Whole transcriptome profiling and water stress study of *A. donax* by RNA-seq

Yuan Fu
Department of Biotechnology
University of Verona

A thesis submitted for the degree of
Doctor of Philosophy
May 2017

UNIVERSITA' DEGLI STUDI DI VERONA

DEPARTMENT OF
Biotechnology

GRADUATE SCHOOL OF
Life and Health Sciences

DOCTORAL PROGRAM IN
Molecular Industrial And Environmental Biotechnologies

WITH THE FINANCIAL CONTRIBUTION OF
Fondazione Edmund Mach

Cycle XXVIII – Year 2013

Whole transcriptome profiling and water stress study of *A. donax* by RNA-seq

S.S.D. AGR/07

Coordinator: Prof. Roberto Bassi

Tutor: Prof.ssa Antonella Furini

Doctoral Student: Dott.ssa Yuan Fu

To my lovely Gaia Fusu,

my dear Michele

and my parents

INDEX

Chapter 1 Introduction....1

1.1 Bioenergy, *A. donax* and water stress 1

1.2 Transcriptome and RNA-seq.... 4

1.3 Thesis overview 7

1.4 Reference.... 8

Chapter 2 Fuelling genetic and metabolic exploration of C3 bioenergy crops through the first reference transcriptome of *Arundo* donax L. 11

2.1 Abstract 11

2.2 Introduction 12

2.3 Results and Discussion.... 14

2.3.1 Organ- specific transcriptome assembly of *A. donax* 14

2.3.2 Sequence similarity comparison across biofuel and other Poaceae species 18

2.3.3 Functional Annotation, relative expression and GO enrichment.... 20

2.3.4 Transcriptome assisted metabolic mining: Implications for engineering *A. donax* as model C_3 biomass species 29

2.3.5 Mining of stress-related genes in *A. donax* transcriptome.... 39

2.4 Conclusions 41

2.5 Materials and Methods.... 42

2.5.1 Plant material, RNA extraction and sequencing 42

2.5.2 Quality control and filtering of Illumina transcriptomic read data set 42

2.5.3 454 reads trimming and assembly 43

2.5.4 Construction of organ-specific transcriptomic catalogs 43

2.5.5 Functional annotation and metabolic mining of *A. donax* 44

2.6 References 45

Chapter 3 Dissection of early transcriptional responses to water stress in *Arundo donax* L. by unigene-based RNA-seq 49

3.1 Abstract... 49

3.2 Introduction 50

3.3 Results and Discussion 54

3.3.1 Identification of differentially expressed genes (DEGs) by RNA-Seq 55

3.3.2 Functional characterization of transcriptional responses to water stress in *A. donax* 57

3.3.3 Metabolic pathways related to water stress in *A. donax* 61

3.3.4 Identification of transcription factors responsive to water stress in *A. donax*...... 63

3.3.5 Characterization of co-regulated gene expression network in *A. donax* 67

3.3.6 Identification of a core set of Poaceae genes differentially regulated upon water stress 70

3.4 Conclusion 74

3.5 Materials and Methods 74

3.5.1 Plant material and application of water limitation stress 74

3.5.2 Next generation RNA sequencing 75

3.5.3 Water stress transcriptome annotation 76

3.5.4 Identification and functional classification of differentially expressed genes 76

3.5.5 GO enrichment 77

3.5.6 Pathway enrichment 78

3.5.7 Real-time validation of selected DEG candidates using qRT-PCR 78

3.6 References 79

Chapter 4 Comparative study on early transcriptional drought response between *A. donax* and *P. australis* 86
4.1 Abstract 86
4.2 Introduction 86
4.3 Results and discussion...... 89
4.3.1 Identification and functional characterization of differentially expressed genes (DEGs) of *P. australis* by RNA-Seq 89
4.3.2 General comparison of differentially expressed genes (DEGs) between *P. australis* and *A. donax* 90
4.3.3 Comparison of functional classification and transcriptional factors responsive to water stress...... 92
4.3.4 Comparison of metabolic pathways related to water stress response 95
4.3.5 Differences in expression of orthologous genes in the two species under water stress...... 103
4.3.6 Conserved drought response genes in related Poaceae species 108
4.4 Conclusions 109
4.5 Materials and Methods 109
4.5.1 Plant material and application of water limitation stress 109
4.5.2 Transcriptome reconstruction, differentially expressed genes (DEGs) identification and functional annotation 110
4.5.3 Classification of gene expression patterns...... 110
4.5.4 Pathway enrichment...... 111
4.5.5 Putative homologs...... 111
4.6 Reference...... 111
Chapter 5 Conclusion 115
Future prospects 116

Acknowledgement 117

Appendix 119

INTRODUCTION 1

"The only true wisdom is in knowing you know nothing."

— Socrates

1.

Biomass and bioenergy productions are becoming crucial issues for sustainable development. While there is considerable technological growth in this field, the biological bases of plant metabolism can also provide ground for exceptional improvement of sustainable biomass and bioenergy production. The European Commission (EC) has indicated that biomass will play an important role in the future. Production of biofuels are attracting growing interest around the world and it is now believed that biofuels will substantially help meeting the world's increasing demand for energy. The EU has emphasized with bio-fuels directives (e.g. 2003/30/EC), green papers (e.g. March 2006) and strategic energy review papers (e.g. January 2007) the need to take effective actions to address climate change (including actions to mitigate greenhouse gas emissions), and enhance security of energy supply.

A. donax as a promising energy plant allows new possibility of agro-industrial exploitation of marginal and semi-marginal lands, therefore attenuating current controversial competition for land use with food crops. *Arundo donax L.* is a perennial, C3 grass, which belongs to the Poaceae family. The origin of this species is still not clear, the newest research is on 2014 proposed by Hardion, showed by Amplified fragment length polymorphism (AFLP) analysis that it originated from Asian. Nowadays it is spread in East/west Asia, Mediterranean, south of North America, middle of South America, and on the border of Australia (geographic map: http://eol.org/pages/1114777/maps). The first description which could be found about *A. donax* would be back to 1913 in *Flore de France*. In the following years, more observations were reported. The study of Hardion that focused on revising the systematics of *A. donax* and offered a detailed description at the taxonomic level. "All

qualitative morphological features mentioned in Floras and other literature were observed, and we retained the most discriminant features delimiting taxonomic clusters." (Hardion et al, 2012)

The study about *A. donax* could trace back to 1958, related to diseases (Perdue 1958). Then in the 70s, two more researches about isolation of chemical natural products from *A. donax* appeared, then one more in 1982, and another in 2008, demonstrating that indoles, bulot enidine, alkaloids, and several other secondary metabolites can be isolated from *A. donax* (Bhattacharya and Sanyal, 1972; Ghosal et al, 1972; Wassel 1982; Jia et al, 2008). However, the reserach hot spots on *A. donax* in the last 4 decades focused on its capability of living in polluted water, contaminated land and high biomass yield (Mirza et al, 2010; Helder et al, 2010; Pilu et al, 2012). The strong vitality, even manifested as invasiveness (*A. donax* is among the top 100 most invasive species in the world), make it very suitable for being used in wastewater and pollution treatment (Ahmad et al, 2008). Though it is an invasive species, with the proper cultivation management it could also be safely used as energy crop (Virtue et al., 2010). The high yield of biomass and the problem of global warming indicated it as a promisiong potential energy crop. Early in 1982, the "energy from biomass" report described a "joint research on *Arundo donax* as an energy crop" from three European partners in five aspects: yield, cultivation, survey, cost and suitable areas. In 1997, "Giant reed (*Arundo donax* L.) Network: improvement, productivity and biomass quality" as a sub-project of FAIR was approved, since then, the biomass content of *A. donax* (lignin, cellulose, lignincellulose) has been characterized in detail (Bacher and Sauerbeck, 1997). Furthermore, the conversion technology, efficiency compared with other energy crop has also been reported. From 2001 to 2005, another project "ENK-CT-2001-00524 Bio-energy chains from perennial crops in south Europe (BIO-ENERGY CHAINS)" was completed. The focus of the project was the assessment of ash content by direct combustion and the impact on atmospheric air quality of using *A. donax* as energy crop. Though until now, plenty of studies has been done on understanding/developing the usage of *A. donax*, the genomic resources related to this crop has not been yet developed, which limits the molecular dissection of *A. donax* strong vitality, as well as its evolution and origin; Most importantly, this lack of knowledge ahs till now prevented any attempt for engineering it as a fully domesticated energy crop.

Furthermore, as *A. donax* is a C3 species with an unusually high photosynthetic capacity (Rossa, 1998), plenty of studies about the biomass yield in the Mediterranean area have been carried out (Lewandowski et al, 2003, Angelini et al, 2009, Mantineo et al, 2009). These studies confirmed in practice the high theoretical yield of dry matter production that can be attained using this species. Meanwhile, the fact that *A. donax* is a best player of marginal land exploitation, given that can adapt to a wide variety of ecological conditions, is well suited to warm-temperate to subtropical climates, and to a variety of soil type, it grows from coarse river sands to heavy clays, no matter how infertile, saline soils or contaminated with arsenic, cadmium and lead makes is one of the most promising energy crops (Papazoglou et al, 2005). The major advantages of its ecploitation are: (1) Low input cultivation condition, i.e. less indirect greenhous gases (GHG) indirect emissions; (2) It offers protection against soil erosion; (3) It does not compete for high-quality arable land, thus preventing the food-energy sources competition and the the decrease of food global availability. Nevertheless, drought is the single most important environmental stress for crops, causing severe effects on growth rate and biomass accumulation of *A. donax* and other plants (Farooq et al, 2009). Drought causes more yield losses than any other single biotic or abiotic factor (Boyer, 1982). The European Commission in 2012 reported the increase of the number of people and areas hit by drought between 1976 and 2006 by almost 20% and the total drought losses over the EU in the past 30 years are estimated to be 100 billion euro (Kindler and Okruszko, 2014). It is thus clear that genetic improvement of crops for drought tolerance becomes a reserach priority; it requires investigation of the possible components of drought tolerance and exploration of the genetic variability of the avaiable accessions/germplasms. Despite the high potential of *A. donax* as prospective bioenergy crop, up to the start of this project, just a few sequences (mostly matK and rbcL and cloned SSR markers for phylogenetic and population studies) have been deposited in GenBank for this species. This nearly complete lack of genetic information is a major obstacle for the genetic improvement of *A. donax* (After our work had been published, also another shoot transcriptome study about *A. donax* has been published (Barrero et al, 2015).)

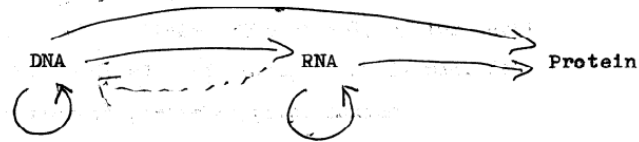
Phragmites australis, as *A. donax*, belongs to the Poaceae family, Arundineae tribe (Linder et al., 1997). [Note: a recent phylogenetic study separated the genera *Arundo* and

Phragmites into two different tribes (Soreng et al, 2015).] *P. australis* is a large perennial grass, 2–6 meters tall, with dry biomass yield 5~20 t h^{-1} y^{-1} (Pilu et al. 2012; Köbbing JF, Thevs N and Zerbe S, 2013). Furthermore, it is also an invasive species. *P. australis* is commonly described as C3 plant; it has been reported also the C3–C4 intermediate ecotype in saline environment (Zheng et al. 2000), which makes it flexible and allows it to compete better to the changing environmental conditions (Srivastava et al. 2012). It perhaps has the largest geographical distribution of any flowering plant in the World: it is occurring on all continents except Antarctica (Brix, 1999 and Clevering and Lissner, 1999). Compared with *A. donax*, *P. australis* prefers wetland areas, such as swamps, lakes and in some cases also the sea. Its roots can grow to extreme lengths, allowing the plant to survive with lower water levels by reaching water deep below ground. Furthermore, *P. australis* reproduces both sexually and asexually and it has a complex population structure (Nguyen et al, 2013). In contrast, *A. donax* has low genetic diversity and it reproduces only via asexual fragments; the molecular study showed that there is only a single genetic clone of *A. donax* in the whole territory of USA.

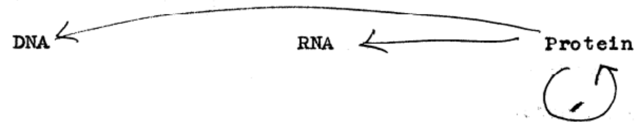
1.2 Transcriptome and RNA-seq

The existence of iscrete inheritable genetic units was demonstrated for the first time in the 19th century by Gregor Mendel with his famous work on pea plants. This started the history of genetics. In 1920, Hans Winkler created the term *genome* to describe a full set of genes and chromosomes in a cell/organism, i.e. the sum of all the genetic information. In 1953, based on X-ray diffraction images of DNA, James Watson and Francis Crick proposed the double helix structure of DNA. From then on, we were lead to another world to biology mechanism in the molecular level. In 1958, Francis Crick first stated the central dogma of molecular biology, as an explanation of the flow of genetic information within a biological system (Crick 1958).

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it. That is, we may be able to have



but never



where the arrows show the transfer of information.

It shows that transcription is the first step of gene expression, in which a particular segment of DNA is copied into RNA by the enzyme RNA polymerase, i.e. RNA is the key intermediate between the genome and the proteome. DNA (Deoxyribonucleic acid) is the carrier of the genetic information in cells. DNA consists of four nucleotides: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T) with fixed base pairing, where A is complementary to T and G is complementary to C. According to central dogma in molecular biology, DNA segments are transcribed to ribonucleic acid (RNA), which is then translated to proteins. RNA also consists of four nucleotides: Adenine (A), Cytosine (C), Guanine (G), and Uracil (U) with fixed base pairing, where A is complementary to U and G is complementary to C. The transcriptome is the set of all RNA molecules, including mRNA, rRNA, tRNA, and other non-coding RNA produced in one or a population of cells. Unlike the genome, which is roughly fixed for a given cell line (excluding mutations), the transcriptome is a dynamic entity which can vary with external environmental conditions. Overall, the key aims of transcriptomics are to catalogue all species of transcripts, the operational units of a genome (Thomas 2007), including mRNAs, non-coding RNAs and small RNAs; to determine the transcriptional structure of genes, in terms of their start sites, 5′ and 3′ ends, splicing patterns and other post-transcriptional modifications; and to quantify the changing expression levels of each transcript during development and under different conditions (Wang et al, 2009).

Thanks to Matthew Meselson and Franklin Stahl work, in the semi conservation of DNA during replicationwas ascertained. This basic mechanism is at the foundation of the majority of the sequencing technologies that would have been developed in the follwoing years. In 1977, the Maxam-Gilbert sequencing and Sanger sequencing technologies allowed for the first time to decode the sequence of bases of any DNA fragment of interest, opening up a new era in the decoding of genomes (Sanger et al, 1977; Maxam and Gilbert, 1977). In the following 30 years, Sanger sequencing, which used the chain terminator method, also known as first generation sequencing now, was broadly used as the golden standard of sequencing. The milestone constituted by the famous draft of the human genome sequence, the first ever decoded, was achieved thanks to Sanger sequencing in 2003. For its outstanding contribution to the development of genomics, Sanger, the inventor of this technology, was awarded the Nobel Prize in 1980. Until now, the Sanger sequencing is still popular, for its accuracy. However, Sanger sequencing has fatal weak points in the limited thorughput and high cost. The human genome sequencing costed several million dollars and took about one decade. Different from Sanger sequencing, the novel parallel, high-throughput sequencing techniques which are also called NGS (next-generation sequencing), provide high speed and reduced costs. To give an example, the human genome sequencing that took several years with the Sanger sequencing can now be completed in a matter of weeks with NGS (Metzker 2010). Nowadays, the main second generation platforms commercially available for NGS are from Illumina/Solexa, Life/APG and Helicos BioSciences. Recently also new technologies known as third generation sequencing appeared, such as Pacific Biosciences SMRT (single-molecule real-time) DNA sequencing method; Life Technologies FRET sequencing platform; The Oxford nanopore sequencing platform (Quail M et al, 2012). (d) The Ion Torrent sequencing platform (Munroe and Harris, 2010), which could generate over 10,000 bp reads or map over 100,000 bp/molecule, unlike second-generation sequencing, which produces short reads a few hundred base-pairs long.

RNA-seq, also known as whole transcriptome shotgun sequencing, uses NGS to show the presence and quantity of RNAs in a biological sample at a certain moment under certain condition (Wang et al, 2009)**.** Compared with microarrays, the prior method for transcriptome study, RNA-Seq is not limeted by the need to know the sequence a priori, therefore, it is suitable also to analysis of species without any genome information. For

instance, many non-model species transcriptome studies on plants have been conducted with RNA-Seq experiment by de novo assembly (e.g. Dorn et al., 2013; Gusberti et al., 2013; Gutierrez-Gonzalez et al., 2013). The RNA-seq approach has reached maturity and several studies showed the manyfold applications it can be used for: (Garcia et al., 2012) quantified mRNA expression in *Fundulus grandis* populations in the marshes and estuaries impacted by Deepwater Horizon oil release; (Der et al., 2011) developed genomic resources for evolutionary studies; (Ashrafi et al., 2012) generated a wealth of genetic information for pepper including thousands of SNPs and Single Position Polymorphic (SPP) markers. Furthermore, RNA-seq is not only a tool for quantitative assessment of RNA but can also be exploratory. A recent study shows that less than 3% of the human genome compose protein coding gene exons, but more than 85% of the genome is transcribed (Hangauer et al, 2013). RNA-seq allows generating a comprehensive catalog of an important class of long non-coding RNA, miRNA, siRNA, and other small RNA classes involved in regulation of RNA stability, protein translation, or the modulation of chromatin states (Trapnell et al, 2010). For instance, Enhancer RNA, a class of short transcript directly transcribed from the enhancer region, have been discovered thanks to RNA-Seq, underlying the contribution that this techniqueis providing, among others, to our knowledge of epigenetic gene regulation (Andersson et al, 2014).

1.3 Thesis overview

In this study, first of all, I characterized the transcriptome of *A. donax* by RNA-Seq, with the primary goal of developing the missing genetic resources that until now hindered the characterization at molecular level and prevented the genetic improvement of *A. donax*. The transcriptome of different tissues of *Arundo* were sequenced (Chapter 2). An independent de-novo assembly was carried out. Secondly, based on the first part of my study, and the existing knowledge of water stress response, I characterized the early water stress transcriptional responses in this species, which as the single most important environmental stress for crops can cause severe effects on plant growth rate and biomass accumulation (Chapter 3). Thirdly, to understand better the water stress responses mechanism of Arundineae tribe, a comparative study with *P. australis* was carried out (Chapter 4).

1.4 Reference

Ahmad, R., Liow, P.-S., Spencer, D. F. & Jasieniuk, M. Molecular evidence for a single genetic clone of invasive Arundo donax in the United States. Aquat. Bot. 88, 113–120 (2008).

Andersson, R. et al. An atlas of active enhancers across human cell types and tissues. Nature 507, 455–461 (2014).

Ashrafi, H. et al. De novo assembly of the pepper transcriptome (Capsicum annuum): a benchmark for in silico discovery of SNPs, SSRs and candidate genes. BMC Genomics 13, 571 (2012).

Bacher, W. & Sauerbeck, G. Giant reed (Arundo donax L.) Network Improvement biomass quality Final report FAIR_CT_96_208. (2001).

Barrero, R. A. et al. Shoot transcriptome of the giant reed, Arundo donax. Data Br. 3, 1–6 (2015).

Bhattacharya, S. K. & Sanyal, A. K. Neuromuscular blocking activity of bufotenidine isolated from Arundo donax L. Naturwissenschaften 59, 650–1 (1972).

Boyer, J. S. Plant Productivity and Environment. Science (80-.). 218, 443–448 (1982).

Brix, H. Genetic diversity, ecophysiology and growth dynamics of reed (Phragmites australis). Aquat. Bot. 64, 179–184 (1999).

Bucci, A., Cassani, E., Landoni, M., Cantaluppi, E. & Pilu, R. Analysis of chromosome number and speculations on the origin of Arundo donax L. (Giant Reed). Cytol. Genet. 47, 237–241 (2013).

Clevering, O. & Lissner, J. Taxonomy, chromosome numbers, clonal diversity and population dynamics of Phragmites australis. Aquat. Bot. 64, 185–208 (1999).

Crick, F.H.C. On Protein Synthesis. Symp. Soc. Exp. Biol. XII, 139-163 (1956).

Der, J. P., Barker, M. S., Wickett, N. J., dePamphilis, C. W. & Wolf, P. G. De novo characterization of the gametophyte transcriptome in bracken fern, Pteridium aquilinum. BMC Genomics 12, 99 (2011).

Dorn, K. M., Fankhauser, J. D., Wyse, D. L. & Marks, M. D. De novo assembly of the pennycress (Thlaspi arvense) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. Plant J. 75, 1028–38 (2013).

Farooq, M., Wahid, A., Kobayashi, N., Fujita, D. & Basra, S. M. A. in Sustainable Agriculture 153–188 (Springer Netherlands, 2009). doi:10.1007/978-90-481-2666-8_12

Fu, Y. et al. Dissection of early transcriptional responses to water stress in Arundo donax L. by unigene-based RNA-seq. Biotechnol. Biofuels 9, 54 (2016).

Garcia, T. I. et al. RNA-Seq reveals complex genetic response to deepwater horizon oil release in Fundulus grandis. BMC Genomics 13, 474 (2012).

Ghosal, S., Chaudhuri, R., Dutta, S. & Bhattacharya, S. OCCURRENCE OF CURARIMIMETIC INDOLES IN THE FLOWERS OF ARUNDO DONAX. Planta Med. 21, 22–28 (1972).

Gingeras, T. R. Origin of phenotypes: genes and transcripts. Genome Res. 17, 682–90 (2007).

Grzelak, M., Gaweł, E., Waliszewska, B., Kaczmarek, Z. & Mocek-Płóciniak, A. Yielding and energy value of reed rush Phragmitetum australis from extensively used meadows. J. Food, Agric. Environ. J. Food Agric. Environ. 1212, 1197–1200 (2005).

Gusberti, M., Gessler, C. & Broggini, G. A. L. RNA-Seq Analysis Reveals Candidate Genes for Ontogenic Resistance in Malus-Venturia Pathosystem. PLoS One 8, e78457 (2013).

Gutierrez-Gonzalez, J. J., Tu, Z. J. & Garvin, D. F. Analysis and annotation of the hexaploid oat seed transcriptome. BMC Genomics 14, 471 (2013).

Hangauer, M. J. et al. Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. PLoS Genet. 9, e1003569 (2013).

Helder, M. et al. Concurrent bio-electricity and biomass production in three Plant-Microbial Fuel Cells using Spartina anglica, Arundinella anomala and Arundo donax. Bioresour. Technol. 101, 3541–3547 (2010).

Jia, A. et al. Journal of Asian Natural Products Research A new indole alkaloid from Arundo donax L. J. Asian Nat. Prod. Res. 10, 105–109 (2008).

Kask, Ü., Kask, L. & Link, S. Combustion characteristics of reed and its suitability as a boiler fuel. 13, 1–10 (2013).

Kindler, J. & Okruszko, T. Integrated Drought Management Programme in Central and Eastern Europe (IDMP CEE). Ann. Warsaw Univ. Life Sci. L. Reclam. 46, 167–179 (2014).
Köbbing, J. F., Thevs, N. & Zerbe, S. The utilisation of reed (Phragmites australis): a review. Mires Peat 13, 1–14 (2013).
Lewandowski, I., Scurlock, J. M. O., Lindvall, E. & Christou, M. The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. Biomass and Bioenergy 25, 335–361 (2003).
Linder, H., Verboom, G. & Barker, N. Phylogeny and evolution in the Crinipes group of grasses (Arundinoideae: Poaceae). Kew Bull. (1997).
Mann, J. J., Barney, J. N., Kyser, G. B. & Di Tomaso, J. M. Miscanthus X giganteus and Arundo donax shoot and rhizome tolerance of extreme moisture stress. GCB Bioenergy 5, 693–700 (2013).
Mantineo, M., D'Agosta, G. M., Copani, V., Patanè, C. & Cosentino, S. L. Biomass yield and energy balance of three perennial crops for energy use in the semi-arid Mediterranean environment. F. Crop. Res. 114, 204–213 (2009).
Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U. S. A. 74, 560–4 (1977).
Metzker, M. L. Sequencing technologies — the next generation. Nat. Rev. Genet. 11, 31–46 (2010).
Mirza, N. et al. Phytoremediation potential of Arundo donax in arsenic-contaminated synthetic wastewater. Bioresour. Technol. 101, 5815–5819 (2010).
Munroe, D. J. & Harris, T. J. R. Third-generation sequencing fireworks at Marco Island. Nat. Biotechnol. 28, 426–428 (2010).
Nguyen, L. X. et al. Photosynthesis of co-existing Phragmites haplotypes in their non-native range: Are characteristics determined by adaptations derived from their native origin? AoB Plants 5, 1–13 (2013).
Papazoglou, E. G., Karantounias, G. A., Vemmos, S. N. & Bouranis, D. L. Photosynthesis and growth responses of giant reed (Arundo donax L.) to the heavy metals Cd and Ni. Environ. Int. 31, 243–249 (2005).
Perdue, R. E. Arundo donax—Source of musical reeds and industrial cellulose. Econ. Bot. 12, 368–404 (1958).
Pilu, R., Bucci, A., Badone, F. C. & Landoni, M. Giant reed (Arundo donax L.): A weed plant or a promising energy crop? African J. Biotechnol. 11, 9163–9174 (2012).
Quail M, Smith ME, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. BMC Genomics, 13:341 (2012).
Rossa, B., Tuffers, a V, Naidoo, G. & von Willert, D. J. Arundo donax L. (Poaceae) - a C-3 species with unusually high photosynthetic capacity. Bot. Acta 111, 216–221 (1998).
Sablok, G. et al. Fuelling genetic and metabolic exploration of C3 bioenergy crops through the first reference transcriptome of Arundo donax L. Plant Biotechnol. J. 12, 554–567 (2014).
Soreng, R. J. et al. A worldwide phylogenetic classification of the Poaceae (Gramineae). J. Syst. Evol. 53, 117–137 (2015).
Srivastava, J., Chandra, H., Nautiyal, A. R. & Kalra, S. J. S. Response of C 3 and C 4 Plant Systems Exposed to Heavy Metals for Phytoextraction at Elevated Atmospheric CO2 and at Elevated Temperature. Environ. Contam. 3–17 (2012). doi:10.5772/1124
Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–5 (2010).
Virtue, J., Reynolds, T. & Malone, J. Managing the weed risk of cultivated Arundo donax L. Seventeenth Australas. Weeds Conf. 176–179 (2010).
Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63 (2009).
Wassel, G. Isolation and Identification of Alkaloids from Arundo donax. Planta Med. 45, 164–164 (1982).
Winkler, H. & Winkler, H. Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche / von Dr. Hans Winkler. (G. FIscher, 1920). doi:10.5962/bhl.title.1460
Zheng, W. J., Zheng, X. P. & Zhang, C. L. A survey

of photosynthetic carbon metabolism in 4 ecotypes of Phragmites australis in northwest China : Leaf anatomy , ultrastructure , and activities of ribulose 1 , 5- bisphosphate carboxylase , phosphoenolpyruvate carboxylase and glycollate oxidase. Physiol. Plant. 110, 201–208 (2000).
Zheng, W. J., Zheng, X. P. & Zhang, C. L. A survey of photosynthetic carbon metabolism in 4 ecotypes of Phragmites australis in northwest China : Leaf anatomy , ultrastructure , and activities of ribulose 1 , 5-bisphosphate carboxylase , phosphoenolpyruvate carboxylase and glycollate oxidase. Physiol. Plant 110, 201–208 (2000).
Zulauf, C. 2012 Drought : Yield Loss , Revenue Loss , and Harvest Price Option. (2012).

FUELLING GENETIC AND METABOLIC EXPLORATION OF C3 BIOENERGY CROPS THROUGH THE FIRST REFERENCE TRANSCRIPTOME OF *ARUNDO* DONAX L.

2

"Life is a system that both stores and processes the information necessary for its own reproduction." — L.L Gatlin

The work described in this chapter was done in collaboration with Dr. Gaurav Sablok. Dr. Sablok performed the further merge analysis with 454 data.

The study has been published: Sablok G, Fu Y, Bobbio V, Laura M, Rotino GL, Bagnaresi P, Allavena A, Velikova V, Viola R, Loreto F, Li M, Varotto C. Fuelling genetic and metabolic exploration of C3 bioenergy crops through the first reference transcriptome of *Arundo donax* L. Plant Biotechnol J. 2014;12:554–67.

2.1 ABSTRACT

The development of inexpensive and highly productive biomass sources of biofuel is a priority in global climate change biology. *Arundo donax*, also known as the giant reed, is recognized as one of the most promising non-food bioenergy crops in Europe. Despite its relevance, to date no genomic resources are available to support the characterization of the developmental, adaptive and metabolic traits underlying the high productivity of this non-model species. We hereby present the first report on the *de novo* assembly of bud, culm, leaf and root transcriptomes of *A. donax*, which can be accessed through a customized BLAST server for mining and exploring the genetic potential of this species (http://10.0.100.66:4567). Based on functional annotation and homology comparison to 48 potential bioenergy plant species, we provide the first genomic view of this so far unexplored crop and indicate the model species with highest potential for comparative genomics approaches. The analysis of the transcriptome reveals strong differences in the enrichment of the gene ontology categories and the relative expression among different organs, which can guide future efforts for metabolic engineering of *A. donax*. A set of homologs to key

genes involved in lignin, cellulose, starch, lipid metabolism and in the domestication of other crops is discussed to provide a platform for possible enhancement of productivity and saccharification efficiency in *A. donax*.

2.2 Introduction

Global warming, increasing energy demands and dwindling fossil energy reserves have presented the scientific community with the urgent need for finding sustainable alternative resources of renewable energy (Ohlrogge et al., 2009). Recently, according to the directive of the European Union Energy Strategy 2020, the intensification of research on biomass, which accounts for 56% of the renewable energy supply (Bentsen and Felby, 2012), has been proposed. Four crop species, namely Miscanthus spp. (C4), Phalaris arundinacea (C4), Arundo donax (C3) and Panicum virgatum (C4), have been proposed as the most promising candidates among biomass herbaceous plants (Lewandowski et al., 2003). *A. donax* L., commonly called giant reed, is a perennial rhizomatous grass belonging to the family Poaceae with cosmopolitan distribution (Pilu et al., 2012). Both phylogenetic placement and geographical origin of *A. donax* are still unclear. The hypothesis that *A. donax* originated in East Asia (Polunin and Huxley, 1965) has been questioned for decades due to the many historical traces of this species in the Mediterranean area (Perdue, 1958). Recently, AFLP data suggested a monophyletic origin of the species and provided support for its Asian origin followed by colonization of the Mediterranean (Mariani et al., 2010). Outside of its native distribution range, *A. donax* was reported to be invasive, because its dense and fast-growing stands outcompete native species (Ahmad et al., 2008). On the other hand, its fast growth is one of the main features, which makes *A. donax* a promising biomass species. Comparative studies, in fact, indicate *A. donax* as the most productive nonfood biomass species reported so far in the Mediterranean area (Lewandowski et al., 2003), with an average aboveground dry matter yield of about 40 tons per hectare, which is comparable or, in some cases, higher than that of C4 species (Angelini et al., 2009). The reported high productivity of *A. donax* has recently been associated with its peculiar canopy architecture that together with a particularly high leaf area index allows a very efficient absorption of incident light (Ceotto et al., 2013). In addition, *A. donax* has low requirements in terms of fertilizer and pesticide input and soil management (Raspolli Galletti et al., 2013) and is highly tolerant to

heavy metals and saline soils (Calheiros et al., 2012; Papazoglou, 2007). Furthermore, the giant reed is perennial and maintains high productivity for about a decade (Pilu et al., 2012). Despite the high potential of *A. donax* as prospective bioenergy crop, to date only 47 sequences (mostly matK and rbcL and cloned SSR markers for phylogenetic and population studies) have been deposited in GenBank for this species. This nearly complete lack of genetic information is a major obstacle for the genetic improvement of *A. donax*.

The recent advent of next-generation sequencing (NGS) has made the development of genomic resources progressively simpler and cheaper (Liu et al., 2012). RNA-Seq is to date by far the most powerful tool for the rapid and inexpensive development of genomic resources for any species of interest, thus practically abolishing the divide between model and nonmodel species (Martin et al., 2013). This is especially relevant for crops like the giant reed that, given their complete lack of sexual reproduction (Balogh et al., 2012), cannot be subjected to conventional breeding. Three main approaches are, in principle, possible to select improved *A. donax* cultivars: (i) identification of natural variation among ecotypes, (ii) large-scale mutagenesis programmes with chemical, physical or targeted methods or (iii) genetic engineering through stable transformation and regeneration. Each of these improvement strategies has constraints limiting their practical application in vivo. *A. donax* natural genetic variation seems to be moderate in the Mediterranean area and higher in Asia, but to date ecotypic variation that could be agronomically exploitable has not been reported (Mariani et al., 2010). Large-scale mutagenesis in *A. donax*, although feasible, is hindered by the sterility and high ploidy level of this species (2n = 108−110, corresponding to a putative pseudo-triploid; Hardion et al., 2012), which would make mutant segregation unfeasible and could constitute a formidable phenotypic buffer in forward genetics screenings, respectively. Even in diploid Arabidopsis thaliana, in fact only a minority of gene knockouts display a phenotype (Bouché and Bouchez, 2001). Given the availability of transformation and regeneration protocols (Dhir et al., 2010; Takahashi et al., 2010), genetic engineering is possibly the most promising option for the improvement of *A. donax* with regard to biomass yield and fermentation efficiency. Both mutagenesis-based reverse genetics (e.g. tilling; Slade et al., 2005) and transformation approaches would greatly benefit from the availability of transcriptomic data sets representative of the entire gene space of this species. The sensitivity of RNA-Seq in resolving transcript isoforms and their expression

levels would be particularly suited also for targeted mutagenesis, which provides an efficient method to generate mutations in duplicate genes (e.g. Curtin et al., 2011). A transcriptome for this species, thus, would enable further functional genomics studies, which, in turn, are the key to dissect the mechanisms underlying the evolution history and the adaptive strategies of *A. donax* (Morozova and Marra, 2008).

We hereby provide the first reference transcript catalog for the biomass and bioenergy plant *A. donax* using tissue-specific NGS of four different organs (leaf, culm, bud and root). Our data represent a comprehensive reference catalog of transcripts in terms of sequence and relative depth of coverage to aid the ongoing elucidation of biomass production in grasses in the omic's era. The availability of the first transcript catalog for *A. donax* will aid in functional and comparative genomics efforts aimed at characterizing and improving the spatial and temporal patterns of expression underlying the high productivity of biomass crops in general and *A. donax* in particular. A BLAST server is made available to the scientific community to support the mining of the gene space in *A. donax* (http://ecogenomics.fmach.it/arundo/).

2.3 Results and Discussion

2.3.1 Organ- specific transcriptome assembly of *A. donax*

Prior to *de novo* assembly, the reads were assessed for quality metrics. A total of 42 806 797 (root), 45 191 660 (leaf), 47 423 847 (culm) and 40 990 813 (bud) reads were trimmed for low-quality bases. Additionally, a quality filtering (threshold = Q20) was performed to eliminate base composition bias and to ensure accurate base calling. Finally, 38 889 012 (root), 39 367 790 (leaf), 44 926 387 (culm) and 38 716 996 (bud) high-quality reads were subsequently used for downstream *de novo* organ-specific transcript assembly. Assembling *de novo* transcriptomes with high breadth of coverage and full-length transcript reconstruction is still a computationally challenging task, especially in case of nonmodel polyploid plant species (Schliesky *et al*., 2012). Among the strategies that have been proposed, de-Bruijn's graph approach is widely accepted as the method of choice for Illumina reads (Pevzner *et al*., 2001). The widely used Trinity assembler has been reported to perform full-length transcripts reconstruction more efficiently than Velvet, another popular assembler based on de-Bruijn's graph theory (Grabherr *et al*., 2011); however, both

theoretical considerations and experimental validation suggest that the restriction of using a single *k-mer* (*k* = 25) may potentially introduce chimeric assemblies and may not be able to cover the whole breadth of expression present in the transcriptome. On the other hand, Velvet/Oases provides a *K-mer* selection criterion, thus allowing a better representation of the diversity of transcript isoforms using the dynamic error removal algorithm implemented for RNA-Seq, especially in cases where noise in the library sequencing could perturb the graph topology (Surget-Groba and Montoya-Burgos, 2010).

Trinity assembly (SK; Figure 2.1) resulted in 222 927 transcripts, 104 323 components in leaf (N50 = 1809 bp); 279 143 transcripts, 122 307 components in culm (N50 = 1791 bp); 254 828 transcripts, 127 878 components in root (N50 = 1607 bp); 259 505 transcripts, 117 103 components in bud (N50 = 1682 bp). Preliminary assessment of the components suggests the presence of high isoform diversity in this species. The ploidy level of *A. donax* in the Mediterranean region is 2n = ca. 108–110 (~18x), corresponding to a proposed pseudotriploid genetic background (Hardion *et al*., 2011, 2012). As each mRNA locus can potentially express multiple allelic isoform transcripts, a parallel transcript assembly using multiple k-mers was also constructed using Velvet/Oases (Schulz *et al*., 2012). Oases was further used to merge the contigs into longer transcripts isoforms (*transfrags*) using paired-end information. On the basis of the observed N50 and to remove false-positive transcript assemblies while preserving long transcripts, k-mer = 51 was used. Several parameters such as number of used reads, nodes, transcripts longer than 200 bp, as well as N50 and of longest contig length, were evaluated to assess the distribution of transcripts assembled at varying k-mer lengths (Supplementary Ch2.1). Previous empirical estimation of the optimization strategies has clearly demonstrated the ability of the Oases-MK strategy to reduce the number of false-positive (chimeric) transcripts (Zhao *et al*., 2011). It is worth to note that we were able to assemble the full-length putative homologs to the auxin transport protein *BIG* (4965 aa, UniProtKB/Swiss-Prot: B9G2A8.1) in all four organ-specific transcriptomes, indicating that the applied strategy was able to assemble and preserve the longer transcripts (Zhao *et al*., 2011).

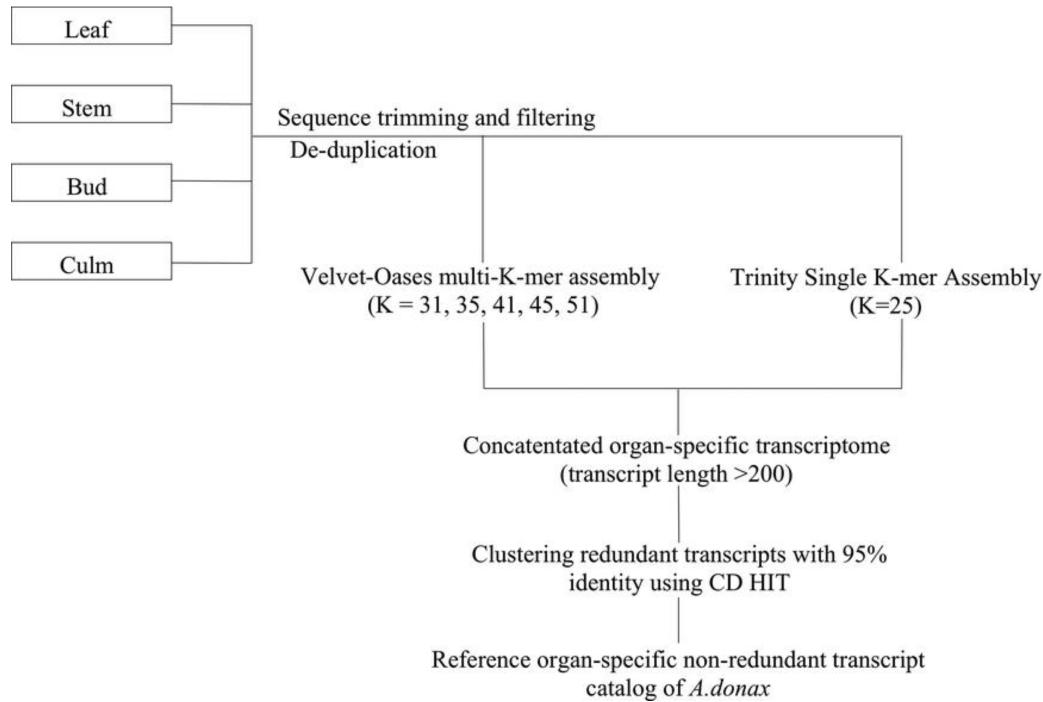


Figure 2.1. Overview of transcript assembly strategy implemented for the organ-specific transcriptome of *Arundo donax*.

To evaluate the effect of putative sequencing errors associated with sequencing, error correction was performed using SEECER (Le *et al*., 2013). For each organ, the set of error-corrected reads were further assembled using Trinity to assess the impact of sequencing errors on transcriptome construction. In total, in root, 38 889 012 reads assembled in 253 530 transcripts with a N50 of 1579 bp, in stem 38 716 996 reads assembled into 256 032 transcripts with a N50 of 1654 bp, in leaf 39 367 790 reads assembled into 220 027 transcripts with a N50 of 1787 bp and in culm 44 926 387 reads assembled into 274 725 transcripts with a N50 of 1765 bp, suggesting that error correction does not have a demonstrated effect on the improvement of transcript assembly in *A. donax*. To remove redundancy and potential artifacts due to the ploidy level of *A. donax*, transcript assemblies (SK and MK) were further merged using CD-HIT-EST ($n = 10$, $c = 95\%$) for each organ-specific transcriptome. A non-redundant data set of 261 130 transcripts in leaf (N50 = 2229 bp), 301 455 in root (N50 = 2125 bp), 328 443 in culm (N50 = 2021 bp) and 304 778 in bud (N50 = 2129 bp) was finally used for functional annotation. The observed N50 is higher as compared to the previously published N50 values in Poaceae like *Zea mays* (1612 bp; Schliesky *et al*., 2012) and *Phragmites australis* (1187 bp; He *et al*., 2012).

