Plant Omics Journal POJ

POJ 8(5):441-448 (2015) ISSN:1836-3644

Analysis of ESTs from the date palm (*Phoenix dactylifera* L.) cv. Sukary, an elite variety

Sulieman A. Alfaifi[1], Muhammad A. Khan[1], Hussein M. Migdadi[1], Jernej Jakse[2], Megahed H. Ammar[1], Ehab H. El-Harty [1], Mohammad I. Althamrah[1], Muhammad Afzal[1], Muhammad M. Javed[1] and Salem S. Alghamdi[1]

[1]Legume Research Group, Plant Production Department, Faculty of Food and Agricultural Sciences, King Saud University, P.O .Box 2460, Riyadh 11451, Saudi Arabia
[2]Agronomy Department, Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, SI-1000 Ljubljana, Slovenia

***Corresponding author: hmigdadi@ksu.edu.sa; h.migdadi@gmail.com**

Abstract

Expressed sequence tags (ESTs) were generated from a normalized cDNA library of the date palm cv. Sukary to uncover the transcriptomic profile of this well-known high-quality variety in Saudi Arabia. The RNA was isolated from leaves and developed fruits at different stages from a single Sukary female tree. Total RNA was pooled and the normalized cDNA were cloned into a pAL17.3 cloning vector. The positives clones were sequenced using BigDye® Terminator chemistry. A total of 6943 high quality ESTs were generated, out of which 6671 ESTs were submitted to Gene bank dbEST (LIBEST_028537). The generated ESTs were assembled into 6362 unigenes, consisting of 494 (14.4%) contigs and 5868 (84.53%) singletons. The functional annotation showed that the majority of the ESTs were associated with binding (44%), catalytic (40%), transporter (5%) and structural molecular activities (5%). The Blastx results showed that 73% of the unigenes have significant similarities to known plant genes and 27% were novel. The latter could be of particular interest in date palm genetic studies. Further analysis showed that some ESTs were categorized as stress/defense and fruit development related genes. These newly generated ESTs could significantly enhance date palm EST databases that are in the public domain and are available to scientists and researchers across the globe.

Key words: ESTs; date palm; Sukary; fruit development; stress/defense.
Abbreviations: ATP_adenosine triphosphate; BLAST_basic local alignment search tool; BP_biological process; cDNA_ complementary DNA; ERF_ethylene response factor; EST_ expressed sequence tag; GO_gene ontology; HSPs_ heat shock proteins; KEGG_ kyoto encyclopedia of genes and genomes; KSU_ king saud university; Mb_ mega base; MF_molecular function; ROS_ reactive oxygen species; TF_transcription factors.

Introduction

The date palm (*Phoenix dactylifera*, L.) is an important member of the Arecaceae family. It is a dioecious, perennial, and monocotyledonous woody fruit tree (Sghaier-Hammami et al., 2009). The date palm has been cultivated for more than 7000 years and is considered among the world's first cultivated fruit trees (Wrigley, 1995). The tree is widely cultivated in Middle East and North Africa, where it is considered as staple food. Every part of the plant, including vegetative and reproductive organs, is of some economic, nutritional, or medicinal importance. The ripened fruit (date) is edible and contains a variety of nutritionally important components, such as sugars, fiber, vitamins and trace amounts of fats and proteins (Al-Shahib and Marshall, 2003). Nearly 2,000 date varieties are grown around the globe, of which more than 340 cultivars are grown in Saudi Arabia (Al-Mssallem, 1996). These date palm varieties differ in color, shape, size, flavor, and ripening time. The date palm genome contains 36 chromosomes (2n=36), and its size is estimated to be between 550 Mbp (Malek, 2010) and 658 Mbp (Al-Dous et al., 2011). Compared with other fruit species, relatively little investment has been made in expanding date palm molecular genetics research. This led to serious constraint on an already under-developed infrastructure of crop genetic and genomic tools. Although some molecular markers were developed and used, including, Inter simple sequence repeats (ISSR), Amplified fragment length polymorphism (AFLP) and simple sequence repeats (SSR) markers, the overall molecular toolbox for the date palm is limited and not as efficient compared with other fruit crops. With the advent of efficient, high throughput and cost effective sequencing technologies, a significant improvement in our understanding of genomics and biology of different plants has already been achieved. Therefore, it is not surprising that initiatives for date palm sequencing have already been conducted. Using data that was generated from the Illumina GAII sequencing platform, the first draft of the date palm genome was published in 2011 (Al-Dous et al., 2011), while the second draft of the date palm genome was published in 2013 (Al-Mssallem, et al., 2013). The complete chloroplast genome sequence was released in 2010 (Yang et al., 2010), and later, the complete mitochondrial genome was uncovered in 2012 (Fang et al., 2012). A transcriptomic profile of the date palm for fruit development has also been reported (Yin et al., 2012). Others studies related to fruits include a comparative analysis of *Elaeis spp.* and *P. dactylifera* oil palms (Bourgis et al., 2011) and in depth transcriptomic sequencing to build *P. dactylifera* gene models based on data retrieved from different tissues and

