Analysis of Leaf and Root Transcriptome of Soil Grown Avena barbata Plants

February, 2011

Correspondence should be addressed to: Stephanie Swarbreck – stephanie.swarbreck@gmail.com

Swarbreck, Stéphanie; Lawrence Berkeley National Laboratory, Earth Sciences Division
Lindquist, Erika; Joint Genome Institute, Genomic Technologies Program
Ackerly, David; University of California Berkeley, Integrative Biology
Andersen, Gary; Lawrence Berkeley National Laboratory, Earth Sciences Division

[1] Earth Sciences Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720

[2] Department of Energy Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA 94598, USA

[3] Department of Integrative Biology, 3060 Valley Life Sciences Building, University of California, Berkeley CA 94720

ACKNOWLEDGEMENT:

The work conducted by the US Department of Energy (DOE) Joint Genome Institute is supported by the Office of Science Biological and Environmental and Research Program, Climate Change Research Division and by the University of California Lawrence Berkeley National laboratory, under Contract Number DE-AC02-05CH11231. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government, or any agency thereof, or the Regents of the University of California.

We would like to thank Calvin Myint for help with samples collection, Christopher Wilks (The Arabidopsis Information Resource) and Robert Bukowski from Cornell's Computational Biology Service Unit (CBSU) for InterProScan analysis. We thank Samuel B. St. Clair, Erika A. Sudderth, Rohit Salve, Kallista Bley, Paul Cook, Melissa Crago, Tara Macomber, Alex Morales, and Margaret S. Torn for their work on the design and execution of the mesocosm experiment. We would also like to thank Todd DeSantis, Ulas Karaoz, and Ping Hu for their support with unix and perl based analyses. Thanks also to David Swarbreck for helpful discussions and suggestions.

Analysis of Leaf and Root Transcriptome of Soil Grown Avena barbata Plants

Journal:	Plant and Cell Physiology
Manuscript ID:	Draft
Manuscript Type:	Regular Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Swarbreck, Stephanie; Lawrence Berkeley National Laboratory, Earth Sciences Division Lindquist, Erika; Joint Genome Institute, Genomic Technologies Program Ackerly, David; University of California Berkeley, Integrative Biology Andersen, Gary; Lawrence Berkeley National Laboratory, Earth Sciences Division
Keywords:	Avena barbata, ESTs, climate change, root

Running Title: Avena root transcriptome

Corresponding author: Stéphanie M. Swarbreck
Adress: 1, Cyclotron Road, Berkeley, CA 94720, USA.
Telephone: 510-486-6125
Fax: 510-486-7152
E-mail: stephanie.swarbreck@gmail.com

Subject areas: (2) Environmental and stress response and (10) genomics, systems biology and evolution.

Two black and white figures, three color figures and five tables.

Title: Analysis of Leaf and Root Transcriptome of Soil Grown Avena barbata Plants

Stéphanie M. Swarbreck[1,3§], Erika A. Lindquist[2], David D. Ackerly[1,3] and Gary L. Andersen[1]

[1] Earth Sciences Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720.

[2] Department of Energy Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA 94598, USA.

[3] Department of Integrative Biology, 3060 Valley Life Sciences Building, University of California, Berkeley CA 94720.

[§]Corresponding author

Email addresses:

SMS: stephanie.swarbreck@gmail.com
EAL: ealindquist@lbl.gov
DDA: dackerly@berkeley.edu
GLA: glandersen@lbl.gov

Abstract (-limit 250 words)

Slender wild oat (Avena barbata) is an annual grass dominant in many grassland ecosystems in Mediterranean climate. This species has been the subject of ecological studies that aim at understanding the effect of global climate change on grassland ecosystems and the genetic basis for adaptation under varying environmental conditions. We present the sequencing and analysis of cDNA libraries constructed from leaf and root samples collected from A. barbata grown on natural soil and under varying rainfall patterns. More than one million expressed sequence tags (ESTs) were generated using both GS 454-FLX pyrosequencing and Sanger sequencing, and these tags were assembled into consensus sequences. We identified numerous candidate polymorphic markers in the dataset, providing possibilities for linking the genomic and the existing genetic information for A. barbata. Using the digital northern method, we showed that genes involved in photosynthesis were down regulated under high rainfall while stress-related genes were up regulated. We also identified a number of genes unique to the root library with unknown function. Real-time RT-PCR was used to confirm the root specificity of some of these transcripts such as two genes encoding O-methyl transferase. Also we showed differential expression under three water levels. Through a combination of Sanger and 454-based sequencing technologies, we were able to generate a large set of transcribed sequences for A. barbata. This dataset provides a platform for further studies of this important wild grass species.

Keywords: Avena barbata, ESTs, climate change, root

Introduction

Slender wild oat (Avena barbata) has been the subject of many ecological studies, as an example of a plant species studied for adaptation and acclimation to changing environment. It is highly abundant in grasslands throughout the world and is well adapted to a range of ecological conditions, encompassing arid sites with shallow soils to wet sites with deep soil. It is a wild relative of the cultivated oat (Avena sativa), and as such, it constitutes a source of alleles that could be used for improvement of this crop and other related crop species.

A. barbata originated from the Mediterranean basin, probably Spain, and has since been a successful invader in areas with a Mediterranean climate. In California, it is an exotic grass that was introduced about 200 years ago (Allard 1996). Hawkes et al. showed that A. barbata increases the gross nitrification rate in the rhizosphere (Hawkes et al. 2005). The increased nitrate availability in the rhizosphere is beneficial to the plants, and it is also more accessible to A. barbata compared to competing plants.

Slender wild oat has been the subject of studies investigating the genetic basis for adaptation in contrasting environments. Allard et al. characterized two ecotypes of A. barbata, specific to mesic (wet) and xeric (dry) environments, in California (Allard et al. 1972). Using allozyme electrophoresis, they demonstrated that the individuals of xeric and mesic population were monomorphic for a balanced opposite set of five alleles at specific loci. Populations in intermediate habitats were usually polymorphic at all five loci. Latta et al. suggested that adaptive evolutionary change is actively occurring in the present in A. barbata (Latta 2009; Latta and Gardner 2009).

Climate change is likely to alter future precipitation in California (Cayan et al. 2008; Hayhoe et al. 2004). However, these changes are difficult to predict with some studies predicting small decline in winter precipitation (Hayhoe et al. 2004) while other projections show small increases depending on the greenhouse gas emission scenarios and global climate model used (Cayan et al. 2008). Drought events occurring during the growing season can lead to sharp decline in carbon assimilation in A. barbata (St Clair et al. 2009b). Investigating the effect of climate change on plant gene expression may bring new mechanistic information regarding acclimation and potential adaptation to new climatic conditions (Leakey 2009; Leakey et al. 2009; Thayer et al. 2008). In particular, genes expressed in the leaves and/or roots that are important for acclimation of grasses to changing water availability may be uncovered. However the lack of genomic data for A. barbata has been a hurdle for such studies.

Avena barbata belong to the Pooideae subfamily of the grass family (Poaceae). Species of this subfamily usually have large and complex genomes, and so far the genome of Brachypodium distachyon is the only fully sequenced example (Vogel et al. 2010). A. barbata is a fully diploidized tetraploid with a large and complex genome. The genome size of tetraploid species from the genus Avena such as A. barbata has been estimated at 9,500 Mbp, which is more than 20 times that of the rice genome (Bennett and Smith 1976; Matsumoto et al. 2005). While whole genome sequencing would require substantial resources, sequencing of the transcriptome is a more attainable goal. The sequencing of ESTs libraries from different plant tissues has been used as an effective method for gene discovery, identification of new molecular markers, and generation of sequence data for analysis of transcripts profiles. The sequencing has traditionally been carried out using the Sanger method, but the development of a new method such as 454-

pyrosequencing (Margulies et al. 2005) offers an alternative method that is cost effective and allows deeper sampling of the transcriptome.

In this study, we discuss the analysis of sequenced cDNA libraries from leaves and roots of A. barbata grown under conditions of varying water availability. The objectives of this EST sequencing project were to generate a large amount of cDNA sequence data that would constitute a strong basis for transcriptomic studies in A. barbata, and to identify the genes differentially expressed in A. barbata leaves under changing rainfall patterns. For this purpose, we collected samples from roots, and leaves from soil-grown plants. For two libraries, one from root and one from leaf, samples from different rainfall and nitrogen-addition treatments were included to capture a wide range of transcripts. Leaf transcriptome varies during development from the vegetative stage to the reproductive stage (Buchanan-Wollaston et al. 2003), so for this study, leaf samples were collected at both developmental stages.

Detection of differentially expressed genes can be achieved by sequencing of non-normalized cDNA libraries (Audic and Claverie 1997). In this study, we use the digital northern method to compare the abundance of transcripts in the leaves of A. barbata plants grown under ambient and high rainfall. The samples were collected from a large greenhouse experiment set up to investigate the effect of variation in rainfall pattern on a grassland ecosystem, from the gene level to ecosystem function (St Clair et al. 2009). The rainfall treatments applied were relatively mild and related to projected climate changes in California.

Results

Features of the ESTs generated using Sanger sequencing and 454

In an effort to generate sequence data for A. barbata, cDNA libraries were generated from leaf and root tissues of plants grown under varying soil moisture content. For each of the four RNA samples, two cDNA libraries were constructed, one with small insert (<2 kb) and one with large inserts (>2 kb). Using the Sanger method, 3,072 clones were sequenced for each leaf library, and 6,912 clones were sequenced for each root library (Table 1). Paired-end transcript sequencing was performed to systematically identify genes of interest. The sequencing success was on average 85%, and the mean length for the reads was 666 bp. In total, 64,512 ESTs were sequenced by the Sanger method, and 54,242 were good-quality ESTs that could subsequently be used for clustering. The percentage of redundancy (number of ESTs in cluster consensus divided by the total number of ESTs) for each library varied between 56.2% and 85.6%.

