MYA.

Figure 5.7 also indicates that the divergence time of the common ancestor between subfamilies Ehrhartoideae and Pooideae is from 38.5 to 46.2 MYA. Those BEP clade subfamilies include economically important food crops, forage crops, and cool-season turfgrasses, such as rice, barley, oat, wheat, tall fescue, Kentucky bluegrass, perennial ryegrass, and orchardgrass. Based on the Ks distribution, two large-scale duplication events were detected in the lineage of tall fescue and wheat (Figure 5.3). ESTs from bread wheat (*T. aestivum*) were exclusively used for our study, which is an allohexaploid (2n = 6x = 42) containing three different genomes (AABBDD). The allohexaploid was formed by chromosome doubling,

between a tetraploid (genome BBAA) and a diploid (genome DD) species ca. 9,500 years ago (Levy & Feldman, 2001). Huang *et al.* (2002) showed that the tetraploidization events occurred relatively recently (2.5 - 4.5 MYA). Thus, the trace of a large-scale duplication event in our study (ca. 7.8 MYA; figure 5.7) may be a mark of large-scale duplications, left by an ancestor of diploid genome progenitors of bread wheat (genome AA, BB, and DD) before the formation of the tetraploid (genome AABB). Sleper (1985) reported that tall fescue is an allohexaploid (2n = 6x = 42) which contains three genomes (PG1G2), the P (2n = 2x = 14) genome from *F. pratensis* and the G1G2 (2n = 4x = 28) genome from *F. arundinacea* 'glaucescens'. Despite the revelation of its allohexaploid progenitors, the age of its formation is still unknown. If the large-scale duplication event found in our study reflects large-scale duplications of an ancestor of diploid genome progenitors, it is likely to occur at ca. 19.2 MYA. Since orthologs analyses between tall fescue and barley (or wheat) indicated that tribes Poeae and Triticeae diverged from 19.2 to 23.1 MYA, the large-scale duplication event of the ancestor is assumed to have occurred immediately after the divergence of those two tribes.

Two subfamilies – Chloridoideae and Panicoideae – included in the PACC clade were also analyzed in our study. Although large-scale duplication events that have formed the genome of the Andropogoneae tribe were proposed in maize (Gaut & Doebley, 1997; Swigonova *et al*., 2004; Wei *et al*., 2007) and sugarcane (Ming *et al*., 1998), no traces of duplication in those two grasses could be found in our study (Table 5.2). In sorghum, an additional secondary peak was found at Ks = 0.30, which may be associated with a possible segmental duplication event between chromosome 5 and 8 (Paterson *et al*., unpublished data). To test whether the peak corresponds to the apparent segmental duplication in sorghum, paralogs with Ks values around 0.30 were extracted and blasted against sorghum predicted proteins. A contingency test showed

that the paralogs were significantly biased to chromosome 5 or 8 ($\chi^2 = 18.13$, df = 1, $P < 0.001$). If the segmental duplication in sorghum is true, it predates the sorghum - maize divergence (ca. 11.5 MYA), which predicts that maize and sugarcane share this segmental duplication; however, no evidence was found in sugarcane and maize in this study. Additionally, this segmental duplication in sorghum is roughly consistent with the segmental duplication in rice based on their Ks values. Wang *et al*. (unpublished data) found that the "segmental duplication" occurs at corresponding locations in sorghum and rice. Further, while comparison of sorghum paralogs to one another, and rice paralogs to one another, each suggest a recent origin of this segment, comparison of sorghum genes to rice genes suggest that it has an ancient origin. Accordingly, Wang *et al*. (unpublished data) suggest that the region could be undergoing concerted evolution, independently in these two taxa (and perhaps also others). However, no traces were found of this phenomenon in the Pooideae and Chloridoideae subfamily in our study (indicated by small circles with question marks in Figure 5.7).

In autopolyploids *Saccharum officinarum* (sugarcane) and *Cynodon dactylon* (Bermudagrass), and the relatively recent paleopolyploid *Zea mays* (maize) (Wei *et al*., 2007), identification of genomic duplication events may be difficult due to their complex genomic structure. Some recent genomic duplication events expected to be found in those grasses were also absent, indicated by circles with question marks in Figure 5.7. While we could identify the ancient whole-genome duplication in Bermudagrass (2n = 4x = 36), a large scale duplication event could not be detected in maize or sugarcane, although maize is a clear paleopolyploid (Wei *et al*., 2007) and sugarcane has duplicated twice since its divergence from sorghum (Ming *et al*., 1998). One possibility is that genomic duplication occurred too recently to detect it with current methods, i.e. causing different paralogs to be collapsed into single unigenes. Since sufficient

sequence data for other Chloridoideae are not available and Bermudagrass is the first Chloridoid for which a large-scale EST project has been performed, the age of its recent genomic duplication cannot yet be estimated.

Methodology to improve

Both methods used for our study have the advantage of requiring only a partial gene sequence for analysis of duplication or speciation. The Ks distributions of orthologous pairs among grasses were quite consistent, whereas those of paralogous pairs within grasses were variable, which may be caused by the nature of ESTs used for our analyses. ESTs are generally partial sequences of expressed genes, allowing an EST database to have multiple entries for the same gene, leading to redundant Ks measures. This defect can be more easily associated with the analysis of paralogous than orthologous sequences. In other words, two ESTs can be derived from nonoverlapping regions of the coding sequence of the same gene because ESTs are partial reads of genes, but potentially are placed into different pairs with paralogous sequences. Another problem associated with ESTs is misassembled consensus sequences in unigene sets. Although TGICL applied to our study has been exclusively designed for ESTs (Pertea *et al*., 2003), it still generates some misassembled consensus sequences. Liang *et al.* (2000) reported that incorrect assemblies can be caused by sequencing error concentrated near the start or end of an EST, and ESTs involved in the same gene family with high sequence identity. For example, an EST sometimes remains as a singleton even if it is homologous to a consensus sequence, or two ESTs are assembled as a consensus sequence even if they are paralogous to each other. In addition, sequencing errors may also slightly overestimate Ks values. As a result, incorrect assembly and sequencing errors can affect the Ks distributions. In our study, to reduce the effect of

misassembled unigenes, paralogous pairs showing no synonymous substitutions (Ks = 0) were discarded before Ks values were plotted. However, any solution for sequencing error was unavailable because it is still beyond our control.

It is impossible to reveal very recent genomic changes with Ks distributions because the changes may be masked by the initial peaks, which is another limitation of Ks approach. The recent genomic change may have the key of genome structures of modern grasses, especially for polyploidy grasses. Among the tested grasses, Bermudagrass, sugarcane, tall fescue, and wheat are known polyploids, but no recent secondary peaks (low Ks values) were found in Bermudagrass or sugarcane. Despite relatively recent secondary peaks in wheat (ca. 7.8 MYA; Figure 5.7) and tall fescue (ca. 19.2 MYA; Figure 5.7), it may not directly reflect their allopolyploidizations but duplication events of common ancestors of their allohexaploid genome progenitors.

The other method, phylogenetic analysis, also uses partial gene sequences such as ESTs but it requires whole genome duplication data for at least one species (e.g. rice in our study). Owing to whole genome sequences of *Arabidopsis* and rice, this method can be widely used for angiosperms. However, it still has some drawbacks. First, the proportion of internal trees seems to be remarkably affected by the number of unigenes for a taxon. According to Table 5.4, Bermudagrass and tall fescue had relatively low proportions compared with those of other grasses, which might result from the small number of unigenes. If the number of unigenes increases, the drawback can be figured out but everyone knows this is not a simple task. Alternatively, the drawback can be solved by increasing the E-value thresholds of TBLASTN search. Table 5.7 indicates a comparison of the proportions by differing E-value thresholds for Bermudagrass and tall fescue. The proportions of internal trees significantly increased in

Bermudagrass along with decreased E-values. Although the proportions in tall fescue did not show significance, they slightly increased along with decreased E-values. Thus, the use of strict thresholds may reduce the problem. Second, the evolutionary distance from tested species to model species (rice) may affect the proportion of internal trees. Despite the higher number of unigenes in sugarcane than in barley, the proportion of barley was higher than that of sugarcane, possibly caused by evolutionary distance. The problem of evolutionary distance will most likely be solved because whole genome sequences of sorghum (closely related to sugarcane) will be available in the near future. Finally, this method needs a more accurate statistical test. The comparison among internal tree ratios for tested species was done using Tukey's studentized range test on the basis of previous reports (Bowers *et al.*, 2003; Paterson *et al.*, 2004). The statistical test can be remarkably influenced by any outliers in certain blocks as well as different numbers of trees in different blocks. For example, the test provided remarkably different results when pairwise comparisons were applied to proportions of total internal trees (Table 5.8) instead of multiple comparisons (Table 5.4). In fact, the pairwise comparison cannot be a better solution because it is also highly influenced by different sample size. Thus, finding an appropriate statistical method remains to be solved for this procedure. However, judging by the overall results, the efficiency of the phylogenetic tree procedure still provided reasonable inference about ancient duplication.

