Plant Biotechnology Journal

Plant Biotechnology Journal (2014) **12**, pp. 554–567 doi: 10.1111/pbi.12159

Fuelling genetic and metabolic exploration of C_3 bioenergy crops through the first reference transcriptome of *Arundo donax* L.

Gaurav Sablok[1], Yuan Fu[1,2], Valentina Bobbio[3,4], Marina Laura[4], Giuseppe L. Rotino[5], Paolo Bagnaresi[6], Andrea Allavena[4], Violeta Velikova[1,7], Roberto Viola[1], Francesco Loreto[8], Mingai Li[1] and Claudio Varotto[1,*]

[1]*Department of Biodiversity and Molecular Ecology, Research and Innovation Centre, Fondazione Edmund Mach, S. Michele all'Adige, TN, Italy*
[2]*Dipartimento di Biotecnologie, Università degli Studi di Verona, Verona, Italy*
[3]*Dipartimento di Scienze della Terra, dell'Ambiente e della Vita, Università degli Studi di Genova, Genova, Italy*
[4]*Unità di Ricerca per la Floricoltura e le Specie Ornamentali, Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Sanremo, IM, Italy*
[5]*Unità di Ricerca per l'Orticoltura, Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Montanaso Lombardo, LO, Italy*
[6]*Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Genomics Research Centre, Fiorenzuola D'Arda, PC, Italy*
[7]*Bulgarian Academy of Sciences, Institute of Plant Physiology and Genetics, Sofia, Bulgaria*
[8]*Dipartimento di Scienze Bio-Agroalimentari (DISBA), Consiglio Nazionale delle Ricerche (CNR), Roma, Italy*

Received 9 August 2013;
revised 13 November 2013;
accepted 11 December 2013.
**Correspondence* (fax +39 0461 650 956;
email claudio.varotto@fmach.it)

Keywords: *Arundo donax*, cell wall, domestication genes, lipid biosynthesis, lignin and starch biosynthesis, RNA-Seq.

Summary

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 nonfood 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 nonmodel 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 (http://ecogenomics.fmach.it/arundo/) for mining and exploring the genetic potential of this species. Based on functional annotation and homology comparison to 19 prospective biofuel Poaceae 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 functional genomics or genetic improvement 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*.

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. (C_4), *Phalaris arundinacea* (C_4), *Arundo donax* (C_3) and *Panicum virgatum* (C_4), 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* is 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 C_4 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/).

Results and discussion

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 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 pseudo-triploid 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 (Table S1). 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

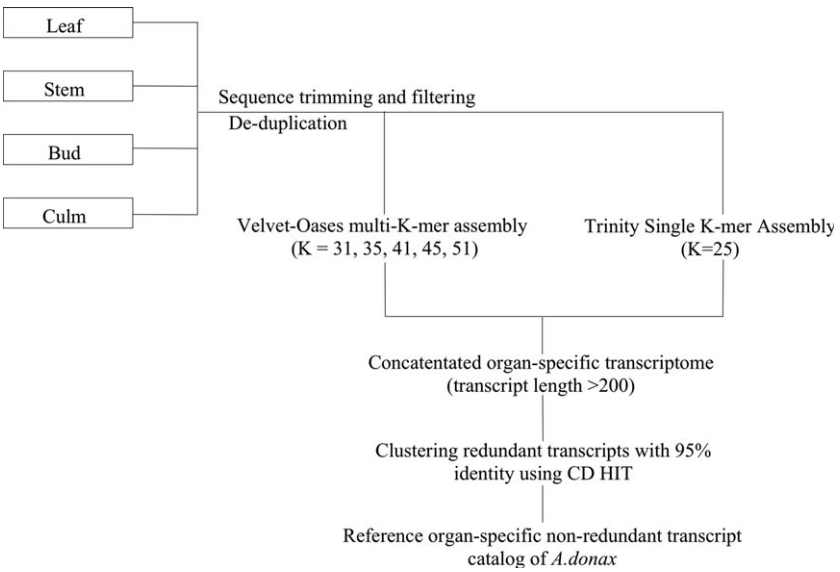

Figure 1 Overview of transcript assembly strategy implemented for the organ-specific transcriptome of *A. donax*.

strategy was able to assemble and preserve the longer transcripts (Zhao *et al.*, 2011).

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 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).

Table 1 Percentage distribution of the reads mapped back to the transcriptome (RMBT) in organ transcriptome of *A. donax*.

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

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 the improvement of high-value traits. To assess sequence conservation of the *A. donax* transcript assembly across phylogenetically related biomass and biofuel species, BLASTx homology searches were run against available sequences of Poaceae species from the BFGR database (http://bfgr.plantbiology.msu.edu/). The highest number of homologs (Figure 2) was retrieved in *Saccharum officinarum* (285 296), followed by *Triticum aestivum* (239 411 transcripts), *P. virgatum* (192 248) and *Hordeum vulgare* (110 954). *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 transcriptomic and phenotypic similarity among the two species, we suggest that sugarcane domestication and yield-related traits could provide a good model for functional studies and improvement of *A. donax*. 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 homologs in this C_4 species (Swaminathan *et al.*, 2012). A total of 24 180 (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. To identify coding sequence conservation, CDS were systematically downloaded for *O. sativa* (39 049), *Sorghum bicolor* (27 607) and *Z. mays* (39 656) from Phytozome (http://www.phytozome.net/). Reciprocal best blast hit (RBH; >70% sequence coverage) searches revealed high sequence conservation across *S. bicolor* (20 779 CDS, corresponding to 75.27% of the total), *Z. mays* (25 618, 64.60%) and *O. sativa* (18 839, 48.24%). 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*.