Two additional cDNA libraries were constructed from the same leaf RNA samples collected at ambient and high rainfall (Table 1), and these libraries were sequenced using 454-FLX pyrosequencing. A total of 967,302 ESTs were sequenced by 454, giving 931,602 good-quality ESTs that could be used for clustering. The average read length was 186 bp. Libraries sequenced by 454 showed slightly higher redundancy (75.2 and 84.4%) than libraries sequenced by the Sanger method.

Assembly of ESTs

ESTs were assembled into consensus sequences for each sample type and all together to give a set of unique transcripts (Table 1). A total of 985,844 good-quality ESTs were

assembled into 191,340 consensus sequences. Most consensus sequences (70.8%) were supported by a single EST (called “singlets”) (Table 2). Two consensus sequences were represented by more than 10,000 ESTs and showed similarity to a fructose 1, 6 bisphosphate aldolase (10,255 ESTs) and rubisco activase (10,180 ESTs).

Out of the 191,340 consensus sequences assembled, 183,800 (96.1% of all consensus sequences) included at least one EST from 454, while the remaining consensuses (3.9%) were solely derived from Sanger sequencing. Some of these sequences originate from libraries that were not sequenced by 454 and therefore may represent transcripts absent from the leaf libraries. Interestingly, similar analysis carried out on the leaf libraries sequenced by both methods indicated that some consensus sequences were built with ESTs solely sequenced by the Sanger method. A total of 1,293 final consensus sequences were exclusively based on Sanger reads and were not built using 454 reads, even though they originated from libraries sequenced by both methods. Of these consensuses, 1,076 were singlet and 217 were supported by at least two Sanger reads. Although the percentage of consensus without ESTs from 454 was small (less than 1% of all sequences), this indicates that sequencing methods may have bias and that using a combination of sequencing method may be necessary to get all transcripts sequenced.

Microsatellites analysis

All cluster consensus sequences were searched for simple sequence repeats (SSRs) using the misa PERL script (Thiel et al. 2003). Among the 191,340 sequences examined, a total of 8,319 cDNA contained at least one SSR (Table 3). Analysis conducted on samples-specific libraries showed that the percentage of sequences containing at least one SSR was on average lower for hybrid libraries (3%) compared to Sanger-only libraries (16.2%) (Chi-square test, p=0.01). This is due to the longer reads obtained from Sanger sequencing, which are more likely to contain SSRs.

Among the SSRs identified from all consensus sequences, 4,017 (48.3%) were trinucleotides, followed by 3,768 mononucleotides (45.3%) and 1,522 dinucleotides (18.3%). A/T (3,334) was the most frequent repeat motif, followed by CCG/CGG (1,214). The profile of SSRs was different depending on the sequencing method (Chi-square test, p=0.01). The number of mononucleotide SSRs identified tended to be higher than the numbers of trinucleotide SSRs in the libraries sequenced by the Sanger method, and lower in the libraries sequenced by both Sanger and 454. A similar profile of microsatellites markers was found in barley ESTs with trinucleotides being the most abundant (56%) type of repeat, followed by mononucleotides (19%) and dinucleotides (18%) (Thiel et al. 2003).

Since it has been reported that 454 shows reduced accuracy in homopolymer regions (Margulies et al. 2005), SSR analysis was conducted on the subset of consensus that was based exclusively on Sanger reads (Table 3). Overall, the percentage of sequences containing an SSR was slightly higher (20.8%) compared to consensus originating from Sanger reads (between 15.7% and 16.8%) in the two libraries not sequenced by 454. More strikingly, the proportion of SSR that were mononucleotides (74.2%) was almost double the number seen for the consensus sequences in the combined 454 and Sanger method (39%). It is possible that some of these consensus reads failed because of the occurrence of one or more homopolymer regions.

Functional annotation

Cluster consensus sequences from each library, including the hybrid Sanger-454 library, were submitted to a BLASTx search against the Swiss-Prot database using a cutoff value of 1e-5 (Altschul et al. 1997) (Table 4). For the Sanger-sequenced libraries, on average, 71.6% of the cluster consensus sequences showed a significant hit, 70.5% of sequences from the root library and 72.7 % of the leaf library. On average, 21.8% of the consensus sequences assembled using ESTs from both 454 and the Sanger method had a significant hit (e-value < 1e-5), as did 24.1% of the sequences from leaf-ambient rainfall and 19.4% of those from leaf-high rainfall.

The most abundant consensus sequences in the leaf were represented by 1.16–2.53% of all ESTs in the corresponding library (Table 5). In the roots, the most abundant sequences represented a smaller percentage (0.19%) of all ESTs in the libraries. The 454-libraries tended to be less dominated by the most abundant sequences as suggested by the low percentage representing the most abundant consensus sequences. In libraries sequenced more deeply, the chance of sequencing new transcripts and gaining greater diversity is higher.

As expected, the most abundant sequences in the leaf libraries correspond to genes encoding photosynthesis-related proteins such as rubisco activase, fructose bis-phosphate aldolase, chlorophyll a-b binding protein, and photosystem II. Sequences with similarity to a victorin-binding protein, a component of the mitochondrial enzyme glycine decarboxylase involved in photorespiration (Wolpert et al. 1994), were also one of the five most abundant sequences for the leaf two libraries (Table 5).

In the roots, two of the most abundant sequences showed similarity to metallothioneins, which are small ubiquitous cysteine-rich proteins that can bind metals and are involved in abiotic stress response. In addition, two other sequences, elongation factor (EF) 1-alpha and EF2, show similarity to proteins that assist in the elongation of the protein during translation (Andersen et al. 2003). Metallothionein and EF1-alpha were also highly expressed in the maize roots, as shown by transcriptome analysis using serial analysis of gene expression (SAGE) (Poroyko et al. 2005). Interestingly, glyceraldehyde 3-phosphate dehydrogenase, which represents 0.12% of the transcripts in A. barbata roots, was not part of the most abundant transcripts of the root transcriptome of poplar, Arabidopsis, or maize (Fizames et al. 2004; Kohler et al. 2003; Poroyko et al. 2005).

Gene ontology (GO) annotation

Sequences from the four cDNA libraries that were 120 bp or longer were translated using Prot4ESTs (Wasmuth and Blaxter 2004). The translated sequences were submitted to InterProScan to identify signatures representing specific protein families or domains, and to possibly infer biological functions for the protein queries. Gene ontology (GO) terms were assigned to specific InterPro signatures, and these were used to classify the gene products in functional GO categories and in plant GOSlim categories.

In the leaf-ambient rainfall library, 23,911 (25.8%) sequences were assigned an InterPro signature, and from these, 17,866 (19.3%) were assigned at least one GO term. Similar results were obtained for the leaf high-rainfall library; 21,616 (27.1%) sequences were assigned an InterPro signature, with 16,473 (20.6%) having an assignment of at least one GO term. Sequences with GO terms corresponding to biological processes fell

into 38 plant GOSlim classes, with most of the ESTs being dedicated to cellular and metabolic processes (48.1%-48.4%; Figure 1).

In the root library, 7,824 (62.4%) sequences were assigned an InterPro signature, and from these, 6,264 (49.9%) were assigned at least one GO term. Sequences with GO terms corresponding to biological process fell into 36 plant GOSlim classes, with most of the ESTs also being dedicated to cellular and metabolic processes (48.6%; Figure 2).

BLAT analysis

Using BLAT, consensus sequences from A. barbata were aligned against cDNA from Brachypodium distachyon (32,255 transcripts sequences, including spliced variants, downloaded from www.phytozome.net). For each library, 57% showed a significant hit with a cDNA from Brachypodium (63,546 sequences for the ambient and 81,957 sequences for the high rainfall library). Amongst the sequences that did not have a significant hit a large proportion were singlets supported only from 454 pyrosequencing reads (80% for ambient rainfall and 86% for high rainfall library). Among the consensus sequences that were supported by Sanger reads, 89% and 90% had a significant hit with the Brachypodium cDNA, for low rain and high rain libraries, respectively.

Comparison of leaf transcriptome between ambient and high rainfall

The distribution of consensus sequences within each plant GOSlim category for "biological process", "cellular component", and "molecular function" was compared between the leaf ambient-rainfall and leaf high-rainfall libraries. Overall, the same categories were represented, except that "extracellular space" and "cell-cell signaling" were not represented in the leaf high-rainfall library (Figure 1). Some categories of biological processes were significantly over-represented (Chi-square, $p<0.05$) in the leaf high-rainfall library compared to the leaf ambient-rainfall, such as "amino acid and derivative metabolic process", "generation of precursor metabolites and energy", "photosynthesis," and "response to stress."

In each library, the number of ESTs from A. barbata that were associated with each Brachypodium cDNA model was counted. A large proportion of the ESTs from each library was associated with specific Brachypodium cDNA models, 85% for ambient rainfall and 82% for high rainfall. Digital northern analysis identified 2,077 cDNA models out of the 16,423 tested that were differentially expressed ($p<0.01$) between leaf ambient and high rainfall libraries. After applying the Bonferroni correction, 126 differentially expressed sequences remained statistically significant, with 58 down-regulated and 68 up-regulated under high rainfall conditions. On average, the fold difference in normalised EST abundance was 0.38 fold for down regulated and 5.8 for up regulated genes. Using the R statistic test (Stekel et al. 2000), which is less stringent than the Bonferroni correction, 134 additional genes were found to be differentially expressed (See Additional Files 1 and 2 for a list of up- and down-regulated genes).