In conclusion, both methods have several obstacles to overcome; however, they are still a reasonable first evaluation for species which lack whole genome sequence. Since a large body of data generated by EST projects is now publicly released for various plant species, the methods are useful to find the approximate time of a large-scale duplication or speciation.

REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang A, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

Birney E, Clamp M, Durbin R. 2004. GeneWise and GenomeWise. *Genome Research* **14**: 988-995.

Blanc G, Wolfe KH. 2004. Widespread Paleopolyploidy in Model Plant Species Inferred from Age Distributions of Duplicate Genes. *Plant Cell* **16**: 1667–1678.

Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phlogenetic analysis of chromosomal duplication events. *Nature* **422**: 433-438.

Chapman BA, Bowers JE, Schulze SR, Paterson AH. 2004. A comparative phylogenetic approach for dating whole genome duplication events. *Bioinformatics* **20**: 180–185.

Chaw S-M, Chang C-C, Chen H-L, Li W-H. 2004. Dating the Monocot–Dicot Divergence and the Origin of Core Eudicots Using Whole Chloroplast Genomes. *Journal of Molecular Evolution* **58**: 424-441.

Gaut B. 2002. Evolutionary dynamics of grass genomes. *New Phytologist* **154**: 15-28.

Gaut BS, Doebley JF. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proceedings of the National Academy of Sciences, USA* **94**: 3809-6814.

Gaut BS, Morton BR, McCaig BC, Clegg MT. 1996. Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. *Proceedings of the National Academy of Sciences, USA* **93**: 10274-10279.

Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular biology and evolution* **11**: 725–736.

Grass Phylogeny Working Group. 2001. Phylogeny and subfamilial classification of the grasses (Poaceae). *Annals of the Missouri Botanical Garden* **88**: 373-457.

Huang S, Sirikhachornkit A, Su X. Faris J, Gill B, Haselkorn R, Gornicki P. 2002. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploidy wheat. *Proceedings of the National Academy of Sciences, USA* **99**: 8133-8138.

Kellogg EA. 2000. The grasses: a case study of macroevolution. *Annual review of ecology and systematics* **31**: 217-238.

Koch MA, Haubold B, Mitchell-Olds T. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, Arabis, and related genera (Brassicaceae). *Molecular biology and evolution* **17**: 1483-1498.

Levy AA, Feldman M. 2002. The impact of polyploidy on grass genome evolution. *Plant Physiology* **130**: 1587-1593.

Li WH. 1997. *Molecular Evolution*. Sunderland, MA: Sinauer Associates.

Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J. 2000. An optimized protocol for analysis of EST sequences. *Nucleic Acids Research* **28**: 3657-3665.

Masterson J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* **264**: 421-423.

Ming R, Liu S-C, Lin Y-R, da Silva J, Wilson W, Braga D, Van Deynze A, Wenslaff TF, Wu KK, Moore PH, Burnquist W, Sorrells ME, Irvine JE, Paterson AH. 1998.

Detailed alignment of *Saccharum* and *Sorghum* chromosomes: comparative organization of closely related diploid and polyploidy genomes. *Genetics* **150**: 1663-1682.

Ouyang S, Buell CR. 2004. The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Research* **32**: D360-D363.

Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proceedings of the National Academy of Sciences, USA* **101**: 9903-9908.

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* **19**: 651-652.

Prasad V, Stromberg CAE, Alimohammadian H, Sahni A. 2005. Dinosaur coprolites and the early evolution of grasses and grazers. *Science* **310**: 1177-1180.

Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC. 2004. Mining EST databases to resolve evolutionary events in major crop species. *Genome* **47**: 868-876.

Sleper DA. 1985. Breeding tall fescue. *Journal of Plant Breeding Review* **3**: 313–342.

Soltis DE, Soltis PS. 1999. Polyploidy: recurrent formation and genome evolution. *Tree* **14**: 348-352.

Stebbins GL. 1971. *Chromosomal evolution in higher plants.* London, UK: Edward Arnold.

Swigonova Z, Lai J, Ma J, Ramarkrishna W, Llaca V, Bennetzen JL, Messing J. 2004. Close split of sorghum and maize genome progenitors. *Genome Research* **14**: 1916-1923.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673-4680.

Vandepoele K, Simillion C, Van de Peer Y. 2003. Evidence that rice and other cereals are ancient aneuploids. *Plant Cell* **15**: 2192-2202.

Wang X, Shi X, Hao B, Ge S, Luo J. 2005. Duplication and DNA segmental loss in the rice genome: implications for diploidization. *New Phytologist* **165**: 937-946.

Wei F, Coe E, Nelson W, Bharti AK, Engler F, Butler E, Kim H, Goicoechea JL, Chen M, Lee S, Fuks G, Sanchez-Villeda H, Schroeder S, Fang Z, McMullen M, Davis G, Bowers JE, Paterson AH, Schaeffer M, Gardiner J, Cone K, Messing J, Soderlund C, Wing RA. 2007. Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLOS Genetics* **3**: e123.

White SE, Doebley JF. 1999. The molecular evolution of *terminal ear1*, a regulatory gene in the genus *Zea*. *Genetics* **153**: 1455-1462.

Wolfe KH 2001. Yesterday's polyploids and the mystery of diploidization. *Nature reviews Genetics* **2**: 333-341.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences* **13**: 555-556.

Yu, J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C, Zhang J, Zhang Y, Li R, Xu Z, Li S, Li X, Zheng H, Cong L, Lin L, Yin J, Geng J, Li G, Shi J, Liu J, Lv H, Li J, Wang J, Deng Y, Ran L, Shi X, Wang X, Wu Q, Li C, Ren X, Wang J, Wang X, Li D, Liu D, Zhang X, Ji Z, Zhao W, Sun Y, Zhang Z, Bao J, Han Y, Dong L, Ji J, Chen P, Wu S, Liu J, Xiao Y, Bu D, Tan J, Yang L, Ye C, Zhang J, Xu J,

Zhou Y, Yu Y, Zhang B, Zhuang S, Wei H, Liu B, Lei M, Yu H, Li Y, Xu H, Wei S, He X, Fang L, Zhang Z, Zhang Y, Huang X, Su Z, Tong W, Li J, Tong Z, Li S, Ye J, Wang L, Fang L, Lei T, Chen C, Chen H, Xu Z, Li H, Huang H, Zhang F, Xu H, Li N, Zhao C, Li S, Dong L, Huang Y, Li L, Xi Y, Qi Q, Li W, Zhang B, Hu W, Zhang Y, Tian X, Jiao Y, Liang X, Jin J, Gao L, Zheng W, Hao B, Liu S, Wang W, Yuan L, Cao M, McDermott J, Samudrala R, Wang J, Wong GK, Yang H. 2005. The Genomes of *Oryza sativa*: A History of Duplications. *PLOS Biology* **3**: 266-281.