Functional annotation, relative expression and GO enrichment

Functional annotation of the assembled transcripts provides an important way to characterize functionally relevant metabolic pathways and to identify the genes involved in traits of economic and ecological importance. Previous studies have highlighted the effectiveness of this approach to pinpoint the functional relevance of a large number of genes in specific plant traits (e.g. Góngora-Castillo *et al.*, 2012; Gu *et al.*, 2012; Suzuki *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

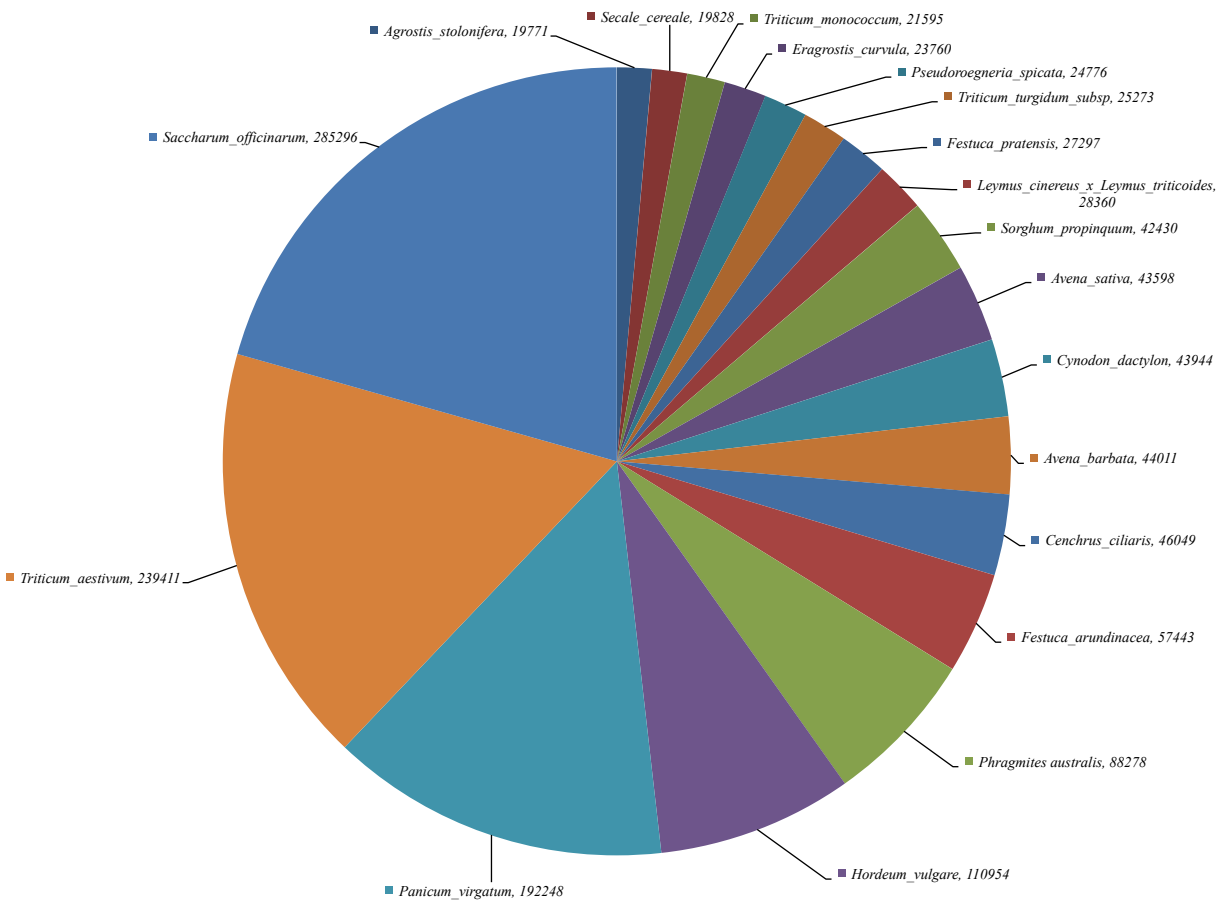

Figure 2 Distribution of transcript homologs across biofuel and other Poaceae species.

searches identified homologs of about 40%–45% of *A. donax* transcripts in SwissProt and GenBank's NR protein databases. Most of the functional hits correspond to *S. bicolor* and *Z. 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 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) and 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, 117 915 transcripts in root, 105 447 in leaf, 127 790 in culm and 118 107 in bud were found to be associated to at least one 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 noncoding RNAs, in line with previous observations in polyploid 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