The MapMan software was used to visualize the results from the digital northern analysis (Thimm et al. 2004). A mapping file was generated using a subset of cDNA from Brachypodium, which is the closest relative with a fully sequenced genome. The 230 differentially expressed genes were mapped to 23 different functional bins, with some genes mapped to 2 or more bins (Figure 2). Photosynthesis is being affected by the treatment, with 19 genes involved in the light reaction down-regulated under high

rainfall. The genes encoding for RuBisCO small subunit and RuBisCO activase were down-regulated under high rainfall while the gene encoding for RuBisCO large subunit was up-regulated. A total of 16 genes assigned to the stress bin were up-regulated under high rainfall conditions. Distinct pathways were affected by the long-term changes in soil water content. As part of the hormone signaling, only the genes involved in ethylene and salicylic acid were affected. A total of 11 genes encoding heat-shock proteins were up regulated under high rainfall. MYB and WRKY were the two main families of transcription factors showing differential regulation under varied rainfall pattern.

Comparison between leaf and root transcriptome

Consensus sequences from roots of *A. barbata* plants grown under different nitrogen and rainfall conditions were compared to the consensus sequences from leaves of plants grown in the same conditions and leaves collected at a later stage of development from ambient and high rainfall treatments. A total of 3,599 consensus sequences from the roots (28.7% of all root consensus) did not have a significant BLAST hit (e-value <1e-10) with sequences from the leaf library.

Fewer GOSlim classes were represented in the root-specific set of transcript (29 compared to 36 in all root). These classes were represented by a small number of sequences (less than 0.2%) in the all-root dataset and were aggregated in the category "Others" in Figure 3. Overall, all GOSlim classes were represented by a similar proportion of sequences in both all-root and root-specific datasets, except for six classes that showed significant difference (Chi-square, $p<0.05$). Two GOSlim classes, "response to external stimulus" and "protein metabolic process", were over-represented in the root-specific dataset. In particular, among the eight sequences from the all-root dataset that were assigned to the "response to external stimulus" class, six sequences belonged to the root-only dataset.

Root-specific sequences supported by more than 5 ESTs are summarized in Additional File 3. Out of the 10 most abundant root-specific sequences, five do not have a significant hit in Swiss-Prot. Transcripts identified as being root specific showed similarity to genes involved in secondary metabolism (Isoflavone-7-O-methyltransferase, cytochrome P450 and (RS)-norcoclaurine 6-O-methyltransferase), protein catabolism (aspartic proteinase nepenthesin-2, vignain cystein endopeptidase, and serine carboxypeptidase), recycling of amino acids (aminotransferase) and stress response (dehydration responsive protein and salt stress-induced proteins). Among the nitrate transporters found in the root, only one was not present in the leaf libraries, perhaps indicating a specific role for this nitrate transporter in the roots.

Real-time RT-PCR analysis of root specific transcripts

The level of transcripts specific to the roots was analyzed using real-time RT-PCR on root samples collected at a different time from those used for cDNA library construction (Figure 4). At the vegetative and reproductive stages, roots from plants grown under three rain treatments (low, ambient and high) were collected eight days after a watering event. The abundance of transcripts for the consensus annotated as a dehydration responsive protein RD22 was lower under low rain compared to ambient and high rain, both at the vegetative (1.7 fold) and reproductive stage (1.4 fold). Transcripts were also 0.4 fold lower at the reproductive stage vs. the vegetative stage. Transcript levels for the

consensus annotated as an xanthoxin dehydrogenase were 0.5 fold lower at the reproductive stage compared to the vegetative stage. Transcript levels for the aspartic proteinase nepenthesin were higher under low rain vs. ambient and high rain both at the vegetative (5.3 fold) and the reproductive stages (4.9 fold). There was no significant effect of plant developmental stages on the abundance of transcripts for this aspartic proteinase. The abundance of transcripts for the consensus sequence annotated as a norcoclaurine 6-O-methyltransferase (OMT1043) was unchanged in roots grown under different rain treatment and under two vegetative stages. However, the mRNA levels for the consensus sequence annotated as an isoflavone-7-O-methyltransferase (OMT66) were 0.8 fold lower at the reproductive stage vs. the vegetative stage ($p<0.01$). The abundance of transcript for OMT1043 was on average 4.5 fold higher than those for OMT66. Real-time RT-PCR also confirmed the absence of expression for these genes in the leaves.

Transcription factors

A total of 9,327 distinct transcription factor protein sequences were downloaded from the Plant Transcription Factor Database (http://plntfdb.bio.uni-potsdam.de/v2.0/downloads.php), in May 2009. These sequences were used in BLASTp analyses against A. barbata protein sequences obtained from the Prot4EST pipeline. Between 332 and 3,416 consensus sequences per library showed a significant hit (e-value <1e-05) with transcription factor sequences. On average, 5.2% of all sequences in a library had a significant hit, with this value being higher for the mixed leaf and root samples (7.92% and 7.64%).

Out of the 68 families of transcription factors and other transcriptional regulators, 63 were represented in the leaf-ambient rainfall library, 64 in the leaf-high rainfall, 44 in the leaf-mixed rain and nitrogen levels, and 54 in the root-mixed rain and nitrogen levels (Figure 5). This is most likely due to the much higher depth of coverage in the libraries sequenced by 454. The difference between libraries could also be indicative of the different developmental stages, since the leaves sampled from different nitrogen and rain levels were younger than those collected at ambient and high rainfall.

Analysis of the abundance of ESTs per consensus sequences for each library can give an indication of the level of expression of particular genes within these families. The family containing forkhead associated domain (FHA) showed a high ESTs per consensus ratio (32 ESTs/consensus on average in leaf-ambient and leaf-high rainfall libraries). PBF-2 like and S1Fa-like also showed a high ESTs/consensus ratio. Interestingly, these two families were not represented in the Sanger-only libraries of leaf and root in our study.

Discussion

We present here the analysis of root and leaf transcriptomes from A. barbata plants grown in soil under varied conditions. Sequences were obtained using both the Sanger method and 454 pyrosequencing methods (Margulies et al. 2005). The aims of this project were to generate a large amount of cDNA sequence data that would constitute a good basis for future more detailed transcriptomic studies in A. barbata, and to identify the genes differentially expressed in A. barbata leaves under changing rainfall patterns. The availability of transcriptomic data for A. barbata is especially important, because this species is the focus of a series of experiments aiming at understanding the response of

grassland plant species to changes in environmental conditions (St Clair et al. 2009a; St Clair et al. 2009b), and adaptive responses to environmental gradients in rainfall and other variables (Latta 2009; Latta and Gardner 2009).

The combined use of pyrosequencing with the GS 454-FLX system and Sanger sequencing was successful in generating a high number of good-quality ESTs (985,844). ESTs were assembled into 191,340 consensus sequences, though a large number of singlets (135,561) remained. Singlets may correspond to reads that include sequencing errors or to non-overlapping sections of a transcript. Studies in Arabidopsis have shown that 454 reads tend to cover the full-length of transcripts, with a slight bias for the 5' end (Weber et al. 2007). Alternatively, some reads may represent novel transcripts that are expressed at very low level.

The presence of longer Sanger-read ESTs in the dataset greatly improved the clustering analysis, leading to fewer consensus sequences once Sanger reads were included for clustering (see Table 1). The inclusion of Sanger sequences in the assembly was useful in closing the gaps for the sequencing of barley BAC (bacterial artificial chromosome) (Wicker et al. 2006). Advances in 454-pyrosequencing technologies, such as the development of the GS FLX Titanium system that can generate reads of over 400 bp, may reduce the need for Sanger reads in the future.

One main advantage of the 454 is the depth of coverage that can be obtained in a single run. Weber et al. (Weber et al. 2007) estimated that with two 454-pyrosequencing runs, about 90% of the transcripts in a tissue could be sequenced. The analysis of TF families clearly demonstrated an increased depth of coverage using 454. In particular, transcripts for TF expressed at a low level were included in the leaf transcriptomes sequenced by 454, but not in Sanger-sequenced libraries.

While some consensus sequences could be annotated on a functional level, a large proportion did not show significant similarity with entries in Swiss-Prot. Perhaps their short length impaired the search for a significant hit in the databases. This was especially salient, as the proportion of sequences with significant BLAST hits was almost four times higher in Sanger-only libraries, compared to the hybrid Sanger/454 libraries. Additionally, some of the short read may correspond to UTR and could not have a hit in protein database. Because we used both sequencing methods, we were able to generate a large amount of high-quality sequences that could, in some cases, assemble into full-length cDNA transcripts.

Sanger and 454 methods differ in the type of sequences that can be processed, as suggested by the analysis of microsatellites (Table 3) and GC content of coding regions (data not shown). The frequency of mononucleotides and trinucleotides differ between sequencing methods. It seems that the presence of stretches of mononucleotides may hinder the sequencing process when using 454. However, other parameters may lead to a stretch of DNA not being sequenced. An analysis of the set of consensus sequences not supported by 454 reads showed that some sequences did not contain mononucleotides repeats. In this case, perhaps the pieces of DNA were not properly sheared at the earlier stage of the preparation. A hybrid assembly of Pythium ultimum transcriptome also included a set of sequences (3.2% of all sequences) that were supported only by Sanger reads (Cheung et al. 2008). Further analysis of a larger set of sequences not supported by 454 may give more clues.