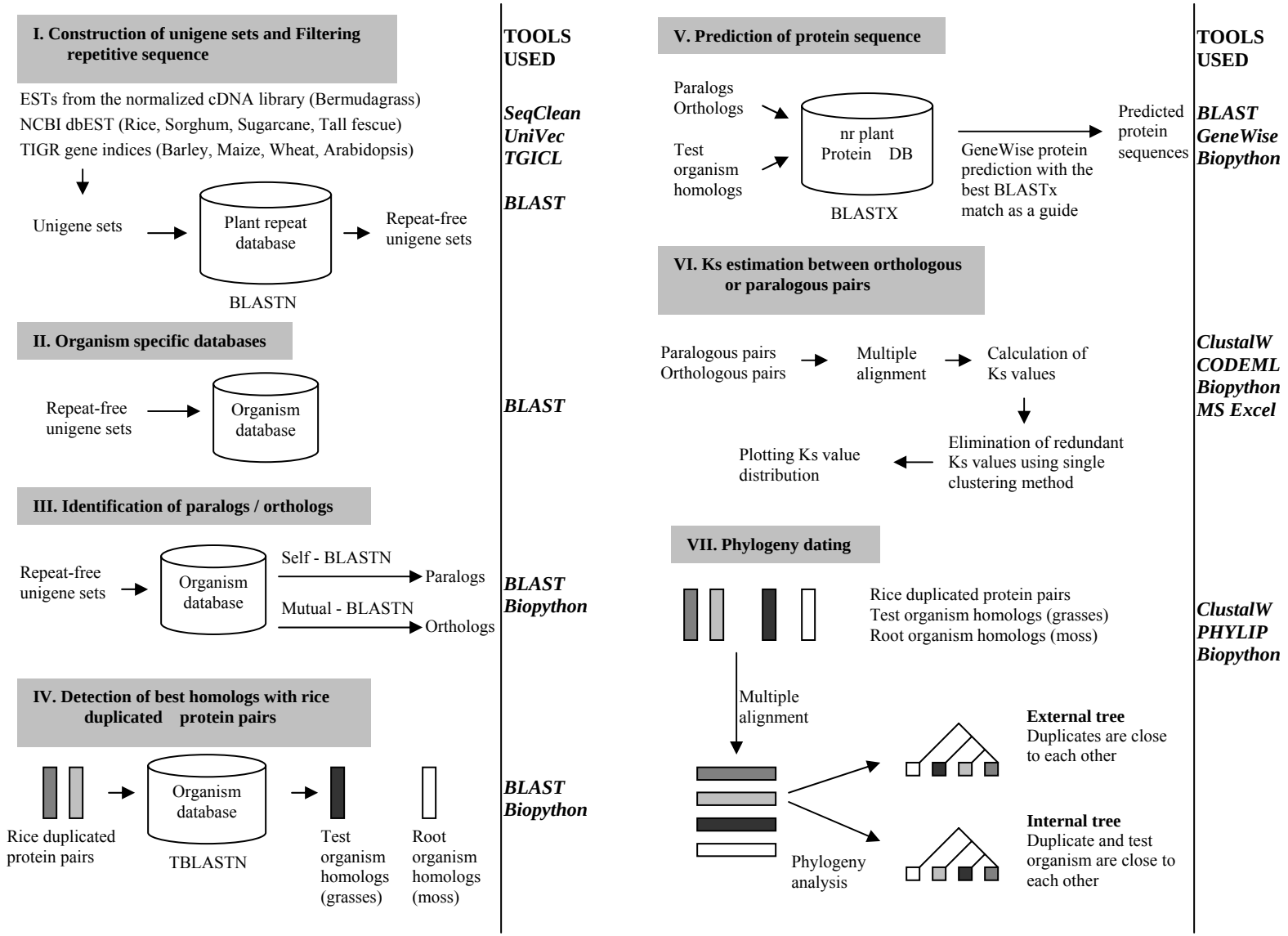


Figure 5.1. Step-by-step procedures and related bioinformatics tools used for this study.

Table 5.1. Number of sequences and paralogs used to analyze the distribution of Ks values in eight tested grasses.

Species	Initial ESTs	Unigenes	Putative Paralogs[a]	Paralogous proteins[b]	Gene Families[c]	Gene Family Size[d]	Duplication Event with Median Ks[e]
O. sativa spp. *japonica* (Rice)	713,222	279,058	199,960	51,393	42,432	1.2	42.073
C. dactylon (Bermudagrass)	19,373	13,082	2,234	981	273	3.6	249
S. bicolor (Sorghum)	205,225	38,546	8,111	5,920	1,339	4.4	1,128
S. officinarum (Sugarcane)	246,318	74,900	38,497	22,084	7,099	3.1	6,770
F. arundinacea (Tall fescue)	44,377	12,205	721	453	195	2.3	181
H. vulgare (Barley)	N/A[f]	50,453	15,169	9,295	4,927	1.9	4,506
Z. mays (Maize)	N/A	115,744	67,917	43,635	28,047	1.6	26,814
T. aestivum (Wheat)	N/A	122,282	55,623	28,663	19,683	1.5	18,782

[a] Number of paralogous sequences based on a self-BLASTN search after eliminating repeated sequences.
[b] Number of translated proteins from putative paralogs using Genewise program with the corresponding best match protein from BLASTX search.
[c] Number of gene families identified from paralogous proteins using the single linkage clustering method.
[d] Number of genes per family.
[e] Number of duplication events for which median Ks values are less than 2.5.
[f] Not available since the unigene sets of those four species were downloaded from TIGR gene indices (http://compbio.dfci.harvard.edu/tgi/plant.html).

Table 5.2. All possible secondary Ks peaks formed by paralogous pairs for the analyzed grasses. Note that Ks values are converted to time ($T = Ks / 2\lambda$) using $\lambda = 6.5\times10^{-9}$ (Gaut *et al.*, 1996).

Species[a]	Representing Subfamily	Representing Tribe	Secondary Ks peak (s) formed by paralogous pairs[b]	Estimated divergent time in MYA[c]
PACC clade (Fig. 5.2)				
C. dactylon (Bermudagrass)	Chloridoideae	Cynodonteae	0.70 to 0.75	53.8 to 57.7
S. bicolor (Sorghum)	Panicoideae	Andropogoneae	0.30 / 0.85	23.1 / 65.4
S. officinarum (Sugarcane)	Panicoideae	Andropogoneae	No traces found	
Z. mays (Maize)	Panicoideae	Andropogoneae	No traces found	
BEP clade (Fig. 5.3)				
O. sativa (Rice)	Ehrhartoideae	Oryzeae	No traces found	
H. vulgare (Barley)	Pooideae	Triticeae	0.75	57.7
F. arundinacea (Tall fescue)	Pooideae	Poeae	0.25 / 0.60 to 0.65	19.2 / 46.2 to 50.0
T. aestivum (Wheat)	Pooideae	Triticeae	0.10	7.8

[a] Arundinoideae, Centothecoideae, and Bambusoideae subfamilies could not be analyzed due to the lack of EST sequence data in public databases.
[b] Only secondary peaks less than Ks = 1.00 were considered due to the uncertainty of Ks values greater than 1.00 (Li, 1997).
[c] Million Years Ago.