several developmental stages (Zhang et al., 2011). Recently, the first genetic map of the date palm was reported (Mathew et al., 2014). The production and life span of different date palm cultivars are affected by both biotic and abiotic factors (Jain, et al., 2011). Among abiotic factors, salinity and drought severely affect date palm cultivation, especially in the Arabian Peninsula countries, including Saudi Arabia. Breeding and conventional approaches were used to improve production costs, but these are laborious due to long generation time and difficulty in gender discrimination, especially in the first five years of tree's lifetime. Functional genomics is essential for identifying the responsible genes, encoding transcripts and to understand the mechanisms behind the drought tolerance. The development of complementary DNA (cDNA) library for the date palm will facilitate genetic and breeding studies, gene discovery and allele mining and comparative genomics. This study focused on analyzing the potential linkage between the transcriptome of the elite date palm cv. Sukary and its high-quality, agronomically important traits, by identifying potential ESTs that encode enzymes and proteins, related to fruit development and abiotic/biotic stresses.

Results

ESTs sequencing and assembly

More than 8000 positive clones were isolated and sequenced by Sanger sequencing platform from a normalized Sukary cDNA library. After removal of the vector, poly-A, and contaminating microbial and short read sequences (less than 100 bp), 6943 high quality EST cleaned sequences were retrieved, with the longest read length of 965 bp. The ESTs were assembled into 6362 unique sequences (unigenes), consisting of 494 (7.05%) contigs and 5868 (84.53%) singletons. The average contig length was 675 bp, while the average singleton length was 477 bp. The average number of ESTs per contig was 2.2, with a maximum of 7 sequences. The redundancy of this library was 9% [(1−Number of Unigenes/Number of ESTs) ×100%], which was relatively low, as expected for a normalized library. The detailed length distributions of the ESTs, unigenes, contigs, and singletons are shown in Table 1. After removing ribosomal and mitochondrial-like sequences, EST sequences were submitted to dbEST at NCBI under the library name (LIBEST_028537).