Transcription factors regulate many aspects of plant metabolism. The number of consensus sequences for each TF family was in general similar to the number of TF gene models in Arabidopsis and rice (http://plntfdb.bio.uni-potsdam.de/v2.0/), with some exceptions. This is interesting, since only TF expressed in two tissue types are included in this study. Some TF families were significantly enriched in A. barbata libraries. For example, between 46 and 56 consensus sequences from the CAMTA family could be found in A. barbata ambient and high leaf libraries, respectively. In both Arabidopsis and rice, the number of gene models belonging to the CAMTA family was seven. Similar results were obtained with the DBP and RWP-RK families.

The FHA family showed a high EST per consensus ratio for both leaf libraries (Figure 5). This is an indication of high abundance of transcripts for genes that belong to this TF family. Two other TF families showed a high EST per consensus ratio, PBF2-like and S1Fa-like. Interestingly, these were represented by only 2-4 consensus sequences in the dataset, which indicates small families of transcription factors compared to the average 46-54 consensus sequences per TF family in the hybrid Sanger-454 libraries. This is also consistent with the number of Whirly or PBF2 sequences identified in other plant species (Desveaux et al. 2005). PBF2-like has been described as a plant defense transcription factor that is composed of p24 proteins that bind the elicitor response element (ERE) of the promoter region of the pathogenesis related gene PR-10a thus eliciting transcription (Despres et al. 1995; Matton et al. 1993).

In this study, we asked whether changing a rainfall pattern by increasing the amount of water throughout a growing season would affect plant gene expression and tested the use of pyrosequencing technology to investigate plant gene expression response to aspects of climate change. Gene expression studies have often focused on acute treatments such as severe drought or heat shock. In the context of climate change studies, it is also important to understand the pattern of plant gene expression in response to long-term relatively mild changes in environmental conditions (Leakey et al. 2009). The rain treatment affected the total amount of water per year as well as the number of wet and dry days. At the time of sampling soil moisture content at 10 cm depth was ca. 20% higher under high rain vs. ambient rain (Data not shown). In this study we used digital northern analysis of EST library as a discovery tool to identify genes that were differentially expressed in response to season-long change in rainfall pattern. In particular, many genes were down-regulated under high-rainfall treatment. No measures of leaf level photosynthesis were taken at the time when the samples for cDNA libraries construction were collected. However, leaf level photosynthesis showed a slightly higher level under ambient rainfall (15.2 μmol m^{-2} s^{-1} +/- 1.3) compared to high rainfall (12.7 μmol m^{-2} s^{-1} +/- 1.0), when measured at a wet time point, four days before sampling (E. Sudderth et al., in preparation). These results are consistent with the decrease in abundance for photosynthesis-related transcripts in the high-rainfall leaf library.

Plant response to changing water availability begins at the root level. Before any changes occur in the shoot due to the lack of water or nutrients, signals (e.g., hormones) are sent from the roots. Soil is a highly heterogeneous media with microsites that are water and nutrients-rich. How plants detect a decline of water in the soil may differ from other media that are more homogeneous such as sand or vermiculite. Within this study, we collected roots grown in natural soil, thus allowing us to identify transcripts that are expressed in field-conditions. To our knowledge, this is the first report of sequencing

soil-grown root transcriptome. Previous reports of transcriptome analysis using SAGE analyses focused on maize, Arabidopsis, or poplar plants grown hydroponically, in vermiculite or peat-vermiculite, respectively (Fizames et al. 2004; Kohler et al. 2003; Poroyko et al. 2005).

Among transcripts identified as being abundant in roots but absent from the leaf libraries we selected five candidates for quantitative RT-PCR studies. In the roots, we studied the expression of genes encoding enzymes involved in abscisic acid synthesis, dehydration-responsive protein RD22, aspartic proteinase, and two genes encoding O-methyl transferase. Transcripts for the dehydration-responsive protein RD22 and aspartic proteinase were significantly down-regulated under the low rain treatment. Microarray analysis of transcript changes in the roots of Populus euphratica also identified dehydration-responsive protein RD22 as being down-regulated in response to dehydration occurring over a long period (Bogeat-Triboulot et al. 2007). Bogeat-Triboulot et al. (2007) also found that most transcripts were down regulated in the roots during a four-week long water deprivation experiment.

There was no interaction between rain treatment and developmental stages on the abundance of transcripts studied here. Due to our sampling technique, the root samples were a mixture of different root sections (e.g., root tip, lateral root zone, mature root). However, we can assume that mature roots represented the largest section types in a root system. Therefore, the changes in transcript abundance in the overall samples are likely to be due to changes in transcript abundance in the mature roots. This suggests that the effect of different rain treatments was similar on mature roots from plants at different developmental stages.

Wild relatives of crops are well-known source of allelic variation for desirable agronomic traits. As a wild relative of the cultivated oat, A. barbata has already been of particular interest since an ecotype of this species from northern Africa was used as a source of gene for resistance to powdery mildew. The resistance to this pathogen was then successfully transferred to the cultivated oat (Aung et al. 1977). The sequencing of cDNA libraries from A. barbata constitutes a resource to the many researchers interested in this species and its relatives.

Materials and Methods

Plant growth conditions

A. barbata seeds were collected at the Hopland Field Station, Northern California (USA). Annual mean precipitation at this location exceeds 500 mm and is therefore described as a mesic environment. In addition, A. barbata plants grown in our experiments showed the light-colored lemma and hairy leaf sheath associated with the mesic ecotype previously described (Allard 1996). Plants were grown from December 2006 to June 2007, in mesocosm pots filled with three horizons of natural soil collected at the Hopland Field station (St Clair et al. 2009a; St Clair et al. 2009b). These were placed in greenhouses, with supplementary lighting supplied during daytime by high-pressure sodium bulbs. Mean temperature in the greenhouses ranged from 10°C to 19°C, while air relative humidity varied between 48% and 88%. Rainfall treatment was applied in cycles of 21 days, with an 11-day wet period and 10-day dry period (St Clair et al. 2009b). The

different precipitation treatments received a varying number of 15 mm watering events during the 11-day wet period (high rain=8, ambient rain=6, low rain=3). Nitrogen treatment consisted of four applications of 2 $g.m^{-2}$ of ammonium nitrate during the season.

Sample collection and RNA extraction

Samples from leaves and roots were collected at the vegetative stage from monocultures of A. barbata. Samples were taken eight days after the last rain event, from mesocosms with combinations of low, ambient, and high rainfall, and low and high nitrogen treatments. At the beginning of the reproductive stage (called the booting stage), leaf samples from A. barbata were collected from the ambient and high-rainfall treatment pots, between two and four days after the last rainfall event. At the time when the leaf samples were collected, the plants grown under ambient rainfall had received 657 mm, while those growing under high rainfall had received 30% more (867 mm) over a period of four months. All samples were harvested between 10.30 am and 12.30 pm. Leaf samples were snap frozen in liquid nitrogen and kept at -80°C until further processing. To collect roots, two soil cores (2.5 cm diameter) was taken at the base of an A. barbata plant and to a depth of 10 cm. Root samples were first washed with water to remove the bulk of soil particles, blotted dry on paper towels, and then frozen in liquid nitrogen. Total RNA was isolated from A. barbata root and leaf tissues using the TRIZOL Reagent (Invitrogen, Carlsbad, CA) according to manufacturer's instructions. Extractions were performed separately for each sample, and RNA was subsequently pooled. RNA quality was checked by spectrophotometry, while its integrity was verified on agarose gel. A. barbata mRNA was isolated from total RNA for four RNA samples using the Absolutely mRNA Purification kit and following manufacturers instructions (Stratagene, La Jolla, CA).

Clone library construction and sequencing

cDNA synthesis and cloning was carried out following a modified procedure based on the "SuperScript plasmid system with Gateway technology for cDNA synthesis and cloning" (Invitrogen, Carlsbad, CA). A total of 1-2 μg of mRNA, reverse transcriptase SuperScript II (Invitrogen) and oligo(dT)-NotI primer (5'

GACTAGTTCTAGATCGCGAGCGGCCGCCCT15VN 3') were used to synthesize first-strand cDNA. Second-strand synthesis was performed with E. coli DNA ligase, polymerase I, and RNaseH followed by end repair using T4 DNA polymerase. The SalI adaptor (5' TCGACCCACGCGTCCG and 5' CGGACGCGTGGG) was ligated to the cDNA, digested with NotI (New England Biolabs, Ipswich, MA), and subsequently size-selected by gel electrophoresis (1.1% agarose). Small and large insert size ranges (0.6-2kb and >2kb, respectively) of cDNA were cut out of the gel for subsequent processing. The cDNA inserts were directionally ligated into the SalI and NotI digested vector pCMVsport6 (Invitrogen) and subsequently transformed into ElectroMAX T1 DH10B cells (Invitrogen).

Library quality was first assessed by randomly selecting 24 clones and PCR amplifying the cDNA inserts with the primers M13-F (5' GTAAAACGACGGCCAGT) and M13-R (5' AGGAAACAGCTATGACCAT) to determine the fraction of insertless clones. Colonies from each library were plated onto agarose plates (Teknova, Hollister, CA) at a density of approximately 1,000 colonies per plate. Plates were grown at 37ºC for 18 hours, then individual colonies were picked and each used to inoculate a well containing LB media with appropriate antibiotic in a 384 well plate (Nunc, Rochester, NY). Clones in 384 well plates were grown at 37ºC for 18 hours. Contained plasmid DNA for sequencing was produced by rolling circle amplification (Templiphi, GE Healthcare, Piscataway, NJ). Subclone inserts were sequenced from both ends using primers complimentary to the flanking vector sequence (Fw: 5' ATTTAGGTGACACTATAGAA, Rv: 5' TAATACGACTCACTATAGGG) and Big Dye terminator chemistry, and then run on ABI 3730 instruments (Applied Biosystems, Foster City, CA).