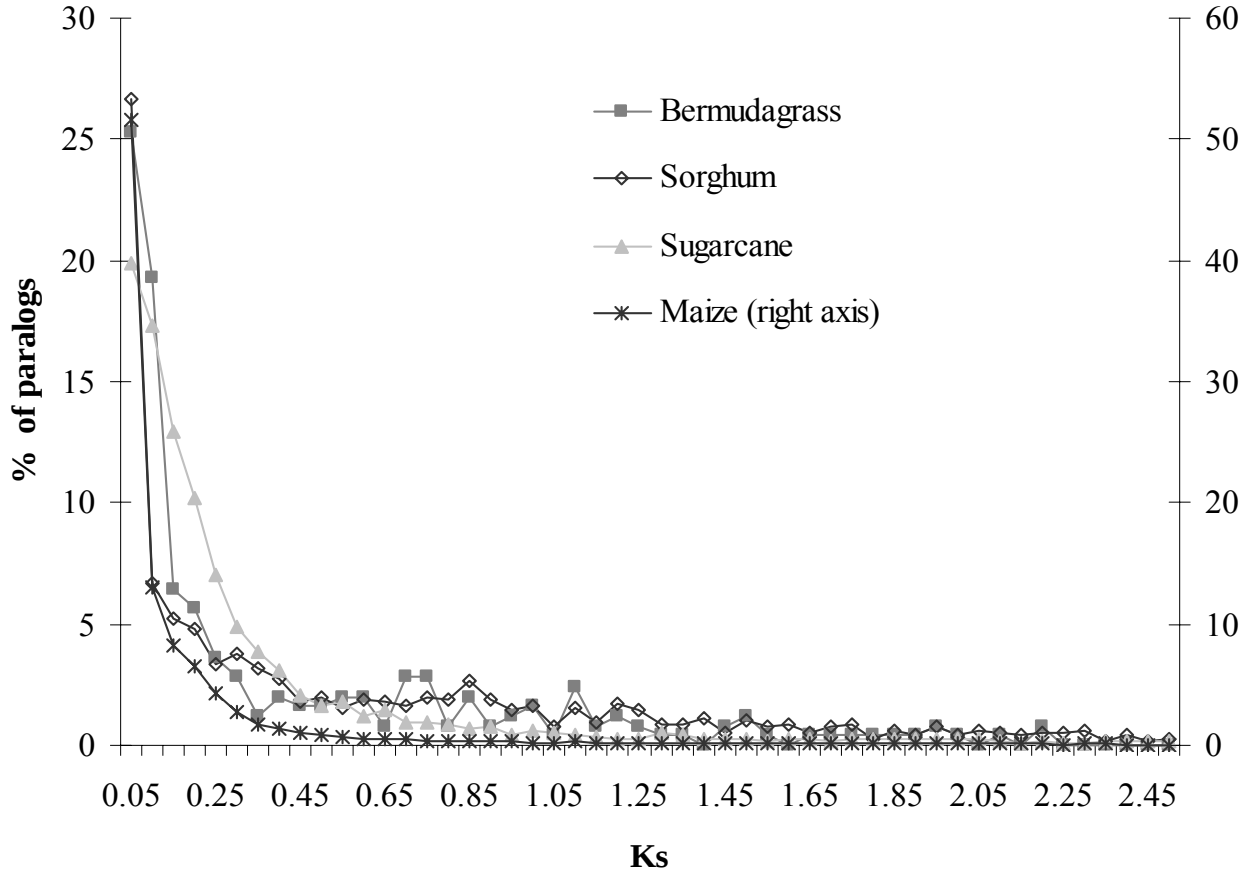


Figure 5.2. Age distribution of paralogous sequences from grass subfamilies included in the PACC clade.

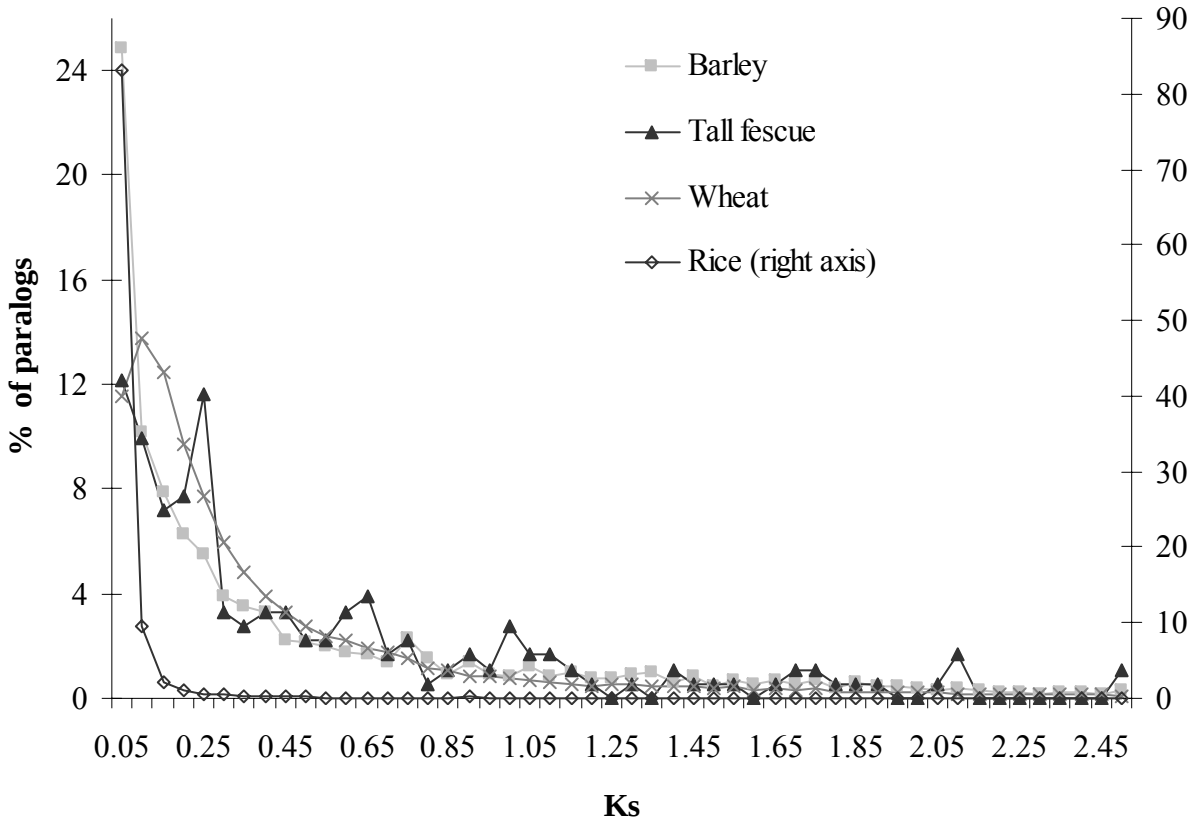


Figure 5.3. Age distribution of paralogous sequences from grass subfamilies included in the BEP clade.

Table 5.3. Ks values representing speciation events of grass subfamilies by the analysis of orthologous pairs between grasses. Note that Ks values are converted to time ($T = Ks / 2\lambda$) using $\lambda = 6.5\times10^{-9}$ (Gaut *et al.*, 1996).

Divergent Events	Number of orthologous pairs*	Ks peak formed by orthologous pairs	Estimated divergent time in MYA**
PACC clade (Fig. 5.4)			
Bermudagrass - Maize	3,622	0.50	38.5
Bermudagrass - Sorghum	3,197	0.45	34.6
Bermudagrass - Sugarcane	3,519	0.45 to 0.50	34.6 to 38.5
Maize - Sorghum	17,064	0.15	11.5
Maize - Sugarcane	29,221	0.15	11.5
Sorghum - Sugarcane	17,213	0.10	7.8
BEP clade (Fig. 5.5)			
Barley - Wheat	23,730	0.15	11.5
Rice - Barley	8,846	0.50 to 0.60	38.5 to 46.2
Rice - Tall fescue	1,087	0.50 to 0.55	38.5 to 42.3
Rice - Wheat	14,678	0.50 to 0.55	38.5 to 42.3
Tall fescue - Barley	2,788	0.25 to 0.30	19.2 to 23.1
Tall fescue - Wheat	3,182	0.25 to 0.30	19.2 to 23.1
PACC vs. BEP clade (Fig. 5.6)			
Bermudagrass - Barley	2,869	0.60	46.2
Bermudagrass - Rice	2,393	0.60 to 0.65	46.2 to 50.0
Sorghum - Barley	9,011	0.55 to 0.60	42.3 to 46.2
Sorghum - Rice	17,604	0.55 to 0.65	42.3 to 50.0

* Based on a BLASTN search over 300 base pairs aligned with each other (note that the actual number of Ks values was less than this number because of Genewise's protein sequence prediction procedure).