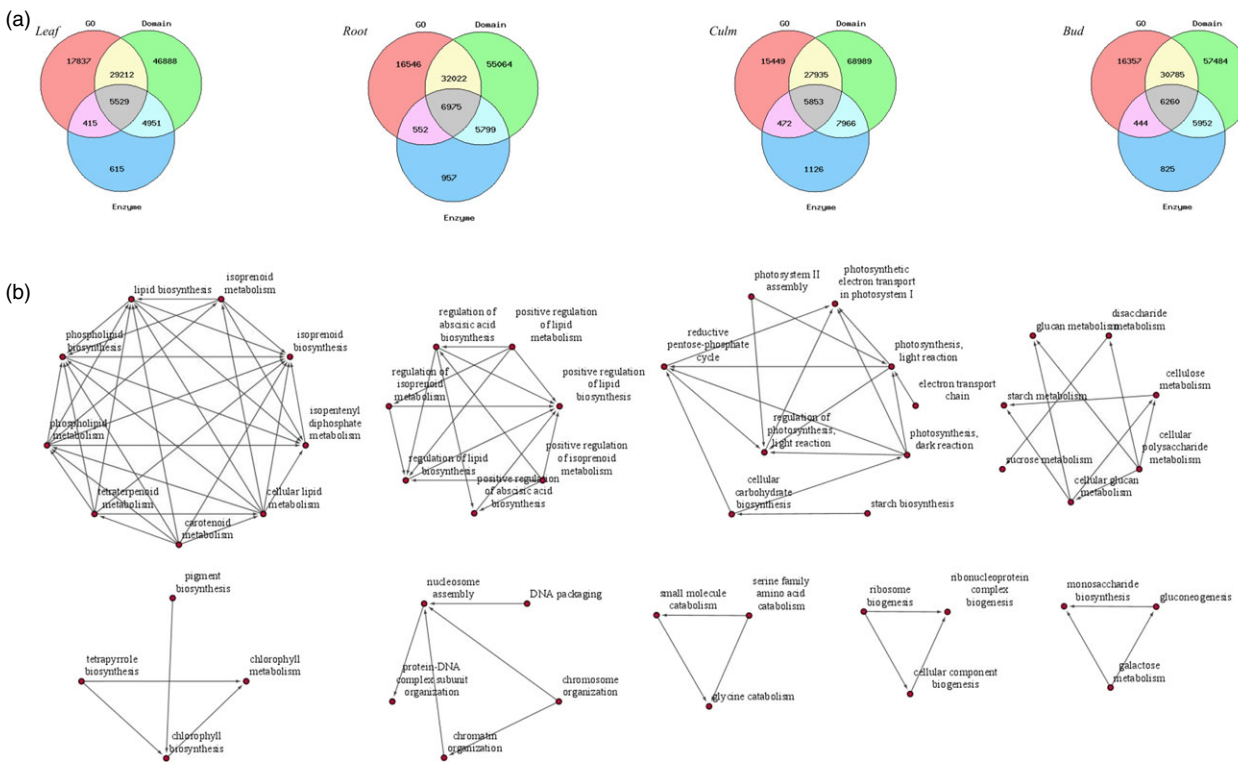

Figure 3 Gene Ontology classification of *Arundo donax* transcripts: (a) functionally annotated GO categories, enzymes and domains in each organ-specific transcriptome. (b) Attribute circle layout representation of the merged over-represented GO categories across the four organs based on the Fischer test and Benjamini and Hochberg false discovery rate correction (0.05%).

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 (Table S2). 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 (Table S3). 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 3b; Table S3). 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) (Table S4). 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 4a,b. 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

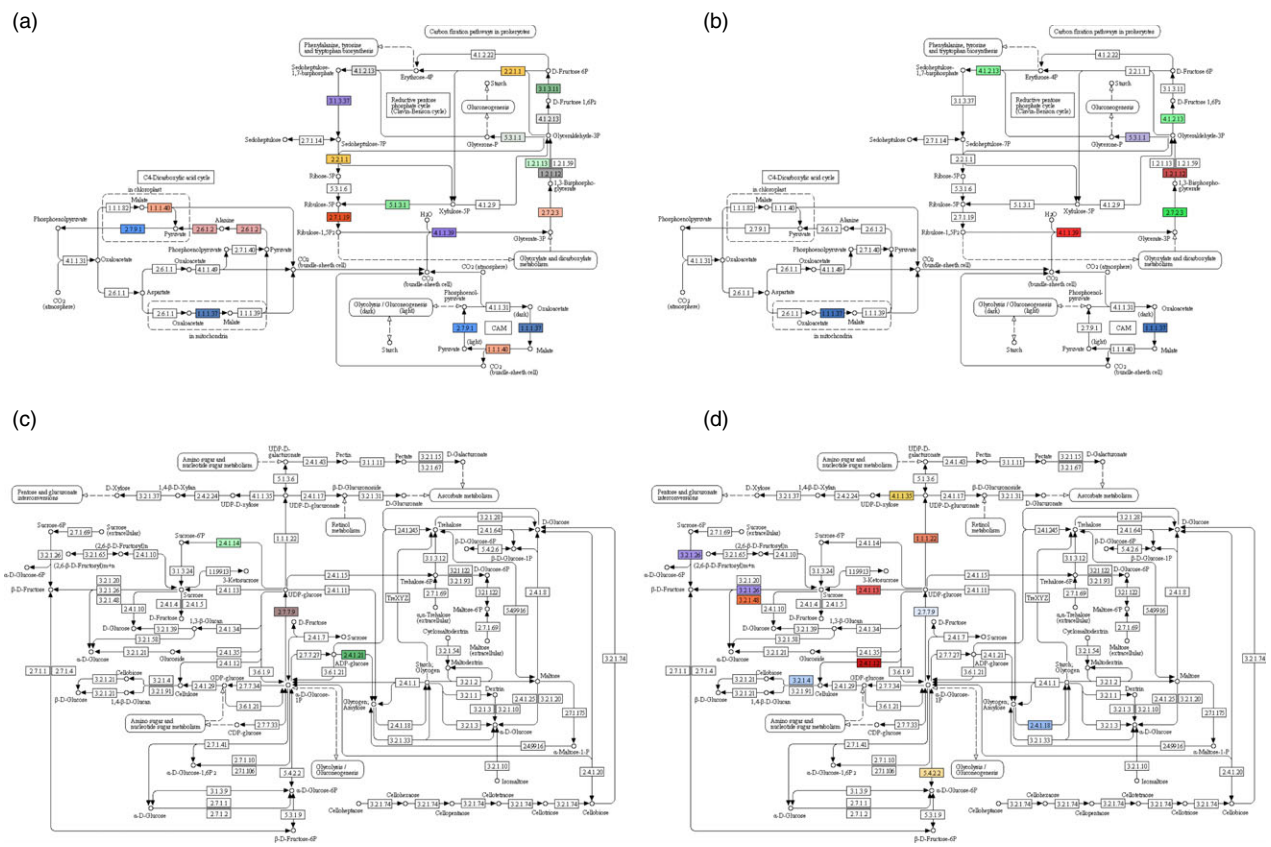

Figure 4 Selected KEGG pathways illustrating transcripts involved in photosynthetic carbon fixation and starch/sucrose metabolism which are highly expressed in leaf (a), (c) and culm (b), (d), respectively.

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.

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).

Transcriptome assisted metabolic mining and identification of candidate genes 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 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 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).