454-FLX cDNA library construction and sequencing

The previously purified mRNA was reversed transcribed using SuperScript III (Invitrogen) and a dT15VN2 primer. Second-strand cDNA was synthesized by nick translation with E. coli DNA ligase, E. coli DNA polymerase I, and RNase H and blunt end repaired using T4 polymerase (Invitrogen). The dscDNA was fragmented and 300-800 base pair fragments were gel purified using a 2% agarose gel. The purified fragments were then used to create the 454 single-stranded cDNA library as described below (454 library preparation kit, Roche, Indianapolis, IN).

The fragment ends were polished using T4 ligase and T4 polynucleotide kinase (Roche, Indianapolis, IN). Adaptors containing primer sequences and a biotin tag were ligated to the fragment ends (Roche). The fragments with properly ligated adapters were immobilized onto magnetic streptavidin coated beads (Roche). Nicks or gaps between the adapters and the dscDNA fragments were repaired using the fill-in polymerase (Roche). The non-biotinylated strands of the immobilized dscDNA fragments were melted off to generate the single-stranded cDNA library for 454 sequencing.

EST sequence processing and assembly

For clone libraries, ESTs were generated in pairs, a 5' and 3' end read from each cDNA clone and for 454 FLX libraries, ESTs represented shotgun fragments from the transcripts. Clone library EST processing included vector and adaptor trimming, as well as quality trimming (trimmed to where the average base quality fell below Q15, based on

a sliding window of 11 bases). ESTs from clones identified as insertless or containing rRNA inserts were also removed. Next, we tested ESTs for the presence of poly(A) or poly(T) tails (which if present were removed) evaluating ESTs for length and removing ESTs with fewer than 100 high-quality bases remaining. Additionally, ESTs consisting of more than 50% low complexity sequence were also removed from the final set of "good ESTs."

For the 454 generated ESTs, the screening was the same as above, with the following exceptions: There was no vector, quality trimming, nor insertless clone identification, and ESTs shorter than 50 base pairs were removed (rather than 100).

For clustering, ESTs were evaluated with malign, a kmer-based alignment tool (Chapman, unpublished) that clusters ESTs based on sequence overlap (kmer = 16, seed length requirement = 32 alignment ID >= 98%). For clone libraries only, clusters of ESTs were further merged based on sister ESTs (5' and 3' from the same clone) using double linkage. Double linkage requires that two or more matching sister ESTs exist in both clusters to be merged. EST clusters were then each assembled using CAP3 (Huang and Madan 1999) to form consensus sequences. ESTs from each separate sample type were clustered and assembled separately, and subsequently the entire set of ESTs for all cDNA libraries were clustered and assembled together. The consensus sequences were then compared to Swiss-Prot using BLASTx, and the hits with the lowest expected value were reported.

Functional annotation

Nucleotides sequences were translated using prot4EST (Wasmuth and Blaxter 2004). Prot4EST is a pipeline that uses BLAST information, ESTscan, and DECODER to generate translation. BLASTx searches were previously carried out against the Swiss-Prot database. Matrices were generated for ESTscan using sequences from the Poaceae available in Genbank. Prot4EST was run using default parameters. InterProScan version 4.4 was run using databases version 19.0 (non-PANTHER) and 14.0 (PANTHER), and including noncommercial applications BLASTprodom, coils, gene3D, hmmpanther, hmmpir, hmmpfam, hmmsmart, hmmtigr, fpri ntscan, patterscan, profilescan, superfamily and seg (Quevillon et al. 2005). GO terms were associated with InterPro signatures and mapped using InterPro2GO. CateGOrizer (Hu 2008) was used to classify GO terms in plant GOSlim classes and give a broad overview of the ontology content without the specific fine grained terms.

Digital northerns

To identify specific genes that may be differentially expressed depending on the rainfall treatment, the number of supporting ESTs was counted for consensus sequences annotated using the same NCBI description. In principle, the abundance of ESTs supporting a particular consensus sequence is an indication of transcripts abundance in the sampled tissue. Differential expression data may therefore be obtained by comparing the abundance of ESTs supporting a particular consensus sequence between libraries (Audic and Claverie 1997). Differential expression levels were tested using the Audic and Claverie statistical test as part of IDEG6 (Romualdi et al. 2003). The Bonferroni correction was used to lower the significance threshold since a statistical test was repeatedly applied to the same dataset and this increases the chance of type one error. As

Bonferroni correction may be too conservative, the R statistical test was also applied (Stekel et al. 2000). Genes with R>8 (true positive rate of ~98%) were considered differentially expressed. Genes identified as significantly differentially regulated using the Audic and Claverie test were confirmed by the R statistic, additional differentially expressed genes were also identified (Additional file 1 and 2).

Real-time reverse transcription-PCR (RT-PCR) experiment and analysis

Total RNA was extracted from A. barbata roots collected at different developmental stages and under low, ambient and high rainfall treatment, using the RNAqueous kit (Ambion, Austin, TX). After DNAse treatment (Invitrogen, Carlsbad, CA), the synthesis of cDNA was carried out using random primers and Superscript II (Invitrogen).

Real-time RT-PCR experiments were performed on the MyIQ cycler (Bio-Rad, Hercules, CA) and using SYBR GreenI supermix (Bio-Rad). Normalization was done against a calibrator sample and two internal control genes (APT1 and PP2A). The expression of these two genes was not affected by the treatments; the arithmetic mean of each internal control gene pairwise variations with the other internal control gene was within the expected range for stable expression (Vandesompele et al. 2002).

PCR Miner was used for analysis of the raw data (Zhao and Fernald 2005); http://www.miner.ewindup.info/Miner_version2/) and calculation of the cycle threshold (Ct value) and PCR efficiency. Relative expression of the gene of interest was then normalized to the two internal control genes using GeNorm (Vandesompele et al. 2002).

Sequence availability

Sanger sequences were deposited in GenBank dbEST under GenBank accession numbers GR313001-GR367242. The 454 reads were sent to the NCBI short-read archive.

Acknowledgments

We would like to thank Calvin Myint for help with samples collection, Christopher Wilks (The Arabidopsis Information Resource) and Robert Bukowski from Cornell's Computational Biology Service Unit (CBSU) for InterProScan analysis. We thank Samuel B. St. Clair, Erika A. Sudderth, Rohit Salve, Kallista Bley, Paul Cook, Melissa Crago, Tara Macomber, Alex Morales, and Margaret S. Torn for their work on the design and execution of the mesocosm experiment. We would also like to thank Todd DeSantis, Ulas Karaoz, and Ping Hu for their support with unix and perl based analyses. Thanks also to David Swarbreck for helpful discussions and suggestions. This work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental and Research Program, Climate Change Research Division, and by the University of California, Lawrence Berkeley National Laboratory, under Contract No. DE-AC02-05CH11231.

References:

Allard, R.W. (1996) Genetic basis of the evolution of adaptedness in plants. Euphytica 92: 1-11.

Allard, R.W., Babbel, G.R., Kahler, A.L. and Clegg, M.T. (1972) Evidence for

coadaptation in Avena Barbata. Proceedings of the National Academy of Sciences of the United States of America 69: 3043-3048.
Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.
Andersen, G.R., Nissen, P. and Nyborg, J. (2003) Elongation factors in protein biosynthesis. Trends in Biochemical Sciences 28: 434-441.
Audic, S. and Claverie, J.M. (1997) The significance of digital gene expression profiles. Genome Research 7: 986-995.
Aung, T., Thomas, H. and Jones, I.T. (1977) Transfer of gene for mildew resistance from Avena barbata (4X) into cultivated oat Avena sativa by an induced translocation. Euphytica 26: 623-632.
Bennett, M.D. and Smith, J.B. (1976) Nuclear-DNA amounts in angiosperms. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 274: 227-274.
Bogeat-Triboulot, M.B., Brosche, M., Renaut, J., Jouve, L., Le Thiec, D., Fayyaz P., et al. (2007) Gradual soil water depletion results in reversible changes of gene expression, protein profiles, ecophysiology, and growth performance in Populus euphratica, a poplar growing in arid regions. Plant Physiology 143: 876-892.
Buchanan-Wollaston, V., Earl, S., Harrison, E., Mathas, E., Navabpour, S., Page, T., et al. (2003) The molecular analysis of leaf senescence - a genomics approach. Plant Biotechnology Journal 1: 3-22.
Cayan, D.R., Maurer, E.P., Dettinger, M.D., Tyree, M. and Hayhoe, K. (2008) Climate change scenarios for the California region. Climatic Change 87: S21-S42.
Cheung, F., Win, J., Lang, J.M., Hamilton, J., Vuong, H., Leach, J.E., et al. (2008) Analysis of the Pythium ultimum transcriptome using Sanger and Pyrosequencing approaches. BMC Genomics 9:542.
Despres, C., Subramaniam, R., Matton, D.P. and Brisson, N. (1995) The activation of the potato PR-10a gene requires the phosphorylation of the nuclear factor PBF-1. Plant Cell 7: 589-598.
Desveaux, D., Marechal, A. and Brisson, N. (2005) Whirly transcription factors: defense gene regulation and beyond. Trends in Plant Science 10: 95-102.
Fizames, C., Munos, S., Cazettes, C., Nacry, P., Boucherez, J., Gaymard, F., et al. (2004) The Arabidopsis root transcriptome by serial analysis of gene expression. Gene identification using the genome sequence. Plant Physiology 134: 67-80.
Hawkes, C.V., Wren, I.F., Herman, D.J. and Firestone, M.K. (2005) Plant invasion alters nitrogen cycling by modifying the soil nitrifying community. Ecology Letters 8: 976-985.
Hayhoe, K., Cayan, D., Field, C.B., Frumhoff, P.C., Maurer, E.P., Miller N.L., et al. (2004) Emissions pathways, climate change, and impacts on California. Proceedings of the National Academy of Sciences of the United States of America 101: 12422-12427.