To evaluate assembly consistency, we assessed the number of reads mapped back to transcriptome (RMBT; Zhao *et al.*, 2011). In each organ, high percentages of RMBT were found to be in proper orientation (Table 2.1). Only a fraction of the aligned reads were found to be improper pairs. The observed number of mapped reads is in accordance with previous transcriptomics reports in polyploid species and supports effective transcript assembly (Nakasugi *et al.*, 2013). Due to the lack of a reference genomic sequence in *A. donax*, putative homologs were searched in the recently sequenced rhizome transcriptome of *P. australis*, the closest Poaceae species for which RNA-Seq data are available (He *et al.*, 2012; Linder *et al.*, 1997). Of the 118 327 transcripts available for *P. australis*, 88 278 (74.60%) showed putative hits in the *A. donax* transcriptome. Additionally, 454 reads from normalized culm (825 369) and root (776 467) libraries were quality-filtered and assembled using MIRA and CAP3. In total, 154 338 (114 306 126 bp) and 129 404 unigenes (86 170 851 bp) were obtained in culm and root, respectively. Using BLAST, we observed that a high fraction (83.9% for root and 85.5% for culm) of 454 unigenes mapped onto the Illumina transcriptome (454 unigene coverage >95%). Finally, an evaluation of the GC content of the assembled transcripts was conducted. The average GC content of *A. donax* transcripts (45.6%) is higher than in *A. thaliana* (42.5%). However, the GC content was relatively low as compared to *Oryza sativa* (51.30%), *Z. mays* (51.14) and *Setaria italica* (52.75%), but perfectly in line with the GC content of *P. australis* (45.1%). Taken together, these results indicate that in case of *A. donax*, high stringent filtering of the reads and merging assemblies from the SK and MK strategies provided a comprehensive transcriptomic view of *A. donax*, as previously proposed in other polyploid species (Duan *et al.*, 2012; Surget-Groba and Montoya-Burgos, 2010).

Organ	**Proper pairs**	**Left read only**	**Right read only**	**Improper pairs**	**Total aligned reads**
Bud	53 602 220 (84.97%)	3 140 412 (4.98%)	4 173 499 (6.62%)	2 166 866 (3.43%)	63 082 997

Culm	62 983 056 (85.19%)	3 513 923 (4.75%)	4 738 637 (6.41%)	2 696 488 (3.65%)	73 932 104
Leaf	56 738 414 (86.47%)	2 780 752 (4.24%)	3 868 534 (5.90%)	2 229 532 (3.40%)	65 617 232
Root	54 092 932 (85.73%)	2 914 700 (4.62%)	4 037 667 (6.40%)	2 049 054 (3.25%)	63 094 353

Table 2.1. Percentage distribution of the reads mapped back to the transcriptome (RMBT) in organ transcriptome of *A. donax*.

2.3.2 Sequence similarity comparison across biofuel and other Poaceae species

Comparative approaches are a powerful tool to pinpoint analogies and differences in the molecular bases of adaptive traits across different plant species and families (Weckwerth, 2011). In the case of *A. donax*, these approaches could provide valuable sets of candidate genes to be exploited for targeted reverse genetics or for genetic engineering of high value traits. To assess sequence conservation of the *A. donax* transcript assembly across biomass and biofuel species, BLASTx homology searches were run against the BFGR database available at (http://bfgr.plantbiology.msu.edu/). The highest number of homologs (Figure 2.2) was retrieved in *Saccharum officinarum* (285296), followed by *Triticum aestivum* (239411 transcripts), *Panicum virgatum* (192248) and *Hordeum vulgare* (110954). *S. officinarum* (sugarcane) has been proposed as model bioenergy species because of its efficient use of solar energy, carbon fixation and ethanol yield per hectare (Tammisola, 2010). Considering the high transcriptome and phenotypic similarity among the two species, we suggest that sugarcane domestication and yield-related traits could provide a rich source of transgenes and indicate genes of molecular and physiological relevance to be used as targets for genetic engineering of *A. donax*.

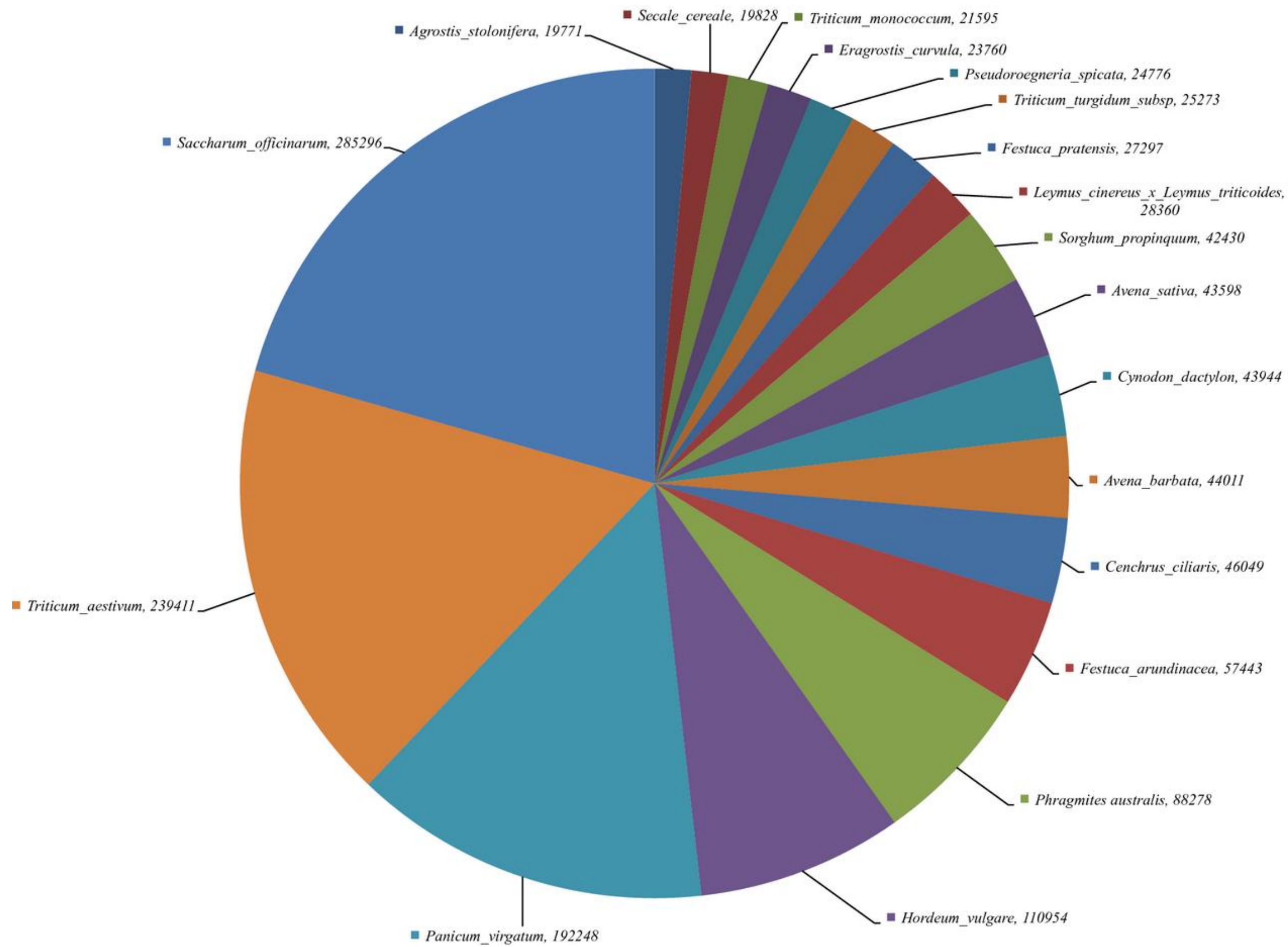


Figure 2.2. Distribution of transcript homologs across biofuel and other Poaceae species.

Previously, it has been demonstrated that C_4 crops like *Miscanthus* represent an ideal source of biomass on the basis of key traits such as efficient conversion of solar light into biomass and energy, high water and nitrogen usage efficiency (Taylor *et al*., 2010). We performed BLASTx searches against the recently published transcriptome of *Miscanthus sinensis* to identify putative homologs in C_4 species (Swaminathan *et al*., 2012). A total of 24180 (80.77%) transcripts of *M. sinensis* showed putative hits in the *A. donax* transcript catalog, thus laying the foundations for a comparative exploration of the high biomass production of giant reed. In the long run, we envision that a deeper understanding of *A. donax* physiology and the availability of a reference transcriptome could support attempts for the engineering of C_4 traits in this C_3 species to further enhance its biomass potential (Gowik and Westhoff, 2011). To identify the coding sequence conservation, CDS were systematically downloaded for *O. sativa* (39049), *S. bicolor* (27607) and *Z. mays* (39656) from Phytozome (http://www.phytozome.net/). Reciprocal best blast hit (RBH) searches revealed high sequence conservation across *S. bicolor* (25475, 92.27%), *Z. mays* (24351, 61.40%), and *O. sativa* (23076, 59%). Compared to the lower conservation from previous reports using transcriptomics reads for gene discovery (28.4% and 29.5%; Garg *et al*., 2011), the high coding sequence conservation in our study indicates that it will be possible to use the extensive resources available for closely related model crop species (in particular *S. bicolor* and *Z. mays*) to define the patterns of gene splicing and its functional relevance in *A. donax*.

2.3.3 Functional Annotation, relative expression and GO enrichment

Functional annotation of the assembled transcripts provide an important way to characterize functionally relevant metabolic pathways and to identify genes involved in traits of economic and ecological importance. Previous studies have highlighted legions of genes and their functional relevance to the plant traits (e.g. Suzuki *et al*., 2012; Gu *et al*., 2012; Góngora-Castillo *et al*., 2012). The complete lack of functional information about metabolic pathways in *A. donax* motivated us to perform an in-depth functional annotation of its transcriptome. BLASTx searches identified homologues of about 40-45% of *A. donax* transcripts in SwissProt and Genbank's NR protein databases. Most of the functional hits correspond to *Sorghum bicolor* and *Zea mays* proteins, two Poaceae species with fully

sequenced genomes. The classification of Gene Ontology obtained from annotation of all transcripts associated with at least one ontology term is shown in Figure 2.3a. GO categories were assigned to root (Biological Process, 39,811; Cellular Component, 44,199; Molecular Function, 43,971), bud (Biological Process, 38,869; Cellular Component, 42,526; Molecular Function, 42,549), leaf (Biological Process, 37,667; Cellular Component, 41,915; Molecular Function, 41,408), culm (Biological Process, 35,666; Cellular Component, 38,659; Molecular Function, 39,159). A high percentage of transcripts contained domains representing more than 50% of the transcript length. In total, 99,860 domains in root, 100,481 in bud, 86,580 in leaf, and 39,159 in culm were functionally categorized. In total, 117915 transcripts in root, 105447 in leaf, 127790 in culm and 118107 in bud were found to have at least one functional associated GO annotation. Previously, similar observations on the relatively high number of transcripts without functional assignments have been reported in several polyploid species such as *Nicotiana benthamiana* (Faino *et al*., 2012; Nakasugi *et al*., 2013), *Nicotiana tabacum* (Bombarely *et al*., 2012), *Castanea dentata* and *C. mollissima* (Barakat *et al*., 2009), and *Solanum tuberosum* (Massa *et al*., 2011). Analysis of functionally annotated transcripts indicates that most of the transcripts putatively associated with one of the functional domains and GO were >500 nt in length, suggesting that shorter transcripts contributed only marginally to the functional proteome diversity in *A. donax* and might represent non-coding RNAs, in line with previous observations in polyploidy species (Nakasugi *et al*., 2013).

Expression levels can be used as predictors of the functional relevance of specific groups of genes in controlling metabolic fluxes, although downstream layers of regulation (e.g. post-transcriptional gene silencing and various types of post-translational modifications in response to stresses; Khraiwesh et al., 2012; Barrero-Gil and Salinas, 2013) can further control the activity of specific gene products. For example, secondary metabolism is controlled at the transcriptional level through the action of transcription factors specialized to concertedly activate enzymes participating in the same biosynthetic pathway (Yang et al., 2012). It has been also widely demonstrated that transcriptional regulation plays a fundamental role in determining the concerted activity of nuclear and chloroplast-encoded components of the photosynthetic machinery through redox-mediated retrograde signalling

(Foyer et al., 2012); especially in polyploid species, a detailed understanding of the relative expression levels of paralogous or homeologous gene copies is also fundamental to maximize the chances to design effective knockout and overexpression strategies for functional studies or metabolic engineering (Chandler and Werr, 2003). We therefore identified the most highly expressed transcripts in the *A. donax* transcriptome to provide a ranking of genes with a putatively large effect on the metabolism of this species. We observed a high number of properly paired reads multimapped to the assembled transcriptome in root (97.41%), bud (97.57%), leaf (98.24%) and culm (97.87%). A stringent threshold of FPKM values (cut-off = 1) was implemented to assess the relative number of expressed transcripts in each assembled organ. A total of 173 994 (57.71%) transcripts in root, 143 850 (55.08%) in leaf, 164 044 (49.94%) in culm and 172 281 (56.52%) in bud were expressed with FPKM value greater than 1.00 (Supplementary Ch2.2). The percentage of transcripts having FPKM value below the detection limit (FPKM = 0.00) was 14.76% in bud, 15.22% in culm, 14.6% in leaf and 13.42% in root, suggesting an overall good coverage of the assembled transcriptome.

Additionally, to explore differences in transcriptional patterns across organs, we applied a stringent threshold of FPKM (>100) to have a unimodal calibrated distribution model of enriched GO categories. About 700 transcripts were found to have FPKM values above the threshold and were associated with different GO functional categories depending on the organ (Supplementary Ch2.3). Classification of the organ-specific enrichment of GO functional categories as a whole indicates that the highly expressed transcripts provide a comprehensive representation of the most relevant metabolic pathways active in *A. donax* (Figure 2.3b; Supplementary Ch2.3). In particular, the leaf transcriptome showed the most abundant overenriched GO categories (221; FDR < 0.01). Among the observed categories, a significant amount of GO terms were enriched in the process of carbon fixation and related biosynthetic pathways, namely starch, pigments, lipids and isoprenoids biosynthesis, in agreement with the high photosynthetic activity of this organ reported by previous studies (e.g. Nakasugi et al., 2013) (Supplementary Ch2.4).

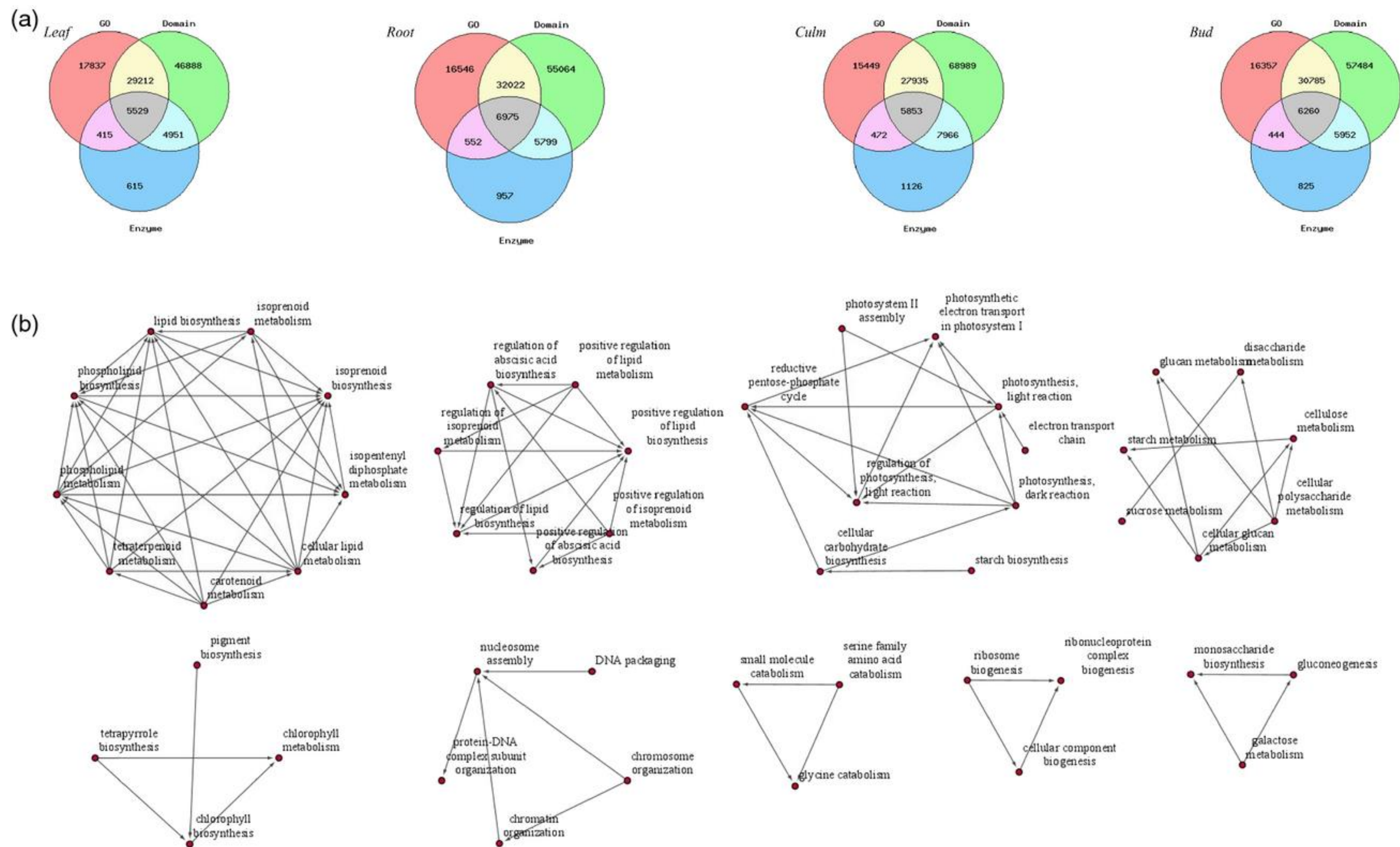
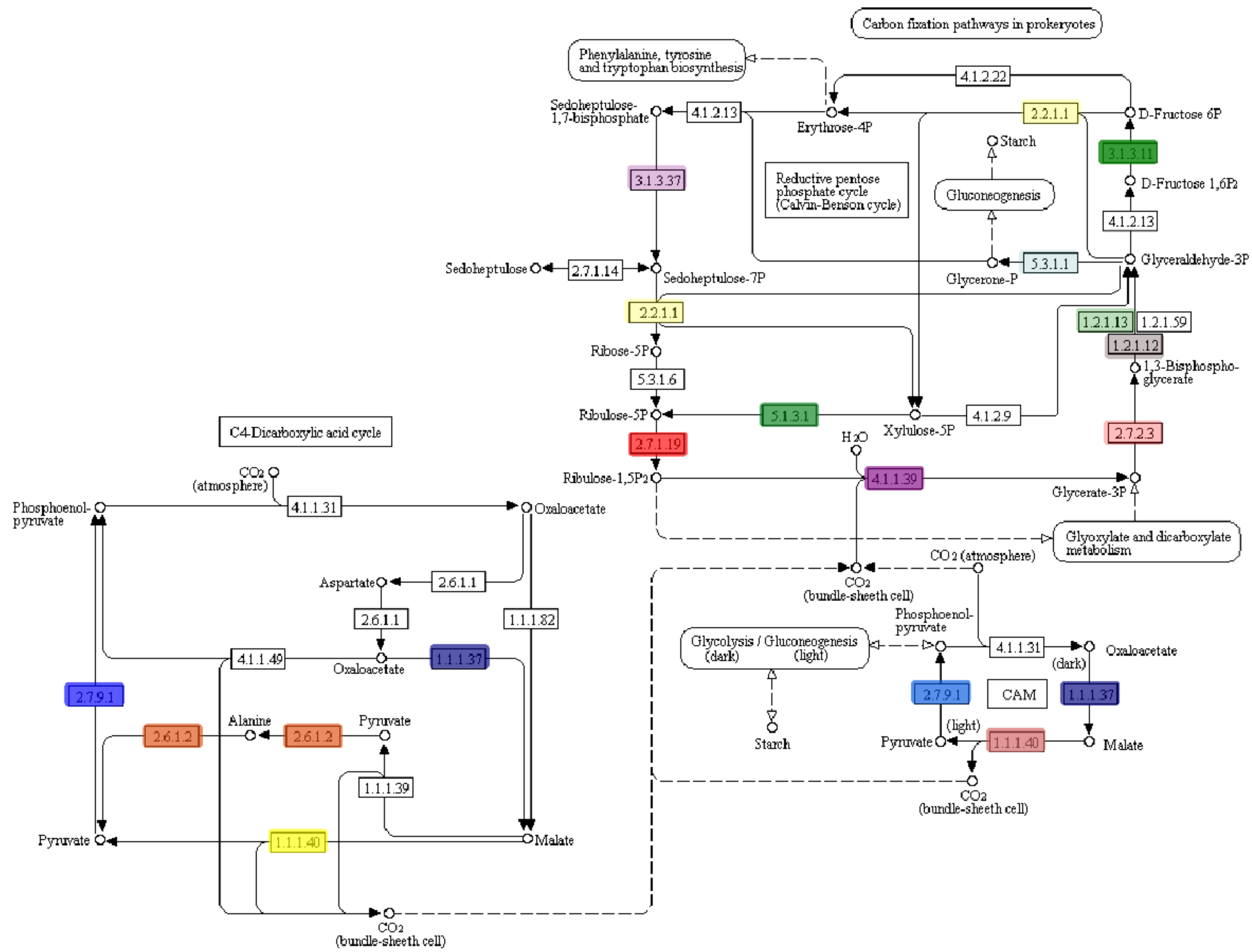


Figure 2.3. Gene Ontology classification of Arundo donax transcripts: (a) functionally annotated GO categories, enzymes and domains in each organ-specific transcriptome. (b) Attribute circle layout representation of the merged over-represented GO categories across the four organs based on the Fischer test and Benjamini and Hochberg false discovery rate correction (0.05%).

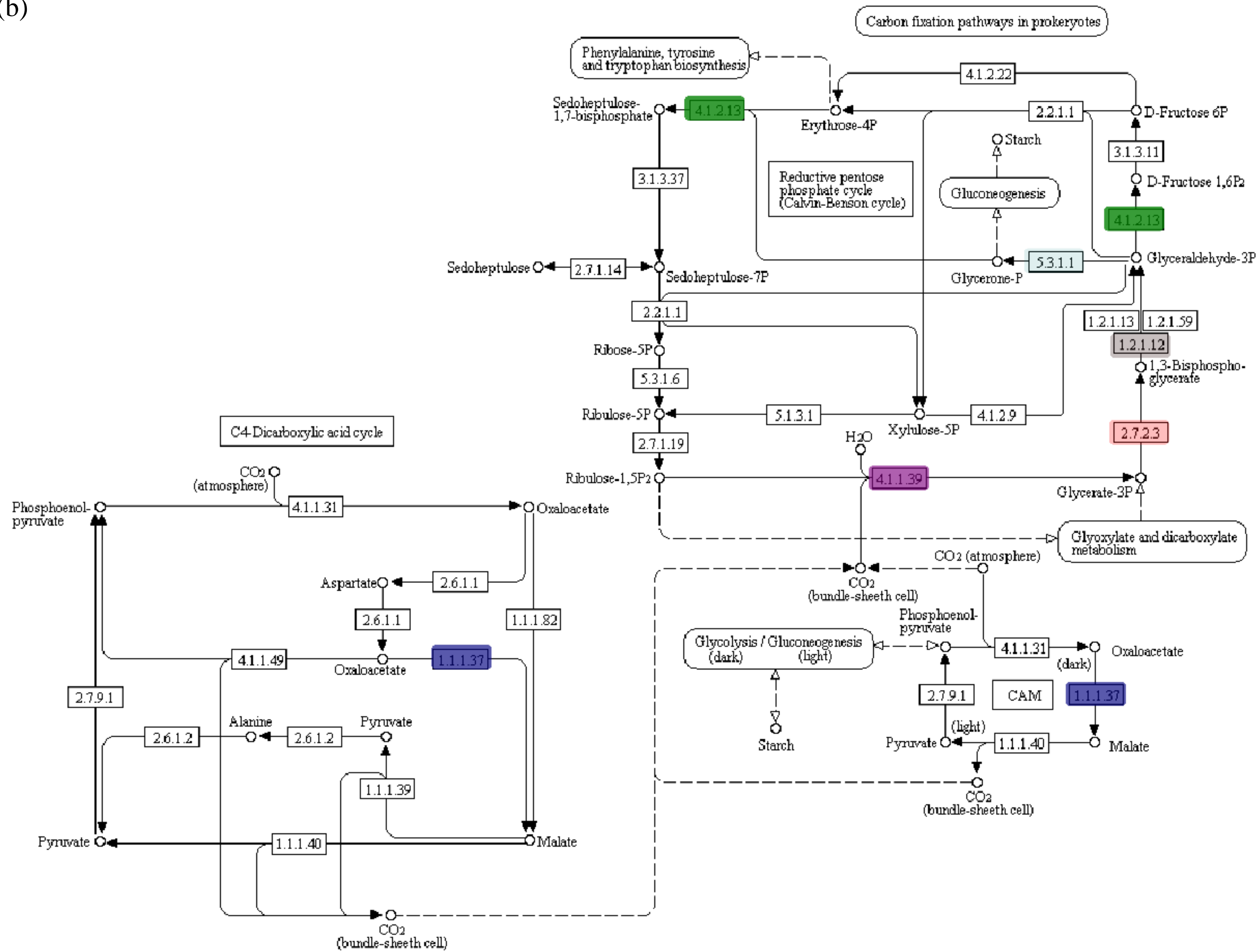
On the contrary, GO terms related to protein synthesis and DNA and amino acids metabolism were significantly under-represented in leaf (59 in total). The bud displayed the lowest number of over- or under-represented GO terms (10 and 4, respectively), with an enrichment of categories associated with active cell replication and protein synthesis typical of developing tissues. In root, only a few GO terms (14, transcription or translation) were found significantly over-represented, while the main feature of this organ is the under-representation of terms associated with photosynthesis and light response. Finally in culm, among the 30 overenriched GO terms, cell wall biogenesis and carbohydrate metabolism (mainly cellulose and sucrose) were the most common. On the contrary, under-represented GO terms (22 in total) were associated with light response and photosynthesis. GO-enriched categories were further mapped in KEGG pathways to identify variations in expression patterns across organs with putatively functional roles in the metabolism of *A. donax*.

A comparative overview of the KEGG maps of the photosynthetic carbon fixation and starch biosynthesis for the leaf and the culm is shown in Figure 2.4. We observed that the majority of metabolic genes involved in C_3 carbon fixation were expressed above the threshold value of FPKM=100 in the leaf transcriptome. In contrast, in the culm transcriptome, the majority of the genes in the Calvin–Benson cycle, responsible for the dark phase of carbon assimilation, were expressed below the threshold, indicating a lower photosynthetic activity in this organ. Also, the expression patterns of the starch biosynthetic pathway support this conclusion, with high expression of starch synthase (EC 2.4.1.21) in leaf, whereas in culm key enzymes involved in sucrose, glucose, fructose and cellulose metabolism were found to be expressed above threshold. Taken together, these results indicate that *A. donax* culm likely acts as sink for the photosynthates produced in leaves, in agreement with previous reports indicating that, before flowering, culm represents one of the major sinks for photosynthetic carbon in grasses (Castleden *et al*., 2004; Okawa *et al*., 2002). Based on the relevance that a high sink strength of culms had in the success of sweet sorghum and sugarcane as bioenergy crops (Calviño and Messing, 2012; Wang *et al*., 2013), enhancing this trait in *A. donax* could be a major progress towards its genetic improvement.

(a)



(b)



(c)

(d)

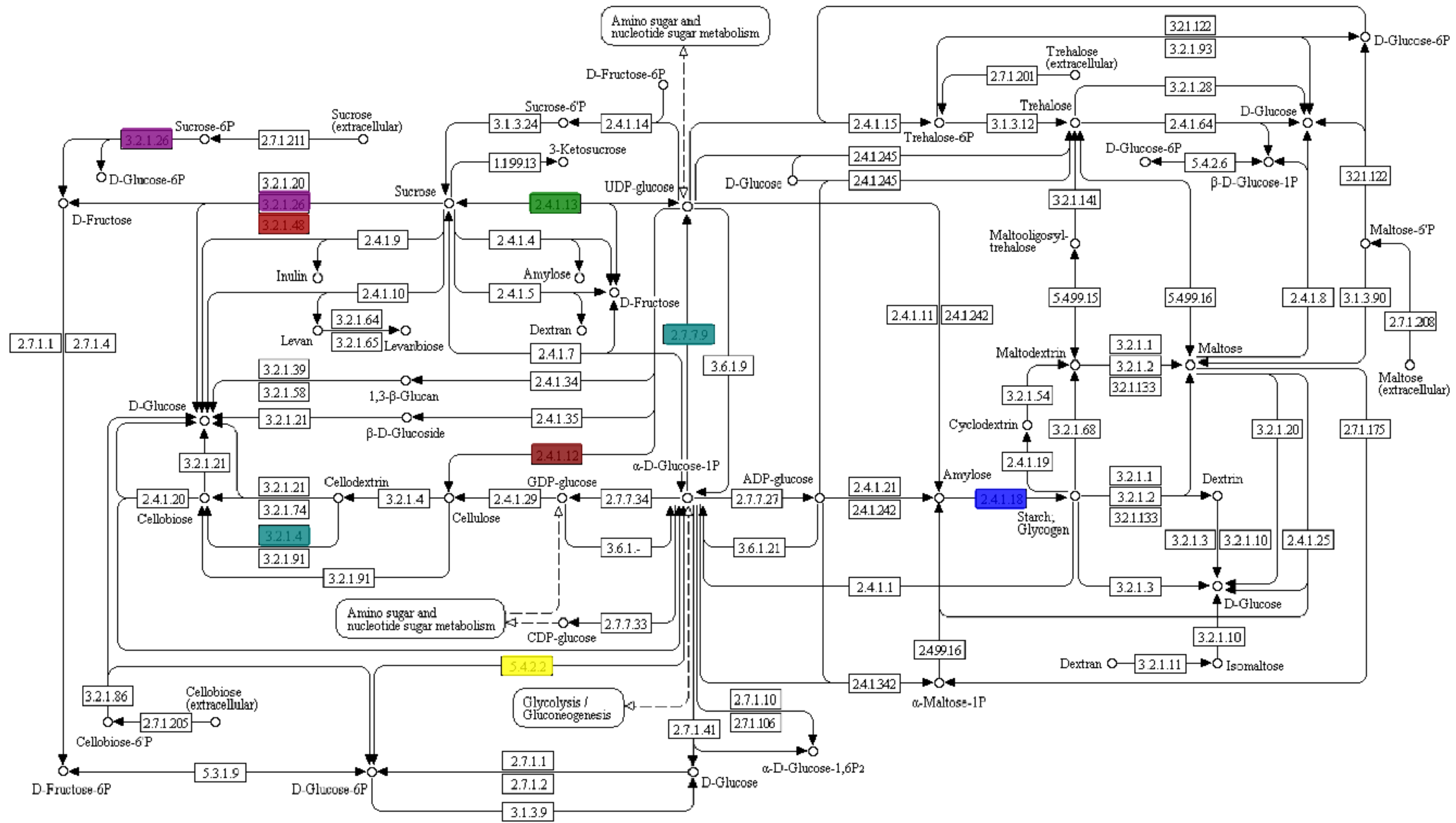


Figure 2.4. Selected KEGG pathways illustrating transcripts involved in photosynthetic carbon fixation and starch/sucrose metabolism which are highly expressed in leaf (a), (c) and culm (b), (d), respectively.

Interestingly, the extremely high content of xylose in cell walls of *A. donax* (up to 27% of the total sugars of nodes, Neto et al., 1997) correlates well with the up-regulation in culms of UDP-glucose 6-dehydrogenase (EC 1.1.1.22) and UDP-glucuronate carboxy-lyase (EC 4.1.1.35), the two enzymes committing UDP-glucose to hemicellulose biosynthesis. The different pattern of expression with respect to leaves suggests that transcriptional regulation can play an important role in the regulation of UDP-glucuronate carboxy-lyase, in addition to the allosteric regulation previously reported (John et al., 1977). It is also worth to note that among the most highly expressed transcripts identified, several code for rate-limiting enzymes of different metabolic pathways (e.g. phytoene synthase, EC 2.5.1.32, for carotenoids; Lindgren et al., 2003; NADH-dependent glutamate synthase, EC 1.4.1.14, for nitrogen metabolism; Baron et al., 1994; ADP-glucose pyrophosphorylase, EC 2.7.7.27, in starch synthesis; Thorbjornsen et al., 1996).

2.3.4 Transcriptome assisted metabolic mining: Implications for engineering *A. donax* as model C_3 biomass species

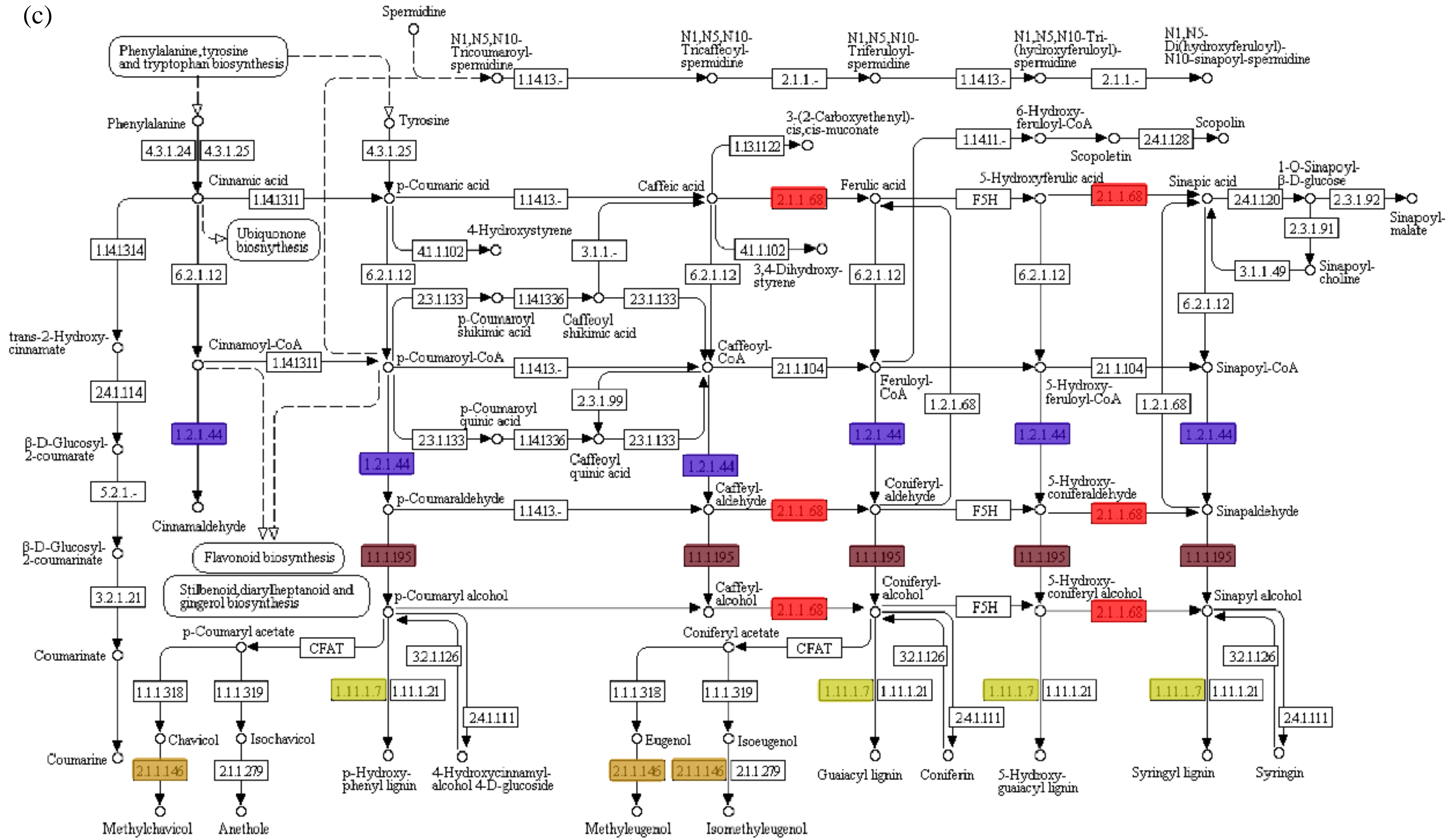
A pathway of high economical relevance in bioenergy crops is lignin biosynthesis, as content of this polymer in cell walls is inversely related to yield and conversion efficiency of polysaccharides into ethanol (Jung et al., 2012). Down-regulation of lignin content by transgenic approaches indeed resulted in improved cell wall digestibility in both monocotyledonous and dicotyledonous plant species (e.g. Brachypodium distachyon, Bouvier d'Yvoire et al., 2013; alfalfa, Chen and Dixon, 2007). In the *A. donax* transcriptome, several homologs of phenylpropanoid biosynthetic genes were highly expressed in the four organs studied (Figure 2.5). Interestingly, despite the overall similarity of the expression patterns, organ-specific up-regulation of different biosynthetic steps was observed (e.g. cinnamyl-alcohol dehydrogenase, CAD, EC 1.1.1.195, was specifically up-regulated in culm and leaf, while cinnamoyl-CoA reductase, CCR, EC 1.2.1.44, was up-regulated in leaf and root). It might be possible that the higher fraction of p-hydroxyphenyl units reported in *A. donax* culms as compared to leaves is at least in part due to the higher expression levels in the former organ of caffeoyl-CoA O-methyltransferase, COMT (EC 2.1.1.104), that redirects the flux of precursors towards syryngil (S) and guaiacyl (G) units (Figure 2.5). Indeed, COMT mutants of S. bicolor display a decrease in S residues content (Bout and Vermerris,

2003). As a further confirmation of the potential utility of the *A. donax* transcriptome as a tool to identify candidate genes for engineering, two of the three most promising mutations affecting lignin content in S. bicolor and B. distachyon (brown midrib mutants; Sattler et al., 2010; Trabucco et al., 2013) are among the most highly expressed genes in the culm data set (COMT and CAD).

(a)

(b)

(c)



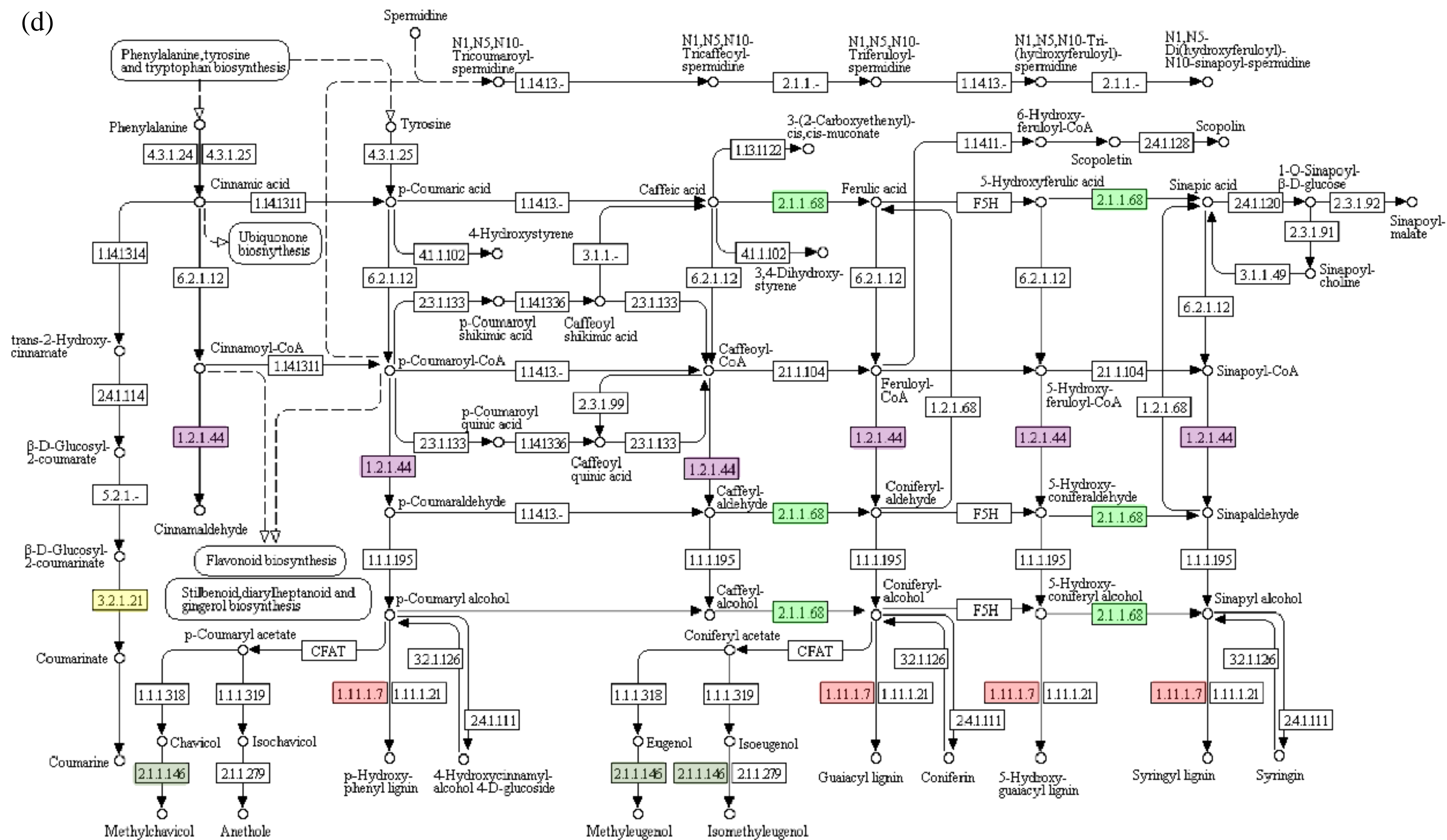


Figure 2.5. Comparative assessment of the abundance of the highly expressed genes in lignin metabolism across the bud (a), culm (b), leaf (c) and root (d).