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

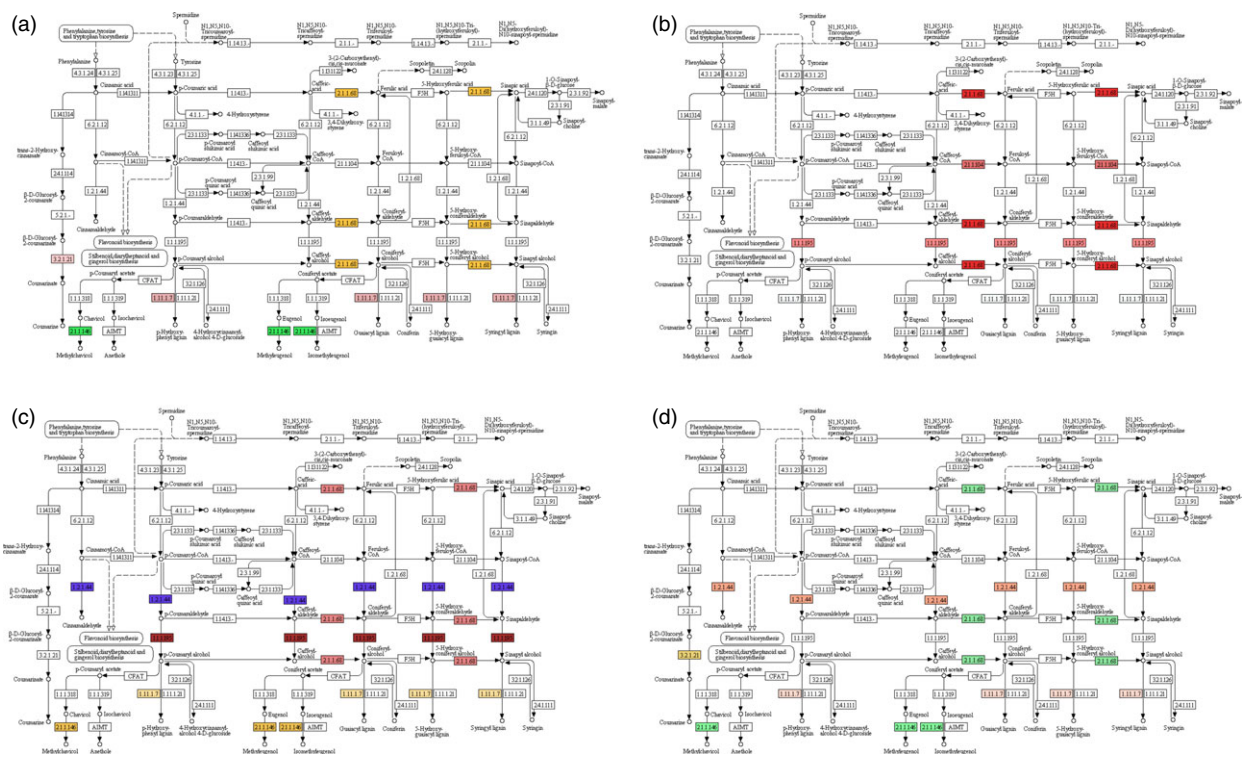

Figure 5 Comparative assessment of the abundance of the highly expressed genes in lignin metabolism across the bud (a), culm (b), leaf (c) and root (d).

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 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.

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). 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.*,

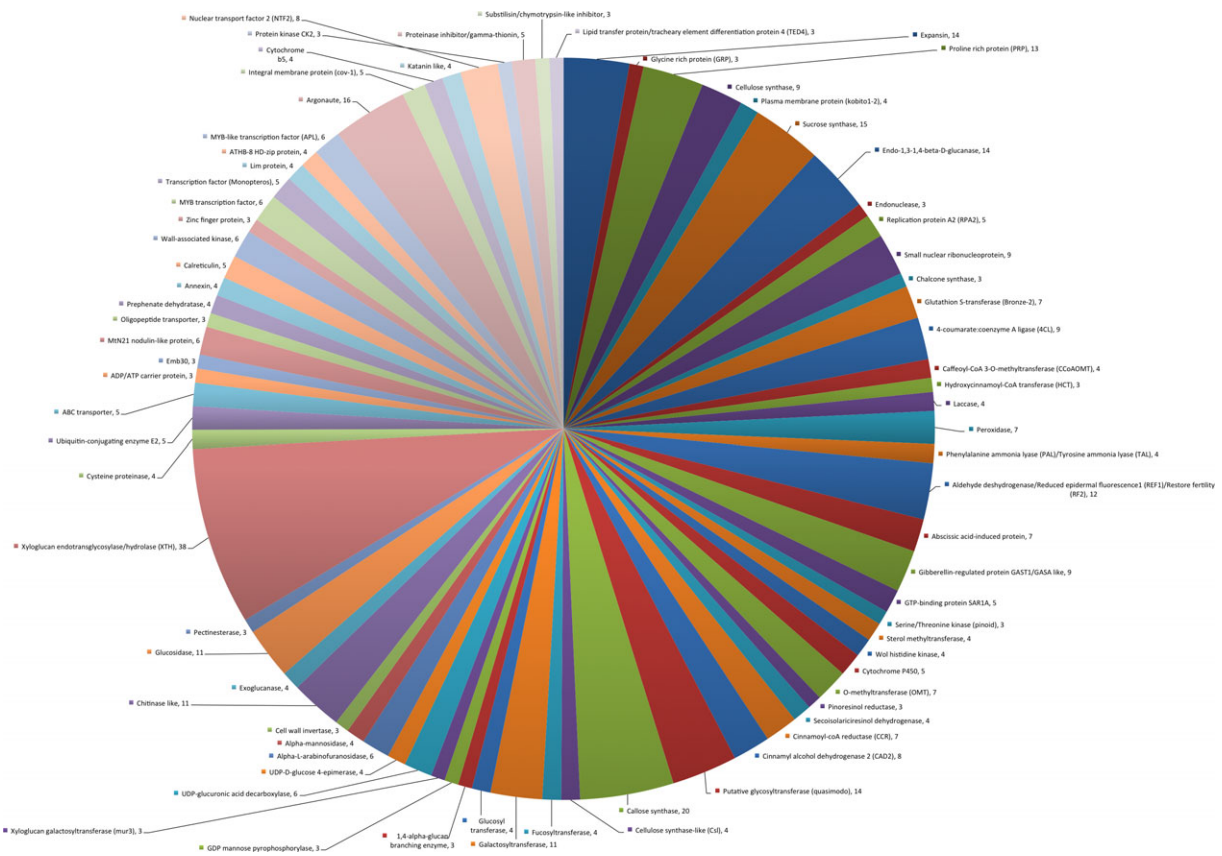

Figure 6 Overview of distribution of *A. donax* transcript homologs associated with cell wall biogenesis and organization.