Functional annotation and classification

Annotation of the unigenes was performed using BLAST2GO based on comprehensive information gathered from sequence similarities against NCBI non-redundant (nr) protein database, Inter ProScan results and plant-related GO terms. The 6362 unigene sequences were used in a blastx search against the non-redundant protein sequence (nr) database in GenBank. The BLAST2GO analysis of these sequences showed 4647 sequences with a blast hit, 1715 sequences without a significant blast hit, 356 sequences with mapping results, and 3690 sequences with annotation (Fig. 1). The unigenes with significant hits (E-value 1-0E-5) are shown in Supplementary Table 1, and those showing no significant similarity to any sequences contained in the nr database are listed in Supplementary Table 2. These unknown transcripts are of special interest and need further analysis to predict their possible function. The blast results of 4647 unigenes showed a high similarity with rice (*Oryza sativa*), followed by grapes (*Vitis vinifera*) and corn (*Zea mays*) and are presented in Supplementary Fig. 1. However, the organism distribution of the unigenes' top hits were as follows: *Vitis vinifera* 578 (12%); *Oryza sativa* 399 (8.5%); *Zea mays* 303 (6.5%); *Amborella trichopoda* 232 (5%); *Oryza brachyantha* 201 (4.3%); *Theobroma cacao* 189 (4%); *Setaria italica* 177 (3.8%); *Citrus sinensis* 154 (3.3%); *Populus trichocarpa* 149 (3.2%); *Sorghum bicolor* 128 (2.75%); *Brachypodium distachyon* 119 (2.5%); *Jatropha ucurcas* 118 (2.56%); *Eucalyptus grandis* 112 (2.4%); *Elaeis guineensis* 107 (2.3%); and *Ricinus communis* 101 (2.1%). The details are given in Fig. 2. At the second GO level, annotated sequences were divided into three principal GO categories: molecular function, biological process and cellular compartment (Fig. 3). In the molecular function (MF) class, (Fig. 3-MF), the majority of the GO terms were grouped into four categories, namely, binding activity (44%), catalytic activity (40%), transporter activity (5%) and structural molecular activity (5%). In binding, heterocyclic compound binding (14%), organic cycling compound binding (14%), ion binding (13%), small molecules binding (9%), protein binding (7%), carbohydrate derivative binding (6%), cofactor binding and lipid binding (each 1%) were the most enriched terms at the third GO level (Supplementary Fig. S2). However, the highly enriched GO terms in catalytic activity included hydrolase activity (8%), transferase activity (8%), and oxidoreductase activity (4%). Based on the biological process (BP) class second level GO terms, (Fig. 3-BP), the major unigenes portions were linked to the cellular process (19%), metabolic process (18%), single-organism process (13%), response to stimuli (9%), cellular component organization (7%), biological regulation (7%), localization (6%), development process (5%), multicellular organismal process (5%), reproduction (3%), signaling (3%), growth (2%) and immune system process (1%). Responses to stress (4%), biotic stimulus (1%), and abiotic stimulus (3%), chemical stimulus (3%), external stimulus (1%), and endogenous stimuli (1%) were included in the response to stimulus (third level GO terms). The cellular metabolic (10%), primary metabolic (10%), and organic substance metabolic processes (10%) were included in the metabolic processes (third level GO terms). Furthermore, for the cellular compartments (CC), class was mostly attributed to the cell (35%), organelle (39%) membrane (16%), and macromolecular complex (8%) (Fig. 3-CC). The function motif/domain for ESTs and unigenes were obtained through the Inter ProScan tool of BLAST2GO, and 3824 unigenes were found with interproscan and 2588 were found without interproscan. The InterPro families and domains are presented in Supplementary Table (S3). In addition, unique sequences were also annotated using KAAS (KEGG Automatic Annotation Server) (Supplementary Table S4). A large number of unigenes were mapped to different KEGG pathways (Supplementary Table 5). The major enzyme commission classes include transferases (480), hydrolases (330), oxidoreductases (218), lyases (67), isomerases (60), and ligases (45). Important KEGG pathways that were obtained were related to plant development, such as glycolysis/gluconeogenesis, the citrate cycle (TCA cycle), photosynthesis, starch and sucrose metabolism, fatty acid biosynthesis and flavonoid biosynthesis. A large number of unigenes were categorized into ''metabolism'', which was further subcategorized into carbohydrate, amino acid, lipid and nucleotide metabolism. Purine metabolism had the highest mapping sequences (84). There were 60 unigenes identified that are involved in starch and sucrose metabolism.

Table 1. Sequence assembly statistics.