Despite its relevance, lignin biosynthesis is only one of the possible targets for improvement of the lignocellulosic biomass in grasses. Recently, cellulose and noncellulosic polysaccharides have been proposed as promising alternatives for improving biomass potential (Carpita, 2012; Pauly and Keegstra, 2010). In order to assess the potential utility of our transcriptome as a resource to characterize cell wall biogenesis and organization in *A. donax*, we carried out a comprehensive identification of *A. donax* homologs of cell wall-related genes previously validated in Z. mays and O. sativa (Guillaumie et al., 2007). Besides the homologs of the phenylpropanoid pathway discussed above, Figure 2.6 provides the overview of several other gene families related to cell wall components. Notably, we identified 15 transcripts homologous to sucrose synthase (EC 2.4.1.13), a key enzyme in sucrose biosynthesis, whose overexpression increases cellulose content without negative effects on growth in P. trichocarpa (Coleman et al., 2009). Cellulose synthase (EC 2.4.1.29) is another enzyme playing a pivotal role in cell wall biogenesis. The recent identification of mutations in subunits of the cellulose synthase complex reducing crystallinity and improving cell wall digestibility with limited effects on plant fitness indicates that this enzyme could be another important target for engineering improved saccharification in *A. donax* (Harris et al., 2009, 2012).

Additional gene families worth of note are those involved in the biosynthesis and modification of hemicellulose, the second most abundant polysaccharide in cell walls (Saha, 2003). Several putative glycosyl transferases have been identified in the transcriptome of *A. donax* that may participate in hemicellulose biosynthesis. The functional characterization of these genes could potentially lead to simplification of the cell wall lignocellulosic network and to improvement of saccharification, as recently demonstrated in both P. trichocarpa and A. thaliana (Lee et al., 2009). We believe that the functional study of the gene families mentioned above and of others equally relevant that were not discussed for brevity could significantly contribute to further our limited understanding of cell wall structure and function in *A. donax* and provide a promising set of candidates for its tailoring for bioenergy production.

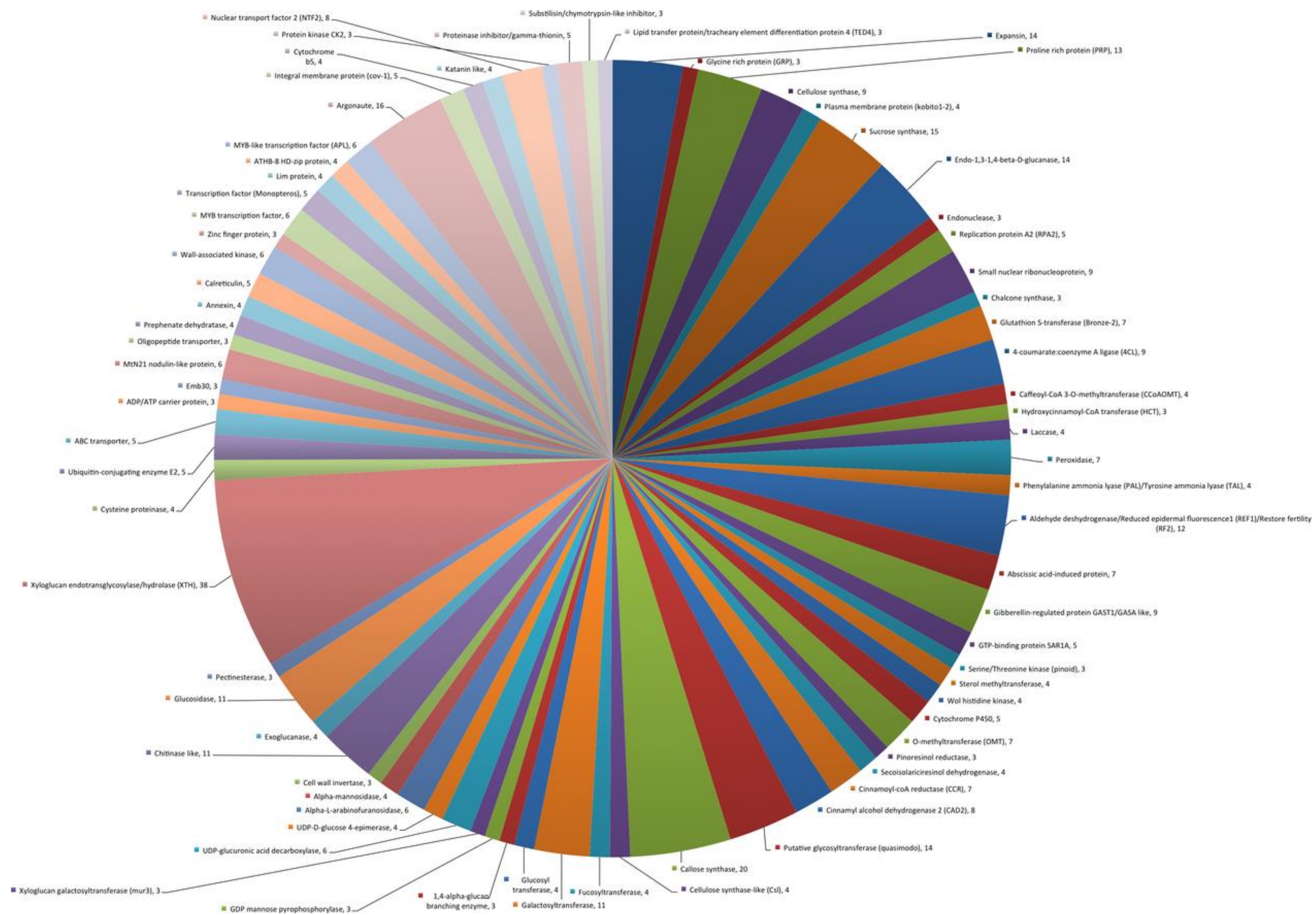


Figure 2.6. Overview of distribution of *A. donax* transcript homologs associated with cell wall biogenesis and organization.

Despite the majority of the studies carried out till now for improvement of bioenergy crops focused on the lignocellulosic component of biomass, the lipidic fraction of green biomass could also be an attractive target as it stores twice as much energy than cellulose per unit of weight (Ohlrogge et al., 2009). We mined the *A. donax* transcriptome by performing BLASTx searches to identify the putative homologs in the entire set of A. thaliana lipid pathway genes available at http://aralip.plantbiology.msu.edu/pathways/pathways. We identified homologs for 91% of the genes (transcript coverage >70%, identity 50%–95%) cataloging the entire lipid-related transcriptome in *A. donax* (Table 2.2). Recently, in vivo packaging of triacylglycerols has been demonstrated to enhance significantly the leaf biomass and energy density profile in A. thaliana (Winichayakul et al., 2013). Several key enzymes such as triacylglycerol lipase (EC 3.1.1.3), diacylglycerol kinase (EC 2.7.1.107), galactolipid galactosyltransferase (EC 2.4.1.184) and N-acetylglucosaminyltransferase (EC 2.4.1.141) involved in glycerophospholipid and lipid metabolism were found to be highly expressed (FPKM > 100). In particular, galactolipid galactosyltransferases play important roles in chloroplast lipid remodelling during various abiotic stresses and ensure optimal functioning of photosynthetic complexes (Moellering and Benning, 2011). Despite the relatively good knowledge available on the lignocellulosic components of *A. donax* aboveground biomass (see e.g. You et al., 2013), to date only one report about the lipid composition of its stems has been published (Coelho et al., 2007). Further studies on the lipid content of all organs for which the transcriptome is now available will constitute a precious basis of information to provide further insights into lipid accumulation and its potential for biofuel production in *A. donax*.

Acyl lipid pathway	**Number of genes in pathway (*A. thaliana*)**	**Number of *A. donax* transcripts**
Plastidial fatty acid synthesis	42	422
Beta oxidation	21	225
TAG synthesis and storage	41	295
Eukaryotic phospholipid synthesis	43	535
Plastidial glycerolipid, galacto-	28	232

lipid and sulfolipid synthesis		
TAG degradation	26	159
Phospholipase; lipase annotation	99	820
Lipid acylhydrolase; lipase	10	81
GDSL; lipase	84	404
Galactolipid degradation; lipase	7	12
Mitochondrial fatty acid and lipoic acid synthesis	13	111
Mitochondrial phospholipid synthesis	8	95
Lipid trafficking	5	44
Sphingolipid synthesis	26	272
Lipid signalling	125	840
Miscellaneous: lipid related	37	225
Fatty acid elongation and cuticular wax synthesis	26	109
Cutin synthesis	26	144
Cuticular wax synthesis	80	401
Aliphatic suberin synthesis	28	248
Aromatic suberin synthesis	6	128

Table 2.2. Distribution of transcript abundance in the lipid biosynthetic pathway available from http://aralip.plantbiology.msu.edu/pathways/pathways. Note: Transcripts having the identity from 50% to 95% with the transcript coverage more than 70% of the query.

Finally, we identified among the genes that played a relevant role in the domestication of other crop species those that could be useful for the domestication of the giant reed (Table 2.3 and references therein). Besides some metabolic genes related to starch and lipid biosynthesis (discussed above), the majority of the other genes are known to affect plant architecture and development. The genes *Corngrass1* (*cg1*), *Heading date 1* (*hd1*), and *Heading date 6* (*hd6*) affect flowering time and could be useful to enhance biomass accumulation by extending the vegetative phase of *A. donax*. In particular, overexpression of *cg1* has recently been demonstrated to have also an effect on starch accumulation in transgenic switchgrass, thus indicating that it could be a promising candidate also for *A.*

donax transformation (Chuck *et al*., 2011). Genes related to *teosinte branched 1* (*tb1*), the most famous example of domestication gene for maize, have also been identified in *A. donax* transcriptome. Together with *barren stalk1* (*ba1*), *tb1* could provide a powerful tool to explore the effect of different plant architectures on biomass yield, while genes like *Reduced height* (*rht*) and *Semidwarf 1* (*sd1*) could be used to modulate plant height optimizing it for maximal productivity.

Gene name	Molecular function	Species	Number of transcripts	Ref
AE1, BT2	Carbohydrate composition. Biosynthetic enzyme	Zea mays	55	Doebley et al. 2006
CG1	Flowering time and carbohydrate composition. microRNA	Zea mays	1	Chuck et al. 2011
EHD1, HD1	Flowering time. Transcription factor	Oryza sativa	25	Doebley et al. 2006
RHT	Plant height. Transcription factor	Triticum aestivum	10	Doebley et al. 2006
SD1	Plant height. Biosynthetic enzyme	Oryza sativa	17	Doebley et al. 2006
SH2, SU1, WAXY	Carbohydrate composition. Biosynthetic enzyme	Zea mays, Oryza sativa	35	Doebley et al. 2006
TAG1	Lipid composition. Biosynthetic enzyme	Arabidopsis thaliana	21	Winichayakul et al. 2013
TB1	Plant and inflorescence structure. Transcription factor	Zea mays	3	Doebley et al. 2006
VRN1, VRN2	Vernalization. Transcription factor	Triticum aestivum	15	Doebley et al. 2006

Table 2.3. Identification of *A. donax* transcript homologs to crop domestication genes

2.3.5 Mining of stress-related genes in *A. donax* transcriptome

To identify transcripts putatively involved in stress in *A. donax*, we used as reference the comprehensive resource provided by the Arabidopsis Stress Responsive Gene Database (ASRGDB; Borkotoky et al., 2013). For reliable identification of putative homologs, BLASTx searches were carried out with sequence identity >70% and sequence coverage >70%, which resulted in the identification of homologs for 148 of the abiotic stress genes in the ASRGDB. The most represented classes of genes having transcript homologs were those for salt, oxidative, metal ion, drought and osmotic stress (Supplementary Ch2.5).

The genes related to salt and heavy metal tolerance are of particular interest, as tolerance of *A. donax* to these stresses holds some promise to exploit the species in marginal soils not suitable for less-tolerant food or bioenergy crops (Ceotto and Di Candilo, 2010). Excess of sodium chloride (NaCl) causes reduced availability of water to plant cells, interferes with the homeostasis of essential elements such as potassium (K) and increases the intracellular formation of reactive oxygen species (ROS), but a broad range of other responses ranging from enhanced protein folding, post-translational modification and protein redistribution have been implicated in salt tolerance as well (Sobhanian et al., 2011). Among the most interesting candidates previously associated with salt stress response, we identified homologs of salt-inducible aquaporins (PLASMA MEMBRANE INTRINSIC PROTEIN 3, SA00108, and PLASMA MEMBRANE INTRINSIC PROTEIN 5, SA0099); several genes involved in the signal transduction cascade (e.g. SNF1-SNF1-RELATED PROTEIN KINASE 2.4, SA0012; CALCIUM-DEPENDENT PROTEIN KINASE 2, SA0024; SHAGGY-LIKE KINASE 42, SA0037); molecular chaperons facilitating correct protein folding (HEAT SHOCK PROTEIN 60-2, SA0059; MITOCHONDRIAL HEAT SHOCK PROTEIN 70-1, SA00111); and enzymes involved in ROS scavenging (e.g. ASCORBATE PEROXIDASE 1, SA0006; MONODEHYDROASCORBATE REDUCTASE 2, SA00113) . We additionally carried out a curated search for genes belonging to the NHX (sodium/hydrogen exchanger) and HKT (high-affinity K+ transporter) gene families, which are the main players in NaCl sequestration and in the maintenance of cytoplasmic ion homeostasis (Cotsaftis et al., 2012; Davenport et al., 2005; Horie et al., 2009; Pires et al., 2013; Platten et al., 2006). This resulted in the identification of transcripts homologs to five NHX and four HKT genes (Supplementary Ch2.5), which are promising candidates for the dissection of the molecular mechanisms underlying salt tolerance in *A. donax*.

We also identified several transcripts homologs to ASRGBD genes involved in metal stress, in particular to cadmium, copper, iron, magnesium and aluminium ions (Supplementary Ch2.5). In addition, we carried out a curated search based on literature of the most important genes involved in zinc and nickel stress response, as well as of metallothionein genes (Supplementary Ch2.5). BLASTx searches identified *A. donax* transcripts homologous to genes involved in the two major mechanisms related to metal stress responses: chelation and transport. In monocots, metal ions can be chelated by a

multiplicity of molecules, like histidine, nicotianamine and its derivative mugineic acid, glutathione, phytochelatins, metallothioneins and organic acids like ascorbate (Sinclair and Krämer, 2012). We found homologs of ATP-phosphoribosyl transferase, the rate-limiting step in histidine biosynthesis, which has been demonstrated to play a pivotal role in nickel tolerance (Ingle et al., 2005) and of PHYTOCHELATIN SYNTHASE, the last step of the phytochelatin biosynthetic pathway (Sinclair and Krämer, 2012). Transcript homologs for metallothioneins, low molecular weight protein-chelating metals that seem to be implicated at the same time in ROS scavenging (Hassinen et al., 2011), were also identified, which might indicate a role of metallothioneins in contributing to *A. donax* metal tolerance. We further mined homologs of several transporters for metal ions complexed with nicotinamide (YELLOW STRIPE1-LIKE genes; Gendre et al., 2007) or citrate (FERRIC REDUCTASE DEFECTIVE 3; Durrett et al., 2007), indicating that chelation with these molecules can also contribute to metal ion homeostasis in *A. donax*. Finally, homologs of the main transport system for cadmium ions across membranes were identified, ranging from HEAVY METAL ATPase 2 and 3 (HMA2, HMA3), responsible, respectively, for cadmium xylematic loading and accumulation in vacuoles, to NATURAL RESISTANCE-ASSOCIATED MACROPHAGE PROTEIN 1 and 5 (NRAMP1, NRAMP5), two cadmium transporters of rice (Sinclair and Krämer, 2012). However, additional analyses of these putative homologs will be required to ascertain their involvement in metal ion homeostasis and their possible relevance for heavy metal tolerance in *A. donax*.

2.4 Conclusions

The present study provides the first comprehensive analysis of the organ-specific transcriptome of *A. donax*, identifying several expressed genes that could be preferential targets for functional studies, for metabolic engineering or for tailoring growth habit/development of the giant reed to higher bioenergy yield. The genomic data developed will provide a much awaited resource to further our limited understanding of the biology of *A. donax* and constitute the first of the three pillars (transcriptomics, proteomics and metabolomics; Liberman et al., 2012) upon which an integrative systems biology characterization of this promising bioenergy crop can be built in the future.

2.5 Materials and Methods

2.5.1 Plant material, RNA extraction and sequencing

Four different organs were collected for Illumina sequencing from mature *A. donax* plants: (i) fully expanded, nonsenescing leaves (4th and 5th from the top), (ii) sections of culm including 4th and 5th nodes and the corresponding lateral buds (each section about 7 cm long), (iii) roots and (iv) dormant buds originating from the pachyrhizome. In addition, 454 sequencing was carried out for culms and root samples. Plant material was ground in liquid nitrogen with precooled mortars and pestles followed by RNA isolation using the Spectrum Plant Total RNA Extraction Kit (Sigma-Aldrich, St. Louis, MO) according to the manufacturer's instructions. Integrity of the isolated RNA was visually checked on agarose gels. Quality checks were carried out using the RNA 6000 Pico kit and the Agilent Bioanalyser 2100 (Agilent, Santa Clara, CA). Four independent paired-end libraries were prepared using the TruSeq RNA Sample Prep V2 kit (Illumina, San Diego, CA), pooled in equimolar ratio and were sequenced on an Illumina HiSeq2000 sequencer (The Genome Analysis Center, Norwich, UK). Two normalized culm and root libraries were prepared by Eurofins MWG Operon (Ebersberg, Germany) using 454 adapters A and B. Sequencing was performed on a Genome Sequencer GS FLX Titanium Instrument (454 Life Sciences, a Roche company, Branford, CT).

2.5.2 Quality control and filtering of Illumina transcriptomic read data set

All sequencing reads were processed further for quality assessment and removal of low-quality bases before transcriptome assembly. In brief, FastQC analysis was performed using the FASTQC version 0.10.1 (downloaded from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). For every organ, sequence reads were trimmed on the basis of the quality score from the 5′ and the 3′ end using fastx_trimmer (http://hannonlab.cshl.edu/fastx_toolkit/index.html). In all the organ-specific libraries, we maintained a quality score of Q20 for downstream analysis. We further excluded reads containing N's using in-house scripts. After de-duplication of the remaining reads using the fastx_collapser (http://hannonlab.cshl.edu/fastx_toolkit/index.html), fastq files were re-synchronized using custom Perl scripts.

2.5.3 454 READS TRIMMING AND ASSEMBLY

Multiplex identifiers (MIDs) and contaminating primer sequences were trimmed from 454 reads using custom Perl scripts. We applied an additional trimming using a cut-off of 2 ambiguous bases in a 50-bp sliding window. Reads shorter than 100 bp were discarded. After cleaning, a total of 766 525 reads in culm and 696 903 in root were used for de novo assembly using the Overlap-Layout-Consensus assembler MIRA (parameters: job = denovo, est, accurate, 454 using the -notraceinfo option) and CAP3 (Chevreux et al., 2004; Huang and Madan, 1999). Parameters not mentioned elsewhere were kept as default parameters.

2.5.4 CONSTRUCTION OF ORGAN-SPECIFIC TRANSCRIPTOMIC CATALOGS

In our assembly pipeline, a combination of single k-mer (SK) and multi-k-mer (MK) assembly was used. In the MK approach, each organ-specific set of reads (from leaf, root, bud and culm) was separately assembled using de-Bruijn's graph based on de novo genome assembler Velvet-1.2.08 (Zerbino and Birney, 2008) using k-mer values of K = 31, 35, 45 and 51. Contigs assembled using Velvet were further constructed into longer transcripts using the insert size as an argument in Oases-0.2.08 (Schulz et al., 2012), which clusters the contigs produced by Velvet into transcript isoforms (transfrags) utilizing the paired-end information. The transcriptomes assembled at different k-mers were further merged using Oases–MK into a non-redundant transcript catalog for each organ. To avoid the creation of the false positives and to increase the sensitivity of the assembly, K = 51 was finally used for merging. In a parallel SK approach, de novo transcripts were also assembled using the Trinity Assembler (Grabherr et al., 2011) using the default k-mer K = 25. MK and SK assemblies were performed on a 64-core AMD-CPUs server with 128 GB of RAM. Assembly contiguity statistics were calculated using custom Perl scripts.

Transcript assemblies from Velvet/Oases and Trinity (minimum transcript length = 200 bp) were clustered at 95% identity with a word size of 10 using CD-HIT v4.5.4 (Fu et al., 2012) merging the clusters with alignment overlap greater than 95% identity to generate an organ-specific transcript catalog of *A. donax*. To evaluate the effect of the sequencing errors on the transcriptome assembly, reads after quality filtering were error-corrected using SEECER, which uses hidden Markov model (HMM)-based probabilistic error corrections

(Le et al., 2013). For each organ, SK assemblies obtained from error-corrected and from the original sets of reads were compared. Assembly evaluation was carried out assessing the percentage of reads mapping back uniquely to the non-redundant set of the organ-specific transcripts using Bowtie (Langmead et al., 2009). Visualization of read mapping was carried out using the Tablet assembly viewer (Milne et al., 2010). For all organs, we quantified the relative expression of each transcript in terms of FPKM (fragments per kilobase of transcript per million mapped reads) using RSEM (Li and Dewey, 2011).

2.5.5 Functional annotation and metabolic mining of *A. donax*

Functional insights into the non-redundant organ-specific transcriptomes of *A. donax* were obtained by performing BLASTx searches (E-value, $1 \times 10{-}3$) against the publicly available GenBank nr (www.ncbi.nlm.nih.gov), UniProt (www.uniprot.org) and TAIR10 (www.tair.org) databases. Functional annotation and Gene Ontology were obtained using FastAnnotator, which employs a four-way classification approach utilizing Blast2GO and additionally sequence homology search by BLAST against NCBI nr, Gene Ontology (GO) term assignment with default annotation rule parameters, InterProScan (IPS) identification of functional motifs, merging of BLAST-based and IPS-based GO annotations and augmentation by Annex (Götz et al., 2008), PRIAM and RPS BLAST (Ashburner et al., 2000; Chen et al., 2012). Additionally, for each organ, the GO categories significantly enriched among highly expressed transcripts based on the Fischer test and Benjamini and Hochberg false discovery rate correction (0.05%) were selected. cytoscape (version 3.01; Smoot et al., 2011) was used to create attribute circle layout merged GO annotation maps for the highly expressed transcripts (FPKM > 100) displaying over-represented GO terms in each organ. We further created custom databases from A. thaliana, O. sativa, S. bicolor, H. vulgare and Z. mays of genes, which are specifically involved in the biosynthesis of lipids, storage polysaccharides (starch) and cell wall components (lignin, cellulose, pectin) and could be potentially used for *A. donax* domestication and metabolic engineering. To assess transcript homolog coverage and to understand sequence conservation across potential and established biofuel species, we retrieved the reference non-redundant transcript catalogs of 19 Poaceae bioenergy species from the BFGR database (Childs et al., 2012). In addition, transcripts of M. sinensis (Swaminathan et al., 2012) and *P. australis* (He et al., 2012) were

also retrieved. All putative homologs were searched using BLASTx with an E-value cut-off of 1 × 10−3 and with sequence coverage greater than 50% of the subject length. Domestication genes were identified using TBLASTN with an E-value cut-off of 1 × 10−3 and with sequence identity greater than 50% against the *A. donax* transcript assembly.

To identify transcript homologs putatively involved in stress, customized databases were made by retrieving all the proteins corresponding to the HKT1, NHX, GST, metallothioneins and heavy metal-associated gene families from UniProt, NCBI, TrEMBL. In addition, we also retrieved the experimentally characterized stress response genes from the Arabidopsis Stress Responsive Gene Database available at http://srgdb.bicpu.edu.in/ (Borkotoky et al., 2013). BLASTx searches were made with an e-value cut-off of 1 × 10−3, and all hits with at least 70% identity and coverage to the query were retained.

2.6 References

Ahmad, R., Liow, P.-S., Spencer, D. F. & Jasieniuk, M. Molecular evidence for a single genetic clone of invasive Arundo donax in the United States. Aquat. Bot. 88, 113–120 (2008).

Ahmad, R., Liow, P.-S., Spencer, D. F. & Jasieniuk, M. Molecular evidence for a single genetic clone of invasive Arundo donax in the United States. Aquat. Bot. 88, 113–120 (2008).

Angelini, L. G., Ceccarini, L., Nassi o Di Nasso, N. & Bonari, E. Comparison of Arundo donax L. and Miscanthus x giganteus in a long-term field experiment in Central Italy: Analysis of productive characteristics and energy balance. Biomass and Bioenergy 33, 635–643 (2009).

Ashburner, M., Ball, C., Blake, J. & Botstein, D. Gene Ontology: tool for the unification of biology. Nat. … 25, 25–29 (2000).

Balogh, E., Herr, J. M., Czakó, M. & Márton, L. Defective development of male and female gametophytes in Arundo donax L. (POACEAE). Biomass and Bioenergy 45, 265–269 (2012).

Barakat, A. et al. Comparison of the transcriptomes of American chestnut (Castanea dentata) and Chinese chestnut (Castanea mollissima) in response to the chestnut blight infection. BMC Plant Biol. 9, 51 (2009).

Barrero-Gil, J. & Salinas, J. Post-translational regulation of cold acclimation response. Plant Sci. 205–206, 48–54 (2013).

Bentsen, N. S. & Felby, C. Biomass for energy in the European Union - a review of bioenergy resource assessments. Biotechnol. Biofuels 5, 25 (2012).

Bouché, N. & Bouchez, D. Arabidopsis gene knockout: phenotypes wanted. Curr. Opin. Plant Biol. 4, 111–7 (2001).

Bout, S. & Vermerris, W. A candidate-gene approach to clone the sorghum Brown midrib gene encoding caffeic acid O-methyltransferase. Mol. Genet. Genomics 269, 205–14 (2003).

Bouvier d'Yvoire, M. et al. Disrupting the cinnamyl alcohol dehydrogenase 1 gene (BdCAD1) leads to altered lignification and improved saccharification in Brachypodium distachyon. Plant J. 73, 496–508 (2013).

Calheiros, C. S. C. et al. Use of constructed wetland systems with Arundo and Sarcocornia for polishing high salinity tannery wastewater. J. Environ. Manage. 95, 66–71 (2012).

Calviño, M. & Messing, J. Sweet sorghum as a model system for bioenergy crops. Curr. Opin. Biotechnol. 23, 323–329 (2012).

Carpita, N. C. Progress in the biological synthesis of the plant cell wall: new ideas for improving biomass

for bioenergy. Curr. Opin. Biotechnol. 23, 330–337 (2012).

Ceotto, E. et al. Comparing solar radiation interception and use efficiency for the energy crops giant reed (Arundo donax L.) and sweet sorghum (Sorghum bicolor L. Moench). F. Crop. Res. 149, 159–166 (2013).

Chandler, J. W. & Werr, W. When negative is positive in functional genomics. Trends Plant Sci. 8, 279–285 (2003).

Chen, F. & Dixon, R. a. Lignin modification improves fermentable sugar yields for biofuel production. Nat. Biotechnol. 25, 759–761 (2007).

Chen, T.-W. et al. FastAnnotator--an efficient transcript annotation web tool. BMC Genomics 13, S9 (2012).

Childs, K. L., Konganti, K. & Buell, C. R. The Biofuel Feedstock Genomics Resource: a web-based portal and database to enable functional genomics of plant biofuel feedstock species. Database (Oxford). 2012, bar061 (2012).

Chuck, G. S. et al. Overexpression of the maize Corngrass1 microRNA prevents flowering, improves digestibility, and increases starch content of sw

Chuck, G. S. et al. Overexpression of the maize Corngrass1 microRNA prevents flowering, improves digestibility, and increases starch content of switchgrass. Proc. Natl. Acad. Sci. U. S. A. 108, 17550–5 (2011).

Coelho, D., Marques, G., Gutiérrez, A., Silvestre, A. J. D. & del Río, J. C. Chemical characterization of the lipophilic fraction of giant reed (Arundo donax) fibres used for pulp and paper manufacturing. Ind. Crops Prod. 26, 229–236 (2007).

Coleman, H. D., Yan, J. & Mansfield, S. D. Sucrose synthase affects carbon partitioning to increase cellulose production and altered cell wall ultrastructure. Proc. Natl. Acad. Sci. U. S. A. 106, 13118–23 (2009).

Curtin, S. J. et al. Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. Plant Physiol. 156, 466–73 (2011).

Dhir, S., Knowles, K., Pagan, C. L. & Mann, J. Optimization and transformation of Arundo donax L. using particle bombardment. African J. Biotechnol. 9, 6460–6469 (2010).

Duan, J., Xia, C., Zhao, G., Jia, J. & Kong, X. Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. BMC Genomics 13, 392 (2012).

Faino, L., de Jonge, R. & Thomma, B. P. H. J. The transcriptome of Verticillium dahliae-infected Nicotiana benthamiana determined by deep RNA sequencing. Plant Signal. Behav. 7, 1065–9 (2012).

Foyer, C. H., Neukermans, J., Queval, G., Noctor, G. & Harbinson, J. Photosynthetic control of electron transport and the regulation of gene expression. J. Exp. Bot. 63, 1637–61 (2012).

Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–2 (2012).

Garg, R., Patel, R. K., Tyagi, A. K. & Jain, M. De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Res. 18, 53–63 (2011).

Góngora-Castillo, E. et al. Development of Transcriptomic Resources for Interrogating the Biosynthesis of Monoterpene Indole Alkaloids in Medicinal Plant Species. PLoS One 7, e52506 (2012).

Götz, S. et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420–35 (2008).

Gowik, U. & Westhoff, P. The path from C3 to C4 photosynthesis. Plant Physiol. 155, 56–63 (2011).

Grabherr, M. et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat. Biotechnol. 29, 644–652 (2011).

Guillaumie, S. et al. MAIZEWALL. Database and developmental gene expression profiling of cell wall biosynthesis and assembly in maize. Plant Physiol. 143, 339–63 (2007).

Hardion, L., Verlaque, R., Fridlender, A. & Vila, B. in Taxon (eds. Marhold, K. & Breitwieser, I.) 60, 1220–1223 (2011).

Hardion, L., Verlaque, R., Baumel, A., Juin, M. & Vila, B. Revised systematics of Mediterranean Arundo (Poaceae) based on AFLP fingerprints and morphology. Taxon 61, 1217–1226 (2012).

Harris, D. M. et al. Cellulose microfibril crystallinity is reduced by mutating C-terminal transmembrane region residues CESA1A903V and CESA3T942I of cellulose synthase. Proc. Natl. Acad. Sci. U. S. A. 109, 4098–103 (2012).

Hardion L, Verlaque R, Saltonstall K, Leriche A, Vila B. Origin of the invasive Arundo donax (Poaceae): a trans-Asian expedition in herbaria. Ann Bot. 2014;114:455–62.

Harris, D., Stork, J. & Debolt, S. Genetic modification in cellulose-synthase reduces crystallinity and improves biochemical conversion to fermentable sugar. GCB Bioenergy 1, 51–61 (2009).

He, R. et al. Next-generation sequencing-based transcriptomic and proteomic analysis of the common reed, Phragmites australis (Poaceae), reveals genes involved in invasiveness and rhizome specificity. Am. J. Bot. 99, 232–47 (2012).

Huang, X. & Madan, A. CAP3: A DNA sequence assembly program. Genome Res. 9, 868–77 (1999).

Jung, H.-J. G., Samac, D. A. & Sarath, G. Modifying crops to increase cell wall digestibility. Plant Sci. 185–186, 65–77 (2012).

Jung, H.-J. G., Samac, D. A. & Sarath, G. Modifying crops to increase cell wall digestibility. Plant Sci. 185–186, 65–77 (2012).

Khraiwesh, B., Zhu, J.-K. & Zhu, J. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. Biochim. Biophys. Acta 1819, 137–48 (2012).

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25 (2009).

Le, H.-S., Schulz, M. H., McCauley, B. M., Hinman, V. F. & Bar-Joseph, Z. Probabilistic error correction for RNA sequencing. Nucleic Acids Res. 1–11 (2013). doi:10.1093/nar/gkt215

Lee, C., Teng, Q., Huang, W., Zhong, R. & Ye, Z.-H. Down-regulation of PoGT47C expression in poplar results in a reduced glucuronoxylan content and an increased wood digestibility by cellulase. Plant Cell Physiol. 50, 1075–89 (2009).

Lewandowski, I., Scurlock, J. M. O., Lindvall, E. & Christou, M. The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. Biomass and Bioenergy 25, 335–361 (2003).

Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323 (2011).

Liberman, L. M., Sozzani, R. & Benfey, P. N. Integrative systems biology: an attempt to describe a simple weed. Curr. Opin. Plant Biol. 15, 162–167 (2012).

Linder, H., Verboom, G. & Barker, N. Phylogeny and evolution in the Crinipes group of grasses (Arundinoideae: Poaceae). Kew Bull. (1997).

Lindgren, L. O., Stålberg, K. G. & Höglund, A.-S. Seed-Specific Overexpression of an Endogenous Arabidopsis Phytoene Synthase Gene Results in Delayed Germination and Increased Levels of Carotenoids, Chlorophyll, and Abscisic Acid. PLANT Physiol. 132, 779–785 (2003).

Liu, L. et al. Comparison of next-generation sequencing systems. J. Biomed. Biotechnol. 2012, 251364 (2012).

Mariani, C. et al. Origin, diffusion and reproduction of the giant reed (Arundo donax L.): a promising weedy energy crop. Ann. Appl. Biol. 157, 191–202 (2010).

Martin, L. B. B., Fei, Z., Giovannoni, J. J. & Rose, J. K. C. Catalyzing plant science research with RNA-seq. Front. Plant Sci. 4, 66 (2013).

Massa, A. N. et al. The transcriptome of the reference potato genome Solanum tuberosum Group Phureja clone DM1-3 516R44. PLoS One 6, e26801 (2011).

Milne, I. et al. Tablet--next generation sequence assembly visualization. Bioinformatics 26, 401–2 (2010).

Moellering, E. R. & Benning, C. Galactoglycerolipid metabolism under stress: a time for remodeling. Trends Plant Sci. 16, 98–107 (2011).

Morozova, O. & Marra, M. a. Applications of next-generation sequencing technologies in functional genomics. Genomics 92, 255–64 (2008).

Nakasugi, K. et al. De novo transcriptome sequence assembly and analysis of RNA silencing genes of Nicotiana benthamiana. PLoS One 8, e59534 (2013).

Neto, C. P. et al. Variations in chemical composition and structure of macromolecular components in different morphological regions and maturity stages of Arundo donax. Ind. Crops Prod. 6, 51–58 (1997).

Ohlrogge, J. et al. Driving on biomass. Science (80-.). 324, 1019–1020 (2009).

Okawa, S., Makino, A. & Mae, T. Shift of the major sink from the culm to the panicle at the early stage of grain filling in rice (Oryza sativa L. cv. Sasanishiki). Soil Sci. Plant Nutr. 48, 237–242 (2002).

Papazoglou, E. G. Arundo donax L. stress tolerance under irrigation with heavy metal aqueous solutions. Desalination 211, 304–313 (2007).

Perdue, R. Arundo donax—source of musical reeds and industrial cellulose. Econ. Bot. 12, 368–404 (1958).

Pevzner, P. A., Tang, H. & Waterman, M. S. An Eulerian path approach to DNA fragment assembly. Proc. Natl. Acad. Sci. U. S. A. 98, 9748–53 (2001).

Pilu, R., Bucci, A., Badone, F. C. & Landoni, M. Giant reed (Arundo donax L.): A weed plant or a promising energy crop? African J. Biotechnol. 11, 9163–9174 (2012).

Polunin, O. & Huxley, A. Flowers of the Mediterranean. London Hogarth Press (1987).
Raspolli Galletti, A. M. et al. From giant reed to levulinic acid and gamma-valerolactone: A high yield catalytic route to valeric biofuels. Appl. Energy 102, 157–162 (2013).
Saha, B. C. Hemicellulose bioconversion. J. Ind. Microbiol. Biotechnol. 30, 279–91 (2003).
Schliesky, S., Gowik, U., Weber, A. P. M. & Bräutigam, A. RNA-Seq Assembly - Are We There Yet? Front. Plant Sci. 3, 220 (2012).
Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28, 1086–92 (2012).
Slade, A. J., Fuerstenberg, S. I., Loeffler, D., Steine, M. N. & Facciotti, D. A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat. Biotechnol. 23, 75–81 (2005).
Surget-Groba, Y. & Montoya-Burgos, J. I. Optimization of de novo transcriptome assembly from next-generation sequencing data. Genome Res. 20, 1432–40 (2010).
Suzuki, N. et al. Construction and analysis of EST libraries of the trans-polyisoprene producing plant, Eucommia ulmoides Oliver. Planta 236, 1405–17 (2012).
Swaminathan, K. et al. A framework genetic map for Miscanthus sinensis from RNAseq-based markers shows recent tetraploidy. BMC Genomics 13, 142 (2012).
Takahashi, W., Takamizo, T., Kobayashi, M. & Ebina, M. Plant regeneration from calli in giant reed (Arundo donax L.). Grassl. Sci. 56, 224–229 (2010).
Tammisola, J. Towards much more efficient biofuel crops - can sugarcane pave the way? GM Crops 1, 181–98 (2010).
Taylor, S. H. et al. Ecophysiological traits in C3 and C4 grasses: a phylogenetically controlled screening experiment. New Phytol. 185, 780–91 (2010).
Thorbjørnsen, T. & Villand, P. Distinct isoforms of ADPglucose pyrophosphorylase occur inside and outside the amyloplasts in barley endosperm. Plant … 10, (1996).
Wang, J., Nayak, S., Koch, K. & Ming, R. Carbon partitioning in sugarcane (Saccharum species). Front. Plant Sci. 4, 201 (2013).
Weckwerth, W. Green systems biology - From single genomes, proteomes and metabolomes to ecosystems research and biotechnology. J. Proteomics 75, 284–305 (2011).
Yang, C.-Q. et al. Transcriptional regulation of plant secondary metabolism. J. Integr. Plant Biol. 54, 703–12 (2012).
You, T.-T., Mao, J.-Z., Yuan, T.-Q., Wen, J.-L. & Xu, F. Structural Elucidation of the Lignins from Stems and Foliage of Arundo donax Linn. J. Agric. Food Chem. 61, 5361–5370 (2013).
Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–9 (2008).
Zhao, Q.-Y. et al. Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics 12 Suppl 1, S2 (2011).

DISSECTION OF EARLY TRANSCRIPTIONAL RESPONSES TO WATER STRESS IN *ARUNDO DONAX* L. BY UNIGENE-BASED RNA-SEQ

3

"Biology is the study of complicated things that give the appearance of having been designed for a purpose." —Richard Dawkins, The Blind Watchmaker

The work described in this chapter was done in collaboration with Michele Poli, who is going to complete a Ph.D at Bologna University. Michele performed all the experimental work in the wet lab.