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*.

Table 2 Distribution of transcript abundance in the lipid biosynthetic pathway available from http://aralip.plantbiology.msu.edu/pathways/pathways

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, galactolipid and sulfolipid synthesis	28	232
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

*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 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 *early heading date 1* (*EHD1*) 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). Homologous transcripts of *teosinte branched 1* (*TB1*), the most famous example of domestication gene for maize, were found in culm (1) and bud (2) transcriptome. The *tb1* gene could provide a powerful tool to explore the effect of different plant architectures on biomass yield, while genes such as *reduced height* (*RHT*) and *semidwarf 1* (*SD1*) could be used to modulate plant height optimizing it for maximal productivity.

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 (Table S5). 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

Table 3 Identification of *A. donax* transcript homologs to crop domestication genes

Gene name	Molecular function	Accession no.	Species	Number of transcripts	Ref.
AE1	Carbohydrate composition. Biosynthetic enzyme	NP_001105316.1	*Zea mays*	25	Doebley *et al.* (2006)
BT2	Carbohydrate composition. Biosynthetic enzyme	AAN39328.1	*Zea mays*	30	Doebley *et al.* (2006)
CG1	Flowering time and carbohydrate composition. microRNA	EF541486.1	*Zea mays*	1	Chuck *et al.* (2011)
EHD1	Flowering time. Transcription factor	BAC77078.1	*Oryza sativa*	21	Doebley *et al.* (2006)
HD1	Flowering time. Transcription factor	BAB17628.1	*Oryza sativa*	4	Doebley *et al.* (2006)
RHT	Plant height. Transcription factor	AGG68565.1	*Triticum aestivum*	10	Doebley *et al.* (2006)
SD1	Plant height. Biosynthetic enzyme	NP_001045014.1	*Oryza sativa*	17	Doebley *et al.* (2006)
SH2	Carbohydrate composition. Biosynthetic enzyme	AFP90368.1	*Zea mays*	13	Doebley *et al.* (2006)
SU1	Carbohydrate composition. Biosynthetic enzyme	AAB97167.1	*Zea mays*	5	Doebley *et al.* (2006)
TAG1	Lipid composition. Biosynthetic enzyme	AT2G19450.1	*Arabidopsis thaliana*	21	Winichayakul *et al.* (2013)
TB1	Plant and inflorescence structure. Transcription factor	AFI70999.1	*Zea mays*	3	Doebley *et al.* (2006)
VRN1	Vernalization. Transcription factor	AAZ76883.1	*Triticum aestivum*	14	Doebley *et al.* (2006)
VRN2	Vernalization. Transcription factor	AAS60240.1	*Triticum aestivum*	1	Doebley *et al.* (2006)
WAXY	Carbohydrate composition. Biosynthetic enzyme	ABC17717.1	*Oryza sativa*	17	Doebley *et al.* (2006)

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 (Table S5), 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 (Table S5). 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 (Table S5). 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*.

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.

Experimental procedures

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).

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.

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.

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).

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, 1×10^{-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.

Acknowledgements

We thank E. Barbaro for sample collection and RNA preparation. This work was supported by the Autonomous Province of Trento (Italy) through core funding of the Ecogenomics group (GS, ML and CV) and the MAN-VIP project (Team 2011 Call approved with provincial government resolution no. 2902 on 14 December 2010; FY and VV); the Ministero per le Politiche Agricole, Alimentari e Forestali (Italy) in the frame of the project CANADAIR (VB, ML, GLR, LC and AA). VB thanks the Università di Genova (DISTAV; Italy) for funding her Ph.D. fellowship. We thank two anonymous reviewers for constructive comments, which helped improving of the manuscript.

References

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

Angelini, L.G., Ceccarini, L., Nassi o Di Nasso, N. and Bonari, E. (2009) 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 Bioenergy*, **33**, 635–643.

Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29.

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

Barakat, A., DiLoreto, D.S., Zhang, Y., Smith, C., Baier, K., Powell, W.A., Wheeler, N., Sederoff, R. and Carlson, J.E. (2009) 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.

Baron, A.C., Tsbin, T.H., Wallsgrove, R.M. and Tobin, A.K. (1994) A metabolic control analysis of the glutamine synthetase/glutamate synthase cycle in isolated barley (*Hordeum vulgare* L.) chloroplasts. *Plant Physiol.* **105**, 415–424.

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

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

Bombarely, A., Edwards, K.D., Sanchez-Tamburrino, J. and Mueller, L.A. (2012) Deciphering the complex leaf transcriptome of the allotetraploid species *Nicotiana tabacum*: a phylogenomic perspective. *BMC Genomics*, **13**, 406.

Borkotoky, S., Saravanan, V., Jaiswal, A., Das, B., Selvaraj, S., Murali, A. and Lakshmi, P.T. (2013) The *Arabidopsis* stress responsive gene database. *Int. J. Plant Genomics*, **2013**, 949564.

Bouché, N. and Bouchez, D. (2001) Arabidopsis gene knockout: phenotypes wanted. *Curr. Opin. Plant Biol.* **4**, 111–117.

Bout, S. and Vermerris, W. (2003) A candidate-gene approach to clone the sorghum *Brown midrib* gene encoding caffeic acid O-methyltransferase. *Mol. Genet. Genomics*, **269**, 205–214.

Bouvier d'Yvoire, M., Bouchabke-Coussa, O., Voorend, W., Antelme, S., Cézard, L., Legée, F., Lebris, P., Legay, S., Whitehead, C., McQueen-Mason, S.J., Gomez, L.D., Jouanin, L., Lapierre, C. and Sibout, R. (2013) Disrupting the *cinnamyl alcohol dehydrogenase 1* gene (*BdCAD1*) leads to altered lignification and improved saccharification in *Brachypodium distachyon*. *Plant J.* **73**, 496–508.