Hu, Z.H., Jie Bao and James M. Reecy (2008) CateGOrizer: A Web-Based Program to Batch Analyze Gene Ontology Classification Categories. Online Journal of Bioinformatics 9: 108-112.
Huang, X.Q. and Madan, A. (1999) CAP3: A DNA sequence assembly program. Genome Research 9: 868-877.
Kohler, A., Delaruelle, C., Martin, D., Encelot, N. and Martin, F. (2003) The poplar root transcriptome: analysis of 7000 expressed sequence tags. Febs Letters 542: 37-41.
Latta, R.G. (2009) Testing for local adaptation in Avena barbata: a classic example of ecotypic divergence. Molecular Ecology 18: 3781-3791.
Latta, R.G. and Gardner, K.M. (2009) Natural selection on pleiotropic quantitative trait loci affecting a life-history trade-off in Avena barbata. Evolution 63: 2153-2163.
Leakey, A.D.B. (2009) Rising atmospheric carbon dioxide concentration and the future of C4 crops for food and fuel. Proceedings of the Royal Society B-Biological Sciences 276: 2333-2343.
Leakey, A.D.B., Ainsworth, E.A., Bernard, S.M., Markelz, R.J.C., Ort, D.R., Placella, S.A., et al. (2009) Gene expression profiling-opening the black box of plant ecosystem responses to global change. Global Change Biology 15:1201-1213.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 376-380.
Matsumoto, T., Wu, J.Z., Kanamori, H., Katayose, Y., Fujisawa, M., Namiki, M., et al. (2005) The map-based sequence of the rice genome. Nature 436: 793-800.
Matton, D.P., Prescott, G., Bertrand, C., Camirand, A. and Brisson, N. (1993) Identification of cis-acting elements involved in the regulation of the pathogenesis-related gene STH-2 in Potato. Plant Molecular Biology 22: 279-291.
Poroyko, V., Hejlek, L.G., Spollen, W.G., Springer, G.K., Nguyen, H.T., Sharp, R.E., et al. (2005) The maize root transcriptome by serial analysis of gene expression. Plant Physiology 138: 1700-1710.
Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R. and Lopez, R. (2005) InterProScan: protein domains identifier. Nucleic Acids Research 33: W116-W120.
Romualdi, C., Bortoluzzi, S., D'Alessi, F. and Danieli, G.A. (2003) IDEG6: a web tool for detection of differentially expressed genes in multiple tag sampling experiments. Physiological Genomics 12: 159-162.
St Clair, S.B., Sudderth, E.A., Castanha, C., Torn, M.S. and Ackerly, D.D. (2009a) Plant responsiveness to variation in precipitation and nitrogen is consistent across the compositional diversity of a California annual grassland. Journal of Vegetation Science 20: 860-870.
St Clair, S.B., Sudderth, E.A., Fischer, M.L., Torn, M.S., Stuart, S.A., Salve, R., et al. (2009b) Soil drying and nitrogen availability modulate carbon and water exchange over a range of annual precipitation totals and grassland vegetation types. Global Change Biology 15: 3018-3030.
Stekel, D.J., Git, Y. and Falciani, F. (2000) The comparison of gene expression from multiple cDNA libraries. Genome Research 10: 2055-2061.

Thayer, S.S., St Clair, S.B., Field, C.B. and Somerville, S.C. (2008) Accentuation of phosphorus limitation in Geranium dissectum by nitrogen: an ecological genomics study. Global Change Biology 14: 1877-1890.
Thiel, T., Michalek, W., Varshney, R.K. and Graner, A. (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theoretical and Applied Genetics 106: 411-422.
Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Kruger, P., et al. (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant Journal 37: 914-939.
Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology 3: 0034.0031-0034.0011.
Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D., Bevan M.W. et al. (2010) Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463: 763-768.
Wasmuth, J.D. and Blaxter, M.L. (2004) Prot4EST: Translating Expressed Sequence Tags from neglected genomes. BMC Bioinformatics 5:187.
Weber, A.P.M., Weber, K.L., Carr, K., Wilkerson, C. and Ohlrogge, J.B. (2007) Sampling the arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiology 144: 32-42.
Wicker, T., Schlagenhauf, E., Graner, A., Close, T.J., Keller, B. and Stein, N. (2006) 454 sequencing put to the test using the complex genome of barley. BMC Genomics 7.
Wolpert, T.J., Navarre, D.A., Moore, D.L. and Macko, V. (1994) Identification of the 100-Kd victorin binding-protein from oats. Plant Cell 6: 1145-1155.
Zhao, S. and Fernald, R.D. (2005) Comprehensive algorithm for quantitative real-time polymerase chain reaction. Journal of Computational Biology 12: 1047-1064.

Figures

Figure 1 – Functional classification of the leaf transcriptome of A. barbata. Plant GOSlim terms were used for classification for GO cellular component, GO molecular function and GO biological process. Solid bars correspond to the leaf collected under ambient rainfall while striped bars correspond to leaves collected under high rainfall. * indicates statistical difference between libraries at the p=0.05 significance level.

Figure 2 – Mapping and visualisation using MapMan of the genes differentially expressed in the leaves of A. barbata grown under varied water availability. A. Black bars indicate down-regulated genes while grey bars indicate genes that were up-regulated under high rainfall. B. In the photosynthesis overview window, each square corresponds to a gene that is differentially regulated. Red indicates significant up-regulation while blue indicates down-regulation under the high rain treatment. Only the genes that were significantly differentially expressed are represented in the MapMan figure.

Figure 3 – Functional classification of the root transcriptome of A. barbata. Plant GOSlim terms were used for classification for GO cellular component, GO molecular function and GO biological process. Shaded area for each bar corresponds to the number of occurrence of a GO term for the root specific dataset. Numbers indicate the proportion of each GO category for the whole root dataset.

Figure 4 – Quantitative RT-PCR for root specific transcripts analyzed in the roots grown under low, ambient and high rain, at both the vegetative and reproductive stages. Transcript abundant was normalized to the abundance of transcripts for two endogenous genes. Data represent the average and standard error for five experimental replicates.

Figure 5 – Relative abundance of consensus sequences and ESTs per consensus for each transcription factors family. Bars in the primary y-axis represent the percentage of consensus sequences showing similarity to a TF that belong to the specific family for the leaf-ambient rainfall (white bar), leaf-high rainfall (grey bar), leaf-mixed (black bar) and root-mixed (hatched bar) libraries. The average number of ESTs per consensus for the leaf-ambient and leaf-high rain libraries is represented in the secondary y-axis.

Tables

Table 1 - Sequencing and clustering statistics of A. barbata ESTs
Table 2 - Clustering of ESTs in consensus sequences
Table 3 - Analysis of microsatellites markers in A. barbata root and leaf libraries
Table 4 - Functional annotation of consensus sequences using BLAST
Table 5 - Most abundant consensus in each library

Additional File 1 - List of genes up regulated in leaves collected under high rainfall treatments. AC.1-2 corresponds to the p value calculated with the Audic and Claverie test. Genes with non-significant p value calculated after applying the Bonferroni correction are reported as ns. R Stat corresponds to the likelihood statistic R.

Additional File 2 - List of genes down regulated in leaves collected under high rainfall treatments. AC.1-2 corresponds to the p value calculated with the Audic and Claverie test. Genes with non-significant p value calculated after applying the Bonferroni correction are reported as ns. R Stat corresponds to the likelihood statistic R.

Additional File 3 - List of root specific transcripts

1 Table 1

Libraries		Total N° of ESTs sequenced	Sequencing Success (%)	ESTs used for Clustering	ESTs in Consensus	ESTs singlet (1)	% Redundancy (2)	Total N° of consensus
Leaves- Ambient rainfall	Sanger Small Inserts	6,144	84.5	5,072	4,134	938	81.5%	2,055
	Sanger Large Inserts	6,144	84.7	5,074	3,219	1,855	63.4%	2,823
	Pyrosequencing	504,934	99.4	487,295	411,315	75,980	84.4%	113,646
	Clustering Sanger / pyrosequencing			497,441	422,174	75,267	84.9%	112,130
Leaves- High rainfall	Sanger Small Inserts	6,144	82.2	4,930	4,010	920	81.3%	2,154
	Sanger Large Inserts	6,144	85.9	5,150	3,291	1,859	63.9%	2,904
	Pyrosequencing	462,368	99.0	444,307	334,311	109,996	75.2%	145,722
	Clustering Sanger / pyrosequencing			454,387	346,008	108,379	76.1%	142,801
Leaves- Mixed rain and N levels	Sanger Small Inserts	6,144	86.1	5,169	4,373	796	84.6%	1,841
	SangerLargeInserts	6,144	85.8	5,132	3,272	1,860	63.8%	2,831
	Clustering Sanger libraries			10,301	7,852	2,449	76.2%	4,345
Roots- Mixed rain and N levels	Sanger Small Inserts	13,824	83.4	11,374	9,733	1,641	85.6%	5,386
	SangerLargeInserts	13,824	90.4	12,341	6,939	5,402	56.2%	7,968
	Clustering Sanger libraries			23,715	17,154	6,561	72.3%	12,543
All libraries		1,031,814	92	985,844	850,283	135,561	86.2%	191,340