The study has been published: Fu Y, Poli M, Sablok G, Wang B, Liang YC, La Porta N, Velikova V, Loreto F, Li M, Varotto C. Dissection of early transcriptional responses to water stress in *Arundo donax* L. by unigene-based RNA-seq. Biotechnology for Biofuels 2016;9:54

3.1 ABSTRACT

Arundo donax is considered one of the most promising energy crops in the Mediterranean region because of its high biomass yield and low input requirements. Despite its potential, little is known about the molecular responses of *A. donax* to abiotic stresses, and especially to water deprivation. In this study, we obtained by RNA-seq the whole root and shoot transcriptomes of young *A. donax* plants subjected to water stress with 10% and 20% polyethylene glycol (PEG), and identified a total of 3034 differentially expressed genes. Mining of stress-related genes indicated the higher responsivity of roots compared to shoots at the early stages of water stress especially under the milder PEG treatment, with a majority of genes responsive to salt, oxidative and dehydration stress. Analysis of gene ontology terms underlined the qualitatively different responses between root and shoot tissues. In root

DEGs were associated mainly to polysaccharide catabolism and biotic stress, in line with responses to osmotic stress and incipient cellular damage. In shoot early water stress was instead dominated by functions related to signal transduction and protein modification

associated to phosphorylation, indicating a milder/delayed onset of stress. Analysis of metabolic pathways highlighted the crucial role played in both shoots and roots by genes involved in the signalling cascade of abscisic acid. We further identified relatively large organ-specific differences in the patterns of drought-related transcription factor AP2-EREBP, AUX/IAA, MYB, bZIP, C2H2 and GRAS families, which may underlie the transcriptional reprogramming differences between organs. Through comparative analyses with major Poaceae species, we finally identified a set of 53 orthologs that can be considered as a core of evolutionary conserved genes important to mediate water stress responses in the family. This study provides the first characterization of *A. donax* transcriptome in response to water stress, thus shedding novel light at the molecular level on the mechanisms of stress response and adaptation in this emerging bioenergy species. The inventory of early-responsive genes to water stress identified could constitute useful markers of the physiological status of *A. donax* and be a basis for the improvement of its productivity under water limitation. The full water-stressed *A. donax* transcriptome is available for Blast-based homology searches through a dedicated web server (http://ecogenomics.fmach.it/arundo/).

3.2 Introduction

Among the different sources of renewable energy, biomass is interesting because it has a nearly neutral carbon balance and the ethanol produced by its fermentation can be blended with petrol-derived fuels giving an important contribution in reducing transport-related CO2 emissions (Acres, 2007). So-called second generation bioethanol (i.e., the one not produced from edible parts of crops) can be obtained from food crop straw, but the yields of such biomass source are expected to be low, as food crops were intentionally selected to maximize photosynthate allocation to edible parts (Doebley, Gaut, & Smith, 2006). In alternative, plant species specifically dedicated to energy production (called bioenergy crops) are normally better biomass producers than food crops, resulting in higher ethanol yields per unit of cultivated area. Arundo donax has been identified among bioenergy crops as the most promising species for the Mediterranean area (Lewandowski et al. 2003). *A. donax*, commonly called giant reed, is a perennial C3, polyploid, bamboo-like grass of the Poaceae family. It favors well-drained soils with abundant moisture, where it can form dense stands up to 6–10 m high with yields of up to 40 tons per hectare each year (comparable, or even

exceeding, those of some C4 species) (Byrt et al, 2011). The origins of the giant reed are still debated, but the latest evidences from plastid DNA sequencing and morphometric parameters data collected from 127 herbarium specimens support a Middle-East origin of *A. donax* (Hardion et al, 2014). Despite the production of panicle-like flowers, no viable seeds from Mediterranean ecotypes have been reported so far (Balogh et al, 2012). Natural propagation exclusively occurs vegetatively by rooting of rhizome and stem fragments originating as a consequence of floodings, followed by a slow colonization through rhizome expansion (Di Tomaso et al, 2003). Consistently, genetic diversity in *A. donax* has been reported to be low, but, possibly due to somatic mutation, detectable (Haddadchi et al, 2013). Possibly because of its high ploidy, the low intraspecific diversity of *A. donax* does not seem to be associated to fitness tradeoffs, as indicated by its high resistance to biotic and abiotic stresses (Mariani et al, 2010). If on one hand this resistance causes the high invasiveness of this plant, on the other hand it makes *A. donax* an excellent bioenergy crop, which can grow with very low management input (e.g., pesticides, fertilization, irrigation) even in marginal lands or in fields irrigated with waste or salty water (Mavrogianopoulos et al, 2002).

The recent advent of Next Generation Sequencing (NGS) has made the development of genomic resources progressively simpler and cheaper (Liu et al. 2012). RNA sequencing (RNA-Seq) is to date by far the most powerful tool for the rapid and inexpensive development of genetic resources for any species of interest. In addition, RNA-Seq allows at once the quantitative determination of the expression levels of virtually all transcribed genes in a specific organ, thus providing an extremely powerful tool for the identification of transcripts differentially expressed in response to the abiotic and biotic stresses which negatively impact crop growth and productivity (Martin, Fei, Giovannoni, & Rose, 2013).

It is widely accepted that global warming will increase the duration and frequency of drought periods over the 21th century (Dai 2012). Many countries already started to develop mitigation strategies to avoid this major threat, which could potentially offset the productivity gains expected from advances in both agricultural and crop breeding techniques (Shanker et al. 2014). Drought is one of the extreme environmental conditions that curtail agricultural crop productivity (Bruce, Edmeades, and Barker 2002). The first response of plants to water limitation is usually avoidance, a strategy that aims at maintaining a neutral

balance between water gained from the root system and lost by transpiration through the stomata. In case of short-term or relatively mild water stress, avoidance can maintain performance and prevent negative effects on plant growth. From a physiological point of view, this is usually achieved by increasing the osmotic potential of root cells, increasing root growth as well as reducing water loss by modulation of stomatal conductance (Verslues et al, 2003). These physiological adjustments are the consequence of complex cellular changes like: (1) the reprogramming of the cellular metabolism, which shifts to polysaccharide degradation and aminoacid biosynthesis to allow for the accumulation of solutes with an osmotic function (e.g., glycine–betaine, proline, mannitol, etc.), (2) the production of abscisic acid (ABA, a phytohormone mainly associated to seed dormancy and water stress, which causes a reduction of stomatal conductance through closure of stomata) and other phytohormones, and (3) an increased synthesis of proteins for cellular protection/detoxification (late-embryogenesis-abundant, LEA; chaperones and heat stress-proteins necessary for proper protein folding), (4) extensive modulation of ribosomal activity to support active cell growth and division in the root system (Verslues et al, 2003 and Claeys and Inzé 2013). When avoidance strategies are not sufficient alone to prevent the onset of water stress, either because of the excessive length or magnitude of the water deficit, tolerance responses become progressively more relevant to limit the damages caused by the reduced availability of water. The same physiological and molecular changes are, however, often shared between the two types of responses, so that a clear-cut distinction between them is not always possible. The medium to long-term adjustments associated to tolerance encompass the development of, e.g., thicker epicuticular waxes to limit water evaporation through epidermal cells, the further decrease of the shoot/root biomass ratio and the allocation of resources to long-term survival organs (e.g., tubers or rhizomes), the enhancement of antioxidant capacity to detoxify the reactive oxygen species (ROS) consequent to photosynthetic limitation, the thickening of xylematic cell walls to prevent collapsing of vasculature, etc. (Verslues et al, 2003 and Claeys and Inzé 2013).

This complex series of cellular responses to water limitation obviously requires also a profound reprogramming of gene expression. Our understanding of the genetic bases of drought resistance largely benefitted from forward and genetic screens in model or crop species (e.g. *Arabidopsis thaliana*, rice, maize, wheat; reviewed by Claeys and Inzé 2013).

In addition, several studies devoted to the dissection of the transcriptional responses to drought stress or water deficit conditions have been carried out for the most common cereal crop species (e.g. rice, Wang et al. 2011; maize, Kakumanu et al. 2012; foxtail millet, Qi et al. 2013; sorghum, Dugas et al. 2011). More recently also Poaceae species used exclusively or partly as energy crops (switchgrass, Xie et al. 2013; miscanthus Lewandowski et al., 2000; sugarcane, Kido et al. 2012) have been object of transcriptomics studies which could provide a robust comparative basis in poorly characterized species like *A. donax*. Unfortunately, the high spatio-temporal complexity of the physiological adaptations to drought and the large number of variables used in different experimental protocols for the application of water stress (methods for induction of water deprivation, combination with other stresses, length of treatment, type of plant materials and their developmental stages) limit the depth of result comparisons across studies (Deyholos, 2010). Polyethylene glycol (PEG) is a high-molecular weight polymer which can be used to induce controlled water deficits in plants by modifying the osmotic potential of water in hydroponic growth media without being absorbed by the root system (Lagerwerff et al., 1961), thus providing an ideal method for water deprivation in RNA-Seq experiments addressing short-term responses of plants to water stress.

Arundo donax is one of the most promising biomass resources for biofuel development but, up to now, little is known at the molecular level on this species' ability to cope with abiotic stresses in general and in particular with water limitation. Leveraging on the recent obtainment of the first reference transcriptome of *A. donax* by RNA-Seq (Sablok et al. 2014) and on the existing knowledge of the genetics of drought responses in plants, in this study we report the characterization of early transcriptional responses to two levels of PEG-induced water deficit in cohorts of young giant reed cuttings. In particular, we addressed the main questions: (1) How many/which genes are differentially expressed during the early phases of water stress in *A. donax*? (2) What are the main biological functions involved? (3) Which are the transcription factors associated to such transcriptional reprogramming? (4) Are the transcriptional responses of *A. donax* conserved/comparable to those of other monocot species, and in particular of rice? The set of about 3000 early-responsive genes to water stress identified in this study are promising reporters of the

physiological status of *A. donax* plantations for the improvement of its management and for a deeper understanding of its biology.

3.3 Results and Discussion

Despite the ability of *A. donax* to withstand prolonged periods of drought, its productivity under water limitation is negatively affected (Lewandowski, Scurlock, Lindvall, & Christou, 2003). Especially during the first year of establishment, *A. donax* growth can be severly retarded and plants damaged for lack of soil moisture (Perdue 1958). An in-depth understanding of the mechanisms involved in water limitation responses in this species is, therefore, an important pre-requisite to improve its management, but till now no characterization of the transcriptional variations of *A. donax* tissues associated to water deprivation is available. To fill this gap of knowledge, we carried out by RNA-Seq a comprehensive identification of early transcriptional responses of shoots and roots to two different levels of PEG-induced water limitation in *A. donax* (details of the experimental design can be found in Materials and Methods and in Supplementary Ch3.1).

Following assembly, we obtained 111,749 transcripts covering 45,821 components. Given the high ploidy of *A. donax*, we chose to use a relatively high Kmer coverage during assembly (min-kmer_cov = 5) to minimize the formation of transcripts with retained introns (Gruenheit, et al. 2012). The observed N50 of the assembled transcriptome is 1826 bp, in line with our previous N50 reports (Sablok et al. 2014), indicating that a good coverage of the transcriptome has been achieved. To eliminate redundant transcripts, we further clustered the transcripts using the CD-HIT software resulting in a total of 80,962 transcripts. The non-redundant transcript set was further assembled into unigenes with MIRA to remove spurious transcripts, resulting in a final set of 80,335 unigenes with an N50 of 1570 bp. Summary statistics results for transcriptome assembly are provided in (Table 3.1).

Assembly*	**Summary statistics**
Total trinity transcripts	111,749
Total trinity components	45,821
Contig N50 (bp)	1826

MIRA unigenes	80,355
Total length of sequence (bp)	75,960,964
Unigene N50 (bp)	1570
GC %	47.70

Table 3.1. Summary statistics of the sequencing reads and the corresponding assemblies. Summary statistics of *A. donax* whole drought transcriptome obtained by Trinity and MIRA. *Trinity assembly: K 25, Kmer coverage 5.

3.3.1 Identification of differentially expressed genes (DEGs) by RNA-Seq

As a first step in the characterization of *A. donax* transcriptional responses to water stress, we carried out the identification of the unigenes whose expression level significantly changed upon PEG-treatments. A total of 3034 genes showed differential expression in at least one of the two stress conditions (mild water stress vs. control, severe stress vs. control and severe vs. mild stress), with roughly the same number of genes being differentially expressed in shoots and roots (1684 and 1712 DEGs, respectively). Validation of expression levels for ten selected DEG candidates was carried out by real-time qRT-PCR, (Supplementary Ch3.2). The high congruence between RNA-Seq and real time PCR results (coefficient of determination R^2= 0.94), indicates the reliability of RNA-Seq quantification of gene expression. Therefore, the selected genes could also constitute useful markers of early water deficit in *A. donax.* DEG identified in biological replicates clustered together in both organs, indicating good reproducibility of treatments. In addition, the heat maps qualitatively indicated the closer similarity of control and mild water stress between each other as compared to severe water stress (Supplementary Ch3.3).

A detailed assessment of the number and the identity of the DEGs between conditions for each organ confirmed this observation: in shoots, only 98 genes were differentially expressed between control condition and mild water stress, *versus* 1572 between control and severe water stress, and 831 between mild and severe water stress. A similar trend, but less marked, characterized also root DEGs (Figure 3.1a), indicating the successful induction of varying degrees of water stress as a function of PEG concentration (Lagerwerff et al., 1961).

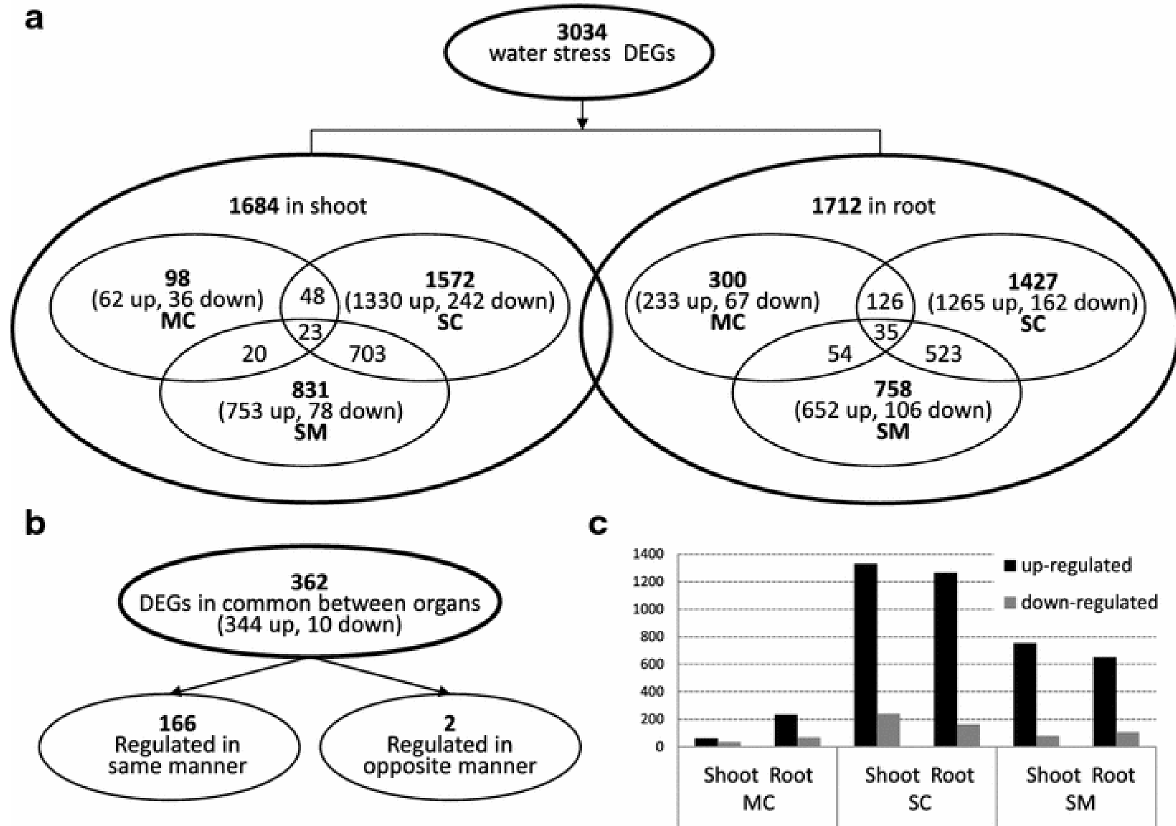


Figure 3.1. Summary of DEGs in shoots and roots of *A. donax* upon drought stress. a Number of genes up-/down-regulated by drought stress under different conditions (MC mild water stress vs. control, SC severe water stress vs. control, SM severe water stress vs. mild water stress.) in root and shoot. b Total number of DEGs in common between root and shoot. c Number of regulated genes between different conditions. Gray bar down-regulated genes; black bar up-regulated genes.

By comparing the 362 DEGs in common between organs, we further observed a general conservation of expression patterns, with 166 of the genes being regulated in the same way in shoot and root and only two genes displaying opposite regulation (Figure 3.1b). In addition, the overall direction of expression variation resulted to be conserved between organs, with the large majority of DEGs being up- rather than down-regulated (Figure 3.1c). A closer analysis of the absolute numbers of DEGs in the two organs, however, highlighted a relatively large difference in gene up-regulation upon mild stress in roots as compared to shoots (300 DEGs in root *vs* 98 in shoot; Figure 3.1a). Given the application of the PEG directly to the root system and the sampling of only one time point, it is possible that, at least in part, these differences could stem from a faster onset of the water stress in roots compared

to shoots. These results, however, are also in line with a transcriptionally higher responsivity of the root system compared to shoots, as previously reported, e.g., in the case of poplar (Cohen, 2010), which could indicate tissue-specific responses.

3.3.2 Functional characterization of transcriptional responses to water stress in *A. donax*

To identify organ-specific differences, stress-related genes were identified based on curated homology searches against genes experimentally characterized in previous studies. The majority of stress-related genes belonged to categories "salt," "oxidative," "dehydration," and "osmotic." This is expected, as water limitation is known to cause reduced turgor and integrity of membranes, increase of intracellular ionic and non-ionic solute concentrations and enhanced production of reactive oxygen species (ROS) that cross-trigger responses to high-salinity, oxidative and osmotic stresses (Huang 2012) (Figure 3.2; Supplementary Ch3.4). Worth of note, the two differentially expressed categories encompassing the largest differences in number of genes between organs are "dehydration" and "osmotic". Both categories are more abundant in shoot than root, but the highest shoot/root ratio (eight times) is found for dehydration-related genes (Figure 3.2).

We next carried out a homology-based annotation specifically for all the 3034 DEGs identified upon PEG treatments, identifying at the same time the GO terms associated to this dataset. (Supplementary Ch3.5). Based on BLASTN searches using a 70 % identity and 50 % query coverage cutoff, we identified a total of 214 differentially expressed *A. donax* unigenes in shoots and 642 in roots, which were not present in the reference transcriptome assembly (Supplementary Ch3.4) (Sablok et al, 2014), thus contributing to ongoing gene discovery and functional annotation in this poorly characterized species.

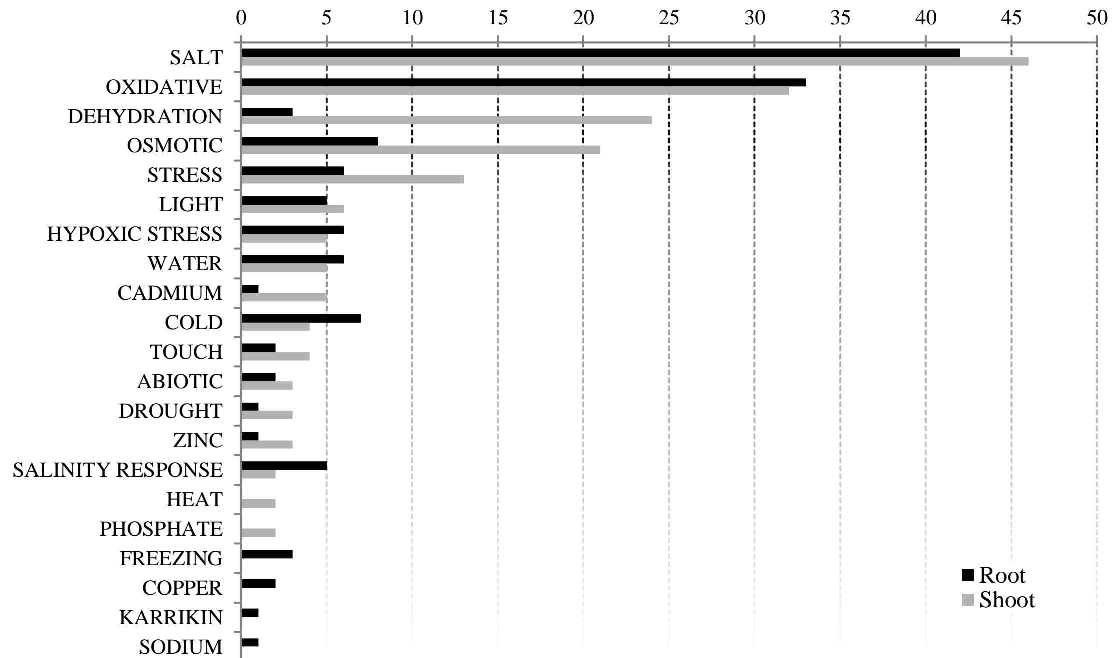


Figure 3.2. Distribution of stress-related functional categories of DEGs. Stress-related functional categories are identified by annotation of *A. donax* putative homologs in of Arabidopsis genes from ASPRGDB. Data are sorted by number of shoot DEGs. *Black bar* root DEGs; *gray bar* shoot DEGs.

To determine the gene functional classes which were chiefly involved in the response to water stress, we carried out an analysis of over/under-representation of GO terms associated to DEGs. A total of three contrasts were carried out: mild water stress vs. control (MC), severe water stress vs. control (SC) and severe water stress vs. mild water stress (SM). We further selected the most significantly enriched GOs using REVIGO (Supek et al, 2011) (Supplementary Ch3.6) and analyzed the number of GO terms in common between contrasts to pinpoint differences and similarities between organs and conditions (Figure 3.3: shoot/SC; root/SC; root/SM for both (1) molecular function and (2) biological process terms).

a. b.

Shoot SC Root SC Shoot SC Root SC

21 4 3 24 4 8

3 5

0 7 3 3

0 1

Root SM Root SM

Figure 3.3. Venn diagram of significantly enriched GOs. The GO terms which were overrepresented under different conditions have been slimmed by REVIGO, and compared by category: a molecular function; b biological process. MC mild water stress vs. control, SC severe water stress vs control; SM severe water stress vs. mild water stress

The seven molecular function and nine biological process GOs consistently over-represented in SC shoot and root contrasts, respectively (see, e.g., GO:0003700 sequence-specific DNA binding transcription factor activity, GO:0030528 transcription regulator activity; GO:0004722 protein serine/threonine phosphatase activity, Supplementary Ch3.6; Figure 3.3), delineated an ongoing reprogramming of cellular transcription and post translational protein modification related to drought recovery and osmotic adjustment in both organs, in line with the severity of the water stress applied (Verslues et al,2006). In particular, the changes observed in proline metabolism in plants (GO:0004657 proline dehydrogenase activity, GO:0006562 proline catabolic process) are a well known response to a multiplicity of abiotic stresses, including drought (Verbruggen & Hermans, 2008). The majority of GO terms were, however, specific for shoot and root, providing compelling evidence of qualitative and quantitative differences in the responses of these organs to water deprivation: despite the highest responsivity of the root transcriptome observed above, in fact, the transcriptional response in the shoot involved 2-3 times more functions than in root (Figure 3.3). Also the GO terms differed significantly in the different organs: in root several DEGs

were associated to functions related to polysaccharide catabolism (e.g. GO:0000272, polysaccharide catabolic process; GO:0016161, beta-amylase activity), indicating extensive osmotic adjustment to reduce the water potential and limit cellular damage (Verslues et al,2006). The enrichment of terms related to biotic stress (e.g. GO:0009816, defense response to bacterium incompatible interaction; GO:0009607, response to biotic stimulus been reported, e.g., in sorghum tissue treated with PEG or ABA (Dugas et al. 2011), thus supporting the conservation of the cross talk between biotic and abiotic stress responses in Poaceae. Also enrichment of gene functions related to translation was observed (e.g., GO:0003735, structural constituent of ribosome; GO:0006412, translation), possibly as a response to the extensive transcriptional reprogramming observed in roots and/or to root cell growth and division. The extension of the root apparatus is indeed a common response in plants to water stress which maximizes the chance of reaching the moisture available in deeper layers of soil (Verslues et al, 2006). Worth of note, in root the only biological function specific to the milder PEG treatment (SM contrast, GO:0009685 gibberellin metabolic process; Figure 3.3b) indicates a possible involvement of gibberellins (GA) in the control of this trait through root growth. Based on the comparison of emmer wheat susceptible and resistant varieties, GA signaling and biosynthesis genes have been associated to resistance to drought in roots (Krugman et al, 2011). These results are consistent with a role of GA in the maintenance of root growth as part of the developmental decrease of the shoot/root biomass ratio usually observed in plants growing under water stress (Colebrook et al, 2014). It is thus possible that the enrichment of functions related to GA observed also in *A. donax* could contribute to the onset of the developmental changes triggered by mild water stress to increase accessibility of roots to soil with higher moisture.

Compared to root, in shoot the pattern of GO terms enrichment in response to water limitation was dominated by functions related to signal transduction and protein modification associated to phosphorylation (Supplementary Ch3.6; Figure 3.3). This result mirrors the dramatic increase in post-translational phosphorylation levels observed in wheat leaves under drought stress (Zhang et al. 2014). Interestingly, several other functional classes specifically enriched in *A. donax* shoot transcriptome corresponded to those of proteins undergoing phosphorylation in wheat (e.g., GO:0009405pathogenesis, GO:0008643carbohydrate transport, GO:0005509calcium ion binding,

GO:0015291secondary active transmembrane transporter activity), indicating that a synergistic effect between transcriptional and post-translation reprogramming may take place in Poaceae shoots during water stress (Zhang et al. 2014).

Taken together, the identified DEGs indicate major differences between organs in the transcriptional responses to water stress: Roots experienced a seemingly more severe/earlier stress, whereas in shoots the transcriptional response was still mainly at the level of signal transduction. Time course analyses will be required to precisely define the relative contribution of stress induction kinetics versus organ-specificity to the patterns of differential expression observed in this study. Given the relevance that the root system plays in both acclimation and adaptation of plants to water stress (Lynch et al 2014), several of the early-responsive genes identified could constitute suitable markers for the detection of early water stress in *A. donax*.

3.3.3 Metabolic pathways related to water stress in *A. donax*

The set of 3034 DEGs was mapped onto KEGG pathways in *Arabidopsis thaliana* and *Oryza sativa*, highlighting the involvement of several drought-related pathways (Figure 3.4).

‘Plant hormone signal transduction’ (ko04075), comprising 11 DEGs in roots and 12 DEGs in shoots, was overrepresented. In this pathway, for both shoots and roots, the transcripts of several hormone-responsive proteins involved in regulation and signal transduction were up-regulated. Plant hormones play crucial roles in a diverse set of developmental processes, as well as in the response to biotic and abiotic stresses (Bari and Jones, 2009). For example, MYC2 is known to function as an activator in ABA signaling and its overexpression in Arabidopsis confers increased tolerance to drought (Abe et al, 2003). As discussed in more detail below, also ABA-activated SnRK2 is required for dehydration stress signaling in Arabidopsis (Yoshida et al, 2002). Previous studies also suggested that in rice OsJAZ1 could connect the jasmonate and drought stress signaling cascades by functionally interacting with OsbHLH148 and OsCOI1 (Seo et al, 2011). Taken together these results confirm the pivotal role played in water stress response by the differential regulation of genes involved in hormone signal transduction (Huang et al, 2012).

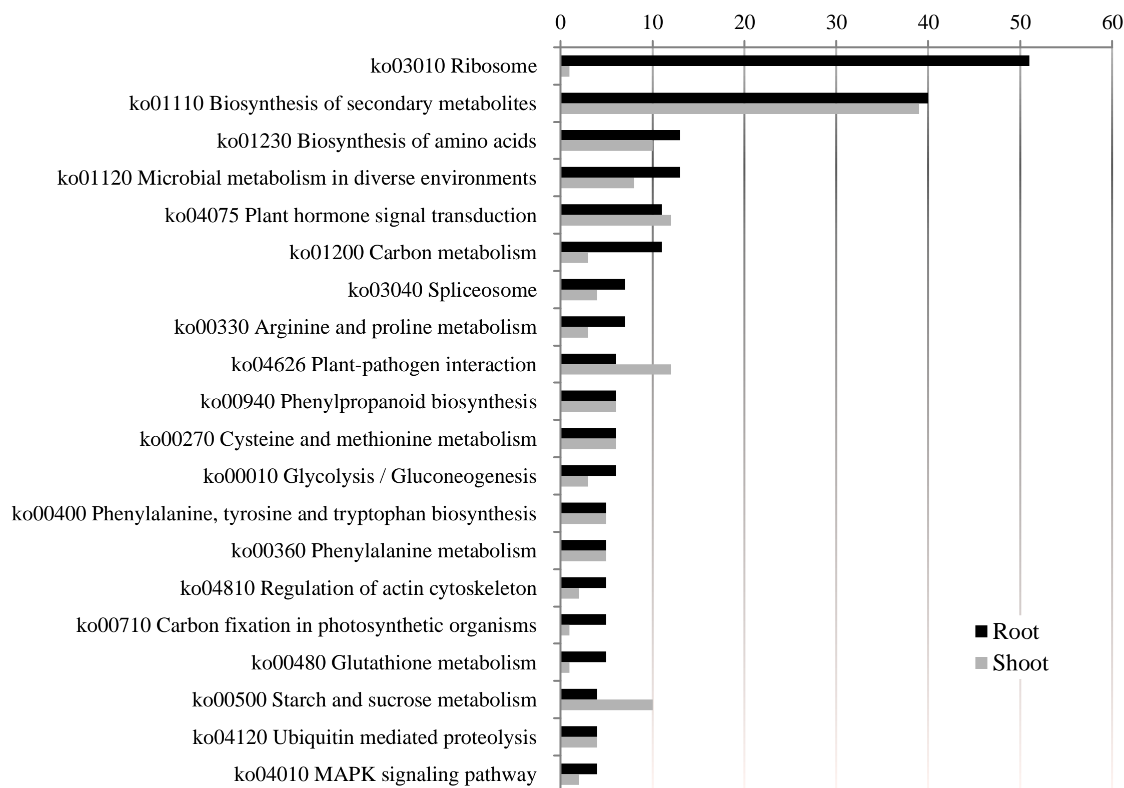


Figure 3.4. Distribution KEGG Pathways for DEGs in shoot and root. Data are sorted by number of root DEGs mapping to KEGG pathways. Black bar root DEGs; gray bar shoot DEGs

Two other important pathways, including 'Phenylalanine metabolism' (ko00360) and 'Plant-pathogen interaction' (ko04626), were also found in our study to be regulated by water stress (Figure 3.4). *CYP73A* (*trans-cinnamate 4-monooxygenase*), *4-coumarate--CoA ligase* and *peroxidase*, associated with 'phenylalanine metabolism', were all highly accumulated in response to water stress, in agreement to the relevance of this pathway in plant responses to drought (Gholizadeh A et al, 2011). Additionally, transcripts from *CALM* (*calmoduli*n), *CML* (*calcium-binding protein: CaM-like protein*), *MYC2*, *RBOH* (*respiratory burst oxidase*), *PR1* (*pathogenesis-related protein 1*) and *JAZ* members of the 'Plant-pathogen interaction' pathway, were also induced by water stress. All these genes have been reported to be involved in response to several stresses. For example, as calcium is one of the most important signaling molecules in plants, the expression of CALM and CMLs is well regulated due to different environmental requirements in Arabidopsis (Fuchs et al. 2011). Finally *RBOH* genes are also commonly expressed in many plants in response to biotic and abiotic stresses (Marino et al., 2012).

Other examples of relevant pathways which are known to be involved in responses to abiotic stresses in general or specifically to drought were 'Starch and sucrose metabolism' (ko00500), 'Arginine and proline metabolism' (ko00330), and 'MAPK signaling pathway' (ko04010) (Huang et al, 2012, Mohammadkhani and Heidari, 2008 and Yoshiba Y et al, 1997).

Strikingly, the biggest difference observed between root and shoot was related to ribosomal DEGs (ko03010 ribosome; Figure 3.4), thus identifying reprogramming of ribosomal translation as one of the largest responses of the root system during the early stages of water stress in *A. donax*. As noted above, it is likely that such large effect to translation could represent the early phases of the modulation of shoot/root resource allocation preluding to root cell growth and division, a typical avoidance responses of the root system during the early phases of water stress. Given the high number of ribosomal subunit genes and the complexity of their regulation as a function of water stress intensity/duration as well as species- and even genotype-dependent variation (Benešová et al, 2012) , the detailed dissection of ribosome-related pathway reprogramming will be relevant for the elucidation of root-specific responses to early water stress in *A. donax*.

3.3.4 Identification of transcription factors responsive to water stress in *A. donax*

Transcription factors (TF) have been identified among the most promising targets for the improvement of plant performance under drought stress. Mining of DEGs for putative TFs and their interactors led to the identification of 238 *A. donax* unigenes, corresponding to 136 high confidence rice homologs previously identified as drought-responsive genes from 37 TF families (Priya & Jain 2013; Supplementary Ch3.7). Because of the altered water potential under salt stress (Huang 2012), the majority of the genes (108) are also responsive to salinity. A total of 18 genes, are, however, specifically responding to drought (Supplementary Ch3.7). The most represented *A. donax* differentially expressed families, constituting alone the majority of the genes, were those of NAC, WRKY, AP2-EREBP, bHLH, bZIP and AUX/IAA, which are known to mediate water stress responses in plants (Hadiarto & Tran 2011). The majority of these families were also among the most

represented in drought-stressed rice (Wang et al. 2011). *A. donax* unigenes from the NAC family are the most common among differentially expressed TF genes (36 in total), matching a total of 14 different rice loci. Six of them (Os03g60080/SNAC1; Os01g66120/SNAC2/OsNAC6; Os11g08210/OsNAC5; Os11g03300/OsNAC10; Os08g06140; Os05g34830) have been previously identified as drought-responsive (Nuruzzaman M et al. 2013). Four of them have been characterized in depth through functional analyses, confirming their pivotal role in water stress-related trasncriptional reprogramming in rice. In particular, all of them have been demonstrated to be ABA-responsive (Hu et al. 2006; Hu et al. 2008; Sperotto et al. 2009; Jeong et al. 2010.), in agreement with the activation of the ABA signal translation cascade observed above.

A total of 67 out of the 150 rice homologs to differentially expressed *A. donax* TF unigenes (45%) were consistently found to be differentially expressed also in rice (Ray et al. 2011; Supplementary Ch3.7.). Not all the TF families, however, were equally represented in both *A. donax* organs, indicating that part of the differences observed between shoot and root transcriptional responses may be mediated by members of these groups. In root among the families encompassing more than 5 differentially expressed unigenes, we found twice as many AP2-EREBP, AUX/IAA and MYB unigenes than in shoot. AP2-EREBP is a superfamily of transcription factors composed by the ERF, AP2 and RAV families (Riechmann & Meyerowitz 1998; Rashid et al. 2012) (Figure 3.5).

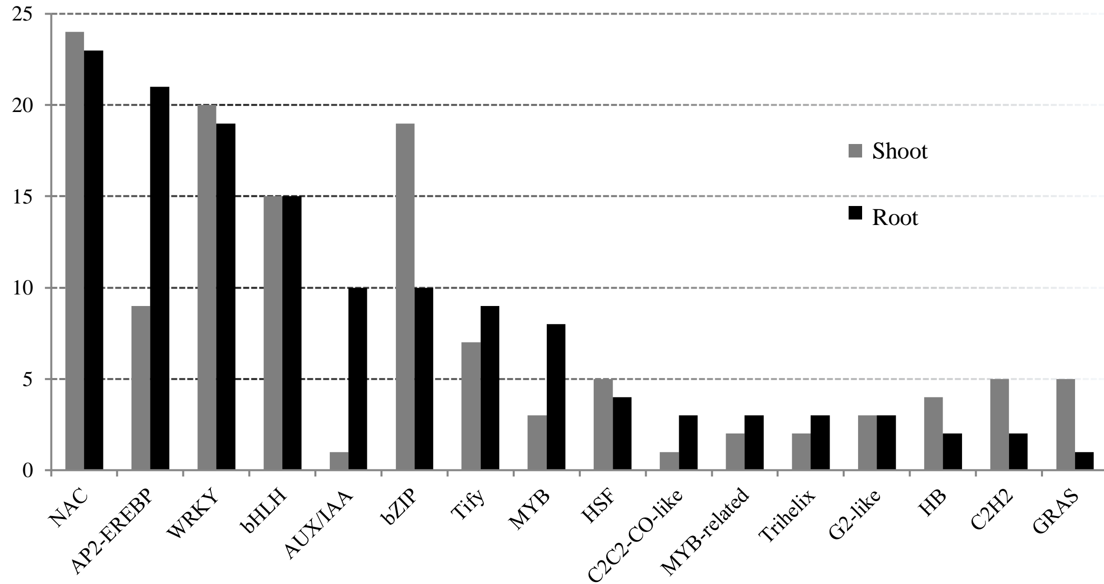


Figure 3.5. Distribution of transcription factors responsive to water stress in *A. donax*. Data are sorted by number of root DEGs. Only categories with more than 3 DEGs identified as transcription factors are shown. *Black bar* root DEGs; *gray bar* shoot DEGs

AP2/EREBP TFs are involved in many fundamental biological processes, ranging from development to response to biotic and abiotic stresses (Nakano et al. 2006). Five of the *A. donax* unigenes, homologous to rice genes Os02g51670 (DREB2B), Os09g20350 (DREBF1), Os04g55520 (DREB2F) and Os06g03670 (OsDREB1C/CBF), belong to the DREB subfamily of ERF TFs, which are known to control expression of several genes in response to dehydration and low temperature (Sakuma Y et al. 2002). All the four rice homologs have been directly involved in the responses to water stress, thus confirming their relevance towards this stress also in *A. donax* (Matsukura et al. 2010; Wang et al. 2008; Dubouzet et al. 2003; Moumeni A et al. 2011). Other *A. donax* unigenes from the AP2/EREBP super-family were homologs of 4 rice ERF genes (Os03g09170; Os08g31580; Os06g10780; Os05g41780). Unlike DREB genes, ERF TFs have been associated mainly to biotic stress responses mediated by ethylene (Ohme-Takagi & Shinshi, 1995). It is, therefore, likely that the *A. donax* ERF TFs identified in this study are at least in part responsible for the enrichment of GO terms related to biotic stresses observed above, possibly with the participation of members of the NAC and WRKY families (Narsai et al. 2013).

The second TF family with most unigenes in roots compared to shoot was that of AUX/IAA. About ten times more *A. donax* unigenes from this family were differentially expressed in root upon water stress compared to shoot (Figure 3.5). AUX-IAA proteins interact with TFs of the ARF (Auxin-Responsive Factors) family, repressing root growth as consequence of the increase of the intracellular levels of the auxin plant hormone indoleacetic acid (IAA; Overvoorde et al. 2010). Since *A. donax* is a perennial plant whose large rhizomes serve as a long-term survival organ, the AUX-IAA genes specifically induced in roots are interesting candidates to dissect the coordination of *A. donax* root and shoot growth under water deprivation.

MYB and MYB-related transcription factors also were more represented in roots as compared to shoots (Figure 3.5). Two of the *A. donax* MYB unigenes were homologs to rice Os12g37690, which had been previously associated to differences among drought-sensitive and drought-tolerant rice cultivars (Degenkolbe et al. 2009). The gene is also upregulated in response to oxidative stress during the early response of japonica rice to chilling (Yun et al. 2010). Interestingly, another *A. donax* unigene was homolog to Os12g37970, a rice MYB TF involved in the coordinate regulation of cellulose and lignin biosynthesis (Ambavaram et al. 2011), indicating that it could contribute to alter the structure of cell walls in response to water stress.

Among the largest TF families displaying a higher number of differentially expressed unigenes in *A. donax* shoots compared to roots we found the bZIP, C2H2 and GRAS families (Figure 3.5). Strikingly, nine of the 21 differentially identified bZIP unigenes were homologous to rice Os02g52780. This rice gene, also called OsbZIP23 (Nijhawan et al. 2008), has been functionally demonstrated to have a relevant role among rice bZIP genes in conferring ABA-dependent drought and salinity tolerance (Xiang et al. 2008). Somehow unexpectedly, none of the rice genes homologs to differentially expressed *A. donax* unigenes (LOC_Os07g39470, LOC_Os01g62460, LOC_Os01g71970, LOC_Os07g36170, LOC_Os11g47870) have been functionally characterized, leaving open their specific role in drought responses. Two of them (Os01g71970 and Os07g36170), however, had already been identified among the few GRAS TFs differentially expressed in rice upon drought stress (Ray et al. 2011). More recently, additional evidences for the involvement of members of

the GRAS family in the responses to water deprivation have been reported for rice (Xu et al. 2015). Consistently with our results in *A. donax*, expression of *OsGRAS23* was significantly induced in rice leaves following treatments with PEG, dehydration, salt, gibberellins and jasmonic acid. In particular, transgenic rice overexpressing *OsGRAS23* (LOC_Os04g50060) was more resistant to drought and tolerant to oxidative stress compared with wild-type, thanks to the upregulation of genes involved in anti-oxidant functions (Xu et al. 2015). Taken together, these results indicate that GRAS genes in general and in particular those identified in *A. donax* represent interesting candidates for increasing water stress tolerance in monocots. Also the majority of the rice homologs of differentailly expressed *A. donax* unigenes from the C2H2 family (LOC_Os03g55540, LOC_Os03g13600, LOC_Os03g60570, LOC_Os09g38340) were have been previously identified as drought responsive (Ray et al. 2011), indicating their conserved role in Poaceae. Two among them and an additional C2H2 gene not previously identified (LOC_Os03g10140, LOC_Os09g38340, LOC_Os09g38790) are known to control the vegetative to floral phase transition in monocots (Colasanti et al. 2006; Higgins et al. 2010), indicating that responsiveness of C2H2 genes to water deprivation may be relictual in *A. donax*: while other species from the *Arundo* genus are fertile and could benefit from accelerating seed setting as a drought-escape strategy, *A. donax* is fully sterile (Hardion et al. 2015) and no clear selective advantage seems to be associated to this trait. Therefore, loss of function mutations of C2H2 or other flowering time TFs could be interesting candidates to extend the vegetative phase and, thus, biomass accumulation in *A. donax* (Sablok et al. 2014).