Total number of EST sequences	>8000
High-quality EST sequences (Q>20)	7225
Total number of EST sequences after removing vector, poly A, short sequences (>100bp)	6943
Number of contigs	494
Number of singletons	5868
Number of unigenes sequences	6362
Average length of contigs (bp)	675
Average length of singletons (bp)	477
Average length of unigenes sequences (bp)	492

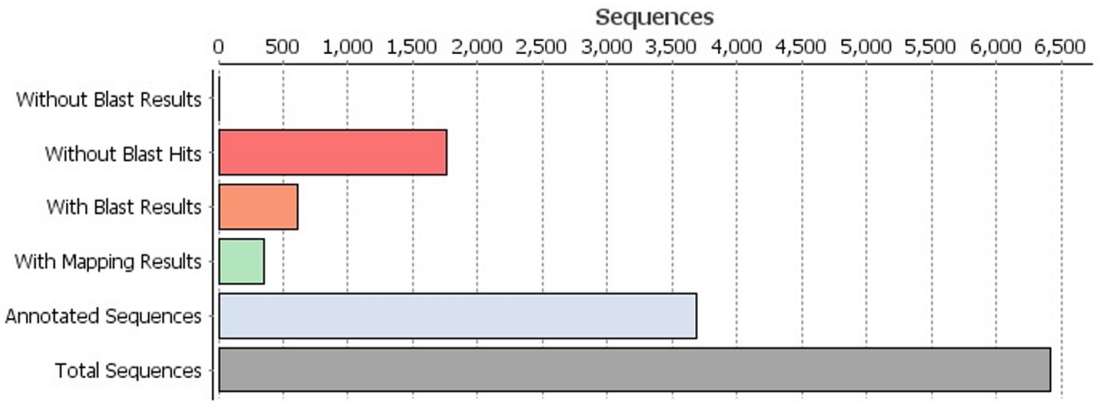

Fig 1. Blastx statistics of 6362 inigenes of Sukary date palm sequences analysed by Blast2go.

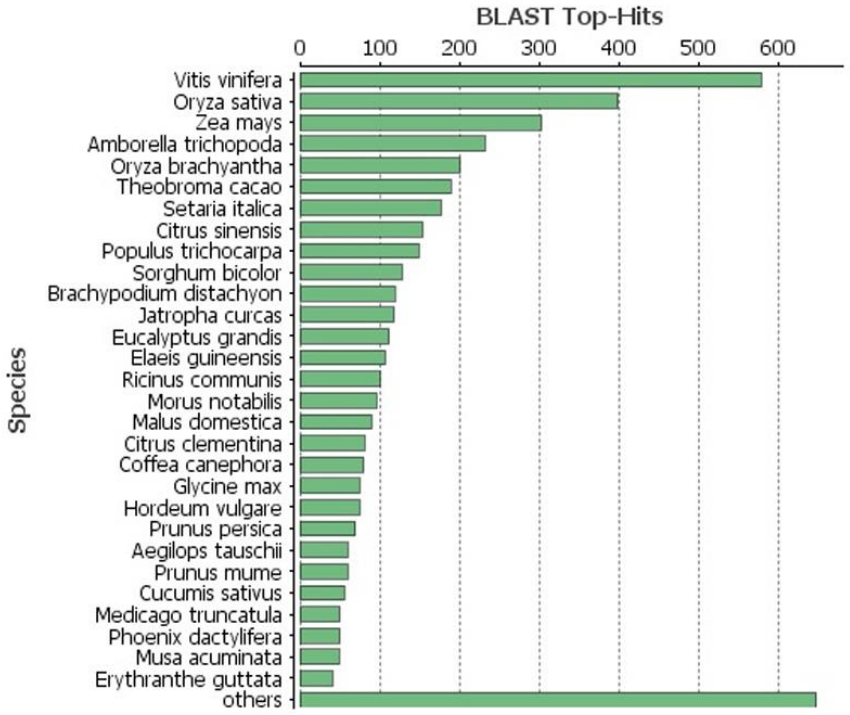

Fig 2. The best hit Species distribution chart of EST of date palm after blastx to NCBI nr. BlastX-based ‘‘top-hit’’ species are ranked by their matched entries. The highest ‘‘hit-species’’ is the *Vitis vinifera* followed by *Oryza sativa*.