Calheiros, C.S.C., Quitério, P.V.B., Silva, G., Crispim, L.F.C., Brix, H., Moura, S.C. and Castro, P.M.L. (2012) Use of constructed wetland systems with *Arundo* and *Sarcocornia* for polishing high salinity tannery wastewater. *J. Environ. Manage.* **95**, 66–71.

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

Carpita, N.C. (2012) Progress in the biological synthesis of the plant cell wall: new ideas for improving biomass for bioenergy. *Curr. Opin. Biotechnol.* **23**, 330–337.

Castleden, C.K., Aoki, N., Gillespie, V.J., Macrae, E.A., Quick, W.P., Buchner, P., Foyer, C.H., Furbank, R.T. and Lunn, J.E. (2004) Evolution and function of the sucrose-phosphate synthase gene families in wheat and other grasses. *Plant Physiol.* **135**, 1753–1764.

Ceotto, E. and Di Candilo, M. (2010) Shoot cuttings propagation of giant reed (*Arundo donax* L.) in water and moist soil: the path forward? *Biomass Bioenergy*, **34**, 1614–1623.

Ceotto, E., Di Candilo, M., Castelli, F., Badeck, F.-W., Rizza, F., Soave, C., Volta, A., Villani, G. and Marletto, V. (2013) Comparing solar radiation interception and use efficiency for the energy crops giant reed (*Arundo donax* L.) and sweet sorghum (*Sorghum bicolor* L. Moench). *Field Crop Res.* **149**, 159–166.

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

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

Chen, T.-W., Gan, R.-C.R., Wu, T.H., Huang, P.-J., Lee, C.-Y., Chen, Y.-Y.M., Chen, C.-C. and Tang, P. (2012) FastAnnotator–an efficient transcript annotation web tool. *BMC Genomics*, **13**, S9.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A.J., Müller, W.E.G., Wetter, T. and Suhai, S. (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Res.* **14**, 1147–1159.

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

Chuck, G.S., Tobias, C., Sun, L., Kraemer, F., Li, C., Dibble, D., Arora, R., Bragg, J.N., Vogel, J.P., Singh, S., Simmons, B.A., Pauly, M. and Hake, S. (2011) Overexpression of the maize *Corngrass1* microRNA prevents flowering, improves digestibility and increases starch content of switchgrass. *Proc. Natl Acad. Sci. USA*, **108**, 17550–17555.

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

Coleman, H.D., Yan, J. and Mansfield, S.D. (2009) Sucrose synthase affects carbon partitioning to increase cellulose production and altered cell wall ultrastructure. *Proc. Natl Acad. Sci. USA*, **106**, 13118–13123.

Cotsaftis, O., Plett, D., Shirley, N., Tester, M. and Hrmova, M. (2012) A two-staged model of Na+ exclusion in rice explained by 3D modeling of HKT transporters and alternative splicing. *PLoS ONE*, **7**, e39865.

Curtin, S.J., Zhang, F., Sander, J.D., Haun, W.J., Starker, C., Baltes, N.J., Reyon, D., Dahlborg, E.J., Goodwin, M.J., Coffman, A.P., Dobbs, D., Joung, J.K., Voytas, D.F. and Stupar, R.M. (2011) Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. *Plant Physiol.* **156**, 466–473.

Davenport, R., James, R.A., Zakrisson-Plogander, A., Tester, M. and Munns, R. (2005) Control of sodium transport in durum wheat. *Plant Physiol.* **137**, 807–818.

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

Doebley, J.F., Gaut, B.S. and Smith, B.D. (2006) The molecular genetics of crop domestication. *Cell*, **127**, 1309–1321.

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

Durrett, T.P., Gassmann, W. and Rogers, E.E. (2007) The FRD3-mediated efflux of citrate into the root vasculature is necessary for efficient iron translocation. *Plant Physiol.* **144**, 197–205.

Faino, L., de Jonge, R. and Thomma, B.P.H.J. (2012) The transcriptome of *Verticillium dahliae*-infected *Nicotiana benthamiana* determined by deep RNA sequencing. *Plant Signal. Behav.* **7**, 1065–1069.

Foyer, C.H., Neukermans, J., Queval, G., Noctor, G. and Harbinson, J. (2012) Photosynthetic control of electron transport and the regulation of gene expression. *J. Exp. Bot.* **63**, 1637–1661.

Fu, L., Niu, B., Zhu, Z., Wu, S. and Li, W. (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**, 3150–3152.

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

Gendre, D., Czernic, P., Conéjéro, G., Pianelli, K., Briat, J.-F., Lebrun, M. and Mari, S. (2007) TcYSL3, a member of the YSL gene family from the hyper-accumulator *Thlaspi caerulescens*, encodes a nicotianamine-Ni/Fe transporter. *Plant J.* **49**, 1–15.

Góngora-Castillo, E., Childs, K.L., Fedewa, G., Hamilton, J.P., Liscombe, D.K., Magallanes-Lundback, M., Mandadi, K.K., Nims, E., Runguphan, W., Vaillancourt, B., Varbanova-Herde, M., DellaPenna, D., McKnight, T.D., O'Connor, S. and Buell, C.R. (2012) Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS ONE*, **7**, e52506.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J. and Conesa, A. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435.

Grabherr, M., Haas, B., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. and Regev, A. (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **29**, 644–652.

Gu, K., Yi, C., Tian, D., Sangha, J.S., Hong, Y. and Yin, Z. (2012) Expression of fatty acid and lipid biosynthetic genes in developing endosperm of *Jatropha curcas*. *Biotechnol. Biofuels*, **5**, 47.

Guillaumie, S., San-Clemente, H., Deswarte, C., Martinez, Y., Lapierre, C., Murigneux, A., Barrière, Y., Pichon, M. and Goffner, D. (2007) MAIZEWALL. Database and developmental gene expression profiling of cell wall biosynthesis and assembly in maize. *Plant Physiol.* **143**, 339–363.