3.3.5 CHARACTERIZATION OF CO-REGULATED GENE EXPRESSION NETWORK IN *A. DONAX*

We compared the distribution of both differentially and non-differentially expressed *A. donax* genes with the 15 drought-responsive modules of rice orthologs recently identified (Zhang et al, 2012). Only Module 7 and Module 14 were over-represented in both shoots and roots, while Module 10 was over-represented only in shoots (Table 3.2).

Rice module	Rice genes	Putative orthologs in *A. donax*	Putative orthologs in *A. donax* shoot DEGs (*p* value)	Putative orthologs in *A. donax* root DEGs (*p* value)	Putative module function
Module 1	303	149	2	8	
Module 2	213	155	5	4	
Module 3	141	61	1	0	
Module 4	134	35	2	3	
Module 5	117	71	2	2	
Module 6	90	29	4	0	
Module 7	77	46	22 (4.89E−15)	12 (3.66E−07)	Hormonal signal transduction
Module 8	48	18	0	0	
Module 9	47	16	3	2	
Module 10	47	27	8 (3.22E−4)	0	Post-translational protein modification
Module 11	46	11	0	0	
Module 12	42	11	0	0	
Module 13	38	13	0	0	
Module 14	28	17	6 (5.42E−4)	8 (3.53E−07)	Stomatal closure
Module 15	21	11	1	0	

Table 3.2. Comparison between *A. donax* water stress response genes and rice drought response network. A total of 56 (in shoot) and 39 (in root) *A. donax* DEGs for which rice orthologs could be identified are mapped onto the 15 co-expression modules previously identified in rice (Zhang et al, 2012). Significance levels for over- and under-representation as compared to rice (p-value) are provided. There are 1392 rice genes in the 15 modules. Based on Blastp reciprocal best hits method, a total of 56 and 39 differentially expressed genes were identified as putative orthologs of rice genes in shoot and root of *A. donax*, respectively.

Module 10 had been identified as a post-translational drought-related signaling/regulation cascade (genes involved in protein amino acid phosphorylation

processes), further confirming the results from GO enrichments discussed above. The functions of Module 7 and Module 14, however, were not reported. Based on the functional mining of rice and *A. donax* orthologs in each module (Supplementary Ch3.8), we found that Module 7 might be related to hormonal signal transduction, since these genes are mapped on JAZ, CML, PTC2_3, and ABF, which all belong to the 'Plant hormone signal transduction' pathway. The observation that PSY (phytoene synthase), which controls metabolic flux through the pathway supplying carotenoid precursors for ABA biosynthesis, is also part of this module further strengthens the identification of Module 7 as likely ABA-related co-expression module. Additionally, promoters of genes from Module 7 in rice were found to be enriched in S-BOX motif, which is the ABI4 binding site. ABI4 is known to be an important link between ABA hormone and glucose signaling pathway, and it has been proposed that in some species carbohydrate metabolism might be the initial response to drought (Acevedo-Hernández, León and Herrera-Estrella, 2005;). Thus, genes belonging to Module 7 likely play a conserved role in Poaceae in the ABA-mediated modulation of carbohydrate metabolism in response to water stress, as observed also in some dicotyledonous species (Pinheiro et al, 2011). Meanwhile, Module 14 comprises several genes related to the ABA hormone: PP2C, which upon inhibition by ABA, is a fundamental trigger in stress-related ABA signaling cascade (Umezawa et al, 2010); raffinose synthase [EC:2.4.1.82], which is involved in the biosynthesis of raffinose, an osmoprotectant associated to drought tolerance (Taji et al, 2002). Taken together, these results support the view that Module 14 is likely involved in a plant hormone transduction pathway related to ABA, necessary for the early onset of stomatal closure. Recent physiological analyses indicate that *A. donax* can fix CO2 at soil water contents close to wilting point, thanks to its ability to effectively control stomatal regulation in relation to soil water content (Cosentino et al, 2016). The association in our transcriptomics data of Module 7 and 14 to ABA-related pathways confirms and further extends this observation, indicating that in *A. donax* such regulation can be activated as early as 1 h after the onset of PEG-induced water stress and that it may contribute to the high adaptability of this species to resource-poor habitats and marginal soils (Lewandowski et al, 2003). Despite the incomplete understanding of the functions of the genes comprised in these clusters, selected members of both Module 7 and

14 could constitute, on one hand, suitable markers to dissect early stress responses and, on the other hand, promising targets to modulate drought tolerance in *A. donax*.

Interestingly, in the afore-mentioned study, only Modules 4, 7 and 14 are significantly associated to rice early responses to drought, where an experimental design similar to ours (3 hours treatment, two tissues) has been used (Zhang et al, 2012). This match supports the conservation of early drought response networks between *A. donax* and rice, two species associated to water-rich environments. A closer examination of the genes belonging to the latter three modules extends the possible conservation of drought-related regulatory networks even further: Among the 53 drought response genes in common among *A. donax*, rice, sorghum and foxtail (see next paragraph), 13.2% of genes (7 genes) are from Module 14. Considering that there are only 28 genes in Module 14, much less than the the others, especially this module seems to capture a particularly important drought-related mechanism across Poaceae species.

3.3.6 Identification of a core set of Poaceae genes differentially regulated upon water stress

The comparison of transcriptomes across different species can provide information about conservation of gene functions over evolutionary time. We, therefore, identified the subsets of water stress-related DEGs in common between *A. donax,* foxtail, sorghum, and rice. When the *A. donax* DEGs were compared with drought-responsive genes reported in previous studies (Qi et al. 2013; Dugas et al. 2011; Zhang et al. 2012), a total of 343, 496 and 143 putative orthologs were identified from foxtail, sorghum and rice, respectively (Figure 3.6; Supplementary Ch3.9). In total 53 groups of putative orthologs present in all species were identified, which constitute a core of evolutionarily conserved genes associated to early responses to water deprivation.

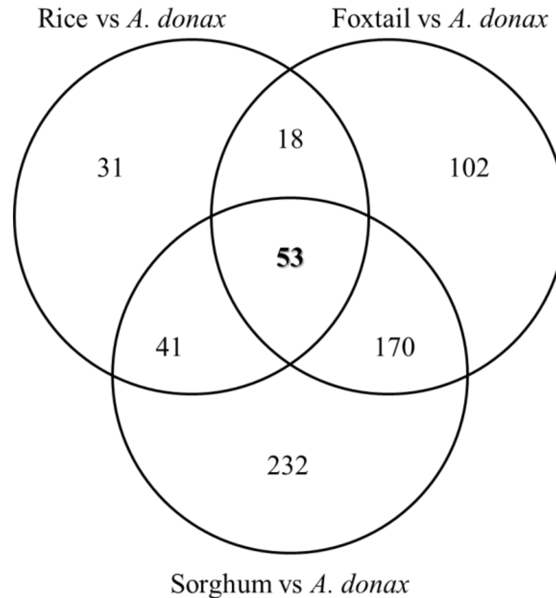


Figure 3.6. Drought response genes comparison across *A. donax*, rice, foxtail, and sorghum. The Venn diagram represents putative orthologs of *A. donax* stress-responsive genes identified by OrthoMCL in at least two species

Some of these genes are involved in the quality control and targeted degradation of proteins. Possibly the most striking example among them is a putative ortholog of *Arabidopsis* AT5G51070 gene, which codes for the ClpD subunit of the plastidial Hsp100/Clp complex, a caseinolytic protease (Clp) necessary for chloroplast biogenesis and protein homeostasis (Supplementary Ch3.9) (Bruch et al. 2012). This gene, also known as *Early Responsive to Dehydration 1* (*ERD1*), is an ATP-dependent molecular chaperone that likely directs unfolded polypeptides to the Clp complex for degradation (Colombo et al. 2014), suggesting that also in *A. donax* it could help eliminate damaged/misfolded proteins and aid proteome reprogramming upon water stress. This observation is supported by the fact that in rice the ClpD protein has been reported to be preferentially upregulated along with several other proteases in response to conditions of active drought signaling but water availability, which in shoots mimicks the early water stress of our study (Mirzaei et al. 2014).

Several of the other conserved DEGs participate in the biosynthesis of different metabolites, ranging from sugars to hormones, lipids and flavonoids (Supplementary Ch3.9). Among the most interesting *A. donax* candidates involved in sugar metabolism there is a

putative raffinose synthase protein homolog to *Arabidopsis* AT5G40390, the only isoform reported to be responsive to a wide array of abiotic stresses, including water deficit (Egert et al. 2013). Knockout mutants of AT5G40390 have reduced amounts of verbascose, sucrose and mannitol, but, surprisingly, their shoots are slightly less susceptible to prolonged drought than WT (Anderson et al. 2001). The *Arundo* homolog of this gene, however, is upregulated in roots, indicating that significant regulatory differences exist between species and suggesting that in the latter species this gene could contribute to short-term osmotic adjustment in roots.

In agreement with previous studies, also a certain number of membrane transporters are among the conserved genes involved in the early response to water stress(Supplementary Ch3.9). For instance, a homolog of *Arabidopsis* gene AT3G20300 is upregulated in both shoots and roots. In *Arabidopsis* this poorly characterized gene belongs to the monosaccharide transporter-like (MST-like) superfamily and codes for a predicted polyol/cyclitol/monosaccharide-H+-symporter of the mitochondrion, indicating that in *Arundo* its ortholog could be involved in the response to water deprivation by redistributing small organic solutes between cytosol and mitochondria. Another transporter which was upregulated in both shoots and roots is the homolog of *Arabidopsis* gene AT1G15520, coding for ABCG40, a plasma membrane ABA uptake transporter. In *Arabidopsis*, stomata of *abcg40* mutants respond more slowly to ABA and are less drought tolerant than WT plants (Kang et al. 2010). Interestingly, the putative ortholog of AT1G78390, nine-cis-epoxycarotenoid dioxygenase 9, a key enzyme in ABA biosynthesis (Lefebvre et al. 2006), is strongly upregulated in both *Arundo* shoots and roots during the early responses to water stress. Taken together, these results are in line with the established role of ABA as the main plant hormone in the early responses to water stress (Huang 2012). The identification of several differentially expressed ABA-responsive kinases and phosphatases allowed also the definition of a conserved core of signalling genes shared between *Arundo* and the other Poaceae considered. The putative ortholog of *Arabidopsis* AT4G33950 gene is strongly upregulated in water-stressed *Arundo* shoots. This gene is a member of SNF1-related protein kinases (SnRK2) responsive to both ionic and non-ionic osmotic stresses. Among the SnRK2 paralogs, AT4G33950 (also called SnRK2.6) is the most important for overall stomatal control (Virlouvet and Fromm 2015). In *Arabidopsis*, loss of function mutations of this gene

completely abolish ABA-mediated stomatal responses, leaving unaffected the ABA-independent reactions and resulting in increased drought susceptibility (Mustilli et al. 2002). Given the proposed involvement of this gene in the early phases of ABA perception before the development of reactive oxygen species associated to cell damage (Mustilli et al. 2002), the ortholog of AT4G33950 could constitute an interesting candidate to modulate the responsiveness to water stress responses in *Arundo* and be used as a sensitive marker for shoot drought stress. Two additional kinases are specifically upregulated in *Arundo* leaves subjected to water stress. The first one, homologous to gene AT1G70520, encodes a cystein-rich receptor-like protein kinase which in *Arabidopsis* has been shown to respond only weakly to ABA and other hormone treatments, but to be upregulated shortly after ozone treatment (Wrzaczek et al. 2010). The second one is a poorly characterized protein kinase homolog to AT1G56130, one of the four paralogous loci present as tandem duplications in the *Arabidopsis* genome. Possibly due to redundancy, very limited functional information is available about this small gene family. However, the early response and high levels of upregulation in water-stressed shoots of *Arundo* makes it an interesting candidate deserving further characterization. In addition to protein kinases, also a phosphatase homologous to *Arabidopsis* AT2G29380 gene, also called *Highly ABA-Induced1* (*HAI1*), is among the conserved Poaceae DEGs. Like several other water-stress clade A protein phosphatase 2Cs (PP2Cs), *HAI1* acts as a negative regulator of osmoregulatory solute accumulation. Unlike the majority of its closest paralogs, however, the HAI1 protein is largely insensitive to inhibition by members of the ABA receptors family (Bhaskara et al. 2012). The concomitant expression of *Arundo* homologs of HAI1 and SnRK2.6 (the latter acting downstream of the other ABA-receptor repressible PP2Cs; Soon et al, 2012) raises the interesting possibility that HAI1 may act antagonistically to SnRK2.6 to prevent excessive osmoregulatory solute accumulation. This hypothesis is supported by the fact that *Arabidopsis hai1* mutants accumulate higher amounts of proline and other osmoregulatory solutes than wild type plants (Bhaskara et al. 2012). Strikingly, among the conserved early water stress DEGs, the only transcription factor is a homolog of *ATHB7* (Arabidopsis gene AT2G46680), a member of class I plant-specific homeodomain-leucine zipper family (Johannesson et al. 2001). In *Arabidopsis*, *ATHB7* and its paralog *ATHB12* modulate abscisic acid signalling by regulating protein phosphatase 2C and abscisic acid receptor gene activities (Valdés et al. 2012).

Despite direct regulation of *HAI1* has not been tested, several other clade A *PP2Cs* are under positive transcriptional regulation by *ATHB7* / *ATHB12*, which at the same time repress transcription of genes from the PYR/PYL family of ABA receptors (Valdés et al. 2012). As, both in *Arabidopsis* and rice, paralogs with different tissue specific and developmental expression patterns have been implicated in different aspects of ABA-mediated growth responses to water stress (Harris et al. 2011), the characterization of *Arundo*'s homeodomain-leucine zipper family members seems to be a promising starting point to dissect the details of abscisic acid signaling modulation and stomatal control in this species.

3.4 Conclusion

The lack of information available about the molecular mechanisms involved in stress responses in *A. donax* is currently a major constraint for the further development of this semi-wild species into a fully fledged bionergy crop. To fill at least in part this gap, we hereby provided the first characterization of *A. donax* shoot and root transcriptomes in response to water stress, one of the factors of highest concern for its productivity. Given the commonality of the responses to water limitation and other stresses, in addition to providing a general overview of the early transcriptional responses to simulated drought, our results shed also light at the molecular level on the general mechanisms of stress response and adaptation in *A. donax*. Upon functional validation, thus, many of the unigenes identified in the present study have the potential to be used for the development of novel *A. donax* varieties with improved productivity and stress tolerance. In addition, the inventory of early-responsive genes to water stress provided in this study could constitute useful markers of the physiological status of *A. donax* plantations to deepen our understanding of its productivity under water limitation.

3.5 Materials and Methods

3.5.1 Plant material and application of water limitation stress

In the present study, we applied a water stress by treating cohorts of *A. donax* cuttings (collected in Sesto Fiorentino, Florence, Italy 43°49'01.8"N 11°11'57.0"E) with two different concentrations of polyethylene glycol 6000 (PEG; 10% and 20% w/w, referred to as mild and severe water stress conditions, corresponding to osmotic potentials of -1.54 bars and -

5.04 bars, respectively; Michel & Kaufmann 1973). Briefly, *A. donax* cuttings were let rooting in tap water, then transferred to 1% Hoagland solution and grown in a growth chamber with day-length of 16 hrs, light intensity of 150 µmol of photons m^{-2} s^{-1}, 24°C and 60% RH. At the three-leaves stage, two cohorts of cuttings were transferred to 1% Hoagland solution containing either 10% or 20% PEG, while a third cohort used as control was transferred to 1% Hoagland solution without PEG. After 1 hr the tissue above the third leaf and roots were separately collected from each cohort, quickly rinsed in distilled water and snap-frozen in liquid nitrogen. A total of 18 samples (three biological replicates from both shoot and root for each of the three conditions) were sampled.

3.5.2 Next generation RNA sequencing

Frozen root and the shoot tissues were grounded in liquid nitrogen with pre-cooled mortars and pestles. RNA isolation was carried out using the Spectrum Plant Total RNA Extraction Kit (Sigma) and Rneasy® Plant Mini Kit, respectively, for shoots and roots according to the manufacturer’s instructions. The quantity and the quality of the isolated RNA was evaluated using agarose gel electrophoresis and spectrophotometric measurements. Additionally, the isolated RNA was subjected to quality checks using the RNA 6000 Pico kit and the Agilent Bioanalyser 2100 (Agilent). Paired-end RNA-Seq libraries were prepared using the TruSeq RNA Sample Prep V2 kit (Illumina, San Diego, CA), pooled in equimolar ratio and sequenced on an Illumina HiSeq2000 sequencer (CIBIO NGS Facility, Povo (TN), Italy).

A minimum of 694 million reads were obtained from each of the 18 libraries sequenced (Supplementary Ch3.1). RNA-Seq data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number [ArrayExpress:E-MTAB-3769]. Assessment of read quality metrics was carried out using the FastQC software (available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) after which stringent quality filtering, removal of reads containing Ns and de-duplication was carried out as previously described (Sablok et al, 2014). For transcript reconstruction, we concatenated all read pairs passing the quality checks and assembled them using the Trinity software with-group-pair-distance=500 and −min-cov=5 to limit intron retention in reconstructed

transcripts (Grabherr et al, 2011). After discarding transcripts shorter than 200 bp, transcript redundancy was reduced using CD-HIT-EST with 95 % identity and a word size of 8 (Li and Godzik, 2006). The resulting non-redundant transcript dataset was further assembled into unigenes using the Overlap-Layout-Consensus assembler MIRA (parameters: job = denovo, est, accurate, 454 using the notraceinfo option) (Chevreux et al, 2004). All final *A. donax* unigenes are available for homology searches and download through a dedicated web Blast server (http://ecogenomics.fmach.it/arundo/) (Priyam et al, 2015).

3.5.3 WATER STRESS TRANSCRIPTOME ANNOTATION

Following the assembly, transcriptome curation was carried out by perfoming BLASTx searchers with E-value threshold 1E^{-5} against NCBI non-redundant (www.ncbi.nlm.nih.gov), UniProt (www.uniprot.org) and TrEMBL (Bairoch A, and Apweiler R., 2000) plant databases. Additionally, we also curated the unigenes using the FastAnnotator program, which integrates the functionality of BLAST2GO, PRIAM, domain identification and Gene Ontology classification (Chen et al. 2011). We further slimmed the obtained Gene Ontology (GO) categories using the PlantGO Slim categories available from the Gene Ontology consortium (www.geneontology.org) by collapsing the small child categories into broader classification of gene ontologies. Protein domains of transcriptome unigenes were identified using InterPro (https://www.ebi.ac.uk/interpro/). To identify the putative homologs of stress responsive genes characterized so far in *Arabidopsis thaliana*, BLASTx searches were performed with an E-value threshold of 1E^{-5} against the ASPRGDB database (Subhomoi Borkotoky et al, 2013), retaining only hits with query sequence coverage and identity higher than 50%. Additionally, we mined functionally relevant genes involved in drought stress by creating a customized, manually curated database from *Sorghum bicolor*, *Zea mays*, *Arabidopsis thaliana*, and *Oryza sativa*.

3.5.4 IDENTIFICATION AND FUNCTIONAL CLASSIFICATION OF DIFFERENTIALLY EXPRESSED GENES

To identity genes which are differentially expressed upon water stress, reads from each of the 18 libraries were individually mapped on the unigene assembly and fragments per kilobase of exon per million fragments mapped (FPKM) values were estimated as a

measure of the expression using RSEM (Bo and Li, 2011). For the identification of differentially expressed genes, we used EdgeR (R version: 3.0.1, edgeR version: 3.4.2; Robinson et al, 2010), implementing the Generalized Linear Model (GLM) (Smyth and Verbyla, 1996) approach. For the normalization of the read count, we applied the trimmed mean of M-values (TMM) normalization method (Robinson and Oshlack, 2010). Additionally contrasts were made to identify the set of differentially expressed genes between mild and severe water stress among the induced treatments. A false discovery rate (FDR) cutoff of 0.001 and a log-fold change (LogFC) threshold of 2 was implemented to filter the statistically significant up- and down-regulated genes between the treatment and the control. The genes with logFC>=2 and logFC<=-2 with a FDR cutoff of FDR=0.001 between two treatment conditions were determined to be up-regulated and down-regulated, respectively. All the statistically significant up- and down-regulated differentially expressed genes were custom annotated against the functionally identified drought- responsive genes in model grass clade. To select the most interesting candidates for functional studies, genes were first ranked all DEGs according to absolute difference of FPKM values and logFC between control and each of the stress conditions. To further select genes with the highest consistency in DE among biological replicates, only genes with CV < 0.7 and |logFC| >= 2 were retained, and the top 20 genes from either top- or down-regulated lists were highlighted in the results. For the identification of transcription factors responsive to water stress in *A. donax* we mined the Stress Responsive Transcriptio Factor Database of rice (SRTFDB; Priya et al, 2013) by Blastn searches with an E-value cutoff of 1e-5. For the identification of the subsets of water stress-related DEGs in common between *A. donax*, foxtail, sorghum and rice, OrthoMCL software V5 was used with default settings (Li et al, 2003). For the comparative study between *A. donax* and rice co-regulation network, we identificated putative orthologs between the *A. donax* and rice with Blastp Reciprocal Best hits method (RBH, E-value $1e^{-6}$; Moreno-Hagelsieb and Latimer, 2008).

3.5.5 GO ENRICHMENT

Blast2GO was also used for a GO functional enrichment analysis of certain genes, by performing Fisher's exact test with a robust FDR (<0.05) correction to obtain an adjusted p-value between certain test gene groups and the whole transcriptome annotation. To provide

a more comprehensive interpretation and of GO data we have used the freely available web-based software REVIGO (Supek et al, 2011). We uploaded the lists of over-represented GO ID along with the p-value from the result of the fisher's exact test. The analysis was run by selecting the small size of the resulting list, with the numbers associated to GO categories p-values, with the Oryza sativa database, and the SimRel as the semantic similarity measure. The GO terms which were over-represented under different conditions were slimmed by REVIGO (Supplementary Ch3.6). The number of over-represented terms in common between conditions was displayed as Venn diagram (Figure 3.3: a. molecular function; b. biological process. MC: mild water stress *vs* control, SC: severe water stress *vs* control; SM: severe water stress *vs* mild water stress).

3.5.6 Pathway enrichment

To identify functionally relevant patterns associated to water stress in shoot and root DEGs, we created a unigene dataset for each organ discarding genes with FPKM ≤1. Each dataset was subsequently used as background to identify over- and under-represented GO categories among DEGs using the fischer's test with a p-value cut-off of 0.05. In addition, pathway enrichment analysis of DEGs were carried out with the KOBAS software (Xie et al, 2011) using BLASTx searches against the *Oryza sativa* var. *japonica* proteins.

3.5.7 Real-time validation of selected DEG candidates using qRT-PCR

Each RNA sample was treated with DNase I (Sigma-Aldrich) and 1 µg of total RNA was reversed transcribed using the SuperScript® III Reverse Transcriptase (Life Technologies), according to the manufacturer's instructions. Real time qRT-PCR was performed for a total of 10 DEGs with Platinum® SYBR® Green qPCR SuperMix-UDG and carried out in the Bio-Rad C1000 Thermal Cycler detection system according to the manufacturer's instructions. All the genes were normalized with putative *A. donax* actin protein with highest homology to sorghum AC1 gene (GenBank accession no. P53504). Each PCR reaction (12,5 µL) contained 11 µL real-time PCR Mix, 0.25 µM of each primer and 1 µl of a 1:5 dilution of cDNA. The thermal cycling conditions were 50 °C for 2 min, 95 °C for 2 min, followed by 40 cycles of 15 s at 95 °C and 30 s at 60 °C. All reactions were performed in triplicate and fold change measurements calculated with the $2^{-\Delta\Delta CT}$ method

(Livak & Schmittgen, 2001). Sequences of primers used for real time PCR are provided in Supplementary Ch3.2.

3.6 REFERENCES

Abe H, Urao T, Ito T, Seki M, Shinozaki K. Transcriptional activators in abscisic acid signaling. Plant Cell. 2003;15:63–78.

Acevedo-Hernández GJ, León P, Herrera-Estrella LR. Sugar and ABA responsiveness of a minimal RBCS light-responsive unit is mediated by direct binding of ABI4. Plant J. 2005;43:506–19.

Acres G. Alcoholic Fuels. Platin Met Rev. 2007;51:34–5.

Ambavaram MMR, Krishnan A, Trijatmiko KR, Pereira A. Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. Plant Physiol. 2011;155:916–31.

Anderson CM, Kohorn BD: Inactivation of Arabidopsis SIP1 leads to reduced levels of sugars and drought tolerance. J Plant Physiol 2001, 158:1215–9.

Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 2000;28:45–8.

Balogh E, Herr JM, Czakó M, Márton L. Defective development of male and female gametophytes in Arundo donax L. (POACEAE). Biomass and Bioenergy. 2012;45:265–9.

Bari R, Jones JDG. Role of plant hormones in plant defence responses. Plant Mol Biol. 2009;69:473–88.

Benešová M, Holá D, Fischer L, Jedelský PL, Hnilička F, Wilhelmová N, Rothová O, Kočová M, Procházková D, Honnerová J, Fridrichová L, Hniličková H. The physiology and proteomics of drought tolerance in maize: early stomatal closure as a cause of lower tolerance to short-term dehydration? PLoS One. 2012;7:e38017.

Bhaskara GB, Nguyen TT, Verslues PE. Unique drought resistance functions of the highly ABA-induced clade A Protein Phosphatase 2Cs. Plant Physiol. 2012;160:379–95.

Borkotoky S, Saravanan V, Jaiswal A, Das B, Selvaraj S, Murali A, Lakshmi PT V. The Arabidopsis stress responsive gene database. Int J Plant Genomics. 2013;2013:949564.

Bruce WB, Edmeades GO, Barker TC. Molecular and physiological approaches to maize improvement for drought tolerance. J Exp Bot. 2002;53:13–25.

Bruch EM, Rosano GL, Ceccarelli EA. Chloroplastic Hsp100 chaperones ClpC2 and ClpD interact in vitro with a transit peptide only when it is located at the N-terminus of a protein. BMC Plant Biol. 2012;12:57.

Bucci A, Cassani E, Landoni M, Cantaluppi E, Pilu R. Analysis of chromosome number and speculations on the origin of Arundo donax L. (Giant Reed). Cytol Genet. 2013;47:237–41.

Byrt CS, Grof CPL, Furbank RT. C4 plants as biofuel feedstocks: optimising biomass production and feedstock quality from a lignocellulosic perspective. J Integr Plant Biol. 2011;53:120–35.

Chen TW, Gan RCR, Wu TH, Huang PJ, Lee CY, Chen YYM, Chen CC, Tang P. FastAnnotator--an efficient transcript annotation web tool. BMC Genomics. 2012;13:S9.

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 2004;14:1147–59.

Claeys H, Inzé D. The agony of choice: how plants balance growth and survival under water-limiting conditions. Plant Physiol. 2013;162:1768–79.

Cohen D, Bogeat-Triboulot MB, Tisserant E, Balzergue S, Martin-Magniette ML, Lelandais G, Ningre N, Renou J-P, Tamby J-P, Le Thiec D, Hummel I. Comparative transcriptomics of drought responses in Populus: a meta-analysis of genome-wide expression profiling in mature leaves and root apices across two genotypes. BMC Genomics. 2010;11:630.

Colasanti J, Tremblay R, Wong AYM, Coneva V, Kozaki A, Mable BK. The maize INDETERMINATE1 flowering time regulator

defines a highly conserved zinc finger protein family in higher plants. BMC Genomics. 2006;7:158.
Colombo C V, Ceccarelli E a, Rosano GL. Characterization of the accessory protein ClpT1 from Arabidopsis thaliana: oligomerization status and interaction with Hsp100 chaperones. BMC Plant Biol. 2014;14:228.
Dai A. Increasing drought under global warming in observations and models. Nat Clim Chang. 2012;3:52–8.
Daszkowska-Golec A, Szarejko I. Open or close the gate - stomata action under the control of phytohormones in drought stress conditions. Front Plant Sci. 2013;4:138.
Davidson RM, Gowda M, Moghe G, Lin H, Vaillancourt B, Shiu S-H, Jiang N, Robin Buell C. Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. Plant J. 2012;71:492–502.
Degenkolbe T, Do PT, Zuther E, Repsilber D, Walther D, Hincha DK, Köhl KI. Expression profiling of rice cultivars differing in their tolerance to long-term drought stress. Plant Mol Biol. 2009;69:133–53.
Deyholos MK. Making the most of drought and salinity transcriptomics. Plant, Cell Environ. 2010;33:648–54.
Di Tomaso JM., Healey EA. Aquatic and riparian weeds of the west. University of California – Agriculture and Natural Resources Publication; 2003. p. 254–62.
Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. Cell. 2006;1309–21.
Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. OsDREB genes in rice, Oryza sativa L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. Plant J. 2003;33:751–63.
Dugas DV, Monaco MK, Olsen A, Klein RR, Kumari S, Ware D, Klein PE. Functional annotation of the transcriptome of Sorghum bicolor in response to osmotic stress and abscisic acid. BMC Genomics. 2011;12:514.
Edenhofer O, Pichs-Madruga R, Sokona Y, Farahani E, Kadner S, Seyboth K, Adler A, Baum I, Brunner S, Eickemeier P, Kriemann B, Savolainen J, Schlömer S, von Stechow C, Zwickel T, Minx JC. Climate change 2014: Mitigation of climate change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press; 2014.
Egert A, Keller F, Peters S: Abiotic stress-induced accumulation of raffinose in Arabidopsis leaves is mediated by a single raffinose synthase (RS5, At5g40390). BMC Plant Biol 2013, 13:218.
European Commission (2009) Directive 2009/28/EC of the European Parliament and of the Council of 23 April 2009 on the promotion of the use of energy from renewable sources and amending and subsequently repealing directives 201/77/EC and 2003/30/EC (Text with EEA relevance). In European Parliament, C.o.t.M.S (ed) European Union
Ferreira de Carvalho J, Poulain J, Da Silva C, Wincker P, Michon-Coudouel S, Dheilly A, Naquin D, Boutte J, Salmon A, Ainouche M. Transcriptome de novo assembly from next-generation sequencing and comparative analyses in the hexaploid salt marsh species Spartina maritima and Spartina alterniflora (Poaceae). Heredity (Edinb). 2013;110:181–93.
Fuchs M, Vothknecht UC. Calmodulin and Calmodulin-like proteins in Arabidopsis thaliana. J Endocytobiosis Cell Res. 2011;21:114–21.
Gholizadeh A. Effects of drought on the activity of Phenylalanine ammonia lyase in the leaves and roots of maize inbreds. Aust J Basic Appl Sci. 2011;5: 952–6.
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.
Gruenheit N, Deusch O, Esser C, Becker M, Voelckel C, Lockhart P. Cutoffs and k-mers: implications from a transcriptome study in allopolyploid plants. BMC Genomics. 2012;13:92.
Guerra D, Crosatti C, Khoshro HH, Mastrangelo AM, Mica E, Mazzucotelli E. Post-transcriptional and

post-translational regulations of drought and heat response in plants: a spider's web of mechanisms. Front Plant Sci. 2015;6:57.

Haddadchi A, Gross CL, Fatemi M. The expansion of sterile Arundo donax (Poaceae) in southeastern Australia is accompanied by genotypic variation. Aquat Bot. 2013;104:153–61.

Hadiarto T, Tran L-SP. Progress studies of drought-responsive genes in rice. Plant Cell Rep. 2011;30:297–310.

Hardion L, Verlaque R, Rosato M, Rosselló JA, Vila B. Impact of polyploidy on fertility variation of Mediterranean Arundo L. (Poaceae). C R Biol. 2015;338:298–306.

Hardion L, Verlaque R, Saltonstall K, Leriche A, Vila B. Origin of the invasive Arundo donax (Poaceae): a trans-Asian expedition in herbaria. Ann Bot. 2014;114:455–62.

Harris JC, Hrmova M, Lopato S, Langridge P. Modulation of plant growth by HD-Zip class I and II transcription factors in response to environmental stimuli. New Phytol. 2011;190:823–37.

Higgins JA, Bailey PC, Laurie DA. Comparative genomics of flowering time pathways using Brachypodium distachyon as a model for the temperate grasses. PLoS One. 2010;5:e10065.

Hu H, Dai M, Yao J, Xiao B, Li X, Zhang Q, Xiong L. Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. Proc Natl Acad Sci U S A. 2006;103:12987–92.

Hu H, You J, Fang Y, Zhu X, Qi Z, Xiong L. Characterization of transcription factor gene SNAC2 conferring cold and salt tolerance in rice. Plant Mol Biol. 2008;67:169–81.

Huang GT, Ma SL, Bai LP, Zhang L, Ma H, Jia P, Liu J, Zhong M, Guo Z-F. Signal transduction during cold, salt, and drought stresses in plants. Mol Biol Rep. 2012;39:969–87.

Jeong JS, Kim YS, Baek KH, Jung H, Ha S-H, Do Choi Y, Kim M, Reuzeau C, Kim J-K. Root-specific expression of OsNAC10 improves drought tolerance and grain yield in rice under field drought conditions. Plant Physiol. 2010;153:185–97.

Johannesson H, Wang Y, Engström P. DNA-binding and dimerization preferences of Arabidopsis homeodomain-leucine zipper transcription factors in vitro. Plant Mol Biol. 2001;45:63–73.

Jørgensen U. Benefits versus risks of growing biofuel crops: the case of Miscanthus. Curr Opin Environ Sustain. 2011;3:24–30.

Kakumanu A, Ambavaram MMR, Klumas C, Krishnan A, Batlang U, Myers E, Grene R, Pereira A. Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-Seq. Plant Physiol. 2012;160:846–67.

Kang J, Hwang JU, Lee M, Kim YY, Assmann SM, Martinoia E, Lee Y. PDR-type ABC transporter mediates cellular uptake of the phytohormone abscisic acid. Proc Natl Acad Sci U S A. 2010;107:2355–60.

Kido ÉA, Neto JRCF, Silva RLD, Pandolfi V, Guimarães ACR, Veiga DT, Chabregas SM, Crovella S, Benko-Iseppon AM. New insights in the sugarcane transcriptome responding to drought stress as revealed by superSAGE. ScientificWorld Journal. 2012;2012:821062.

Lagerwerff JV, Ogata G, Eagle HE. Control of osmotic pressure of culture solutions with polyethylene glycol. Science. 1961;133:1486–7.

Lal R. Soil quality impacts of residue removal for bioethanol production. Soil Tillage Res. 2009;102:233–41.

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10:R25.

Lefebvre V, North H, Frey A, Sotta B, Seo M, Okamoto M, Nambara E, Marion-Poll A. Functional analysis of Arabidopsis NCED6 and NCED9 genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. Plant J. 2006;45:309–19.

Lewandowski I, Clifton-Brown JC. Miscanthus: European experience with a novel energy crop. Biomass and Bioenergy. 2000;19:209–27.

Lewandowski I, Scurlock JMO, Lindvall E, Christou M. The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. Biomass and Bioenergy. 2003;25:335–61.

Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-seq data with or without a

reference genome. BMC Bioinformatics. 2011;12:323.

Li F, Tsfadia O, Wurtzel ET. The phytoene synthase gene family in the Grasses: subfunctionalization provides tissue-specific control of carotenogenesis. Plant Signal Behav. 2009;4:208–11.

Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003;13:2178–89.

Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9.

Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012:251364.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods San Diego Calif. 2001;25:402–8.

Lynch JP, Chimungu JG, Brown KM. Root anatomical phenes associated with water acquisition from drying soil: targets for crop improvement. J Exp Bot. 2014;65:6155–66.

Mariani C, Cabrini R, Danin A, Piffanelli P, Fricano A, Gomarasca S, Dicandilo M, Grassi F, Soave C. Origin, diffusion and reproduction of the giant reed (Arundo donax L.): a promising weedy energy crop. Ann Appl Biol. 2010;157:191–202.

Marino D, Dunand C, Puppo A, Pauly N. A burst of plant NADPH oxidases. Trends Plant Sci. 2012;17:9–15.

Martin LBB, Fei Z, Giovannoni JJ, Rose JKC. Catalyzing plant science research with RNA-seq. Front Plant Sci. 2013;4:66.

Matsukura S, Mizoi J, Yoshida T, Todaka D, Ito Y, Maruyama K, Shinozaki K, Yamaguchi-Shinozaki K. Comprehensive analysis of rice DREB2-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. Mol Genet Genomics. 2010;283:185–96.

Mavrogianopoulos G, Vogli V, Kyritsis S. Use of wastewater as a nutrient solution in a closed gravel hydroponic culture of giant reed (Arundo donax). Bioresour Technol. 2002;82:103–7.

Michel BE, Kaufmann MR. The osmotic potential of polyethylene glycol 6000. Plant Physiol. 1973;51:914–6.

Mirzaei M, Soltani N, Sarhadi E, George IS, Neilson KA, Pascovici D, Shahbazian S, Haynes PA, Atwell BJ, Salekdeh GH. Manipulating root water supply elicits major shifts in the shoot proteome. J Proteome Res. 2014;13:517–26.

Mohammadkhani N, Heidari R. Drought-induced accumulation of soluble sugars and proline in two maize varieties. World Appl Sci J. 2008;3:448–53.

Moreno-Hagelsieb G, Latimer K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatics. 2008;24:319–24.

Moumeni A, Satoh K, Kondoh H, Asano T, Hosaka A, Venuprasad R, Serraj R, Kumar A, Leung H, Kikuchi S. Comparative analysis of root transcriptome profiles of two pairs of drought-tolerant and susceptible rice near-isogenic lines under different drought stress. BMC Plant Biol. 2011;11:174.

Mustilli A-C, Merlot S, Vavasseur A, Fenzi F, Giraudat J. Arabidopsis OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. Plant Cell. 2002;14:3089–99.

Nakano T, Suzuki K, Fujimura T, Shinshi H. Genome-wide analysis of the ERF gene family in Arabidopsis and rice. Plant Physiol. 2006;140:411–32.

Narsai R, Wang C, Chen J, Wu J, Shou H, Whelan J. Antagonistic, overlapping and distinct responses to biotic stress in rice (Oryza sativa) and interactions with abiotic stress. BMC Genomics. 2013;14:93.

Nijhawan A, Jain M, Tyagi AK, Khurana JP. Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. Plant Physiol. 2008;146:333–50.

Nuruzzaman M, Sharoni AM, Kikuchi S. Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. Front Microbiol. 2013;4:248.

Ohme-Takagi M, Shinshi H. Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. Plant Cell. 1995;7:173–82.

Overvoorde P, Fukaki H, Beeckman T. Auxin control of root development. Cold Spring Harb Perspect Biol. 2010;2:a001537.

Perdue R. Arundo donax—source of musical reeds and industrial cellulose. Econ Bot. 1958;12:368–404.

Pilu R, Cassani E, Landoni M, Badone FC, Passera A, Cantaluppi E, Corno L, Adani F. Genetic characterization of an Italian Giant Reed (Arundo donax L.) clones collection: exploiting clonal selection. Euphytica. 2013;196:169–81.

Pinheiro C, António C, Ortuño MF, Dobrev PI, Hartung W, Thomas-Oates J, Ricardo CP, Vanková R, Chaves MM, Wilson JC. Initial water deficit effects on Lupinus albus photosynthetic performance, carbon metabolism, and hormonal balance: metabolic reorganization prior to early stress responses. J Exp Bot. 2011;62:4965–74.

Polunin O, Huxley A. Flowers of the Mediterranean. London: Hogarth Press; 1987.

Priya P, Jain M. RiceSRTFDB: a database of rice transcription factors containing comprehensive expression, cis-regulatory element and mutant information to facilitate gene function analysis. Database (Oxford). 2013;2013:bat027.

Priyam A, Woodcroft BJ, Rai V & Wurm Y (in prep) SequenceServer: BLAST search made easy. http://sequenceserver.com

Qi X, Xie S, Liu Y, Yi F, Yu J. Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. Plant Mol Biol. 2013;83:459–73.

Rashid M, Guangyuan H, Guangxiao Y, Hussain J, Xu Y. AP2/ERF Transcription Factor in Rice: Genome-Wide Canvas and Syntenic Relationships between Monocots and Eudicots. Evol Bioinform Online. 2012;8:321–55.

Ray S, Dansana PK, Giri J, Deveshwar P, Arora R, Agarwal P, Khurana JP, Kapoor S, Tyagi AK. Modulation of transcription factor and metabolic pathway genes in response to water-deficit stress in rice. Funct Integr Genomics. 2011;11:157–78.

Riechmann JL, Meyerowitz EM. The AP2/EREBP family of plant transcription factors. Biol Chem. 1998;379:633–46.

Robinson MD, McCarthy DJ and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139-40.

Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11:R25.

Sablok G, Fu Y, Bobbio V, Laura M, Rotino GL, Bagnaresi P, Allavena A, Velikova V, Viola R, Loreto F, Li M, Varotto C. Fuelling genetic and metabolic exploration of C3 bioenergy crops through the first reference transcriptome of Arundo donax L. Plant Biotechnol J. 2014;12:554–67.

Sakuma Y, Liu Q, Dubouzet JG, Abe H, Shinozaki K, Yamaguchi-Shinozaki K. 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. 2002;290:998–1009.