(1) A singlet is defined as sequence consensus based on a single EST

(2) Redundancy is defined as number of ESTs in cluster / total number of ESTs

2
3
4

1 Table 2

Number of ESTs per consensus	Number of consensus	% of total number of consensus
1	135,561	70.85%
2	22,939	11.99%
3-9	21,871	11.43%
10-49	8,267	4.32%
50-199	2,190	1.14%
200-999	434	0.23%
>1,000	78	0.04%
Total	191,340	100%

2
3

4
5
6
7

1 Table 3

	Leaves- Ambient rainfall	Leaves- High rainfall	Leaves- Mixed rain and N levels	Roots- Mixed rain and N levels	All consensus	Subset no pyrosequencing
	All	All				
Total number of sequences examined	112,130	142,801	4,345	12,543	191,340	1,293
Total of sequences with SSR	4,519	3,961	728	1,973	8,319	269
Number of sequences containing more than 1 SSR	447	348	135	364	1,067	86
TotalnumberofSSR	5,090	4,405	889	2,422	9,650	396
Mononucleotides SSR	1,866	1,393	400	1,088	3,768	294
Dinucleotides SSR	888	686	123	323	1,522	27
Trinucleotides SSR	2,167	2,162	344	909	4,017	70
Tetranucleotides SSR	136	127	20	69	253	3
Other(5-6nucleotidesSSR)	33	37	2	33	90	2

2

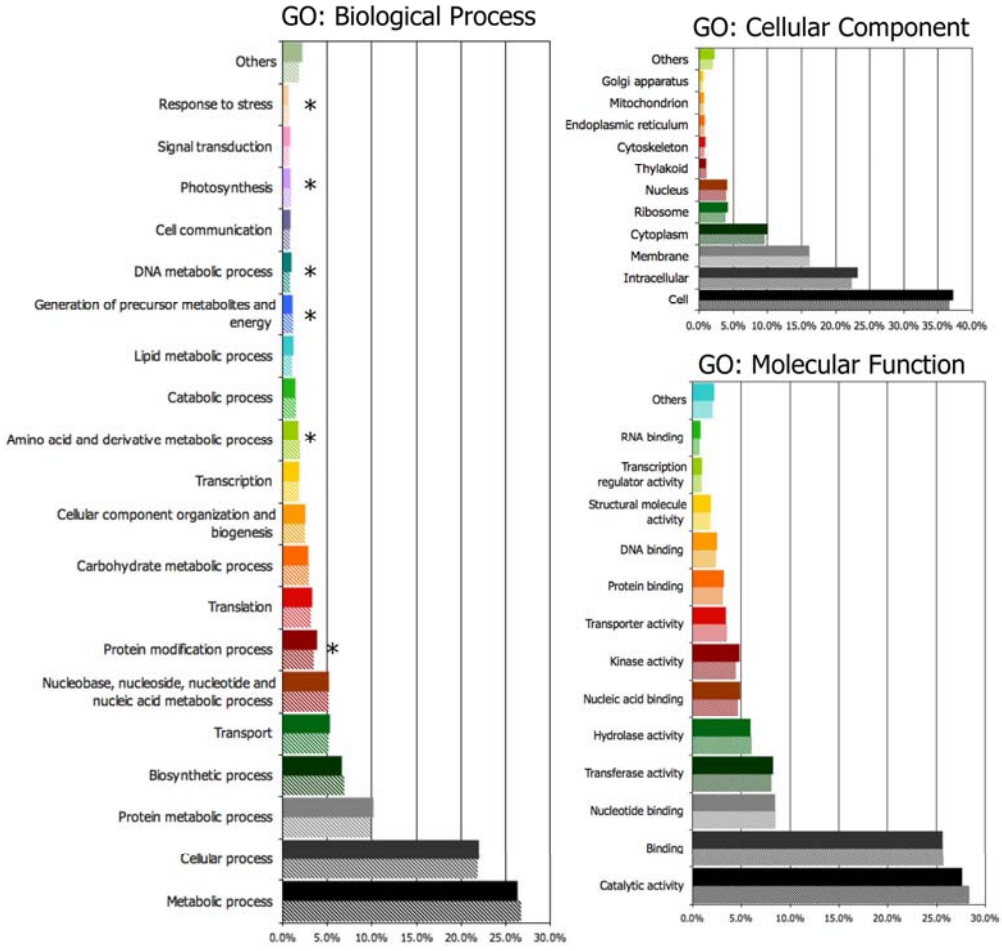
1 Table 4
2

	Total Consensus Number	Number of consensus with significanthit	% Consensus with hit
Leaves- Ambient rainfall	112,130	27,054	24.1
Leaves- High rainfall	142,801	27,703	19.4
Leaves- Mixed rain and N levels	4,345	3,157	72.7
Roots-MixedrainandNlevels	12,543	8,837	70.5

Table 5

	Consensus Name	ESTs per consensus	% of all ESTs	Annotation
Leaves-Ambient rainfall	2_159_CCGS_CCOO_CCOP	6237	1.25%	RuBisCO activase, chloroplast precursor
	2_211_CCGS_CCOO_CCOP	4109	0.83%	Victorin binding protein (component of the mitochondrial enzyme glycine decarboxylase)
	2_525_CCGS_CCOO_CCOP	4032	0.81%	Chloroplast fructose-bisphosphate aldolase
	2_227_CCGS_CCOO_CCOP	2942	0.59%	Chlorophyll a-b binding protein 1
	38_3_CCGS_CCOO_CCOP	2457	0.49%	PhotosystemII10kDapolypeptide,chloroplastic
Leaves-High rainfall	4_619_CCGT_CCOS_CCOT	5261	1.16%	RuBisCO activase, chloroplast precursor
	4_498_CCGT_CCOS_CCOT	4562	1.00%	Chloroplast fructose-bisphosphate aldolase
	4_159_CCGT_CCOS_CCOT	2329	0.51%	Victorin binding protein (component of the mitochondrial enzyme glycine decarboxylase)
	4_331_CCGT_CCOS_CCOT	2203	0.48%	Chlorophyll a-b binding protein 1
	490_3_CCGT_CCOS_CCOT	1577	0.35%	Serinehydroxymethyltransferase,mitochondrial
Leaves-Mixed rain and N levels	20_1_CCOU_CCOW	261	2.53%	Fructose-bisphosphate aldolase 2, chloroplastic
	49_1_CCOU_CCOW	130	1.26%	Ribulose bisphosphate carboxylase/oxygenase activase B, chloroplastic
	52_2_CCOU_CCOW	75	0.73%	Chlorophyll a-b binding protein 1B-21, chloroplastic
	49_3_CCOU_CCOW	60	0.58%	RuBisCO activase alpha form precursor
	319_2_CCOU_CCOW	58	0.56%	PhotosystemII10kDapolypeptide,chloroplastic
Roots-Mixed rain and N levels	2269_1_CCOX_CCOY	44	0.19%	Metallothionein-like protein 1
	40_3_CCOX_CCOY	38	0.16%	Elongation factor 1-alpha
	2242_2_CCOX_CCOY	31	0.13%	Metallothionein-like protein 1
	256_2_CCOX_CCOY	30	0.13%	Elongation factor 2
	200_1_CCOX_CCOY	28	0.12%	Glyceraldehyde-3-phosphate dehydrogenase, cytosolic

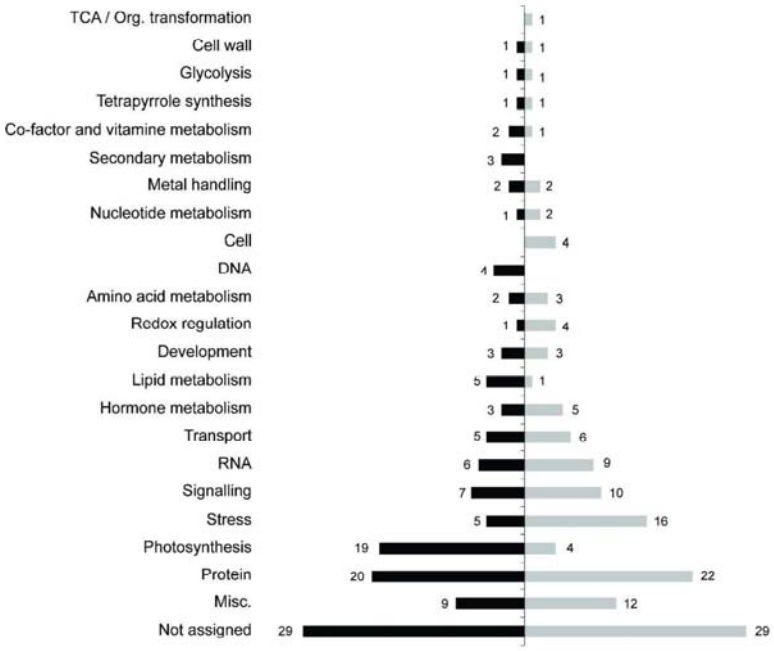
Figure 1



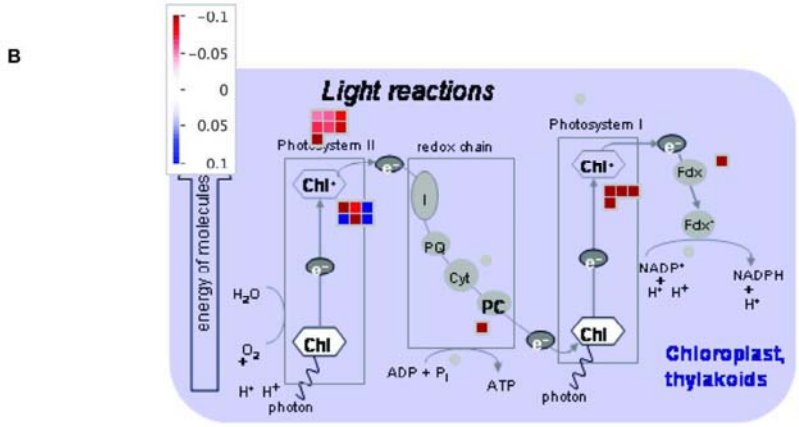
Functional classification of the leaf transcriptome of A. barbata. Plant GOSlim terms were used for classification for GO cellular component, GO molecular function and GO biological process. Solid bars correspond to the leaf collected under ambient rainfall while striped bars correspond to leaves collected under high rainfall. * indicates statistical difference between libraries at the p=0.05 significance level.
169x187mm (400 x 400 DPI)