Hardion, L., Verlaque, R., Fridlender, A. and Vila, B. (2011) Arundo. In *IAPT/IOPB Chromosome Data 11* (Marhold, K., ed.), *Taxon*, **60**, 1221–1223.

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

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

Harris, D.M., Corbin, K., Wang, T., Gutierrez, R., Bertolo, A.L., Petti, C., Smilgies, D.-M., Estevez, J.M., Bonetta, D., Urbanowicz, B.R., Ehrhardt, D.W., Somerville, C.R., Rose, J.K.C., Hong, M. and Debolt, S. (2012) Cellulose microfibril crystallinity is reduced by mutating C-terminal transmembrane region residues CESA1A903V and CESA3T942I of cellulose synthase. *Proc. Natl Acad. Sci. USA*, **109**, 4098–4103.

Hassinen, V.H., Tervahauta, A.I., Schat, H. and Kärenlampi, S.O. (2011) Plant metallothioneins–metal chelators with ROS scavenging activity? *Plant Biol (Stuttg)*, **13**, 225–232.

He, R., Kim, M.-J., Nelson, W., Balbuena, T.S., Kim, R., Kramer, R., Crow, J.A., May, G.D., Thelen, J.J., Soderlund, C.A. and Gang, D.R. (2012) 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**, 232–247.

Horie, T., Hauser, F. and Schroeder, J.I. (2009) HKT transporter-mediated salinity resistance mechanisms in Arabidopsis and monocot crop plants. *Trends Plant Sci.* **14**, 660–668.

Huang, X. and Madan, A. (1999) CAP3: a DNA sequence assembly program. *Genome Res.* **9**, 868–877.

Ingle, R.A., Mugford, S.T., Rees, J.D., Campbell, M.M. and Smith, J.A.C. (2005) Constitutively high expression of the histidine biosynthetic pathway contributes to nickel tolerance in hyperaccumulator plants. *Plant Cell*, **17**, 2089–2106.

John, K.V., Schutzbach, J.S. and Ankel, H. (1977) Separation and allosteric properties of two forms of UDP-glucuronate carboxy-lyase. *J. Biol. Chem.* **252**, 8013–8017.

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

Khraiwesh, B., Zhu, J.-K. and Zhu, J. (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim. Biophys. Acta*, **1819**, 137–148.

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

Le, H.-S., Schulz, M.H., McCauley, B.M., Hinman, V.F. and Bar-Joseph, Z. (2013) Probabilistic error correction for RNA sequencing. *Nucleic Acids Res.* **41**, e109.

Lee, C., Teng, Q., Huang, W., Zhong, R. and Ye, Z.-H. (2009) Down-regulation of PoGT47C expression in poplar results in a reduced glucuronoxylan content and an increased wood digestibility by cellulase. *Plant Cell Physiol.* **50**, 1075–1089.

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

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

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

Linder, H.P., Verboom, G.A. and Barker, N.P. (1997) Phylogeny and evolution in the Crinipes group of grasses (Arundinoideae: Poaceae). *Kew. Bull.* **52**, 91–110.

Lindgren, L.O., Stålberg, K.G. and Ho, A. (2003) 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.

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

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

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

Massa, A.N., Childs, K.L., Lin, H., Bryan, G.J., Giuliano, G. and Buell, C.R. (2011) The transcriptome of the reference potato genome *Solanum tuberosum* Group Phureja clone DM1-3 516R44. *PLoS ONE*, **6**, e26801.

Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F. and Marshall, D. (2010) Tablet–next generation sequence assembly visualization. *Bioinformatics*, **26**, 401–402.

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

Morozova, O. and Marra, M.A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics*, **92**, 255–264.

Nakasugi, K., Crowhurst, R.N., Bally, J., Wood, C.C., Hellens, R.P. and Waterhouse, P.M. (2013) *De novo* transcriptome sequence assembly and analysis of RNA silencing genes of *Nicotiana benthamiana*. *PLoS ONE*, **8**, e59534.

Neto, C.P., Seca, A., Nunes, A.M., Coimbra, M.A., Domingues, F., Evtuguin, D., Silvestre, A. and Cavaleiro, J.A.S. (1997) 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.

Ohlrogge, J., Allen, D., Berguson, B., DellaPenna, D., Shachar-Hill, Y. and Stymne, S. (2009) Driving on biomass. *Science*, **324**, 1019–1020.

Okawa, S., Makino, A. and Mae, T. (2002) 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.

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

Pauly, M. and Keegstra, K. (2010) Plant cell wall polymers as precursors for biofuels. *Curr. Opin. Plant Biol.* **13**, 305–312.

Perdue, R.E. (1958) *Arundo donax*—Source of musical reeds and industrial cellulose. *Eco. Bot.* **12**, 368–404.

Pevzner, P.A., Tang, H. and Waterman, M.S. (2001) An Eulerian path approach to DNA fragment assembly. *Proc. Natl Acad. Sci. USA*, **98**, 9748–9753.

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

Pires, I.S., Negrão, S., Pentony, M.M., Abreu, I.A., Oliveira, M.M. and Purugganan, M.D. (2013) Different evolutionary histories of two cation/proton exchanger gene families in plants. *BMC Plant Biol.* **13**, 97.

Platten, J.D., Cotsaftis, O., Berthomieu, P., Bohnert, H., Davenport, R.J., Fairbairn, D.J., Horie, T., Leigh, R.A., Lin, H.X., Luan, S., Mäser, P., Pantoja, O., Rodríguez-Navarro, A., Schachtman, D.P., Schroeder, J.I., Sentenac, H., Uozumi, N., Véry, A.A., Zhu, J.K., Dennis, E.S. and Tester, M. (2006) Nomenclature for HKT transporters, key determinants of plant salinity tolerance. *Trends Plant Sci.* **11**, 372–374.

Polunin, O. and Huxley, A.J. (1965) ARUNDO. In *Flowers of the Mediterranean*, pp. 196, 199. London: Chatto and Windus.