Seo JS, Joo J, Kim MJ, Kim YK, Nahm BH, Song SI, Cheong J-J, Lee JS, Kim J-K, Choi Y Do. OsbHLH148, a basic helix-loop-helix protein, interacts with OsJAZ proteins in a jasmonate signaling pathway leading to drought tolerance in rice. Plant J. 2011;65:907–21.

Shanker AK, Maheswari M, Yadav SK, Desai S, Bhanu D, Attal NB, Venkateswarlu B. Drought stress responses in crops. Funct Integr Genomics. 2014;14:11–22.

Smyth GK, Verbyla AP. A conditional likelihood approach to REML in Generalized Linear Models. J R Stat Soc Ser B. 1996;58:565–72.

Soon FF, Ng LM, Zhou XE, West GM, Kovach A, Tan MH, Suino-Powell KM, He Y, Xu Y, Chalmers MJ, Brunzelle JS, Zhang H, Yang H, Jiang H, Li J, Yong EL, Cutler S, Zhu JK, Griffin PR, Melcher K, Xu HE. Molecular mimicry regulates ABA signaling by SnRK2 kinases and PP2C phosphatases. Science 2012;335:85-8.

Sperotto RA, Ricachenevsky FK, Duarte GL, Boff T, Lopes KL, Sperb ER, Grusak MA, Fett JP. Identification of up-regulated genes in flag leaves during rice grain filling and characterization of OsNAC5, a new ABA-dependent transcription factor. Planta. 2009;230:985–1002.

Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of Gene Ontology terms. PLoS One. 2011;6: e21800.

Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K. Important roles of drought- and cold-inducible genes for galactinol synthase in stress tolerance in Arabidopsis thaliana. Plant J. 2002;29:417–26.

Umezawa T, Nakashima K, Miyakawa T, Kuromori T, Tanokura M, Shinozaki K, Yamaguchi-Shinozaki K. Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. Plant Cell Physiol. 2010;51:1821–39.

Valdés AE, Övernäs E, Johansson H, Rada-Iglesias A, Engström P. The homeodomain-leucine zipper (HD-Zip) class I transcription factors ATHB7 and ATHB12 modulate abscisic acid signalling by regulating protein phosphatase 2C and abscisic acid receptor gene activities. Plant Mol Biol. 2012;80:405–18.

Valliyodan B, Nguyen HT. Understanding regulatory networks and engineering for enhanced drought tolerance in plants. Curr Opin Plant Biol. 2006;9:189–95.

Verbruggen N, Hermans C. Proline accumulation in plants: A review. Amino Acids. 2008;35:753–9.

Verslues PE, Agarwal M, Katiyar-Agarwal S, Zhu J, Zhu JK. Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. Plant J. 2006;45:523–39.

Virlouvet L, Fromm M. Physiological and transcriptional memory in guard cells during repetitive dehydration stress. New Phytol. 2015;205:596–607.

Wang D, Pan Y, Zhao X, Zhu L, Fu B, Li Z. Genome-wide temporal-spatial gene expression profiling of drought responsiveness in rice. BMC Genomics. 2011;12:149.

Wang Q, Guan Y, Wu Y, Chen H, Chen F, Chu C. Overexpression of a rice OsDREB1F gene increases salt, drought, and low temperature tolerance in both Arabidopsis and rice. Plant Mol Biol. 2008;67:589–602.

Wrzaczek M, Brosché M, Salojärvi J, Kangasjärvi S, Idänheimo N, Mersmann S, Robatzek S, Karpiński S, Karpińska B, Kangasjärvi J. Transcriptional regulation of the CRK/DUF26 group of receptor-like protein kinases by ozone and plant hormones in Arabidopsis. BMC Plant Biol. 2010;10:95.

Xiang Y, Tang N, Du H, Ye H, Xiong L. Characterization of OsbZIP23 as a key player of the basic leucine zipper transcription factor family for conferring abscisic acid sensitivity and salinity and drought tolerance in rice. Plant Physiol. 2008;148:1938–52.

Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li CY, Wei L. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;39:W316–22.

Xie F, Stewart CN, Taki F a, He Q, Liu H, Zhang B. High-throughput deep sequencing shows that microRNAs play important roles in switchgrass responses to drought and salinity stress. Plant Biotechnol J. 2014;12:354–66.

Xu K, Chen S, Li T, Ma X, Liang X, Ding X, Liu H, Luo L. OsGRAS23, a rice GRAS transcription factor gene, is involved in drought stress response through regulating expression of stress-responsive genes. BMC Plant Biol. 2015;15:141.

Yadav NR, Taunk J, Rani A, Aneja B, Yadav RC. Role of Transcription Factors in Abiotic Stress Tolerance in Crop Plants. In: Tuteja N, Gill SS, editors. Climate change and plant abiotic stress tolerance. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA; 2013..

Yoshiba Y, Kiyosue T, Nakashima K, Yamaguchi-Shinozki K, Shinozaki K. Regulation of levels of proline as an osmolytes in plants under water stress. Plant Cell Physiol.1997;38:1095-102.

Yoshida R, Hobo T, Ichimura K, Mizoguchi T, Takahashi F, Aronso J, Ecker JR, Shinozaki K. ABA-activated SnRK2 protein kinase is required for dehydration stress signaling in Arabidopsis. Plant Cell Physiol. 2002;43:1473–83.

Yun KY, Park MR, Mohanty B, Herath V, Xu F, Mauleon R, Wijaya E, Bajic VB, Bruskiewich R, de Los Reyes BG. Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of japonica rice to chilling stress. BMC Plant Biol. 2010;10:16.

Zhang L, Yu S, Zuo K, Luo L, Tang K. Identification of gene modules associated with drought response in

rice by network-based analysis. PLoS One. 2012;7:e33748.
Zhang M, Lv D, Ge P, Bian Y, Chen G, Zhu G, Li X, Yan Y. Phosphoproteome analysis reveals new drought response and defense mechanisms of seedling leaves in bread wheat (Triticum aestivum L.). J Proteomics. 2014;109:290–308.
Zhang Q. Strategies for developing Green Super Rice. Proc Natl Acad Sci U S A. 2007;104:16402–9.
Zheng B, Halperin T, Hruskova-Heidingsfeldova O, Adam Z, Clarke AK. Characterization of Chloroplast Clp proteins in Arabidopsis: Localization, tissue specificity and stress responses. Physiol Plant. 2002;114:92–101.

COMPARATIVE STUDY ON EARLY TRANSCRIPTIONAL DROUGHT RESPONSE BETWEEN *A. DONAX* AND *P. AUSTRALIS*

4

> "Somewhere, something incredible is waiting to be known."
>
> — Carl Sagan

The work described in this chapter was done in collaboration with Michele Poli, who is going to complete a Ph.D at Bologna University. Michele performed all the experimental work in the wet lab.

4.1 ABSTRACT

P. australis is a closely related species of *A. donax*. In a previous study, we reported the first characterization of *A. donax* transcriptome in response to water stress. To understand the commonalities and differences of water stress response mechanisms in Arundineae, a comparative study has been carried out. In this study, the same experimental setting has been conducted; we obtained by RNA-seq the whole root and shoot transcriptome of young *P. australis* plants subjected to water stress with 10% and 20% polyethylene glycol (PEG), and identified a total of 4622 differentially expressed genes. As the first comparative study of *A. donax* and *P. australis* transcriptome in response to water stress, this study shed novel light at the molecular level on general response mechanisms of wild perennial grasses versus those that are specifically used as energy crops. The operational definition of "drought tolerance" can in fact dramatically change meaning when productivity instead of survival is the main parameter taken into consideration. Another interesting aspect of this study is that this is one of the first in-depth transcriptome comparisons between species without genome information and as such it could constitute a useful reference and stimulus for further comparative transcriptome studies on other species.

4.2 INTRODUCTION

Drought is the single most important environmental stress for crops, causing severe effects on plant growth rate and biomass accumulation (Farooq et al, 2009). It causes more yield losses than any other single biotic or abiotic factor (Boyer, 1982). The European Commission in 2012 reported the increase of the number of people and areas hit by drought between 1976 and 2006 by almost 20% and the total drought losses over the EU in the past 30 years are estimated to be 100 billion euro (Kindler and Okruszko, 2014). Also in US, drought caused a 20% decrease of the yield of corn and soybean (Zulauf 2012). In addition, the previous study showed that 2 weeks of drought could cause more than 50% of biomass reduction for *A. donax* (Mann et al, 2013). The severity of drought is hard to predict, because of lacking of direct observations, geographical inconsistencies in the trends and dependencies of inferred trends on the index choice (IPCC2013). However, it is likely that the frequency and intensity of drought has increased in the Mediterranean and West Africa since 1950, and until the end of the 21st century there will be an increased risk of drought linked with decreased soil moisture and increased surface temperatures, i.e. the drought will be more frequent and intense in some seasons and areas, e.g. Mediterranean, Southwest USA, and southern Africa (IPCC2013). In this scenario, genetic improvement of crops for drought tolerance becomes a priority to guarantee constant yields; it requires investigation of the possible components of drought tolerance and exploration of the crop genetic variability associated to this stress. In our previous study (Fu et al, 2016), we reported the early drought response transcriptome of *A. donax*, a promising energy crop with high biomass yield (35~40 t h^{-1} y^{-1}). Energy crops, as one of the sources of renewable energy, can contribute to substitute the fossil energy (petrol and coal). This, in turn, helps reducing the direct Greenhouse Gases (GHG) emission, thus contributing to mitigation of global warming. In this study, we report a comparative study on early drought response between *A. donax* and the closely related species *P. australis*.

P. australis is a large perennial grass, 2–6 meters tall, with dry biomass yield 5~20 t h^{-1} y^{-1} (Pilu et al. 2012; Köbbing JF, Thevs N and Zerbe S, 2013/2014). Commonly described as C3 plant; also a C3–C4 intermediate ecotype has been reported in saline environment (Zheng et al. 2000), which makes it more flexible and allows it to better cope with changing environmental conditions (Srivastava et al. 2012). It perhaps has the largest geographical

distribution of any flowering plant in the World: it is occurring on all continents except Antarctica (Brix, 1999 and Clevering and Lissner, 1999). Compared with *A. donax*, *P. australis* prefers wetland areas, such as swamps, lakes and also the sea shores. Its roots can grow to extreme lengths, allowing the plant to survive with lower water levels by reaching water deep below ground. Furthermore, *P. australis* reproduces both sexually and asexually and it has a complex population structure (Nguyen et al, 2013). In contrast, *A. donax* has low genetic diversity and it reproduces only via asexual fragments; molecular studies, for instance, showed that there is a single genetic clone of *A. donax* in US. Both *A. donax* and *P. australis* have been traditionally included in the Arundineae tribe. While a recent phylogenetic study separated *Arundo* and *Phragmites* genera into two sister tribes (Soreng et al, 2015), these two species are very close both morphologically and phylogenetically. A previous study raised a hypothesis regarding the origin of *A. donax* as interspecific cross between two closely related Arundineae species, *A. plinii* and *P. australis*. In particular, it has been hypothesized that "the tetraploid crossed with *Phragmites australis* (96 chromosomes), another species similar to *A. plinii*, producing a sterile hybrid (allopolyploid) with 120 chromosomes that lost 10 chromosomes (aneuploid) to create *A. donax*" (A. Bucci et al., 2013). Despite this hypothesis can possibly be rejected based on more recent studies (Hardion et al. 2014), the relatively close phylogenetic relationship of these two species makes them an interesting model for the comparative dissection of the common/unique drought response mechanisms in the Arundineae tribe. Taking advantage of Illumina-based RNA-Seq, whole transcriptome comparison between these two species, under the very same experimental design, has been conducted. We raised our hypothesis based on the concepts introduced by Chaves et al. 2002 and Ingram & Bartels 1996 that these two plants may be using different drought response strategies, "avoidance" for *P. australis*, and "tolerance" for *A. donax*. According to this hypothesis, the main *P. australis* response to drought would be that of preventing the exposure to stress and thus avoid the associated damages, e. g., by growing deeper the roots, and trigger leaf senescence to enter a quiescence state until the end of the drought. Instead, *A. donax* would set into action mainly physiological adjustments to allow the plant to withstand stress, e. g., by regulating stomata by the ABA signaling pathway.

The comparative transcriptomic approach we adopted may help to provide a view of the drought general response mechanisms of wild perennial grasses *versus* those that are specifically used as energy crops. The operational definition of “drought tolerance” can in fact dramatically change meaning when productivity instead of survival is the main parameter taken into consideration. Another interesting aspect of this study is that this is one of the first in-depth transcriptome comparisons between species without genome information and as such, it could constitute a useful reference and stimulus for further comparative transcriptome studies on other species.

4.3 Results and discussion

4.3.1 Identification and functional characterization of differentially expressed genes (DEGs) of *P. australis* by RNA-Seq

We applied to *P. australis* the same experimental setup for water stress that we used in our previous study for *A. donax*, i.e., by application of PEG to the root system of hydroponically-grown cohorts of cuttings as in (Fu et al. 2016). The expression profiles of genes involved in the water stress response in shoot and root, especially during the initial stages of the water stress response, were characterized. Like in our previous study (Fu et al. 2016), transcript reconstruction was carried out using Trinity, CD-HIT and MIRA, resulting in a final set of 82623 unigenes with an N50 of 1858bp. Differentially expressed genes (DEGs) identification was carried out using RSEM and edgeR (Li et al. 2011, Robinson et al. 2010), resulting in a total number of 4622 DEGs, 573 in shoots and 4187 roots. A detailed assessment of the number and the identity of the DEGs between conditions for each organ showed that in shoots, only 43 genes were differentially expressed between mild water stress *vs* control, while 504 between severe water stress *vs* control, and 224 between severe *vs* mild water stress. By contrast, a much larger number of DEGs were identified in root (Figure 4.1), but in both organs, the higher level of stress corresponded to more DEGs identified. By comparing the 138 DEGs in common between organs, we further observed a general conservation of expression patterns, with 53 of the genes being regulated in the same way in shoot and root and only one gene displaying opposite regulation (Supplementary Ch4.1).

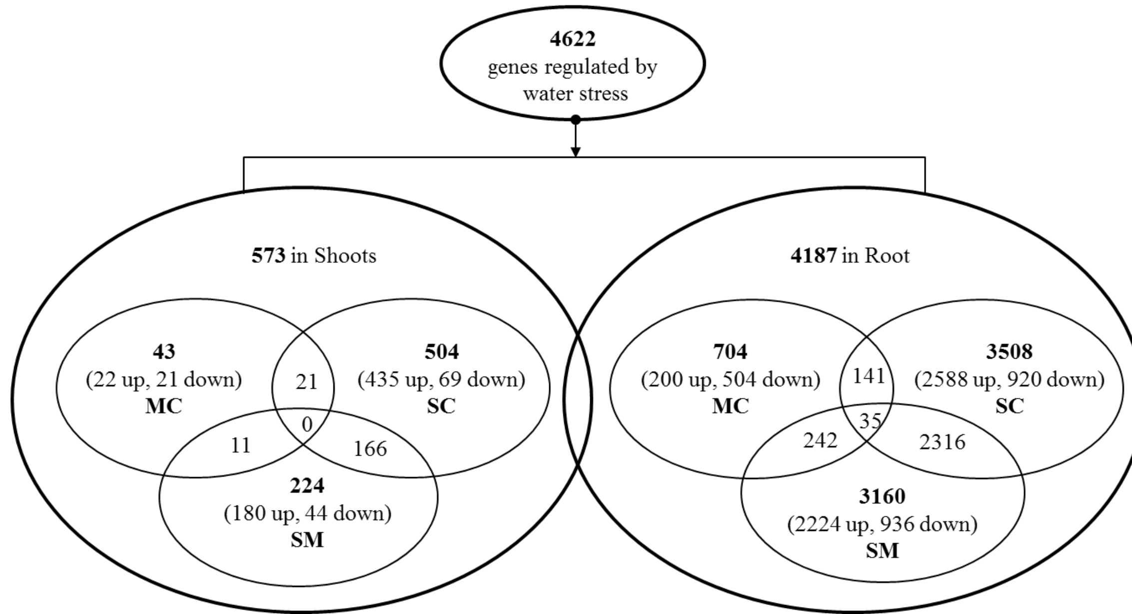


Figure 4.1. Summary of DEGs in shoots and roots of *P. australis* upon drought stress. Number of genes up/down regulated by drought stress under different conditions (MC: mild water stress vs control, SC: severe water stress vs control; SM: severe water stress vs mild water stress) in roots and shoots.

4.3.2 General comparison of differentially expressed genes (DEGs) between *P. australis* and *A. donax*

In order to standardize as much as possible the comparison, we applied exactly the same experimental conditions and the same method/parameters for assembly to both species. In this way, we were able to identify the major interspecific differences in transcriptional responses. We first observed a slightly larger size of *P. australis*' transcriptome compared with *A. donax* (Supplementary Ch4.2), which is possibly the result of a larger amount of species-specific single-copy genes and species-specific gene families (Vandepoele and Van de Peer, 2005). The most evident difference is, however, in the number of DEG for both organs of *P. australis* and *A. donax* (Figure 4.2). Also form a qualitative point of view some differences exist. In particular, while in all *A. donax* organs and conditions the large majority of DEGs are up-regulated, in *P. australis* the overall direction of expression variation resulted to be different between organs, especially for mild drought *vs.* Ctrl (MC), with the large majority of DEGs being down- rather than up-regulated in root (Figure 4.2).

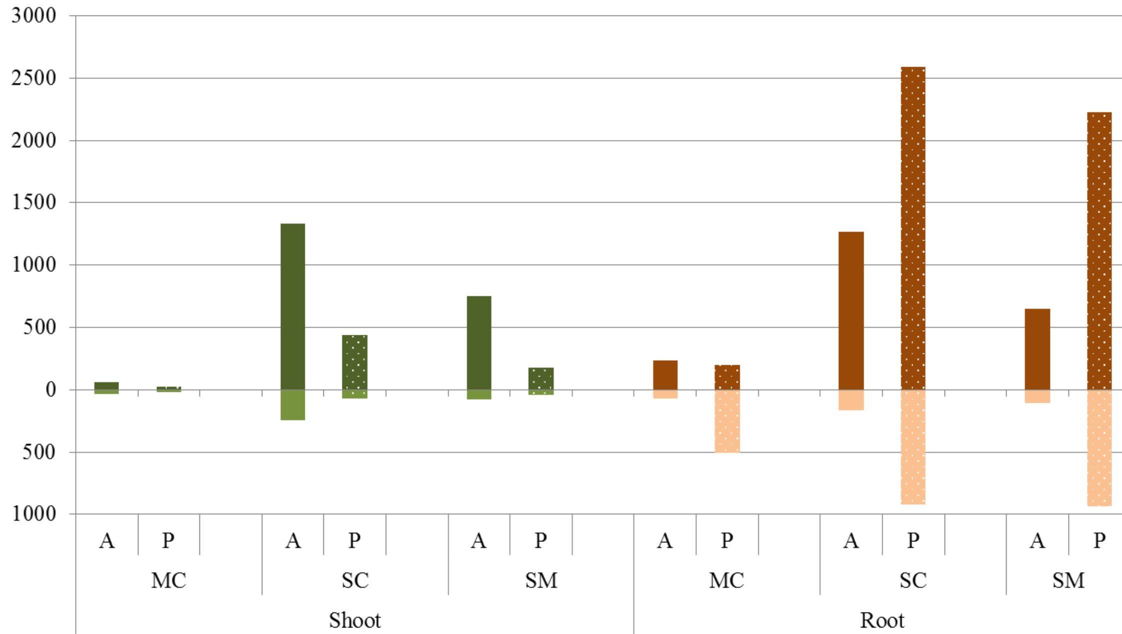


Figure 4.2. Comparison of DEGs in shoots and roots between *P. australis* and *A. donax* upon drought stress. Number of genes differentially regulated between stress conditions. MC: Mild water stress vs. Control, SC: Severe water stress vs. Control, SM: Severe water stress vs. Mild water stress. Solid: *A. donax*, Patterned: *P. australis*; green color refers to shoot, brown to root DEGs; darker and paler colors indicate up-regulated and down-regulated genes, respectively. "A" refers to *A. donax*; "P" refers to *P. australis.*

A closer analysis of the absolute numbers of DEGs in the two organs, however, highlighted a relatively large difference in genes regulated under stress in roots as compared to shoots (Figure 4.1), and the huge difference (nearly one order of magnitude) on the number of DEGs between two species for both organs. In order to understand the functional relevance of such differences, comparison between species of Gene Ontology (GO) enrichment of DEGs under each organ/condition combination was carried out by AgriGO (Du et al. 2010). The result shows that the most common enriched molecular functions across organ, species and conditions are "binding", "transcription regulator activities" and "catalytic activity", in line with a massive reprogramming of the transcriptome of both species. This reprogramming in general is associated to the enrichment of transcription regulator activities in the up-regulated fraction of DEGs, with the only exception of the opposite trend observed for *A. donax* leaves under severe stress (i.e., SC down-regulation of

A. donax in leaf; Supplementary Ch4.3). “Ligase activity” and “transferase activity” (both part of the larger term “catalytic activity”) are only enriched in up-regulated situations, indicating that even at the relatively early stress stage analyzed also transcripts associated to metabolism adjustment are over-represented among DEGs. Besides the general enrichments observed above, also some species/organ/regulation specific enriched terms were identified. For instance, the “structural molecular activity” was only enriched in roots. In addition, some GO terms were specifically enriched to *P. australis’* root, like “antioxidant activity” (only enriched among down-regulated DEGs); “enzyme regulator activity” (enriched in up-regulated DEGs); and “cofactor binding” under the general “binding” term (enriched in general in *P. australis’* root) (Supplementary Ch4.3). The DEGs associated to these functions are interesting candidates for the future dissection of the high responsiveness of *P. australis* root system to drought.

4.3.3 COMPARISON OF FUNCTIONAL CLASSIFICATION AND TRANSCRIPTIONAL FACTORS RESPONSIVE TO WATER STRESS

To identify organ-specific differences, stress-related genes were identified based on curated homology searches against genes experimentally characterized as in *A. donax* water stress study (Fu et al, 2016). For *P. australis* shoot, only 18 DEGs matched known stress-related genes, approximately ten times less than the number that was found in *A. donax* shoot. For root the difference between species was negligible, as 131 and 135 DEGs matched stress-related genes in *P. australis* and *A. donax*, respectively (Figure 4.3). With a detailed comparison, we found that the majority of stress-related genes common in both species belonged to the functional categories “salt”, “cold”, “stress” and “salinity response”, in line with the type of stress applied. Also species-specific stress-response categories were identified. For *P. australis,* more genes than in *A. donax* were related to “oxidative”, “cadmium” and “hypoxic stress”, while for *A. donax,* more DEGs related to “osmotic”, “water”, “hydrogen peroxide” and “metal” were identified (Figure 4.3). These results may indicate either that the two species experience somehow different types of stress consequent to water limitation, or that the levels and/or kinetics of stress response differ between species. While the hypoxic and osmotic stress unique to, respectively, *P. australis* and *A. donax* the

support the first hypothesis, the other stress categories are related to each other and may thus indicate different stages of a common stress response.

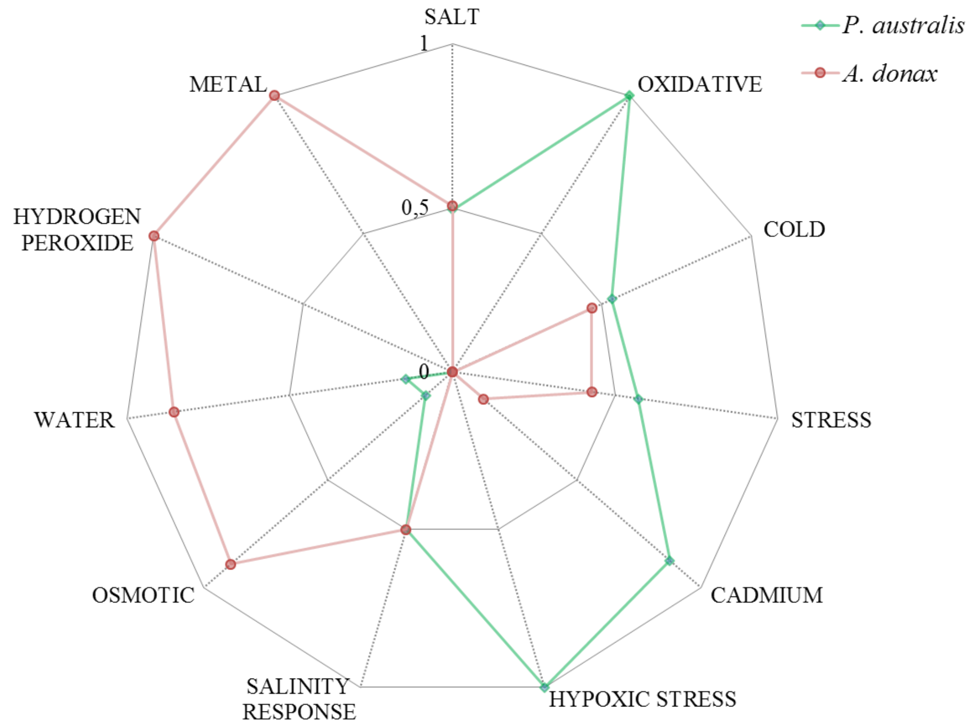


Figure 4.3. Stress-related functional categories of DEGs. Blue for *P. australis*, and red for *A. donax*. Axis present the rate of gene numbers drops inside the corresponding category: one species/both species (Supplementary Ch4.4).

Transcription factors (TFs) have been identified among the most promising targets for the improvement of plant performance under drought stress. Compared to *A. donax*, for which we previously found a total of 145 putative TF unigenes from 37 TF families (Fu et al 2016), in the water-stressed *P. australis* transcriptome we found 57 and 162 unigenes from shoot and root, respectively (Supplementary Ch4.5). Because of the altered water potential under salt stress (Huang et al 2012), compared to the 18 genes for *A. donax*, 32 TF genes were found specifically responding to drought for *P. australis*. Like in *A. donax*, the majority of differentially expressed TFs of *P. australis* belonged to the NAC, WRKY, AP2-EREBP, bHLH and bZIP families, which are known to mediate water stress responses in plants (Hadiarto and Tran, 2011). While in both species the proportions of shoot TF unigenes were

similar in each family, fewer TFs were identified in *P. australis* as compared to *A. donax*. In root, by contrast, the ranking of the most represented TF families two species differed. The AUX/IAA, for instance, was no longer one of the top represented TF families for *P. australis*, while the HB (Homeodomain-leucine zipper (HD-Zip) proteins) and HSF (heat shock transcription factor) families were more represented than in *A. donax* (Figure 4.4).

More in detail, there were 14 DEGs belonging to the HB family in *P. australis* roots, which corresponded to 7 rice homologs. Among them, 4 *P. australis* genes were down-regulated. Three of them, Phr39237, Phr39239 and Phr39241, which are the closest homologs to rice LOC_Os03g07450.1 and Arabidopsis AT1G69780.1, have been reported as negative regulators of early root growth (Riechmann et al, 2000), indicating that root elongation may constitute an avoidance strategy for *P. australis*. This is supported by previous reports indicating that *P. australis* reacts to exposure to water stress by growing deeper its root system (Pagter M., Bragato C. and Brix H., 2005). The HB unigene downregulated in root, Phr27396, is homologous to rice LOC_ Os09g29460.1 and Arabidopsis AT2G22430.1). Consistently to the down-regulation of the *P. australis* homolog, LOC_ Os09g29460.1 has been reported to be down-regulated by drought, even though no detail hypothesis on its mechanism of action has been proposed (Dai et al, 2008). Interestingly, the other up-regulated genes in the HB family have also been reported to be regulated by drought stress. The different regulation of HB genes has not been elucidated in detail, yet. It is, however, known that HB genes encode transcription factors that typically switch on cascades of other genes. Most of the times, in plants homeodomain proteins bind the promoter region of their target genes together as complexes together with other transcription factors in plants (Bürglin and Affolter, 2016). It is, thus, likely that the up- or down-regulation of different HB TFs in response to drought may depend from either different tissue-specific functions and/or by the different binding partners.

In the HSF family, 13 DEGs were identified in *P. australis* roots, which correspond to 7 rice homologs. All *P. australis* DEGs are up-regulated in response to water stress. HSFs have been reported as hub genes for drought response in rice (Zhang et al, 2012) and the co-/overexpression of HSF enhanced drought tolerance (Bechtold et al.,_2013, Scharf et al., 2012, Guo, et al, 2016). Also inhibition of root elongation by NaCl or mannitol was

significantly alleviated in the HsfA2-overexpressed transgenic plants in comparison to wild-type plants (Ogawa, Yamaguchi and Nishiuchi, 2007), thus indicating that these genes may act synergistically to HB TFs to maintain root elongation under water and osmotic stresses as an avoidance mechanism.

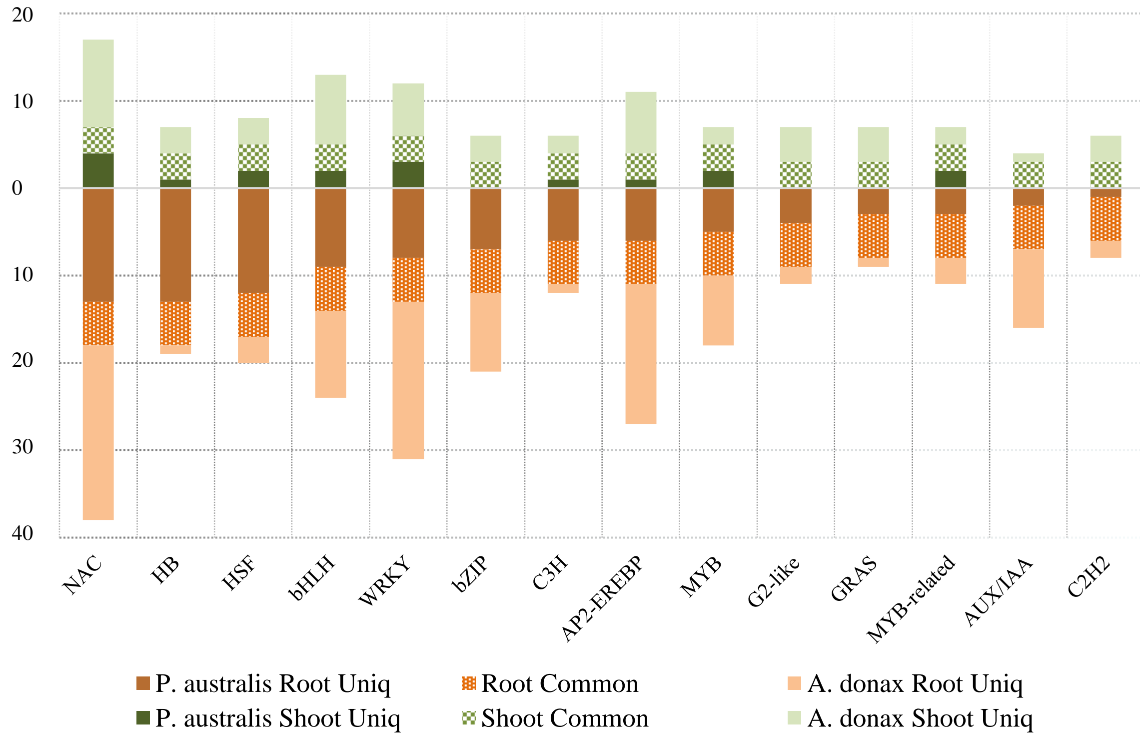


Figure 4.4. Comparison of distribution of transcription factors responsive to water stress between *A. donax* and *P. australis* (Supplementary Ch4.5). Data are sorted by number of *P. australis* root DEGs. Only categories with more than three DEGs identified as transcription factors are shown. Brown bars correspond to root DEGs, green bars to shoot DEGs; darker colors refer to *P. australis* unique DEGs, paler colors to *A. donax* unique DEGs; patterned bars indicate DEGs in common between the two species.

4.3.4 Comparison of metabolic pathways related to water stress response

In order to pinpoint putative metabolic pathways underlying the differences between species, pathway enrichment analysis of DEGs was carried out (Table 4.1).

For shoot, the patterns of pathway enrichment were identical in *A. donax* and *P. australis*, as DEGs belonging to Plant-pathogen interaction and Plant hormone signal transduction pathways were enriched in both species under severe drought *vs.* both control

and mild drought and no pathway was enriched under MWS condition for either species. Also the trend of differential expression condition was identical under SWS, with 95% of DEGs in the plant hormone signal transduction pathway up-regulated in both species. The number of genes regulated in *A. donax*, however, was more than twice than that in *P. australis*. This agrees with previous results showing the dramatic increase of ABA in response to drought for *A. donax* leaves (Velikova et al, 2016). As plant hormone signal transduction pathway, and especially the path for stomatal closure, is known as a basic mechanism for reducing the impact of drought (De Leonardis et al, 2012, taken together the above results indicate that in both *P. australis* and *A. donax* shoots the response to water stress applied to the root is perceived and elicits qualitatively analogous transcriptional responses (Table 4.1).

By contrast, in the root interspecific differences in either significance ranking and/or the enriched pathways were preponderant for all comparisons of stress levels (i.e., MC, SC and SM in Table 4.1). In particular, DEGs belonging to the plant hormone signal transduction and carotenoid biosynthesis pathways were specifically enriched in *A. donax*, while flavonoid biosynthesis and protein processing in endoplasmic reticulum pathways were specifically enriched in *P. australis* roots. Only enrichment of DEGs belonging to the pathway of ribosomal genes (“Ribosome” in Table 4.1) was conserved and had the same rank just in the comparison between severe and mild water stress.

	Shoot		**Root**	
	A. donax	***P. australis***	***A. donax***	***P. australis***
MC				Ribosome
				Flavonoid biosynthesis
SC	Plant hormone signal transduction	Plant hormone signal transduction	Plant hormone signal transduction	Ribosome
	Plant-pathogen interaction	Plant-pathogen interaction	Carotenoid biosynthesis	Protein processing in endoplasmic reticulum
			Ribosome	
SM	Plant hormone signal transduction	Plant hormone signal transduction	Ribosome	Ribosome
	Plant-pathogen interaction	Plant-pathogen interaction	Carotenoid biosynthesis	Protein processing in endoplasmic reticulum
			Plant hormone signal transduction	

Table 4.1. Comparison of DEG pathway enrichment between *P. australis* and *A. donax*.

Taking a more general view at the results one can notice that ribosome was the most significantly enriched pathway under all the conditions for root (Table 4.1). Extremely large differences between *P. australis* and *A. donax* were found in the transcriptional regulation of ribosomal subunits, especially for the number of genes and the expression pattern (Figure 4.5). For *P. australis*, compared with control condition, most of the ribosomal DEGs were down-regulated under MWS condition (MC). Intriguingly, when comparing severe and mild water stress, about one third of the MC down-regulated ribosomal DEGs were not changing expression level, one third of them were up-regulated and the last third of them were down-regulated. Compared with control condition, two thirds of ribosomal DEGs are up-regulated under SWS condition, and one third of them are down-regulated. Among all the up-regulated genes under SWS condition, half are also down-regulated under MWS condition, while the other half of them are not regulated and none of them are up-regulated under MWS. The differences with respect to *A. donax* were very large: For *A. donax*, only one gene was up-regulated under MWS condition compared with control condition. Under SWS condition, there was only one gene which was down-regulated, the other 47 genes were up-regulated. For both species, the number of ribosomal DEGs regulated under SWS were many more than under MWS condition (Fig. 4.5). However, there were much fewer ribosomal DEGs regulated for *A. donax* root compared with *P. australis* root. For *P. australis* a very large number of ribosome genes were down-regulated under the MWS condition. Meanwhile, the pattern of ribosomal DEGs' regulation was also corresponding to the overall expression pattern of DEGs' in root (Figure 4.5), i.e. more down-regulated genes under MWS condition and more up-regulated genes under SWS condition. These results suggest that the regulation of ribosomal proteins may result in a control of translational efficiency, presumably by biased translation (Xue and Barna, 2016). It is, therefore, likely that this massive reprogramming of RPs transcription may be functionally relevant in the response of plants to water stress (Fu et al. 2016). *P. australis* higher sensitivity to water stress correlates with the stronger and qualitatively very different regulation on ribosomal genes compare with *A. donax* especially under MWS and SWS. Unfortunately, the high number of specialized paralogs associated stress response did not to allow the detailed elucidation of which ribosome genes were specifically regulated under specific conditions (Xue and Barna, 2016, Wang et al, 2013). Another limit of the approach we took is that the tissue sampling was too

coarse to allow the dissection of ribosome genes in both different cell types and for the same cell type in different subcellular locations, whose expression regulation is also known to differ under different conditions (Sormani et al, 2011).

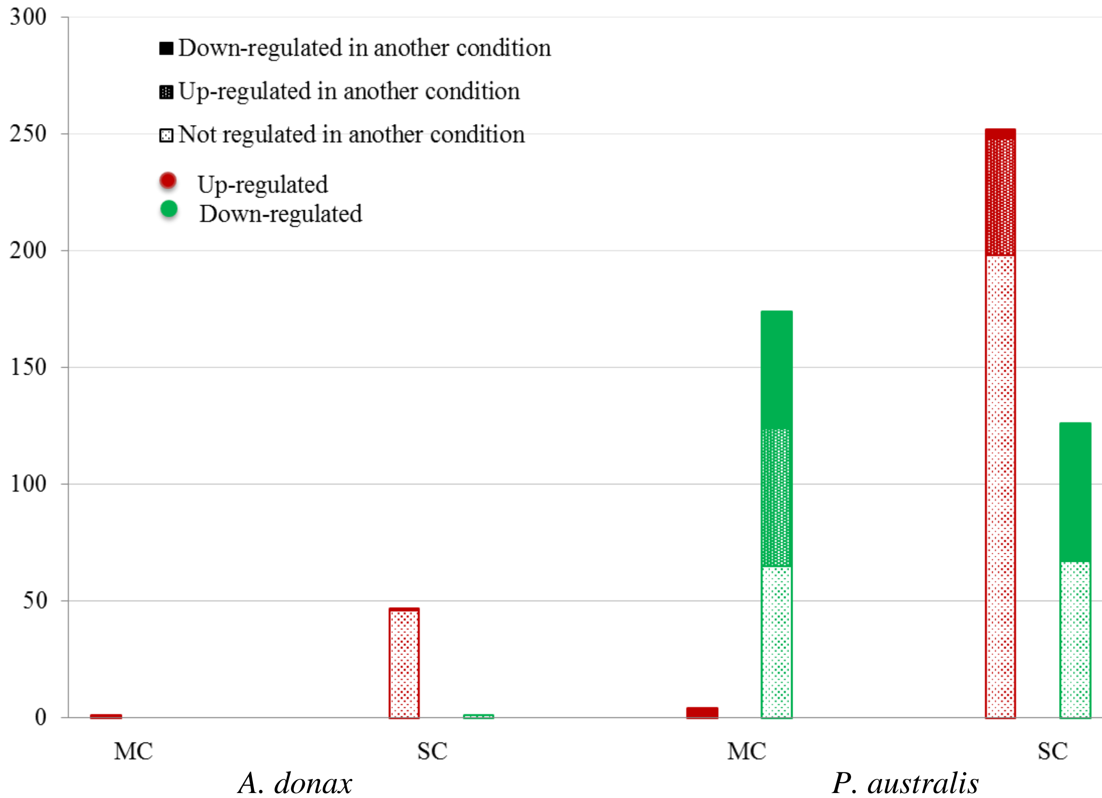


Figure 4.5. Comparison of root ribosomal DEGs regulation responsive to water stress between *A. donax* and *P. australis*. Red bar for up-regulation, green bar for down-regulation; solid for up-regulation in another condition for the same species, darker pattern for down-regulation in another condition for the same species and paler pattern for none regulation in another condition for the same species.