** Million Years Ago.

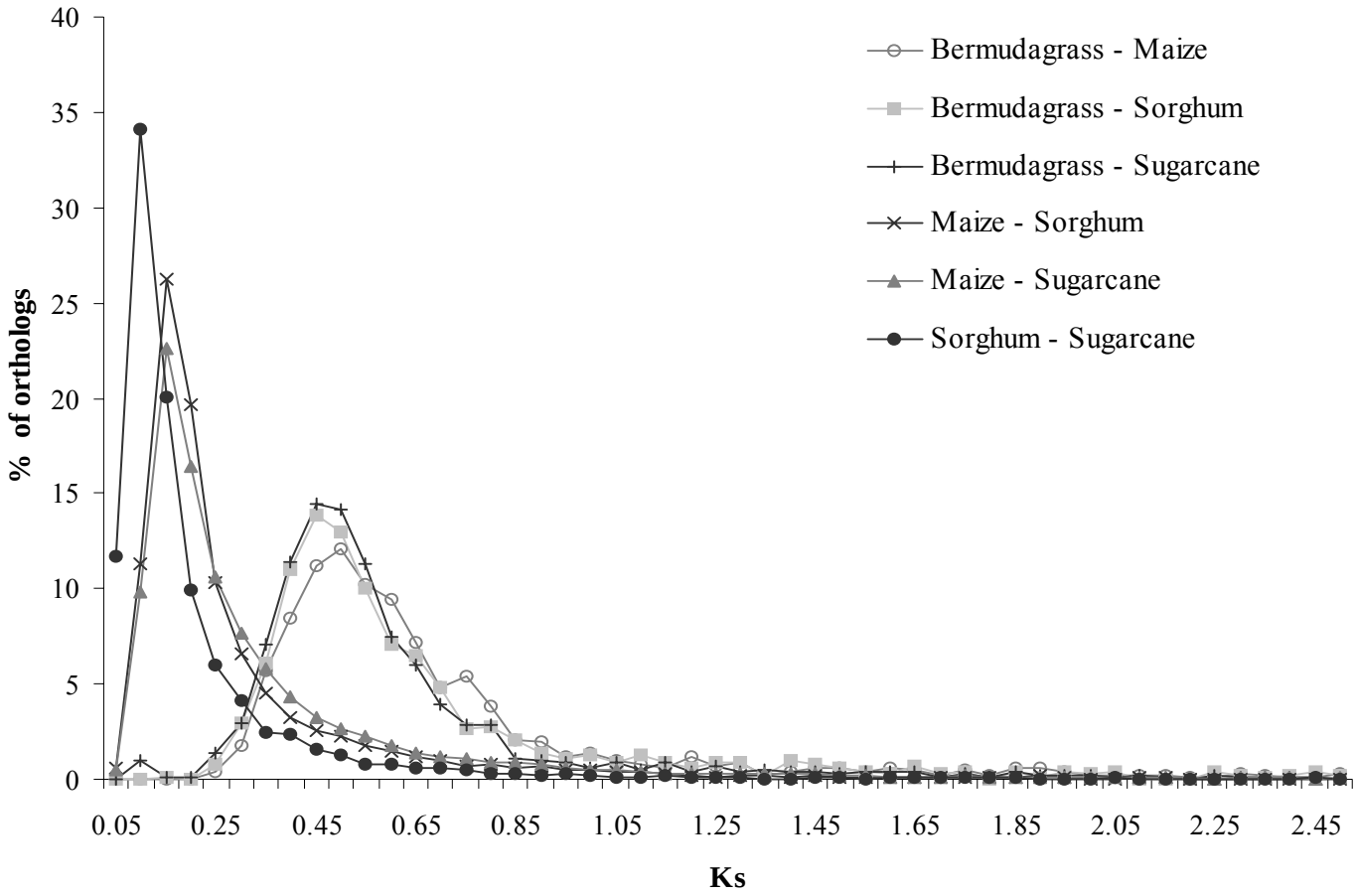


Figure 5.4. Age distribution of orthologous sequences from grass subfamilies included in the PACC clade.

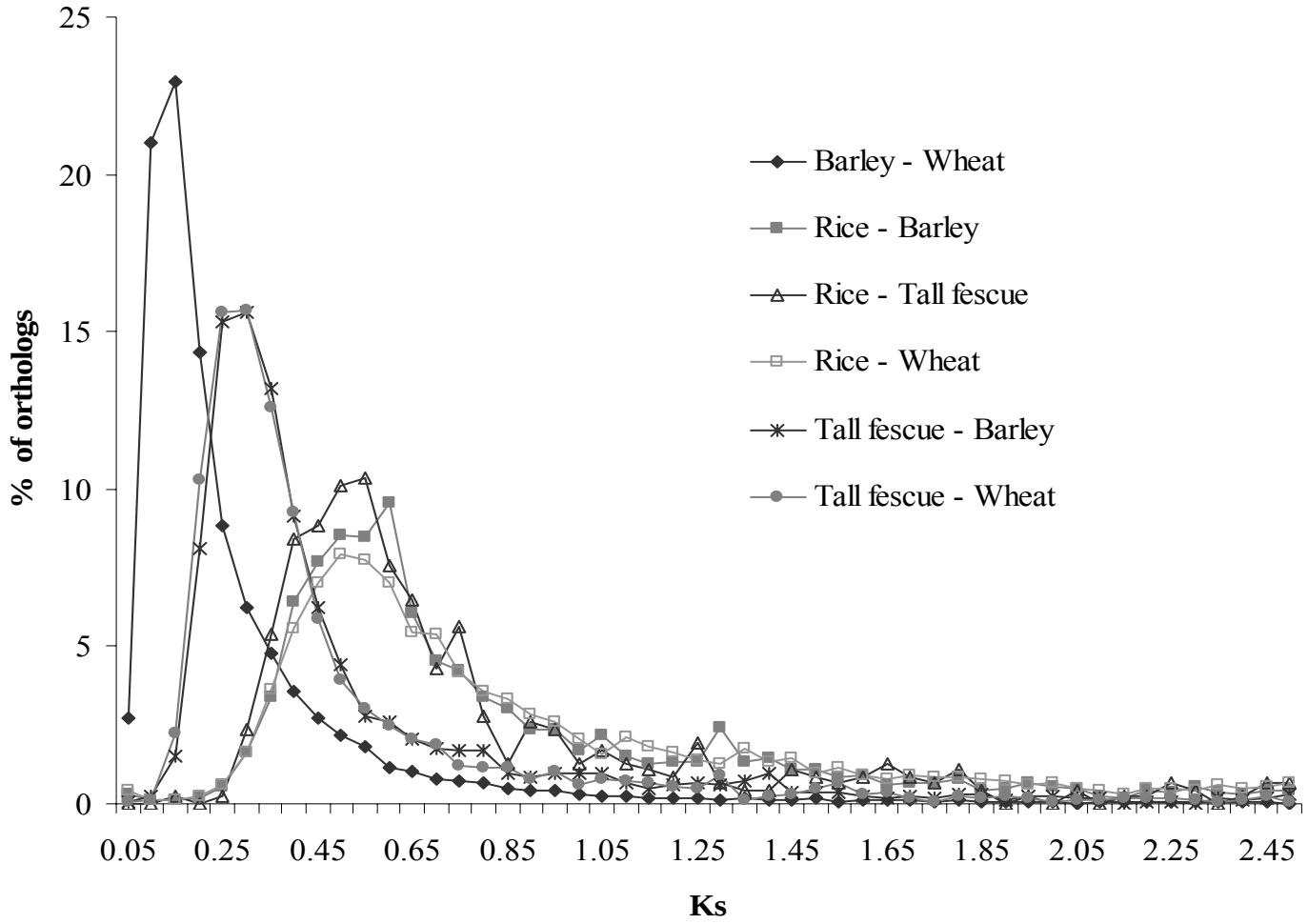


Figure 5.5. Age distribution of orthologous sequences from grass subfamilies included in the BEP clade.