Raspolli Galletti, A.M., Antonetti, C., Ribechini, E., Colombini, M.P., Nassi o Di Nasso, N. and Bonari, E. (2013) From giant reed to levulinic acid and gamma-valerolactone: a high yield catalytic route to valeric biofuels. *Appl. Energy*, **102**, 157–162.

Saha, B.C. (2003) Hemicellulose bioconversion. *J. Ind. Microbiol. Biotechnol.* **30**, 279–291.

Sattler, S.E., Funnell-Harris, D.L. and Pedersen, J.F. (2010) Brown midrib mutations and their importance to the utilization of maize, sorghum and pearl millet lignocellulosic tissues. *Plant Sci.* **178**, 229–238.

Schliesky, S., Gowik, U., Weber, A.P.M. and Bräutigam, A. (2012) RNA-Seq assembly—are we there yet? *Front. Plant Sci.* **3**, 220.

Schulz, M.H., Zerbino, D.R., Vingron, M. and Birney, E. (2012) Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, **28**, 1086–1092.

Sinclair, S.A. and Krämer, U. (2012) The zinc homeostasis network of land plants. *Biochim. Biophys. Acta*, **1823**, 1553–1567.

Slade, A.J., Fuerstenberg, S.I., Loeffler, D., Steine, M.N. and Facciotti, D. (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. *Nat. Biotechnol.* **23**, 75–81.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L. and Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.

Sobhanian, H., Aghaei, K. and Komatsu, S. (2011) Changes in the plant proteome resulting from salt stress: toward the creation of salt-tolerant crops? *J. Proteomics*, **74**, 1323–1337.

Surget-Groba, Y. and Montoya-Burgos, J.I. (2010) Optimization of *de novo* transcriptome assembly from next-generation sequencing data. *Genome Res.* **20**, 1432–1440.

Suzuki, N., Uefuji, H., Nishikawa, T., Mukai, Y., Yamashita, A., Hattori, M., Ogasawara, N., Bamba, T., Fukusaki, E., Kobayashi, A., Ogata, Y., Sakurai, N., Suzuki, H., Shibata, D. and Nakazawa, Y. (2012) Construction and analysis of EST libraries of the *trans*-polyisoprene producing plant, *Eucommia ulmoides* Oliver. *Planta*, **236**, 1405–1417.

Swaminathan, K., Chae, W.B., Mitros, T., Varala, K., Xie, L., Barling, A., Glowacka, K., Hall, M., Jezowski, S., Ming, R., Hudson, M., Juvik, J.A., Rokhsar, D.S. and Moose, S.P. (2012) A framework genetic map for *Miscanthus sinensis* from RNA seq-based markers shows recent tetraploidy. *BMC Genomics*, **13**, 142.

Takahashi, W., Takamizo, T., Kobayashi, M. and Ebina, M. (2010) Plant regeneration from calli in giant reed (*Arundo donax* L.). *Grassland Sci.* **56**, 224–229.

Tammisola, J. (2010) Towards much more efficient biofuel crops—can sugarcane pave the way? *GM Crops*, **1**, 181–198.

Taylor, S.H., Hulme, S.P., Rees, M., Ripley, B.S., Woodward, F.I. and Osborne, C.P. (2010) Ecophysiological traits in C3 and C4 grasses: a phylogenetically controlled screening experiment. *New Phytol.* **185**, 780–791.

Thorbjornsen, T., Villand, P., Denyer, K., Olsen, O.-A. and Smith, A.M. (1996) Distinct isoforms of ADPglucose pyrophosphorylase occur inside and outside the amyloplasts in barley endosperm. *Plant J.* **10**, 243–250.

Trabucco, G.M., Matos, D.A., Lee, S.J., Saathoff, A.J., Priest, H.D., Mockler, T.C., Sarath, G. and Hazen, S.P. (2013) Functional characterization of cinnamyl alcohol dehydrogenase and caffeic acid O-methyltransferase in *Brachypodium distachyon*. *BMC Biotechnol.* **13**, 61.

Wang, J., Nayak, S., Koch, K. and Ming, R. (2013) Carbon partitioning in sugarcane (*Saccharum* species). *Front. Plant Sci.* **4**, 201.

Weckwerth, W. (2011) Green systems biology—from single genomes, proteomes and metabolomes to ecosystems research and biotechnology. *J. Proteomics*, **75**, 284–305.

Winichayakul, S., Scott, R.W., Roldan, M., Hatier, J.H., Livingston, S., Cookson, R., Curran, A.C. and Roberts, N.J. (2013) *In vivo* packaging of triacylglycerols enhances *Arabidopsis* leaf biomass and energy density. *Plant Physiol.* **162**, 626–639.

Yang, C.-Q., Fang, X., Wu, X.-M., Mao, Y.-B., Wang, L.-J. and Chen, X.-Y. (2012) Transcriptional regulation of plant secondary metabolism. *J. Integr. Plant Biol.* **54**, 703–712.

You, T.-T., Mao, J.-Z., Yuan, T.-Q., Wen, J.-L. and Xu, F. (2013) Structural elucidation of the lignins from stems and foliage of *Arundo donax* Linn. *J. Agric. Food Chem.* **61**, 5361–5370.

Zerbino, D.R. and Birney, E. (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829.

Zhao, Q.-Y., Wang, Y., Kong, Y.-M., Luo, D., Li, X. and Hao, P. (2011) Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics*, **12**, S2.

Supporting information

Additional Supporting information may be found in the online version of this article:

Table S1 Summary statistics of transcript assembly at k-mer from 31 to 51 in Velvet–Oases and single k-mer in Trinity (minimum transcript length = 200 bp).

Table S2 Distribution of the FPKM values as a function of the assembled transcripts across the four organs.

Table S3 Functional annotation and FPKM values for the highly expressed genes.

Table S4 Classification of over- and under-represented GO categories across the four organs.

Table S5 Mining of the *A. donax* transcriptome for putative transcript homologs of stress-related genes.