In general, however, it is very likely that the dramatic upregulation of ribosome genes under severe water stress of *P. australis* is associated to the observed up-regulation/enrichment of genes of the protein processing in endoplasmic reticulum as a consequence of ER stress inducing unfolded protein response (UPR). Transcriptional regulation of many ribosomal genes indicate that all major steps of this pathway, namely (1) ER import, (2) folding and recognition, (3) transduction of stress signals, (4) export and (5) degradation (Figure 4.6) are affected during the early phase of severe water stress. The Sec61 complex, which is necessary for binding the ribosomes to ER membrane, is the main

translocon mediating the import of plant proteins in the ER lumen and it was found to be up-regulated in conditions of severe stress, indicating enhancement of ER protein import (Schuck et al, 2009). The second major step of ER processing is proper folding of proteins. The major DEGs identified which are related to this step are OSTs, Bip, GRP94 and CRT transcripts. While the nascent polypetides are still associated to the Sec61 translocon, they can be scanned by specific oligosaccharyltransferases (OSTs) for the presence of sequons or glycosylation site recognition motifs. If sequons are detected, then presynthesized oligosaccharides linked to lipid carriers (dolichol-pyrophosphate) are transferred en bloc by OST to the γ-amido group of Asn residues in the sequon and activate calreticulin-mediated folding (Kornfeld and Kornfeld, 1985; Howell, 2013). Worth to note, Calreticulin protein (CRT), which promotes proper folding by preventing aggregation and premature export from the ER, has only been up-regulated in *P. australis* root, but not in *A. donax*. Previous studies showed that the expression of a CRT3 gene (EF452301) was significantly enhanced by PEG-induced drought stress, and the up-regulation of CRTs is considered as a conserved self-protection mechanism that was acquired during a long-term evolutionary process and likely to facilitate the survival of plants under unfavorable conditions (Jia et al, 2008). Alternatively, Binding Immunoglobulin Protein (BiP) assists in the folding of proteins lacking glycosylation and also acts in ER protein quality control mechanism as well as in the transduction of stress signal to the nucleus (Ma and Hendershot, 2004). Another DEG taking part in the folding process is one of the many ER chaperones, Glucose Regulated Protein 94 (GRP94), which, together with some isoforms of nucleotide exchange factor (NEF) and Hsp40 (a dnaJ protein homolog) optimizes the output of properly folded secretory and membrane proteins (Eletto et al, 2010). In addition, GRP94 stores calcium and assists in the targeting of malfolded proteins to ER associated degradation (ERAD) (Eletto et al, 2010). Three classes of ER DEG transcripts manily associated to ER protein export were significantly up-regulated in our study, namely PDI, TRAP and Sec61. Protein disulfide isomerases (PDI) can be implicated in the folding of both glycosylated and non-glycosylated ER proteins (Howell, 2013). They play important roles in the maturation of secreted plasma membrane and storage proteins (Jacquot et al, 2002). They are involved in the export of proteins from the ER by catalyzing protein disulfide bonding and isomerization in the endoplasmic reticulum and functions as a chaperone to inhibit the aggregation of misfolded

proteins (Zhu et al, 2014). TRAP (translocon-associated protein) localises in the membrane, associates with the Sec61 translocon, and both proteins have been proposed to be involved in export of proteins destined to ERAD outside of the ER (Nagasawa et al, 2007). We found that many transcripts up-regulated in response to water stress correspond to genes involved in degradation by the ERAD pathway (listed in Figure 4.6 under "Degradation"). Like GRP94, several of the chaperones in this group can assist both proper folding or contribute to ERAD. For instance, NEF exchanges ADP for ATP, dissociating the client protein from Hsp70 (Ohta and Takaiwa, 2014). DnaJ homologs assist in the translocation, secretion, retro-translocation, and ER-associated degradation (ERAD) of secretory pathway proteins (Shen et al, 2002). Additionally, cell division cycle protein 48 homolog (p97) helps to direct polyubiquitinated proteins for proteasome-mediated degradation, interacting with an equivalent group of ubiquitin ligases, the hubs in ERAD (Morreale et al, 2009), including SKP1-like protein (Skp1), cullin-1 (CUL1), RING-box protein (RBX1), ubiquitin-conjugating enzyme (UbcH5), etc. These genes are also all up-regulated in *P. australis* root, thus confirming the clear activation of all the major steps involved in ER processing in endoplasmic reticulum associated to ER stress (Table 4.1). Somehow surprisingly, however, despite all these transcriptional changes associated to protein processing in ER, we did not find transcriptional evidences of stress signaling to the nucleus in *P. australis* through the unfolded protein response (UPR) signaling pathway. There are two branches of UPR in plants, working together to perceive the level of protein misfolding in the ER and signaling it to the nucleus for the activation of additional stress-related genes involved in recovery from ER stress. One branch involves membrane-associated transcription factors such as bZIP28, bZIP17; the other one involves a membrane-associated dual-functioning protein kinase/ribonuclease, IRE1, that splices the mRNA encoding bZIP60 (Howell, 2013). For both species, we found that IRE1 was not regulated at the early onset of the water stress. By contrast, *P. australis* SBT6.1, also known as site-1-protease (S1P), which initiates the transduction of stress signals from the endoplasmic reticulum to the nucleus by specific cleavage and activation of bZIP28 transcription factors (Liu et al., 2007) was found to be down-regulated. Thus, it seems that both path leading to Recovery of ER stress were blocked. Instead, as mentioned above, ERAD related genes were clearly enriched. As in plants ER stress has been linked to both autophagy and cell death (Liu et al, 2013), it is possible that

the level of ER misfolded proteins was not sufficient yet in our early stress detection of transcriptional responses to water stress to trigger signal transduction to the nucleus. This is consistent with lack of IRE1 upregulation, which has been suggested as key regulator of both cell-survival and cell death depending on stress levels. According to this view, it seems that the ER-associated degradation of misfolded proteins, an avoidance strategy, is triggered only in *P. australis* but not in *A. donax* under the same water stress treatments.

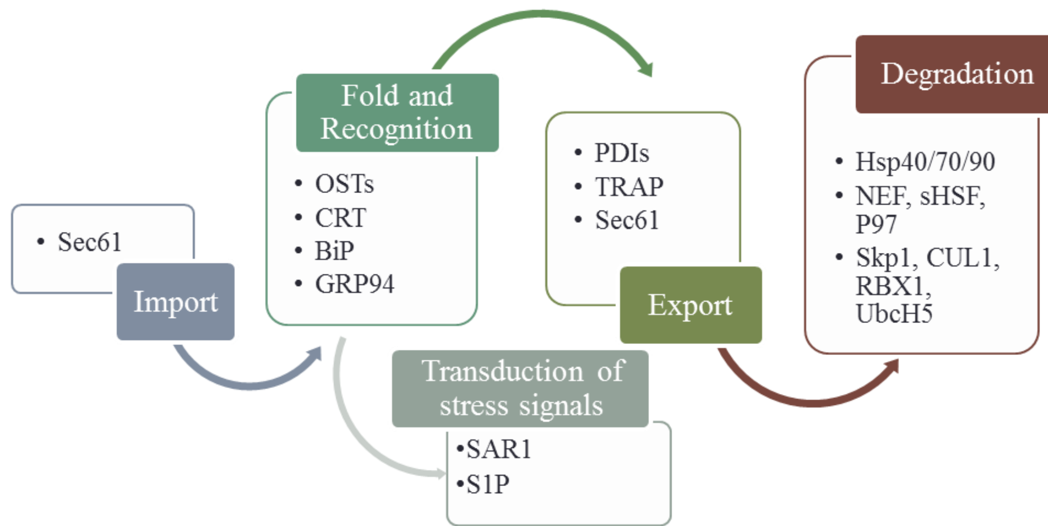


Figure 4.6. Regulated genes in "Protein processing in endoplasmic reticulum" pathway in *P. australis* root.

Also the Flavonoid biosynthesis pathway was enriched specifically in *P. australis* but not *A. donax* root. Flavonoid biosynthesis initiates with the step catalyzed by Chalcone Synthase, leading to production of flavanoids, isoflavonoid-type phytoalexins, and other metabolites with stress protective functions for plants. Flavonoids interfere with hormone signaling by inhibiting polar auxin transport (Dao et al., 2011). Increasing evidence suggests that plants exposed to drought accumulate secondary metabolites, and a plausible explanation could be to protect cells from oxidative stress by consuming $NADPH^+$ and H^+ for the synthesis of highly reduced precursors like aromatic amino acids, monoterpenes, alkaloids (Selmar and Kleinwächter, 2013). Increased abundance under drought of the enzyme (+)-neomenthol dehydrogenase, which participates in monoterpenoid biosynthesis,

observed in this study, could be linked to possible protective functions of monoterpenes. Caffeoyl-CoA 3-O-methyltransferase is engaged in the pathway of lignin biosynthesis. The accumulation of this enzyme under drought could be related to increased lignification of the cell wall—a modification associated to avoidance of water loss by increased vessel impermeabilization (Simova-Stoilova et al, 2015).

A pathway that, by contrast, was enriched only in *A. donax* but not *P. australis* root is that of Carotenoid biosynthesis, which is partly related to Plant hormone signal transduction pathway enriched in leaf of both species. While 9-cis-epoxycarotenoid dioxygenase (NCED), which is thought to be a key enzyme in ABA biosynthesis, was up-regulated in leaf and roots for both species, Phytoene Synthase (PSY) was up-regulated in both shoot and root in *A. donax*, but up- and down- regulated in *P. australis* root under different stress levels and not regulated at all in shoot. It has been proposed that *PSY* is subject to a positive feedback regulation through ABA (Welsch et al, 2008) and an increased PSY activity provides carotenoid precursors for the enhanced production of ABA under drought stress (Ruiz-Sola et al, 2012). One possibility, still highly speculative given the multiple functions and interconnection of this pathway to other fundamental metabolic pathways, is that interspecific regulatory differences in Carotenoid may cause different size of ABA pools in the two species. Further studies will need to test this hypothesis.

4.3.5 Differences in expression of orthologous genes in the two species under water stress

After the high-level comparison on the gene functions, we asked whether part of the observed interspecific differences could be due to variation in expression patterns/levels of orthologous genes. The putative orthologs of *P. australis* were identified by the Reciprocal Best Hits (RBH) method with blastp versus the *A. donax* transcriptome (Moreno-Hagelsieb and Latimer, 2008). A total of 20772 pairs of putative orthologs were identified, of which 17736 genes (21% of the whole transcriptome) from *A. donax* and 18126 genes (22% of the whole transcriptome) from *P. australis*. The different numbers are due to the small fraction of orthologs that have more than one unigene per species (closely related paralogs that have the same score in blast searches). Among them, 970 genes from *A. donax (*551 for shoot and

539 for root*)* and 1441 genes from *P. australis* (233 for shoot and 1274 for root) were DEGs. In total, 419 pairs of RBHs (290 pairs in roots and 112 pairs in shoots) were DEGs in both species. A possible limitation of this method is the possible mis-identificaton of homologs (e.g. when 2 sequences have low identity and coverage but are the best match to each other). The method is also known to have a relatively low sensitivity, as several genes will not be included in any orthologous group because of the constraint imposed on the reciprocity of the blast hits. The putative orthologous pairs are, however, for the same reason, identified with elevated accuracy. Taken together, these results confirm the quantitative differences of the core of conserved genes differentially expressed in response to water stress among species, thus validating the differences in functional DEGs classes reported above. These differences could be due to multiple factors and *per se* the mere assessment of the number of differentially expressed orthologous sets does not provide clues on the processes underlying this observation. In order to get a more accurate understanding of the reasons for this, the analysis of expression patterns, i.e. the trends of variation associated to the different stress conditions was carried out. By grouping the genes according to up- and down regulation in all the pairwise comparisons between conditions, groups of DEGs whose expression changes in root and shoot responding in a consistent manner in the two species were identified (Figure 4.7 and Supplementary Ch4.6) A total of 150 such pairs of differentially expressed orthologs in root, of which 148 from *A. donax* and 149 from *P. australis* , and a total of 59 pairs of orthologs in shoot, of which 57 from *A. donax* and 55 from *P. australis* were found. The overall number of orthologs having different expression patterns is much higher than in the shoot. They encompass both differentially expressed and non-differentially expressed genes. The majority of them are expressed in root (1214 pairs of orthologs in total, of which 1198 from *A. donax* and 1137 from *P. australis*). The remaining ones are expressed in the shoot (a total of 195 pairs of orthologs with different expression patterns in two species, of which 189 from *A. donax* and 182 from *P. australis).* GO annotation was used to assess whether the orthologs with conserved expression pattern have different functions as compared to those with different expression pattern across species. In roots, orthologous genes having different expression patterns in *P. australis* and *A. donax* were mainly associated to molecular functions such as “enzyme regulator”, “molecular transducer”, “translation regulator” and “nutrient reservoir” and the biological process

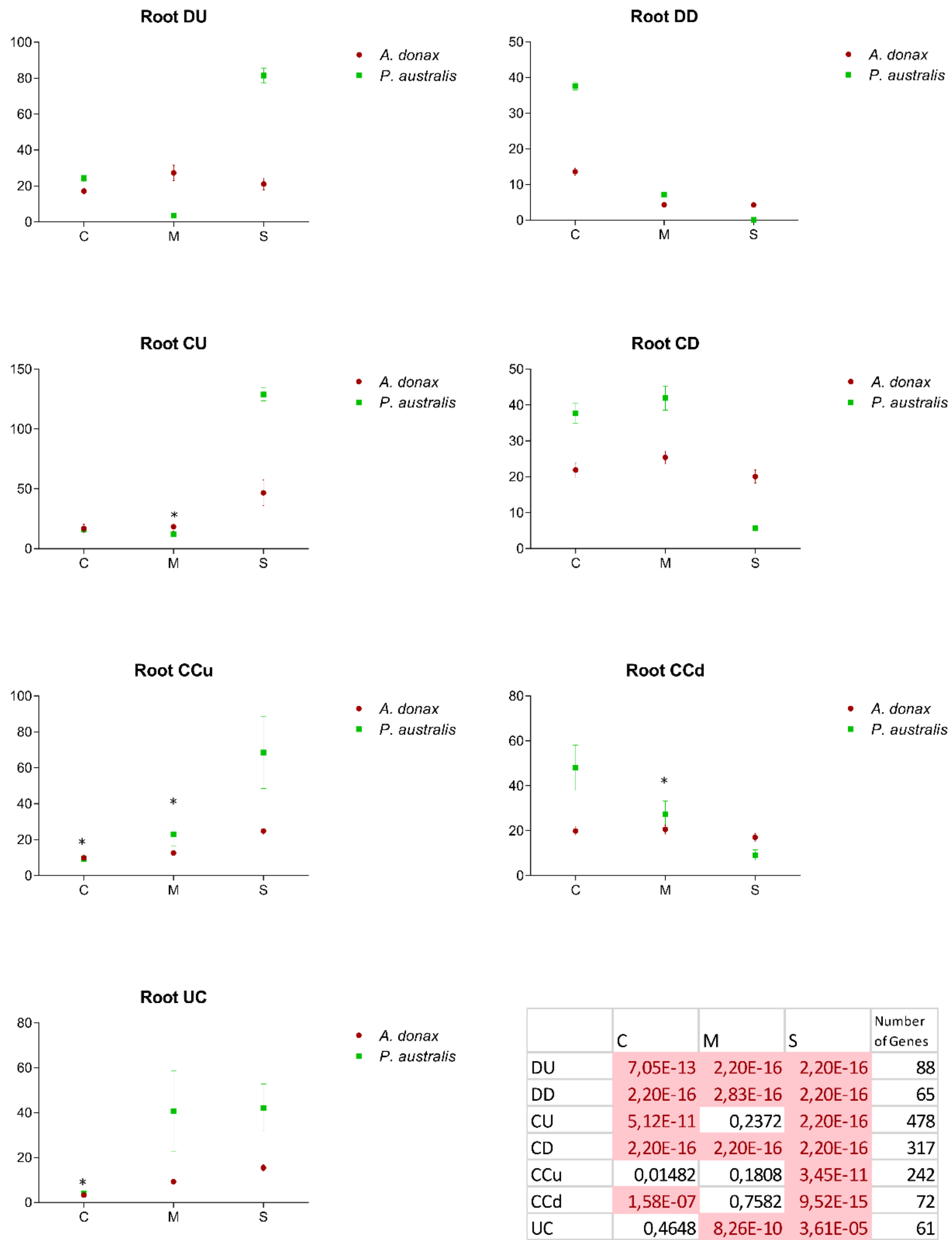
"growth" and "multi-organism process" (Figure 4.7). Instead, in shoots, there is no specific function associated to more than one gene with either conserved or different expression pattern (Supplementary Ch4.6). Taken together, these results further support the view that the major functional differences to water stress between *P. australis* and *A. donax* (1) depend from the root system and (2) are associated to large differences in expression of orthologous genes. Whether the differences in expression patterns of orthologs are due to qualitative (i.e. due to different tissue-specificity) or quantitative factors (i.e. the kinetic of transcriptional change in presence of a conserved pattern) is currently not known. In root, thus, possibly species-specific responses and different kinetics of common responses contribute together to overcoming the stress. Time-resolved and *ad hoc* functional analyses will be required to assess the relative contribution of species-specific versus shared responses to water stress in these species.

In addition to the assessment of the functional differences between orthologs, also the level of conservation of expression levels between orthologs can provide important insights into the adaptive mechanisms that shape the evolution of stress tolerance in related species. Given the low number of DEG orthologs shared between *P. australis* and *A. donax*, the analysis of expression level conservation was carried out only on the orthologs corresponding to *P. australis* DEGs. The results clearly show the low conservation in expression levels among RBH orthologs (Figure 4.7). For roots, all the patterns with down regulation have different, or even opposite regulation. Meanwhile, for the patterns UC, CU and Ccu (associated to overall gene up-regulation), an increase on expression levels under drought stress quicker for *P. australis* than *A. donax* could be observed. For shoots, all the patterns associated to down regulation have only few putative orthologs, and they display different expression levels. At the same time, for patterns, UC, CU and CCu, analogously to the situation in roots, a quicker increase on expression levels under drought stress for *P. australis* than *A. donax* was observed. To understand whether there is statistically significant difference between the expression level under each condition of each organ and pattern, pairwise T-tests were carried out and the results were corrected with the Bonferroni method for multiple testing.

As detailed in Figure 4.7 and Supplementary Ch4.6, all patterns with more than 20 putative homolog pairs had significantly different expression levels in at least one condition. In particular, patterns, DD, DU, AD, CU in roots and pattern CCu in shoots showed significantly different expression levels in all the conditions. Taken together, these results can be interpreted as follows:

(1) the normalization procedure applied to the comparison of expression level of orthologs is effective. In fact, even when the interspecific differences between average expression levels in control condition are statistically significant, they are very small for the majority of the patterns (DU, CU, Ccu, UC versus DD, CD, CCd in Figure 4.7). The large difference in control condition which can be observed, e.g., for root DD, CD and CCd indicate that genes with these patterns of expression are those that are associate with constitutive transcriptional differences between species, but at present it is not possible to establish whether these differences are functionally significant. However, the fact that these are the major categories of down-regulation suggests that these genes may, in general, have a negative impact on the plant's ability to cope with water stress.

(2) *P. australis* orthologs are generally expressed at higher levels than their *A. donax* counterparts. Especially for the patterns associated to gene up-regulation (left side of Figure 4.7), the variation under increasing levels of stress increases and it is most often larger in absolute value for *P. australis* than for *A. donax*. This suggests that the patterns associated to up-regulation of the orthologous sets may be involved in processes that enhance the ability of the plant to withstand water stress. This further indicates that genes in these expression categories respond either more intensely or more quickly to the stress in *P. australis* than in *A. donax*. As before, time-resolved and ad hoc functional analyses will be required to elucidate the relative relevance of these factors.



	C	M	S	Number of Genes
DU	7,05E-13	2,20E-16	2,20E-16	88
DD	2,20E-16	2,83E-16	2,20E-16	65
CU	5,12E-11	0,2372	2,20E-16	478
CD	2,20E-16	2,20E-16	2,20E-16	317
CCu	0,01482	0,1808	3,45E-11	242
CCd	1,58E-07	0,7582	9,52E-15	72
UC	0,4648	8,26E-10	3,61E-05	61

Figure 4.7. Common gene expression patterns in response to different levels between *P. australis* and *A. donax* root. The expression patterns for DEGs have been classified in terms of regulation trends. Up- and down-regulated genes were identified by having |logFC| ≥ 2. (mild water stress *vs* control and severe water stress *vs* mild water stress). Expression levels presented by Fragments Per

Kilobase of exon per Million fragments mapped (FPKM). (C: control; M: mild water stress; S: severe water stress; FPKM: Fragments per kilobase of exon per million fragments mapped; CU: Constant, up-regulated; CD: Constant, down-regulated; UU: Up-regulated, up-regulated; UC: Up-regulated, constant; UD: Up-regulated, down-regulated; DC: Down-regulated, constant; DU: Down-regulated, up-regulated, CCu: up regulated and only regulated by severe water stress, CCd: down-regulated and only regulated by severe water stress.). Red: *A. donax*; green: *P. australis.* Asterisks (*) mark no statistically differences between two species in a specific pattern under each condition. Red color in the table means there is significant differences between two species, with t-test ($p<0.05$) corrected by Bonferroni method.

4.3.6 Conserved drought response genes in related Poaceae species

In order to complete the comparative study carried out in *A. donax* (Fu et al., 2016), We identified the subsets of DEGs in common between *P. australis*, *A. donax*, foxtail, sorghum and rice by OrthoMCL (Qi et al., 2013; Dugas et al. 2011; Zhang et al, 2012). A total of 41 groups of putative orthologs were identified (Supplementary Ch4.7), indicating that they constitute a core of evolutionarily highly conserved genes with relevant function in water stress.

In detail, for all the conserved genes across the species, there was only one *P. australis* gene (Ph49585) regulated by MWS and it was up-regulated in root, while its orthologs in *A. donax* were down-regulated in shoot under SWS. Beside one gene in shoot, and one gene in root, all the other genes (17 in shoot, 39 in root) were up-regulated only under SWS. By contrast, among their orthologs in *A. donax*, 9 genes (2 in shoot, 7 in root) were regulated by MWS, and all of them were also regulated under SWS. This result indicate that in *A. donax* the highly conserved genes involved in water stress tend to have a more sensitive regulation than their *P. australis* counterparts, which are normally not regulated in mild stress condition. Taken together these observations support the hypothesis that *P. australis* may preferentially adopt an avoidance strategy *versus* a putative tolerance strategy of *A. donax*. It is also interesting that, as reported earlier, one of the most striking differences in expression between *P. australis* and *A. donax* are those in root when comparing mild water stress versus control condition. (Supplementary Ch4.7). Analysis of GO enrichment on these 504 down-regulated genes in *P. australis* showed that the most significantly enriched term

in each gene category were “translation”, “structure molecular activity” and “ribosome” in the groups of biological process, molecular function and cellular component respectively. This indicates that ribosomal genes are among the most interesting candidates to elucidate in detail the species-specific transcriptional responses to water stress in *P. australis*, a species normally growing in water-rich environment like ponds, swamps and river sides, as compared to the other Poaceae which are more xerophilous.

4.4 Conclusions

The present study is the first in-depth comparative analysis of the patterns of whole transcriptome regulation upon simulated drought stress in the Arundinoideae clade. As the comparison was carried out between species with different drought tolerance, it provided a robust basis for further hypothesis testing and functional studies by shedding novel light on commonalities/differences of general response mechanisms between wild perennial. Given the relevance of both species analyzed in this study as propsective bionergy crops, the results of this study will hopefully contribute to the further understanding and enhance use of Aundinoidea species as renewable energy sources. Meanwhile, as one of the first in-depth transcriptome comparisons between species without genome information, this study could constitute a useful reference and stimulus for further comparative transcriptome studies in other plant caldes.

4.5 Materials and Methods

4.5.1 Plant material and application of water limitation stress

In the present study, we applied a water stress by treating cohorts of *P. australis* cuttings (collected in San Michele all‘Adige, Trentino, Italy 46°11'18.7"N 11°08'08.3"E) with two different concentrations of polyethylene glycol 6000 (PEG; 10 and 20 % w/w, referred to as mild and severe water stress conditions, corresponding to osmotic potentials of −1.54 and −5.04 bars, respectively). Briefly, *P. australis* cuttings were let rooting in tap water, then transferred to 1 % Hoagland solution and grown in a growth chamber with day-length of 16 h, light intensity of 150 µmol of photons m−2 s−1, 24 °C and 60 % RH. At the three-leaves stage, two cohorts of cuttings were transferred to 1 % Hoagland solution containing either 10 or 20 % PEG, while a third cohort used as control was transferred to 1 %

Hoagland solution without PEG. After 1 h shoots and whole roots were separately collected from each cohort, quickly rinsed in distilled water and snap-frozen in liquid nitrogen. A total of 18 samples (three biological replicates from both shoot and root for each of the three conditions) were sampled.

4.5.2 Transcriptome reconstruction, differentially expressed genes (DEGs) identification and functional annotation

The methods for transcriptome reconstruction, DEGs identification and functional annotation have been described in detail in a previous study (Fu et al, 2016). In summary, a minimum of 38 million reads were obtained from each of the 18 libraries sequenced. Quality control was carried out using the FastQC software (available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and other in-house programs as previously described (Sablok et al. 2014). RSEM (Bo and Li, 2011) was used to estimate the fragments per kilobase of exon per million fragments mapped (FPKM) values, which is a measure of the expression level. To identity genes which are differentially expressed upon water stress, EdgeR (R version: 3.0.1, edgeR version: 3.4.2; Robinson et al, 2010) was used, with a very stringent cutoff, 0.001 false discovery rate (FDR) and log fold change (LogFC) value of 2. The genes with logFC≥2 and logFC≤-2 FDR<0.001 between two treatment conditions were determined to be up-regulated and down-regulated, respectively. Functional annotation was performed by BLASTx searchers with E-value threshold $1E^{-5}$ against NCBI non-redundant (www.ncbi.nlm.nih.gov), UniProt (www.uniprot.org) and TrEMBL (Bairoch and Apweiler, 2000) plant databases.

4.5.3 Classification of gene expression patterns

In order to identify general regulatory trends, DEGs of shoot and root were separately classified based on their expression patterns. Three possible values (down-regulated, constant, up-regulated, abbreviated D, C and U, respectively) of two classification criteria (expression variation of control *versus* mild water stress condition and expression variation of mild *versus* severe water stress condition) resulted in a total of nine expression pattern classes corresponding to all possible combinations of value pairs. A logFC threshold of 2

(FDR<0.001) between conditions was used to assign each differentially expressed gene to one of the classes.

4.5.4 Pathway enrichment

Pathway enrichment analysis of DEGs was carried out with the KOBAS software (Xie et al, 2011) using BLAST searches with default threshold against the *Oryza sativa* var. *japonica* proteins.

4.5.5 Putative homologs

To identify the putative homologs of *P. australis*, Blastp searches were performed with an E-value threshold of 1E^{-5} against *A. donax* database (Fu et al, 2016), with the Reciprocal Best Hits method (RBH, Moreno-Hagelsieb and Latimer, 2008). Additionally, we mined functionally relevant genes involved in water stress by comparing with the water stress database from *Sorghum bicolor*, *Zea mays*, *Arabidopsis thaliana*, and *Oryza sativa* by OrthoMCL (Li et al. 2003).

4.6 Reference

A. Bucci, A., Cassani, E., Landoni, M., Cantaluppi, E. & Pilu, R. Analysis of chromosome number and speculations on the origin of Arundo donax L. (Giant Reed). Cytol. Genet. 47, 237–241 (2013).

Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 28, 45–8 (2000).

Bechtold, U. et al. Arabidopsis HEAT SHOCK TRANSCRIPTION FACTORA1b overexpression enhances water productivity, resistance to drought, and infection. J. Exp. Bot. 64, 3467–81 (2013).

Boyer, J. S. Plant Productivity and Environment. Science (80-.). 218, 443–448 (1982).

Brix, H. Genetic diversity, ecophysiology and growth dynamics of reed (Phragmites australis). Aquat. Bot. 64, 179–184 (1999).

Bürglin, T. R. & Affolter, M. Homeodomain proteins: an update. Chromosoma 125, 497–521 (2016).

Chaves, M. M. et al. How plants cope with water stress in the field. Photosynthesis and growth. Ann. Bot. 89, 907–916 (2002).

Clevering, O. & Lissner, J. Taxonomy, chromosome numbers, clonal diversity and population dynamics of Phragmites australis. Aquat. Bot. 64, 185–208 (1999).

Dai, M., Hu, Y., Ma, Q., Zhao, Y. & Zhou, D.-X. Functional analysis of rice HOMEOBOX4 (Oshox4) gene reveals a negative function in gibberellin responses. Plant Mol. Biol. 66, 289–301 (2008).

Dao, T. T. H., Linthorst, H. J. M. & Verpoorte, R. Chalcone synthase and its functions in plant resistance. Phytochem. Rev. 10, 397–412 (2011).

De Leonardis, A. M., Petrarulo, M., De, P. & M., A. in Advances in Selected Plant Physiology Aspects (InTech, 2012). doi:10.5772/31352

Du, Z. et al. agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res. 38, W64–70 (2010).

Dugas, D. V et al. Functional annotation of the transcriptome of Sorghum bicolor in response to

osmotic stress and abscisic acid. BMC Genomics 12, 514 (2011).

Eletto, D., Dersh, D. & Argon, Y. GRP94 in ER quality control and stress responses. Semin. Cell Dev. Biol. 21, 479–85 (2010).

Farooq, M., Wahid, A., Kobayashi, N., Fujita, D. & Basra, S. M. A. in Sustainable Agriculture 153–188 (Springer Netherlands, 2009). doi:10.1007/978-90-481-2666-8_12

Fu, Y. et al. Dissection of early transcriptional responses to water stress in Arundo donax L. by unigene-based RNA-seq. Biotechnol. Biofuels 9, 54 (2016).

Guo, M. et al. The Plant Heat Stress Transcription Factors (HSFs): Structure, Regulation, and Function in Response to Abiotic Stresses. Front. Plant Sci. 7, 114 (2016).

Hardion, L., Verlaque, R., Saltonstall, K., Leriche, A. & Vila, B. Origin of the invasive Arundo donax (Poaceae): a trans-Asian expedition in herbaria. Ann. Bot. 114, 455–62 (2014).

Hadiarto, T. & Tran, L.-S. P. Progress studies of drought-responsive genes in rice. Plant Cell Rep. 30, 297–310 (2011).

Howell, S. H. Endoplasmic reticulum stress responses in plants. Annu. Rev. Plant Biol. 64, 477–99 (2013).

Huang, G.-T. et al. Signal transduction during cold, salt, and drought stresses in plants. Mol. Biol. Rep. 39, 969–987 (2012).

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

Jacquot, J.-P. et al. Thioredoxins and related proteins in photosynthetic organisms: molecular basis for thiol dependent regulation. Biochem. Pharmacol. 64, 1065–1069 (2002).

Jia, X.-Y. et al. Molecular cloning and characterization of wheat calreticulin (CRT) gene involved in drought-stressed responses. J. Exp. Bot. 59, 739–751 (2008).

Kindler, J. & Okruszko, T. Integrated Drought Management Programme in Central and Eastern Europe (IDMP CEE). Ann. Warsaw Univ. Life Sci. L. Reclam. 46, 167–179 (2014).

Köbbing, J. F., Thevs, N. & Zerbe, S. The utilisation of reed (Phragmites australis): a review. Mires Peat 13, 1–14 (2013).

Kornfeld, R. & Kornfeld, S. Assembly of Asparagine-Linked Oligosaccharides. Annu. Rev. Biochem. 54, 631–664 (1985).

Li, B. et al. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323 (2011).

Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–89 (2003).

Liu, J.-X., Srivastava, R., Che, P. & Howell, S. H. An endoplasmic reticulum stress response in Arabidopsis is mediated by proteolytic processing and nuclear relocation of a membrane-associated transcription factor, bZIP28. Plant Cell 19, 4111–9 (2007).

Ma, Y. & Hendershot, L. M. ER chaperone functions during normal and stress conditions. J. Chem. Neuroanat. 28, 51–65 (2004).

Mann, J. J., Barney, J. N., Kyser, G. B. & Di Tomaso, J. M. Miscanthus × giganteus and Arundo donax shoot and rhizome tolerance of extreme moisture stress. GCB Bioenergy 5, 693–700 (2013).

Masaru Ohta & Fumio Takaiwa. Emerging features of ER resident Jproteins in plants. Plant Signal. & Behav. 9, e28194 (2016).

Moreno-Hagelsieb, G. & Latimer, K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatics 24, 319–24 (2008).

Moreno, A. A. & Orellana, A. The physiological role of the unfolded protein response in plants. Biol. Res. 44, 75–80 (2011).

Morreale, G., Conforti, L., Coadwell, J., Wilbrey, A. L. & Coleman, M. P. Evolutionary divergence of valosin-containing protein/cell division cycle protein 48 binding interactions among endoplasmic reticulum-associated degradation proteins. FEBS J. 276, 1208–1220 (2009).

Nagasawa, K., Higashi, T., Hosokawa, N., Kaufman, R. J. & Nagata, K. Simultaneous induction of the four subunits of the TRAP complex by ER stress accelerates ER degradation. EMBO Rep. 8, 483–9 (2007).

Nguyen, L. X. et al. Photosynthesis of co-existing Phragmites haplotypes in their non-native range: are characteristics determined by adaptations derived from their native origin? AoB Plants 5, plt016–plt016 (2013).

Ohta, M. & Takaiwa, F. Emerging features of ER resident J-proteins in plants. Plant Signal. Behav. 9, e28194 (2014).

Ogawa, D., Yamaguchi, K. & Nishiuchi, T. High-level overexpression of the Arabidopsis HsfA2 gene confers not only increased themotolerance but also salt/osmotic stress tolerance and enhanced callus growth. J. Exp. Bot. 58, 3373–83 (2007).

Pagter, M., Bragato, C. & Brix, H. Tolerance and physiological responses of Phragmites australis to water deficit. Aquat. Bot. 81, 285–299 (2005).

Pilu, R., Bucci, A., Badone, F. C. & Landoni, M. Giant reed (Arundo donax L.): A weed plant or a promising energy crop? African J. Biotechnol. 11, 9163–9174 (2012).

Qi, X., Xie, S., Liu, Y., Yi, F. & Yu, J. Genome-wide annotation of genes and noncoding RNAs of foxtail millet in response to simulated drought stress by deep sequencing. Plant Mol. Biol. 83, 459–473 (2013).

Riechmann, J. L. et al. Arabidopsis Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes. Science (80-.). 290, (2000).

Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–40 (2010).

Ruiz-Sola, M. Á. & Rodríguez-Concepción, M. Carotenoid biosynthesis in Arabidopsis: a colorful pathway. Arab. B. 10, e0158 (2012).

Sablok, G. et al. Fuelling genetic and metabolic exploration of C 3 bioenergy crops through the first reference transcriptome of Arundo donax L. Plant Biotechnol. J. 12, 554–567 (2014).

Scharf, K.-D., Berberich, T., Ebersberger, I. & Nover, L. The plant heat stress transcription factor (Hsf) family: structure, function and evolution. Biochim. Biophys. Acta 1819, 104–19 (2012).

Schuck, S., Prinz, W. A., Thorn, K. S., Voss, C. & Walter, P. Membrane expansion alleviates endoplasmic reticulum stress independently of the unfolded protein response. J. Cell Biol. 187, (2009).

Selmar, D. & Kleinwachter, M. Stress Enhances the Synthesis of Secondary Plant Products: The Impact of Stress-Related Over-Reduction on the Accumulation of Natural Products. Plant Cell Physiol. 54, 817–826 (2013).

Shen, Y., Meunier, L. & Hendershot, L. M. Identification and characterization of a novel endoplasmic reticulum (ER) DnaJ homologue, which stimulates ATPase activity of BiP in vitro and is induced by ER stress. J. Biol. Chem. 277, 15947–56 (2002).

Simova-Stoilova, L. P. et al. 2-DE proteomics analysis of drought treated seedlings of Quercus ilex supports a root active strategy for metabolic adaptation in response to water shortage. Front. Plant Sci. 6, 627 (2015).

Sormani, R., Masclaux-Daubresse, C., Daniel-Vedele, F., Daniele-Vedele, F. & Chardon, F. Transcriptional regulation of ribosome components are determined by stress according to cellular compartments in Arabidopsis thaliana. PLoS One 6, e28070 (2011).

Srivastava, J., Chandra, H., Nautiyal, A. R. & Kalra, S. J. S. Response of C 3 and C 4 Plant Systems Exposed to Heavy Metals for Phytoextraction at Elevated Atmospheric CO2 and at Elevated Temperature. Environ. Contam. 3–17 (2012). doi:10.5772/1124

Vandepoele, K. & Van de Peer, Y. Exploring the plant transcriptome through phylogenetic profiling. Plant Physiol. 137, 31–42 (2005).

Velikova, V. et al. Physiological significance of isoprenoids and phenylpropanoids in drought response of Arundinoideae species with contrasting habitats and metabolism. Plant. Cell Environ. 39, 2185–2197 (2016).

Wang, J. et al. Expression changes of ribosomal proteins in phosphate- and iron-deficient Arabidopsis roots predict stress-specific alterations in ribosome composition. BMC Genomics 14, 783 (2013).

Wang, J. et al. Expression changes of ribosomal proteins in phosphate- and iron-deficient Arabidopsis roots predict stress-specific alterations in ribosome composition. BMC Genomics 14, 783 (2013).

Welsch, R., Wüst, F., Bär, C., Al-Babili, S. & Beyer, P. A third phytoene synthase is devoted to abiotic

stress-induced abscisic acid formation in rice and defines functional diversification of phytoene synthase genes. Plant Physiol. 147, 367–80 (2008).
Xie, C. et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 39, W316–22 (2011).
Xue, S. & Barna, M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. Nat. Rev. Mol. Cell Biol. 13, 355–69 (2012).
Zhang, L. et al. Identification of Gene Modules Associated with Drought Response in Rice by Network-Based Analysis. PLoS One 7, e33748 (2012).
Zheng, W. J., Zheng, X. P. & Zhang, C. L. A survey of photosynthetic carbon metabolism in 4 ecotypes of Phragmites australis in northwest China : Leaf anatomy , ultrastructure , and activities of ribulose 1 , 5-bisphosphate carboxylase , phosphoenolpyruvate carboxylase and glycollate oxidase. Physiol. Plant. 110, 201–208 (2000).
Zhu, C. et al. Molecular Characterization and Expression Profiling of the Protein Disulfide Isomerase Gene Family in Brachypodium distachyon L. PLoS One 9, e94704 (2014).
Zulauf, C. 2012 Drought: Yield Loss, Revenue Loss, and Harvest Price Option. (2012).

CONCLUSION

5

"DNA neither cares nor knows. DNA just is. And we dance to its music."

— Richard Dawkins, River Out of Eden: A Darwinian View of Life

The present study provides the first comprehensive analysis of the organ-specific transcriptome of *A. donax*, identifying several expressed genes that could be preferential targets for functional studies, for metabolic engineering or for tailoring growth habit/development of the giant reed to higher bioenergy yield. The transcriptomics data developed will provide a much awaited resource to further our limited understanding of the biology of *A. donax* and constitute the first of the three pillars (transcriptomics, proteomics and metabolomics) upon which an integrative systems biology characterization of this promising bioenergy crop can be built in the future. Furthermore, we provided the first characterization of *A. donax* shoot and root transcriptomes in response to water stress, one of the factors of highest concern for its productivity. Given the commonality of the responses to water limitation and other stresses, in addition to providing a general overview of the early transcriptional responses to simulated drought, our results shed also light at the molecular level on the general mechanisms of stress response and adaptation in *A. donax*. Upon functional validation, thus, many of the unigenes identified in the present study have the potential to be used for the development of novel *A. donax* varieties with improved productivity and stress tolerance. In addition, the inventory of early-responsive genes to water stress provided in this study could constitute useful markers of the physiological status of *A. donax* plantations to deepen our understanding of its biology and productivity under water limitation. Lastly, the comparative study in the Arundinoideae clade, in particular between species with different drought tolerance, provided the basis for functional studies and shed novel light on commonalities/differences of general response mechanisms between wild perennial grasses and those that are specifically used as energy crops. Meanwhile, as one of the first in-depth transcriptome comparisons between species without genome information, this work could constitute a useful reference and stimulus for further comparative transcriptome studies on other species.

FUTURE PROSPECTS

RNA-seq technology: RNA-seq technology and tools are continuosly evolving. It should be noted that with the development of new technique of long-read sequencing, we could obtain better transcript identification. As we know, the difficulty in accurately reconstructing expressed full-length transcripts from the assembly of reads is the main limitation of short-read RNA-seq, especially for polyploid species like those studied here. The complexity and difficulty of short-read assembly limited our study to provide with absolute accuracy full-length novel transcripts, and to specify the expression level of genes with highly similar CDS. Long-read sequencing provides the possibility of sequencing the full-length transcripts without the need for an assembly step. Nevertheless, long-read sequencing is still immature. High error rate limits the application of this technology. With the development of the sequencing technology, eventually, the sequencing accuracy will be improved, and will allow us to obtain a de novo transcript at isoform-resolution.

Experimental design: With the better sequencing quality and speed and lower cost, we may carry out a series of experiment with time course and different environmental stresses. This may help to build a real stress related co-expression network, to elucidate the relationships between the regulated genes, further helping to understand the metabolic pathways in detail. Meanwhile, the plant cultivation could be improved, since the roots should avoid light as root greening has been noticed in the comparative study.

Analyses: Until now, there is no single optimal pipeline for transcript quantification, normalization and ultimately differential expression analysis which could be suitable for all RNA-seq experimental scenarios. With the development of RNA-seq technologies and corresponding analytical tools, in the future, the development of fully automated or at least semi-automated methods/pipelines is also required, especially for comparative studies.

ACKNOWLEDGEMENT

This project would not have been possible without the support of many people. Many thanks to my supervisor, Claudio Varotto, for the patient guidance, encouragement and advice he has provided throughout my time as his student. Also thanks to my tutor, Antonella Furini and committee members, who offered guidance and support. Thanks to the Fondazione Edmund Mach (The project has been funded with support from Fondazione Edmund Mach. Core funding of the Ecogenomics group supported by the Autonomous Province of Trento (Italy) and MAN-VIP project) and university of Verona for awarding me a fellowship, providing me with the financial means to complete this project.

At last, thanks to my dear husband Michele, my lovely daughter Gaia, my parents, and numerous friends who endured this long process with me, always offering support and love.