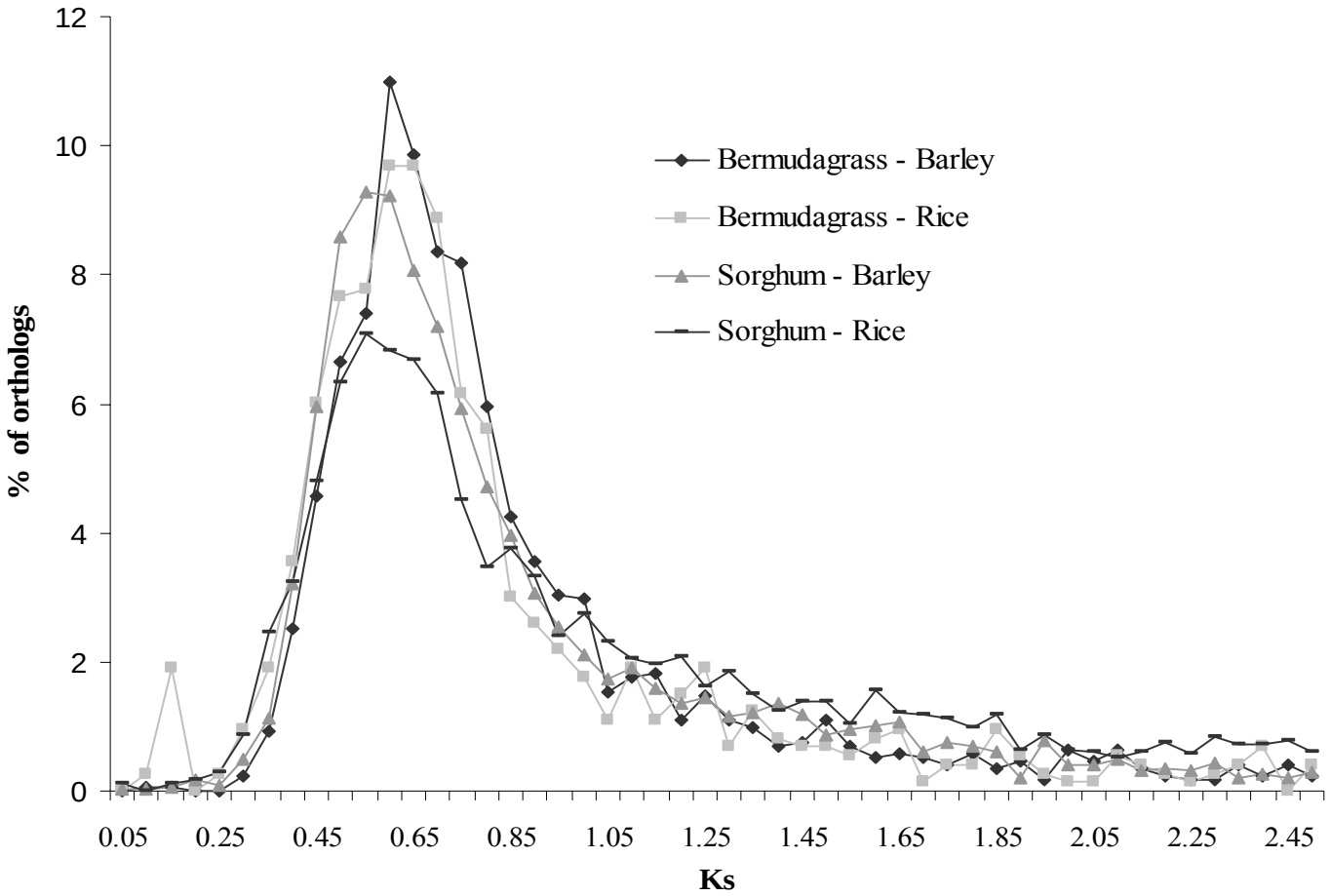


Figure 5.6. Age distribution of orthologous sequences between grass subfamilies included in the PACC and the BEP clade, respectively.

Table 5.4. Phylogenetic dating of a genomic duplication in the rice lineage. Values in parentheses indicate the number of informative trees. Significances with the same letters are not significantly different according to Tukey's studentized range (HSD) test ($P < 0.01$).

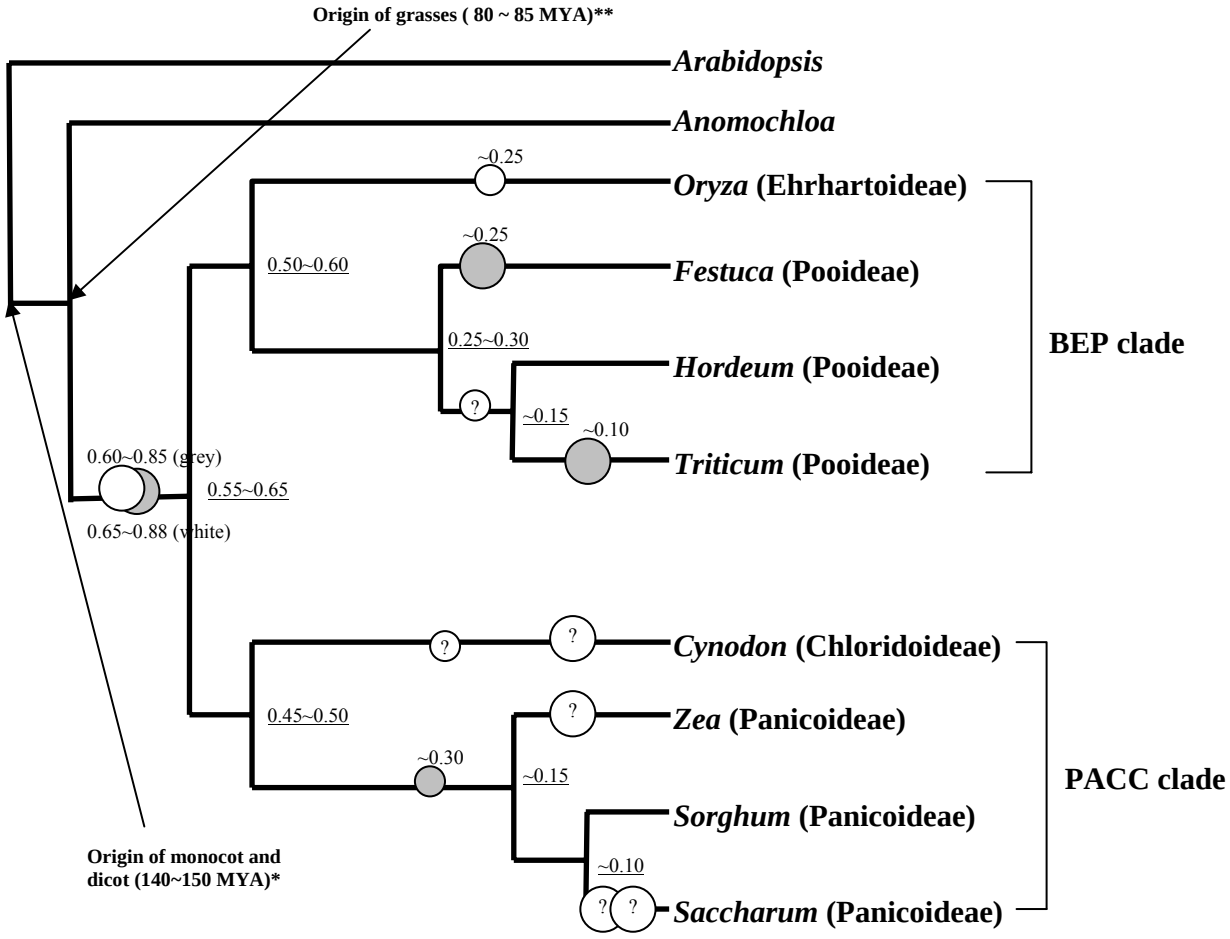
Species	*A. thaliana*	*C. dactylon*	*S. bicolor*	*S. officinarum*	*F. arundinacea*	*H. vulgare*	*Z. mays*	*T. aestivum*
No. of Unigenes	38,462	13,082	38,546	74,900	12,205	50,453	115,744	122,282
Duplicated blocks (chromosomes)								
1 (1-5)	0.099 (485)	0.352 (315)	0.481 (526)	0.497 (475)	0.309 (421)	0.619 (485)	0.680 (485)	0.680 (484)
2 (2-4)	0.105 (237)	0.297 (145)	0.452 (228)	0.470 (230)	0.370 (192)	0.623 (236)	0.671 (234)	0.671 (237)
3 (2-6)	0.158 (298)	0.321 (209)	0.475 (297)	0.478 (295)	0.394 (264)	0.581 (298)	0.664 (298)	0.638 (298)
4 (3-7)	0.115 (217)	0.340 (141)	0.514 (212)	0.558 (215)	0.363 (182)	0.636 (217)	0.765 (217)	0.673 (217)
5 (3-10)	0.108 (139)	0.333 (84)	0.467 (137)	0.493 (138)	0.360 (125)	0.619 (139)	0.734 (139)	0.640 (139)
6 (3-12)	0.221 (68)	0.429 (42)	0.453 (64)	0.500 (66)	0.466 (58)	0.632 (68)	0.750 (68)	0.750 (68)
7 (4-8)	0.086 (58)	0.341 (44)	0.509 (57)	0.611 (54)	0.316 (57)	0.586 (58)	0.707 (58)	0.741 (58)
8 (4-10)	0.268 (41)	0.455 (22)	0.692 (39)	0.625 (40)	0.528 (36)	0.810 (42)	0.905 (42)	0.756 (41)
9 (8-9)	0.087 (196)	0.333 (117)	0.357 (196)	0.482 (195)	0.295 (156)	0.582 (196)	0.672 (195)	0.571 (196)
10 (11-12)	0.030 (203)	0.159 (107)	0.177 (192)	0.184 (196)	0.113 (151)	0.176 (199)	0.184 (201)	0.199 (201)
Total	0.110 (1942)	0.323 (1226)	0.441 (1948)	0.470 (1904)	0.331 (1642)	0.570 (1938)	0.646 (1937)	0.614 (1939)
Significance	C	B	AB	AB	B	A	A	A