Figure 2

A

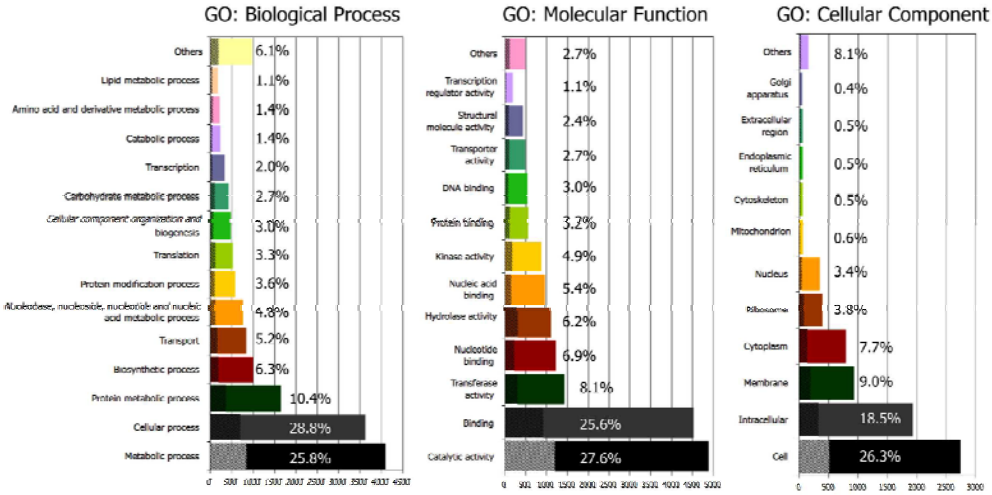


B



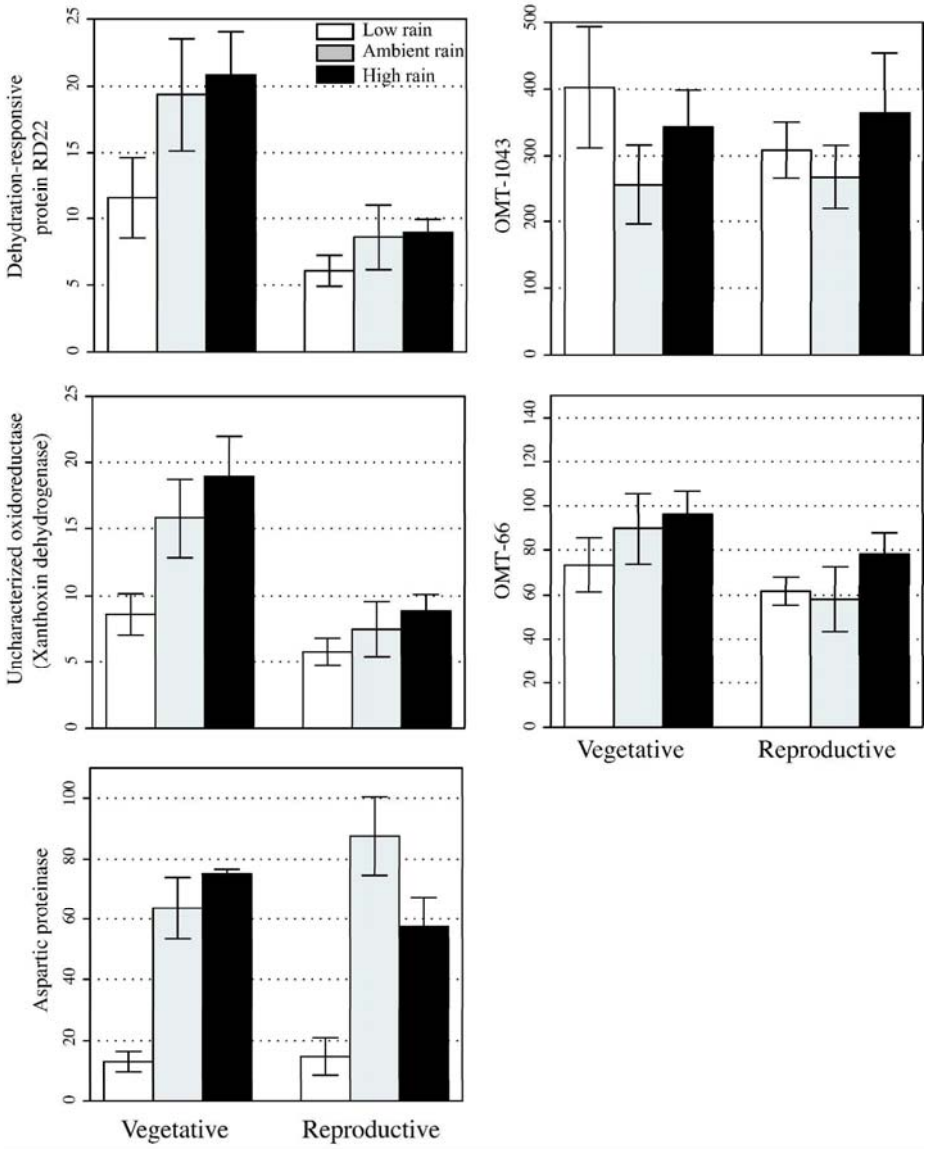
Mapping and visualisation using MapMan of the genes differentially expressed in the leaves of A. barbata grown under varied water availability. A. Black bars indicate down-regulated genes while grey bars indicate genes that were up-regulated under high rainfall. B. In the photosynthesis overview window, each square corresponds to a gene that is differentially regulated. Red indicates significant up-regulation while blue indicates down-regulation under the high rain treatment. Only the genes that were significantly differentially expressed are represented in the MapMan figure.
170x247mm (600 x 600 DPI)

Figure 3



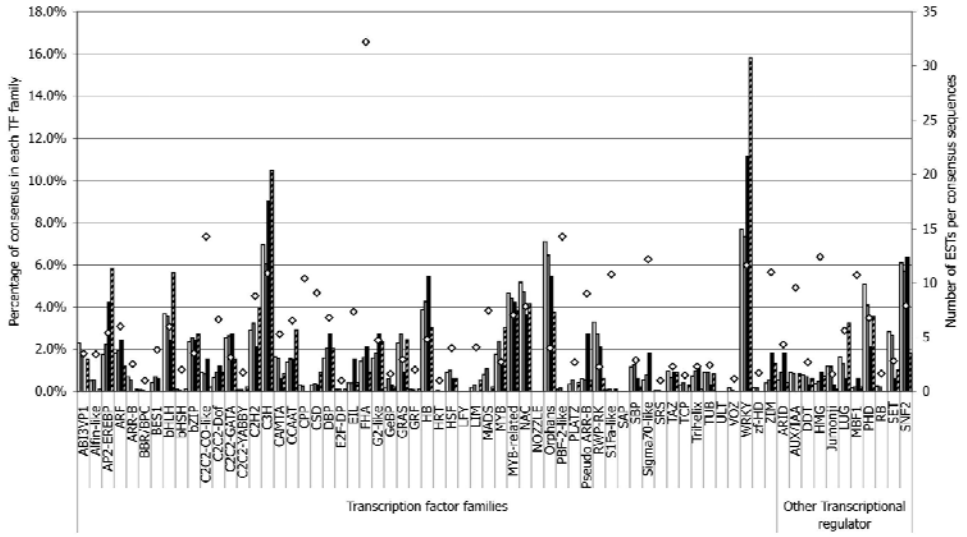
Functional classification of the root transcriptome of A. barbata. Plant GOSlim terms were used for classification for GO cellular component, GO molecular function and GO biological process. Shaded area for each bar corresponds to the number of occurrence of a GO term for the root specific dataset. Numbers indicate the proportion of each GO category for the whole root dataset.
99x58mm (600 x 600 DPI)

Figure 4



Quantitative RT-PCR for root specific transcripts analyzed in the roots grown under low, ambient and high rain, at both the vegetative and reproductive stages. Transcript abundant was normalized to the abundance of transcripts for two endogenous genes. Data represent the average and standard error for five experimental replicates.
152x198mm (600 x 600 DPI)

Figure 5



Relative abundance of consensus sequences and ESTs per consensus for each transcription factors family. Bars in the primary y-axis represent the percentage of consensus sequences showing similarity to a TF that belong to the specific family for the leaf-ambient rainfall (white bar), leaf-high rainfall (grey bar), leaf-mixed (black bar) and root-mixed (hatched bar) libraries. The average number of ESTs per consensus for the leaf-ambient and leaf-high rain libraries is represented in the secondary y-axis.
103x63mm (600 x 600 DPI)