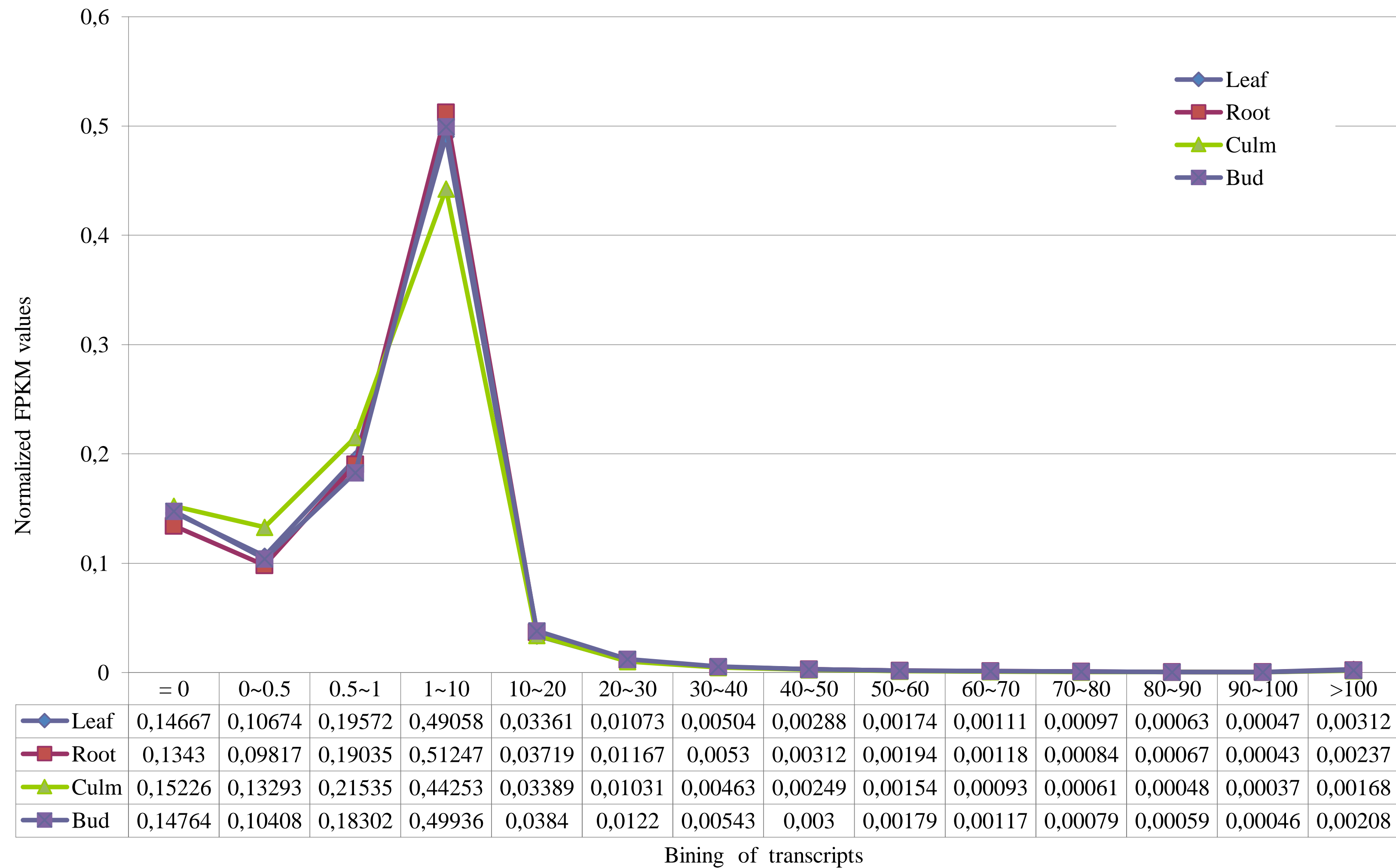
Appendix

Supplementary Ch2.1: Summary statistics of transcript assembly at k-mer from 31 to 51 in Velvet-Oases and single k-mer in Trinity (minimum transcript length =200bp).

	Trinity Assemblies					
	Bud	**Culm**	**Leaf**	**Root**	**Culm (454)**	**Root (454)**
Total number of transcripts	259505	279143	222927	254828	154338	129404
Total number of transcripts bases (bp)	259665320	294922043	233942276	236696470	114306126	86170851
Min sequence length	201	201	201	201	200	200
Max sequence length	12095	13063	12053	15019	8149	7177
Average sequence length	1000,62	1056,53	1049,41	928,85	740,62	665,91
Median sequence length	616	656	624	532	589	533
N50	1682	1791	1809	1607	800	727
As	27,73%	27,36%	27,78%	27,60%	26,78%	26,67%
Ts	27,77%	27,48%	27,83%	27,69%	28,28%	27,65%
Gs	22,39%	22,69%	22,32%	22,49%	23,91%	24,28%
Cs	22,11%	22,48%	22,06%	22,22%	21,03%	21,40%
(A+T)s	55,50%	54,84%	55,61%	55,29%	55,05%	54,32%
(G+T)s	44,50%	45,16%	44,39%	44,71%	44,94%	45,68%
Ns	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%

Velvet-Oases assemblies					
Bud	**k31**	**k35**	**k41**	**k45**	**k51**
Total reads	77433992	77433992	77433992	77433992	77433992
Used reads	63959551	62822469	62009189	60778233	58291813
Unused reads	13474441	14611523	15424803	16655759	19142179
Nodes	13111154	1116365	727369	615353	413539
N50 (bp)	162	186	236	276	340
Largest contig length (bp)	5463	5434	6601	6601	6576
Culm	**k31**	**k35**	**k41**	**k45**	**k51**
Total reads	89852774	89852774	89852774	89852774	89852774
Used reads	73274055	72241579	71555933	70077890	67317968
Unused reads	16578719	17611195	18296841	19774884	22534806
Nodes	1451903	1226928	793696	674872	453020
N50 (bp)	164	191	240	279	345
Largest contig length (bp)	5762	6766	6766	6766	6590
Leaf	**k31**	**k35**	**k41**	**k45**	**k51**
Total reads	78735580	78735580	78735580	78735580	78735580
Used reads	50841477	54349635	53116409	53687815	51686114
Unused reads	27894103	24385945	25619171	25047765	27049466
Nodes	1736555	1031544	807431	552572	418143
N50 (bp)	148	192	233	279	343
Largest contig length (bp)	6496	6600	6399	5953	7007
Root	**k31**	**k35**	**k41**	**k45**	**k51**
Total reads	77778024	77778024	77778024	77778024	77778024
Used reads	59168827	62460222	61106282	60869819	58196404
Unused reads	18609197	15317802	16671742	16908205	19581620
Nodes	1804993	1084642	846698	593566	398605
N50 (bp)	145	184	225	269	332
Largest contig length (bp)	4569	6537	6351	5761	6384

Supplementary Ch2.2: Distribution of the FPKM values as a function of the assembled transcripts across the four organs.



	= 0	0~0.5	0.5~1	1~10	10~20	20~30	30~40	40~50	50~60	60~70	70~80	80~90	90~100	>100
Leaf	0,14667	0,10674	0,19572	0,49058	0,03361	0,01073	0,00504	0,00288	0,00174	0,00111	0,00097	0,00063	0,00047	0,00312
Root	0,1343	0,09817	0,19035	0,51247	0,03719	0,01167	0,0053	0,00312	0,00194	0,00118	0,00084	0,00067	0,00043	0,00237
Culm	0,15226	0,13293	0,21535	0,44253	0,03389	0,01031	0,00463	0,00249	0,00154	0,00093	0,00061	0,00048	0,00037	0,00168
Bud	0,14764	0,10408	0,18302	0,49936	0,0384	0,0122	0,00543	0,003	0,00179	0,00117	0,00079	0,00059	0,00046	0,00208

Supplementary Ch2.3: Functional annotation and FPKM values for the highly expressed genes. (Online)

Supplementary Ch2.4: Classification of over- and under- represented GO categories across the four organs. (Online)

Supplementary Ch2.5: Mining of the *A. donax* transcriptome for putative transcript homologs of stress-related genes. (Detail Online)

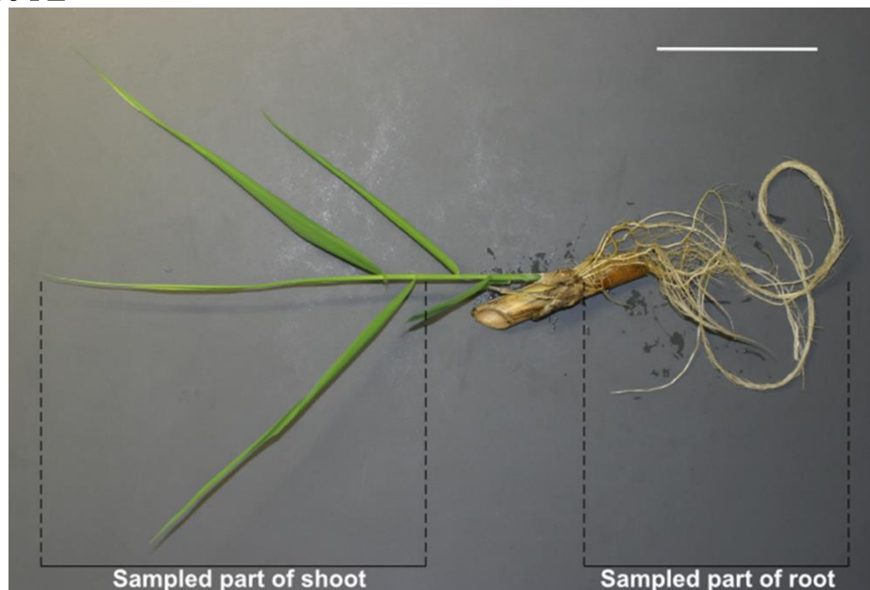
From: Borkotoky, S., Saravanan, V., Jaiswal, A., Das, B., Selvaraj, S., Murali, A. and Lakshmi, P.T. (2013) The Arabidopsis Stress Responsive Gene Database. International Journal of Plant Genomics, 2013:949564

Type of Stress	**ASRGDB genes**	***A. donax* transcript homologs**
ABCISSIC ACID	1	6
ABIOTIC	7	19
ALUMINIUM	1	7
CADMIUM	8	77
COLD	4	18
COPPER	3	6
DEHYDRATION	5	44
DNA DAMAGE	1	21
DROUGHT	10	53
ENDOPLASMIC RETICULUM	2	7
HEAT	6	23
IRON	1	4
KARRIKIN	1	4
LIGHT	2	15
MAGNESIUM	1	1
OSMOTIC	9	222
OXIDATIVE	21	138
PHOTOOXIDATIVE	1	3
REACTIVE OXYGEN SPECIES	1	3
SALT	47	628
STRESS	11	127
TOUCH	1	7
UV	1	3
WATER	3	10

Supplementary Ch3.1: Experimental design and number of reads per library.

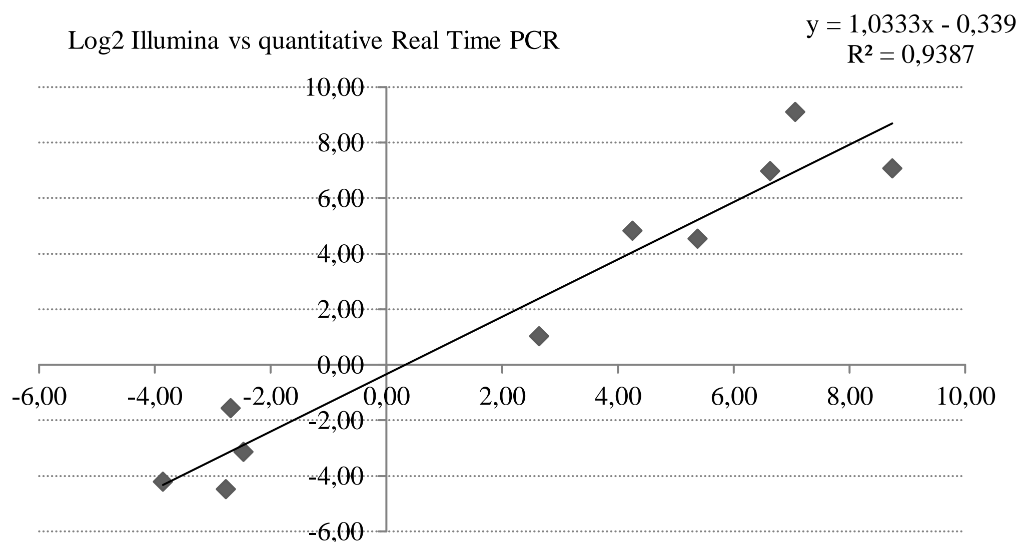
Library name	Number of Sequencing Reads*	Treatment/Condition	Organ
Ad_Control_1	47792849	Untreated control	Shoot
Ad_Control_2	35432010	Untreated control	Shoot
Ad_Control_3	32186149	Untreated control	Shoot
Ad_PEG10%_1	38952041	10% PEG	Shoot
Ad_PEG10%_2	34331086	10% PEG	Shoot
Ad_PEG10%_3	30234460	10% PEG	Shoot
Ad_PEG20%_1	32648859	20% PEG	Shoot
Ad_PEG20%_2	34100100	20% PEG	Shoot
Ad_PEG20%_3	38857109	20% PEG	Shoot
Ad_Control_1	41912793	Untreated control	Root
Ad_Control_2	35027592	Untreated control	Root
Ad_Control_3	35234162	Untreated control	Root
Ad_PEG10%_1	42457897	10% PEG	Root
Ad_PEG10%_2	42854988	10% PEG	Root
Ad_PEG10%_3	43240510	10% PEG	Root
Ad_PEG20%_1	40387040	20% PEG	Root
Ad_PEG20%_2	42169350	20% PEG	Root
Ad_PEG20%_3	46354173	20% PEG	Root

* 2x100 PE



Schematic representation of the portion of *A. donax* shoot and root sampled for RNA-Seq. Bar: 5 cm.

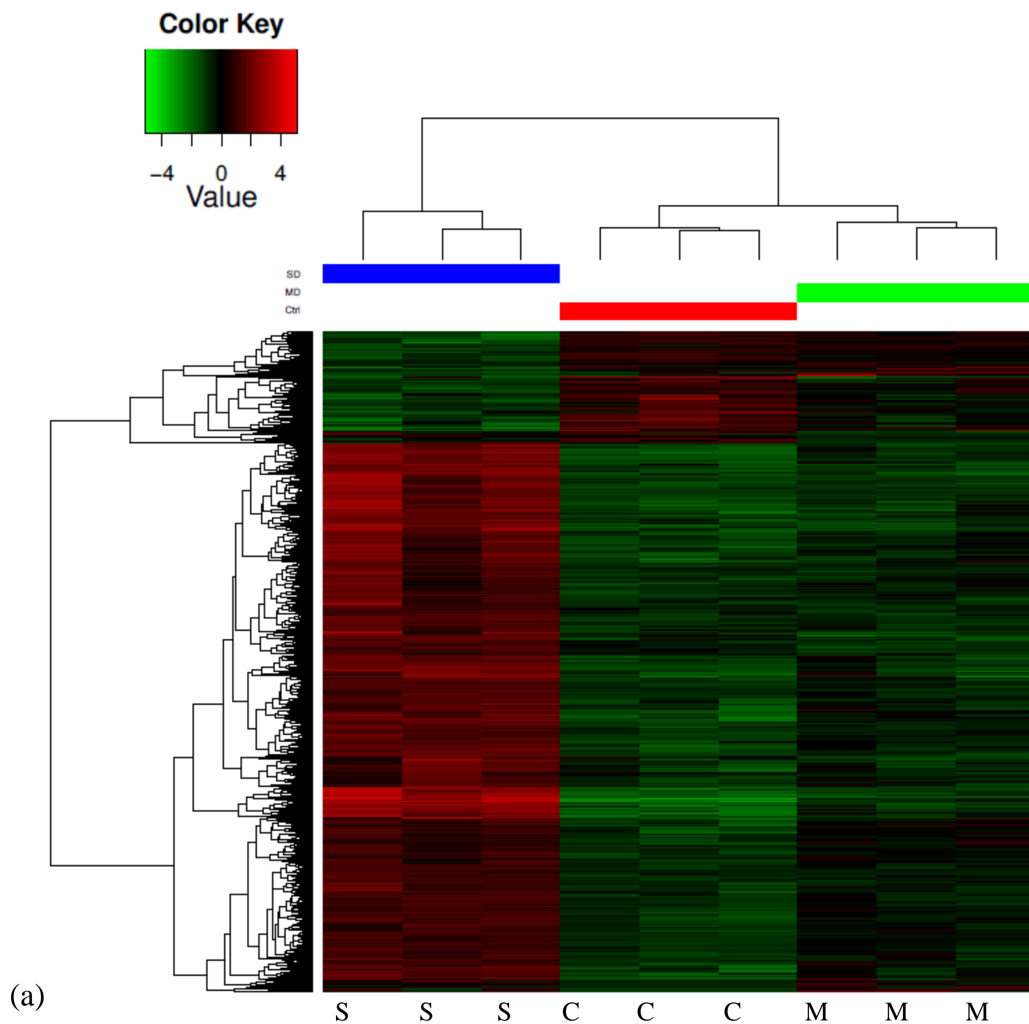
Supplementary Ch3.2: Validation of 10 *A. donax* DEGs by Real Time qrt-PCR.

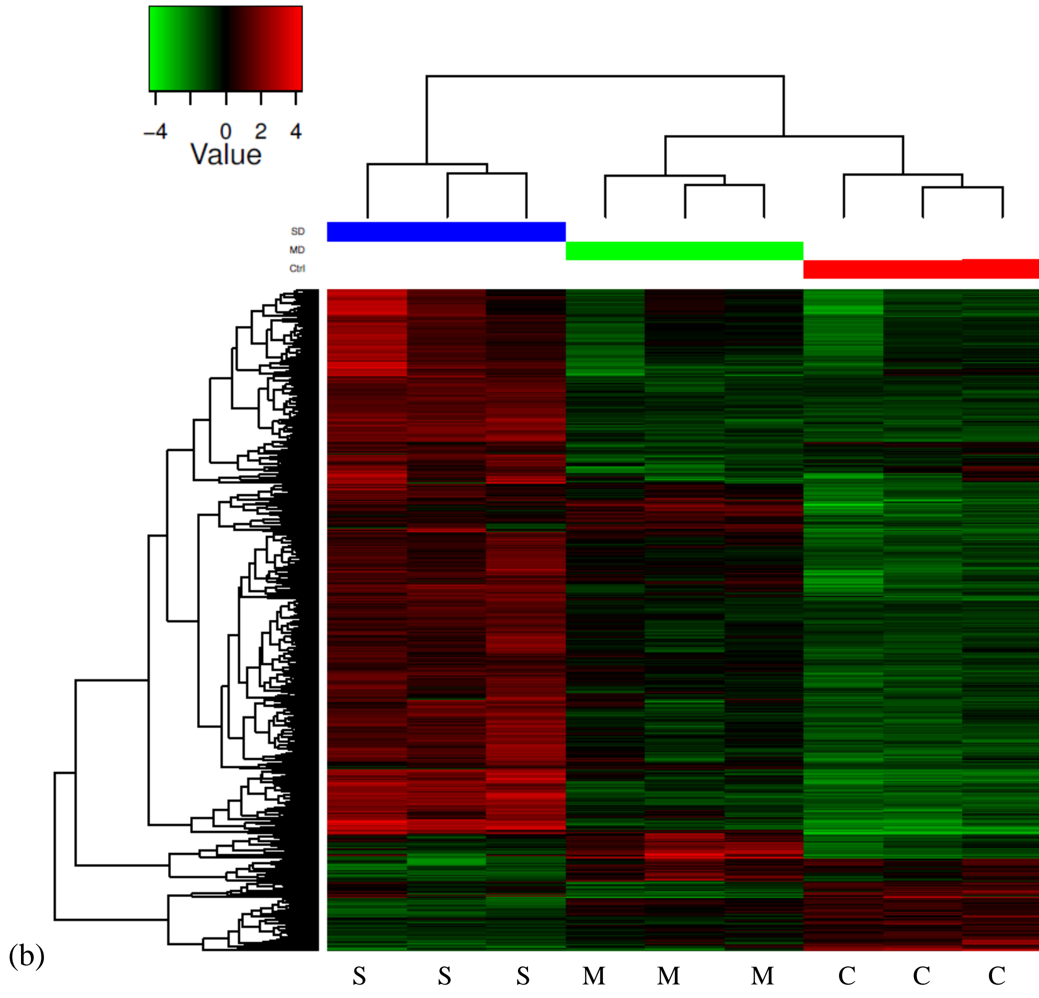


Primer List

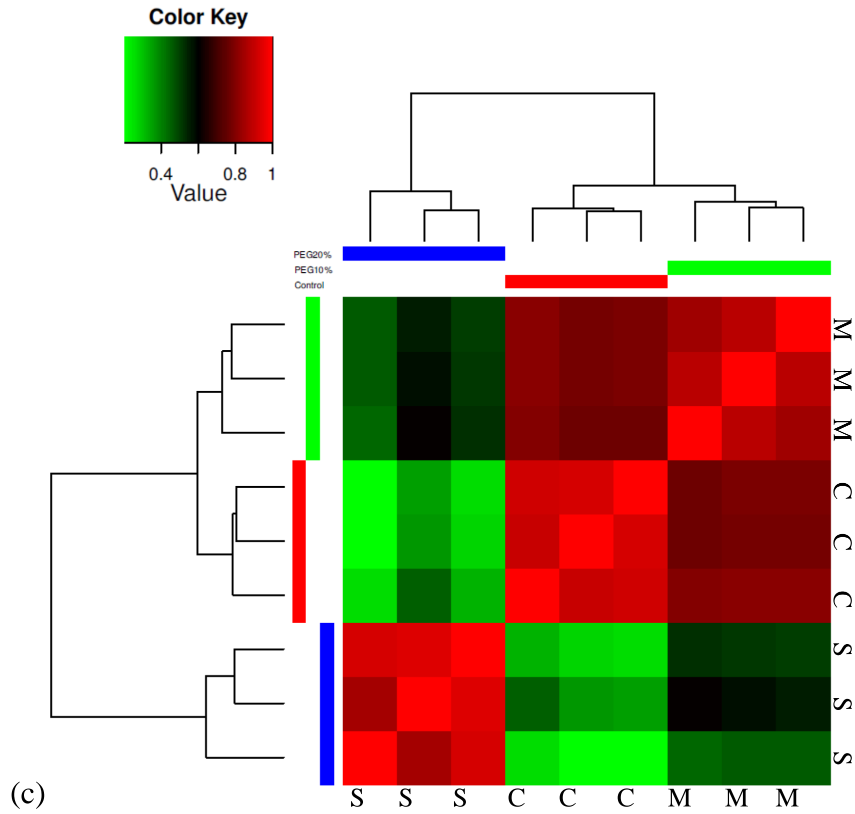
Organ/Regulation	Gene ID	Primer F	Primer R
Shoot_up	Arundo_WS_Unigene028474	AACCTCCCTCATTTCCTCGAAC	TTTTGCGCTGATACGTGTGG
	Arundo_WS_Unigene065161	AACCATCGTTGCTTCGCTTG	TGGCTAACCAACAACAACCG
Shoot_down	Arundo_WS_Unigene045711	ACTGACAGGCTGAGTTCGAC	AGCACAGCTGGCATTCAGAT
	Arundo_WS_Unigene074110	GTCTCTTCTTTGTGCCATCTGC	AGCAGCGATCTCCTTTATCTGG
Root_up	Arundo_WS_Unigene011476	GCAGCACGACTACTACTTTCAC	ATGGCATGCAACAGCAATCC
	Arundo_WS_Unigene053771	AATCGTGGCGTTGTATCGTG	TTGGCGACCAAATTCATGCC
	Arundo_WS_Unigene020231	GATGAAATGGGGATGGGTGAAC	TGCTTTTGCCAGCATTGTCC
	Arundo_WS_Unigene023719	ACAAGGGAACAGGTTCAGGTC	CGCCAACGAAATGCCAAAAG
Root_down	Arundo_WS_Unigene042884	AGGGAAAGCTTGACACGATG	ATGGTACCATTGCGTCTTGG
	Arundo_WS_Unigene077521	GCTCGTCGAACTTCTTGGTG	ACCTTGCTAGTGACCGTATCAG
All	AdoACT	TCTTGGCTTGCATTCTTGGG	TGGATTGCGAAGGCTGAGTAC

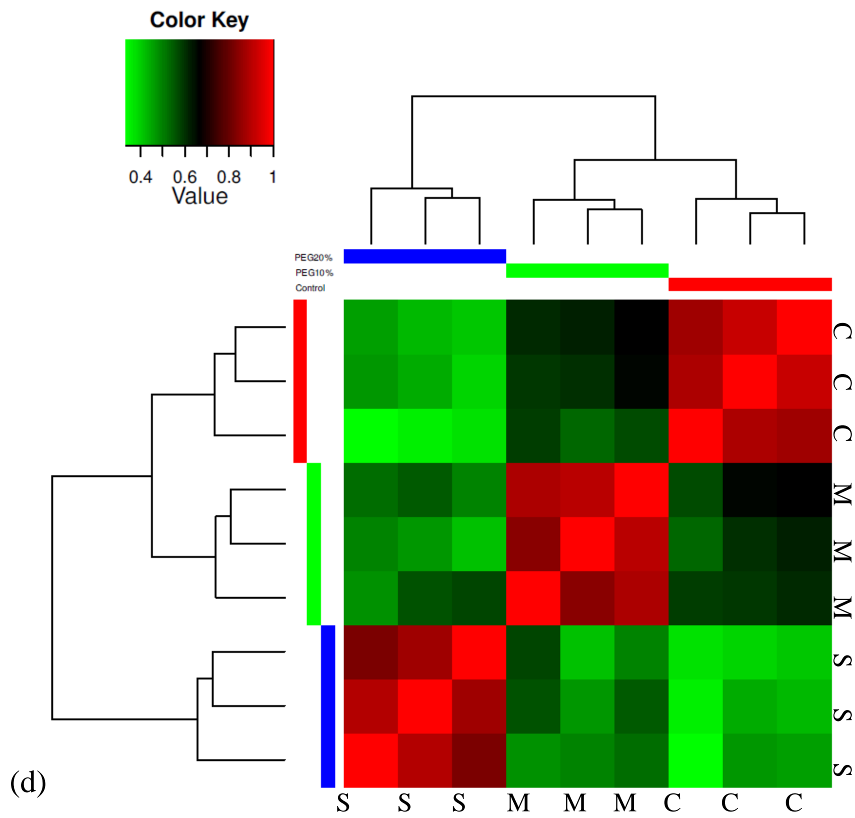
Supplementary Ch3.3: Visualization of differentially expressed genes. Heat maps visualization of the differentially expressed genes between three conditions for shoots (a) and roots (b) using Euclidean distances between TMM normalized expression values. Expression levels for genes in each cDNA library were measured as fragments per kilobase per million reads (FPKM), and color-coded from green (lowly expressed) to red (highly expressed). Hierarchical clustering level at individual libraries is represented by the dendrogram for shoots (c) and roots (d), and color-coded from green (weak correlation) to red (strong correlation). S: Severe water stress (20% PEG) M: Mild water stress (10% PEG) C: Control condition (No PEG)





(b)


Color Key
0.4
0.8
1
Value
PEG20%
PEG10%
Control
M
M
M
C
C
C
S
S
S
S S S C C C M M M
(c)


Color Key
0.4
0.6
0.8
1
Value
PEG20%
PEG10%
Control
C
C
C
M
M
M
S
S
S
S S S M M M C C C
(d)

Supplementary Ch3.4: Stress-related functional categories of DEGs. (Detail Online)

TYPE OF STRESS	ASRGDB GENES	A. DONAX TRANSCRIPT HOMOLOGS	A. DONAX TRANSCRIPT HOMOLOGS DIFFERENTIALLY EXPRESSED IN SHOOT	A. DONAX TRANSCRIPT HOMOLOGS DIFFERENTIALLY EXPRESSED IN ROOT
ABIOTIC	15	35	3	2
ABSCISIC ACID	8	33	0	0
ALUMINIUM	4	10	0	0
CADMIUM	12	68	5	1
CATION	1	4	0	0
CELLULAR	1	5	0	0
CHILLING	1	1	0	0
COLD	11	64	4	7
COPPER	4	16	0	2
DEHYDRATION	13	97	24	3
DNA DAMAGE	1	8	0	0
DROUGHT	20	88	3	1
ENDOPLASMIC RETICULUM	10	19	0	0
FREEZING	1	9	0	3
GENOTOXIC	1	8	0	0
HEAT	16	41	2	0
HYPOXIC STRESS	1	10	5	6
IRON	1	4	0	0
KARRIKIN	1	7	0	1
LIGHT	7	19	6	5
MAGNESIUM	1	3	0	0
MALONDIALDEHYDE	1	2	0	0
METAL	1	4	0	0
OSMOTIC	20	176	21	8
OXIDATIVE	85	482	32	33
PHOSPHATE	1	5	2	0
PHOTOOXIDATIVE	1	1	0	0
POTASSIUM	1	2	0	0
REACTIVE OXYGEN SPECIES	1	1	0	0
SALINITY RESPONSE	2	19	2	5
SALT	95	540	46	42
SODIUM	1	26	0	1
STRESS	39	147	13	6
TEMPERATURE	1	4	0	0
TOUCH	1	20	4	2
UV	1	1	0	0
WATER	6	22	5	6
ZINC	8	16	3	1

Supplementary Ch3.5: Functional annotation, fold change values and statistical significance of DEGs. (Online)

Supplementary Ch3.6: Slimmed GO terms overrepresented under different conditions.

List of over-represented Molecular Function GO terms in common between different combinations of organ/condition comparisons corresponding to the Venn chart in Figure 3.3.

Intersection	Total	Elements
Shoot-SC_MolFunct Root-SC_MolFunct Root-SM_MolFunct	3	GO:0003700sequence-specific DNA binding transcription factor activity GO:0001071nucleic acid binding transcription factor activity GO:0030528transcription regulator activity
Shoot-SC_MolFunct Root-SC_MolFunct	4	GO:0043565sequence-specific DNA binding GO:0004657proline dehydrogenase activity GO:0016645"oxidoreductase activity acting on the CH-NH group of donors" GO:0004722protein serine/threonine phosphatase activity
Root-SC_MolFunct Root-SM_MolFunct	7	GO:0003735structural constituent of ribosome GO:0003677DNA binding GO:0004551nucleotide diphosphatase activity GO:0047884FAD diphosphatase activity GO:0005198structural molecule activity GO:0016161beta-amylase activity GO:0047274galactinol-sucrose galactosyltransferase activity
Root-SC_MolFunct	3	GO:0015035protein disulfide oxidoreductase activity GO:0046983protein dimerization activity GO:0045548phenylalanine ammonia-lyase activity
Shoot-SC_MolFunct	21	GO:0004497monooxygenase activity GO:0005509calcium ion binding GO:0016776"phosphotransferase activity phosphate group as acceptor" GO:0016165linoleate 13S-lipoxygenase activity GO:0004623phospholipase A2 activity GO:0009055electron carrier activity GO:0005506iron ion binding GO:0003883CTP synthase activity GO:00055451-phosphatidylinositol binding

		GO:0015291secondary active transmembrane transporter activity GO:0003824catalytic activity GO:0016301kinase activity GO:0004143diacylglycerol kinase activity GO:0019201nucleotide kinase activity GO:1901476carbohydrate transporter activity GO:0004713protein tyrosine kinase activity GO:0016757"transferase activity transferring glycosyl groups" GO:0016740transferase activity GO:0003978UDP-glucose 4-epimerase activity GO:0030554adenyl nucleotide binding GO:0016758"transferase activity transferring hexosyl groups"

List of over-represented Biological Process GO terms in common between different combinations of organ/condition comparisons corresponding to the Venn chart in Figure 3.3.

Intersection	Total	elements
Shoot-SC_BiolProc Root-SC_BiolProc Root-SM_BiolProc	5	GO:0071704organic substance metabolic process GO:0044260cellular macromolecule metabolic process GO:0065007biological regulation GO:0044237cellular metabolic process GO:0009058biosynthetic process
Shoot-SC_BiolProc Root-SC_BiolProc	4	GO:0008152metabolic process GO:0006971hypotonic response GO:0006562proline catabolic process GO:0009819drought recovery
Root-SC_BiolProc	8	GO:0009310amine catabolic process GO:0009816"defense response to bacterium incompatible interaction" GO:0009605response to external stimulus GO:0014070response to organic cyclic compound GO:0009611response to wounding GO:0009607response to biotic stimulus GO:0016101diterpenoid metabolic process GO:0006807nitrogen compound metabolic process
Root-SC_BiolProc Root-SM_BiolProc	3	GO:0006412translation GO:0010467gene expression GO:0000272polysaccharide catabolic process
Shoot-SC_BiolProc	24	GO:0043412macromolecule modification GO:0016050vesicle organization GO:0048268clathrin coat assembly

		GO:0006796phosphate-containing compound metabolic process
		GO:0043085positive regulation of catalytic activity
		GO:0048518positive regulation of biological process
		GO:0006464cellular protein modification process
		GO:0009405pathogenesis
		GO:0008643carbohydrate transport
		GO:0006793phosphorus metabolic process
		GO:0015837amine transport
		GO:0007205protein kinase C-activating G-protein coupled receptor signaling pathway
		GO:0023052signaling
		GO:0031399regulation of protein modification process
		GO:0010562positive regulation of phosphorus metabolic process
		GO:0023060signal transmission
		GO:0016311dephosphorylation
		GO:0051174regulation of phosphorus metabolic process
		GO:0006470protein dephosphorylation
		GO:0044238primary metabolic process
		GO:0043687post-translational protein modification
		GO:0007154cell communication
		GO:0009987cellular process
		GO:0034637cellular carbohydrate biosynthetic process
Shoot-SC_BiolProc Root-SM_BiolProc	3	GO:1901576organic substance biosynthetic process
		GO:0043170macromolecule metabolic process
		GO:0050794regulation of cellular process
Root-SM_BiolProc	1	GO:0009685gibberellin metabolic process

Settings for REVIGO (http://revigo.irb.hr/)

Allowed similarity: Small (0.5)

Numbers associated to GO categories are p-values

Database with GO term sizes: Oryza sativa

Semantic similarity measure to use SimRel

Guide to names:

S: Severe water stress (20% PEG) M: Mild water stress (10% PEG) C: Control condition (No PEG)

E.g.: The 3 Biological Process GO terms listed under "Root-SC_BiolProc Root-SM_BiolProc" are those significantly over-represented both in the comparison between severe and control condition in root as well as in the comparison between severe and mild water stress in root.

Supplementary Ch3.7: Transcription factors responsive to water stress in *A. donax*. (Detail Online)

TF family	Shoot	Root	Rice homologs
NAC	24	23	14
WRKY	20	19	16
AP2-EREBP	9	21	13
bHLH	15	15	15
bZIP	19	10	6
Tify	7	9	4
AUX/IAA	1	10	6
MYB	3	8	5
HSF	5	4	5
C2H2	5	2	6
TUB	7	0	1
G2-like	3	3	5
GRAS	5	1	5
HB	4	2	3
MYB-related	2	3	2
Trihelix	2	3	2
C2C2-CO-like	1	3	2
C2C2-Dof	1	2	2
C2C2-GATA	3	0	3
TRAF	1	2	2
ABI3VP1	0	2	1
C3H	1	1	2
DBB	0	2	2
DBP	1	1	1
Jumonji	2	0	1
LOB	0	2	1
LUG	2	0	1
TCP	0	2	1
ARF	1	0	1
ARID	1	0	1
BES1	1	0	1
CAMTA	1	0	1
GNAT	1	0	1
MADS	1	0	1
PHD	1	0	1
PLATZ	0	1	1
Pseudo_ARR-B	1	0	1

Supplementary Ch3.8: Comparison of co-regulated expression modules responsive to water stress in *A. donax* and rice. (Online)

Supplementary Ch3.9: Drought responsive orthologs identified among *A. donax*, foxtail, sorghum and rice. (Online)

Supplementary Ch4.1: Functional annotation of DEGs. DE genes (annotation, similarity, fold-change, p-values) (Online)

Supplementary Ch4.2: General Statistics on *P. australis*' assembly compare with *A. donax*

	P. australis	*A. donax*
Total length of sequence:	91526462 bp	75960964 bp
Total number of sequences:	82623	80355
N50 stats:	50% of total sequence length is contained in the 15955 sequences >= 1858 bp	50% of total sequence length is contained in the 15006 sequences >= 1570 bp
N75 stats:	75% of total sequence length is contained in the 32096 sequences >= 1042 bp	75% of total sequence length is contained in the 31553 sequences >= 809 bp
GC %:	47.13%	47.70%

Supplementary Ch4.3: GO enrichment. (Online)

Supplementary Ch4.4: Stress-related functional categories of DEGs. (Detail Online)

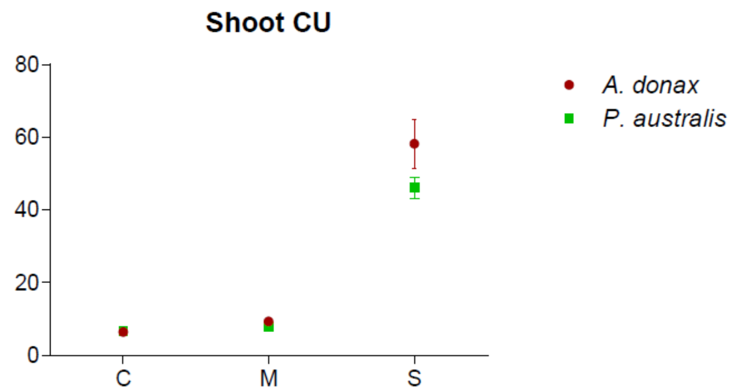
Type of Stress	**ASRGDB genes**	***P. australis* transcript homologs**	***P. austalis* transcript homologs differentially expressed in shoot**	***P. austalis* transcript homologs differentially expressed in root**
ABCISSIC ACID	5	15	1	2
ABIOTIC	15	29	2	2
ALUMINIUM	5	9	0	0
CADMIUM	11	45	0	7
CATION	1	3	0	0
CELLULAR	1	3	0	0
CHILLING	1	1	0	1
COLD	10	31	0	8
COPPER	7	16	0	0
DEHYDRATION	9	22	0	2
DNA DAMAGE	1	4	0	0

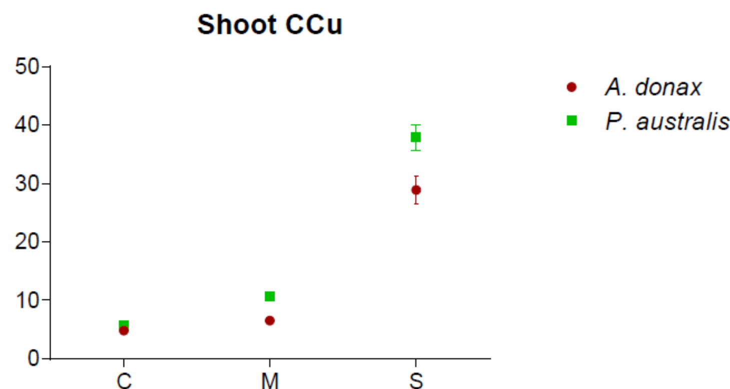
DROUGHT	16	67	2	4
ENDOPLASMIC RETICULUM	9	26	0	3
FREEZING	1	3	0	1
GENOTOXIC	1	9	0	0
HEAT	11	24	0	1
HYDROGEN PEROXIDE	1	1	0	0
HYPOXIC STRESS	1	12	3	7
IRON	1	2	0	0
KARRIKIN	1	2	0	0
LIGHT	6	12	0	4
MAGNESIUM	1	3	0	0
MALONDIALDEHYDE	1	2	0	0
METAL	1	9	0	0
OSMOTIC	17	111	3	4
OXIDATIVE	53	197	2	25
PHOTOOXIDATIVE	1	2	0	2
POTASSIUM	1	2	0	0
REACTIVE OXYGEN SPECIES	1	1	0	0
SALINITY RESPONSE	2	11	0	5
SALT	85	350	4	41
SODIUM	1	12	0	3
STRESS	29	106	0	8
TEMPERATURE	1	9	0	0
TOUCH	0	0	0	0
UV	1	2	0	0
WATER	5	13	0	1
ZINC	8	22	1	0

Supplementary Ch4.5: Transcription factors responsive to water stress in *A. donax* and *P. australis*. (Online)

Supplementary Ch4.6: Common gene expression patterns in response to different levels of between *P. australis* and *A. donax* shoot. The expression patterns for DEGs have been classified in terms of regulation trends. Up – and down-regulated genes were identified by having $|logFC| \geq 2$. (mild water stress vs control and severe water stress vs mild water stress). Expression levels are presented by Fragments Per Kilobase of exon per Million fragments mapped (FPKM). (C: control; M: mild water stress; S: severe water stress; FPKM:

Fragments per kilobase of exon per million fragments mapped; CU: Constant, up-regulated; CD: Constant, down-regulated; UU: Up-regulated, up-regulated; UC: Up-regulated, constant; UD: Up-regulated, down-regulated; DC: Down-regulated, constant; DU: Down-regulated, up-regulated, CCu: up regulated and only regulated by severe water stress, CCd: down-regulated and only regulated by severe water stress.). Red: *A. donax*; green: *P. australis*. Red color in the table means there is significant differences between two species, with t-test ($p<0.05$) corrected by Bonferroni method.





	C	M	S	Number of Genes
CU	0,01382	0,5835	0,001565	96
CCu	5,04E-08	2,06E-14	4,58E-10	110

Supplementary Ch4.7: Common drought response genes across *P. australis*, A. donax, rice, foxtail and sorghum. The putative homologs of *P. australis* stress-responsive genes identified by OrthoMCL. (Online)

Online depository:

https://drive.google.com/open?id=0Bwu-DUfLtRghcFRpd2hGZ092NVU