Table 5.5. Comparison of average Ks values for each rice duplication block and its corresponding time in MYA (million years ago). Note that Ks values are converted to time (T = Ks / 2λ) using $\lambda = 6.5\times10^{-9}$ (Gaut *et al.*, 1996). Ks values with the same letter are not significantly different according to Student-Newman-Keuls (SNK) test ($P < 0.01$).

Duplicated blocks (chromosomes)	Number of gene pairs	Average Ks	Estimated duplication time in MYA
1 (1-5)	579	0.705_{AB}	≈ 54.2
2 (2-4)	301	0.716_{AB}	≈ 55.0
3 (2-6)	355	0.746_{AB}	≈ 57.4
4 (3-7)	282	0.777_{AB}	≈ 59.8
5 (3-10)	170	0.757_{AB}	≈ 58.3
6 (3-12)	84	0.832_{AB}	≈ 64.0
7 (4-8)	69	0.785_{AB}	≈ 60.4
8 (4-10)	50	0.881_{A}	≈ 67.8
9 (8-9)	249	0.646_{B}	≈ 50.0
10 (11-12)	282	0.246_{C}	≈ 18.9
Total (block 1 ~ block 9)	2,139	0.646 ~ 0.881	50.0 ~ 67.8

Table 5.6. Pairwise comparison of the frequency of internal trees for each duplicated block between two different data types (unigenes and proteins predicted by whole genome sequences). The confidence interval method for a binomial distribution was applied to the evaluation of differences (ns, not significant; **, significantly different at $P < 0.05$; ***, significantly different at $P < 0.01$).

Species	*A. thaliana*	*A. thaliana*		*S. bicolor*	*S. bicolor*	
Database type	Unigene	Predicted protein		Unigene	Predicted protein	
No. of sequences	38,462	26,784		38,546	34,784	
Duplicated blocks (chromosomes)			Stat.			Stat.
1 (1-5)	0.099 (485)	0.093 (485)	ns	0.481 (526)	0.858 (485)	***
2 (2-4)	0.105 (237)	0.089 (237)	ns	0.452 (228)	0.840 (237)	***
3 (2-6)	0.158 (298)	0.117 (298)	ns	0.475 (297)	0.872 (298)	***
4 (3-7)	0.115 (217)	0.106 (217)	ns	0.514 (212)	0.871 (217)	***
5 (3-10)	0.108 (139)	0.115 (139)	ns	0.467 (137)	0.835 (139)	***
6 (3-12)	0.221 (68)	0.221 (68)	ns	0.453 (64)	0.897 (68)	***
7 (4-8)	0.086 (58)	0.086 (58)	ns	0.509 (57)	0.897 (58)	***
8 (4-10)	0.268 (41)	0.275 (40)	ns	0.692 (39)	0.881 (42)	**
9 (8-9)	0.087 (196)	0.077 (196)	ns	0.357 (196)	0.867 (196)	***
10 (11-12)	0.030 (203)	0.030 (203)	ns	0.177 (192)	0.227 (203)	ns
Total	0.110 (1942)	0.099 (1941)	ns	0.441 (1948)	0.796 (1943)	***



*** Divergence of monocot and dicot at about 140 ~ 150 MYA (million years ago; Chaw *et al.*, 2004)**
**** Origin of grasses at about 80 ~ 85 MYA (Prasad *et al.*, 2005)**

Figure 5.7. The phylogeny of grass subfamilies analyzed in this study. The phylogenetic tree is based on the phylogeny of the grass family that is currently accepted (Grass Phylogeny Working Group, 2001). The grey circles indicate suspected large-scale duplication events observed in this study based on the Ks distribution of paralogous pairs. The white circles indicate large-scale duplication events based on the rice genome duplication dataset (Plant Genomic Duplication Database, PGDD). The circles with question marks indicate the duplication events expected to be found but not to be recovered in this study. The underlined numbers at the branch points indicate the approximate timing of the speciation event found in this study based on the Ks distribution of orthologous pairs among species. All the numbers imply Ks values in which duplication or speciation events occurred.

Table 5.7. Comparisons of frequencies of internal trees for Bermudagrass and tall fescue among different E-value thresholds of TBLASTN search. Significances with the same letters within each species are not significantly different according to Tukey's studentized range (HSD) test ($P < 0.05$).

Species	***C. dactylon***				
E-value threshold for TBLASTN	$e < 10^{-5}$	$e < 10^{-20}$	$e < 10^{-40}$	$e < 10^{-60}$	$e < 10^{-80}$
Duplicated blocks (chromosomes)					
1 (1-5)	0.352 (315)	0.370 (273)	0.438 (203)	0.439 (155)	0.500 (94)
2 (2-4)	0.297 (145)	0.323 (127)	0.361 (97)	0.400 (60)	0.412 (34)
3 (2-6)	0.321 (209)	0.339 (183)	0.364 (132)	0.395 (86)	0.321 (53)
4 (3-7)	0.340 (141)	0.352 (122)	0.351 (97)	0.370 (73)	0.455 (44)
5 (3-10)	0.333 (84)	0.361 (72)	0.408 (49)	0.400 (35)	0.400 (20)
6 (3-12)	0.429 (42)	0.472 (36)	0.556 (27)	0.611 (18)	0.600 (10)
7 (4-8)	0.341 (44)	0.349 (43)	0.351 (37)	0.393 (28)	0.417 (12)
8 (4-10)	0.455 (22)	0.444 (18)	0.438 (16)	0.500 (10)	0.667 (6)
9 (8-9)	0.333 (117)	0.323 (99)	0.348 (69)	0.368 (38)	0.407 (27)
10 (11-12)	0.159 (107)	0.173 (81)	0.190 (58)	0.216 (37)	0.286 (21)
Total	0.323 (1226)	0.341 (1054)	0.377 (785)	0.400 (540)	0.430 (321)
Significance within species	B	AB	AB	AB	A

Species	***F. arundinacea***				
E-value threshold for TBLASTN	$e < 10^{-5}$	$e < 10^{-20}$	$e < 10^{-40}$	$e < 10^{-60}$	$e < 10^{-80}$
Duplicated blocks (chromosomes)					
1 (1-5)	0.309 (421)	0.332 (340)	0.392 (227)	0.403 (129)	0.441 (59)
2 (2-4)	0.370 (192)	0.404 (161)	0.413 (104)	0.520 (50)	0.429 (21)
3 (2-6)	0.394 (264)	0.413 (218)	0.455 (156)	0.607 (84)	0.500 (38)
4 (3-7)	0.363 (182)	0.378 (156)	0.381 (118)	0.350 (60)	0.360 (25)
5 (3-10)	0.360 (125)	0.379 (103)	0.444 (72)	0.324 (37)	0.563 (16)
6 (3-12)	0.466 (58)	0.490 (49)	0.559 (34)	0.611 (18)	0.667 (6)
7 (4-8)	0.316 (57)	0.314 (51)	0.357 (42)	0.412 (17)	0.333 (6)
8 (4-10)	0.528 (36)	0.594 (32)	0.619 (21)	0.556 (9)	0.667 (3)
9 (8-9)	0.295 (156)	0.289 (128)	0.268 (71)	0.265 (34)	0.267 (15)
10 (11-12)	0.113 (151)	0.125 (120)	0.128 (86)	0.171 (41)	0.091 (11)
Total	0.331 (1642)	0.351 (1358)	0.383 (931)	0.420 (479)	0.425 (200)
Significance within species	A	A	A	A	A

Table 5.8. Pairwise comparisons of internal tree ratio for all the blocks between each species tested. Each number indicates P-value based on the 99 % confidence limit of a binomial distribution (Note that numbers in bold indicate that the frequencies of internal trees for two species are **not** significantly different at the 99 % confident level).

	Frequencies of Internal Trees (number of trees examined)							
Total (block 1-10)	0.110 (1942)	0.323 (1226)	0.441 (1948)	0.470 (1904)	0.331 (1642)	0.570 (1938)	0.646 (1937)	0.614 (1939)
Species	*A. thaliana*	*C. dactylon*	*S. bicolor*	*S. officinarum*	*F. arundinacea*	*H. vulgare*	*Z. mays*	*T. aestivum*
A. thaliana		0.000	0.000	0.000	0.000	0.000	0.000	0.000
C. dactylon			0.000	0.000	**0.651**	0.000	0.000	0.000
S. bicolor				**0.071**	0.000	0.000	0.000	0.000
S. officinarum					0.000	0.000	0.000	0.000
F. arundinacea						0.000	0.000	0.000
H. vulgare							0.000	0.005
Z. mays								**0.039**
T. aestivum								

CHAPTER 6

SUMMARY AND CONCLUSIONS

Common Bermudagrass (*Cynodon dactylon*) was chosen as the subject of this study because molecular information for this species is very limited, despite its economic importance and ecological dominance. Collecting expressed sequence tags (ESTs) from Bermudagrass is a key step to initiate a number of molecular studies such as gene expression profiling and comparative genomics. Thus, we first constructed a cDNA library from Bermudagrass and normalized the library, to increase the chance of collecting as many new genes as possible. We generated a total of 15,588 high-quality ESTs, with which we produced the first unigene set. A total of 9,414 unigenes were generated by TIGR Gene Index Clustering tools (TGICL) and were annotated using comparative genomic tools such as BLAST, InterProScan, and Gene Ontology (GO). A total of 2,882 unigenes could not be annotated by the comparative genomic tools. Annotation data showed that a variety of proteins involved in a number of metabolic pathways were discovered from the cDNA library, indicating that the library can be linked to various Bermudagrass research fields such as breeding, physiology, and evolution. In addition, we searched the ESTs for SSRs (single sequence repeats) and CISPs (Conserved Intron Scanning Primers) as candidate DNA markers. A total of 143 EST-SSRs and 6,014 candidate CISPs were identified from the unigene set, which are open to the public for those who want to test the candidate markers for their purpose. Although the candidate DNA markers found in this study still need to be further verified, they can be applied to other plant species because EST-SSRs and

CISPs are usually abundant and highly conserved across species. Also, CISPs increase the chance to find polymorphic sites because CISPs expand scannable genomic regions and introns are less affected by evolutionary constraints.

Using the Bermudagrass ESTs, we conducted gene expression profiling in order to identify genes up- and down-regulated under drought conditions. Initially, Bermudagrass was grown with nutrient solution and drought stress was imposed using polyethylene glycol to minimize the effect of environmental factors. Drought candidate genes were identified using a macroarray technique. A total of 120 and 69 genes were identified as up- and down-regulated genes under drought conditions, respectively. To further classify drought candidate genes, a gene ontology (GO) search was performed, which provided obvious differences between up- and down-regulated genes. We also conducted analysis of putative cis-regulatory elements with rice proteins homologous to drought candidate genes from Bermudagrass. The analysis was wholly dependent upon rice 1kb-upstream database because no genomic sequences of Bermudagrass are available. The results of both the GO and putative cis-regulatory element searches were quite consistent with drought-responsive genes from other plant species, reflecting that most drought-responsive pathways are shared across different plant species. Therefore, the results from our study may allow Bermudagrass researchers to leverage genes that have been identified and lessons that have been learned in other plant species for enhancement of drought tolerance. Also, the experimental procedures used in this study can be applied to various abiotic or biotic stress conditions other than drought stress. By accumulating data on gene expression by tissue type, developmental stage, hormone and herbicide treatment, genetic background, and environmental conditions, it should be possible to identify genes involved in many important processes of development and responses to environmental conditions in Cynodon.

Finally, ESTs generated from the normalized cDNA library were subjected to evolutionary study, providing the first substantial EST sample from the Chloridoideae subfamily. ESTs from grasses in the Ehrhartoideae (*Oryza sativa*), Pooideae (*Festuca arundinacea*, *Hordeum vulgare*, and *Triticum aestivum*), and Panicoideae (*Zea mays*, *Saccharum officinarum*, and *Sorghum bicolor*) subfamilies were also included in this study. We used two different approaches to re-investigate the evolutionary history of the grass family: 1) synonymous substitution rates of paralogous or orthologous pairs and 2) phylogenetic trees based on rice duplicated proteins. Most duplication or speciation events observed in our study were quite consistent with existing publications, but a few were new findings. The Chloridoideae and Panicoideae subfamilies are estimated to have diverged from a common ancestor ca. 34.6 ~ 38.5 million years ago (MYA) based on Ks distributions of orthologous pairs involved in the two subfamilies. The evolution of the Chloridoideae subfamily, previously with only limited sequence data, was clarified by virtue of Bermudagrass ESTs. We also found that the Poeae and Triticeae tribes (involved in Pooideae subfamily) diverged from a common ancestor ca. 19.2 ~ 23.1 MYA. A segmental duplication event occurring at ca. 23.1 MYA was observed in the lineage of sorghum, which was roughly consistent with the segmental duplication event of rice according to their Ks values. The chromosomal locations of the segmental duplication on the comparative map also support that the two distinct events are similar to each other (sorghum chromosome 5 - 8 and rice chromosome 11 - 12). Other evidence (Wang et al., unpublished data) suggests that those events may have resulted from a common ancestral duplication followed by concerted evolution within specific cereal lineages, but it still needs to be further elucidated with additional evidence. We expected to observe some recent duplication events in the lineage of Bermudagrass, sugarcane, and maize because Bermudagrass and sugarcane are known

autotetraploids and maize a relatively recent paleopolyploid. However, no traces were found in those grasses, which points out a limitation of the Ks distribution approach. Secondary peaks corresponding to large-scale duplication events must be sufficiently dissociated from an initial peak to be observable in the distributions; otherwise, the peaks, especially recent ones, tend to be masked by the ongoing single gene duplication. Both methods have several obstacles to overcome; however, they provide a reasonable first evaluation for species which lack whole genome sequences. Since a large body of data generated by EST projects is now publicly released for various plant species, the methods are useful for estimating the approximate timing of large-scale duplications or speciations.