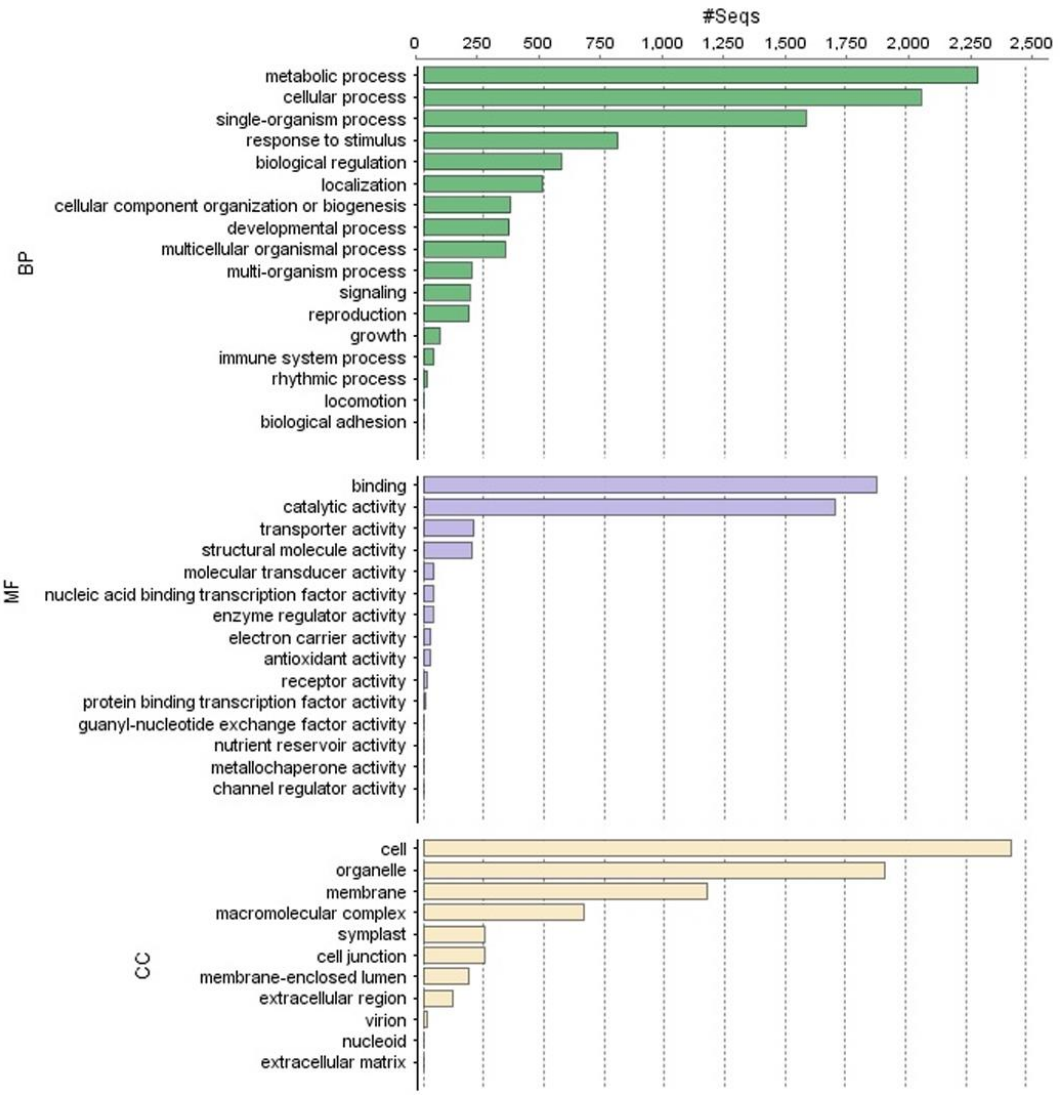

Fig 3. Functional classifications for the 6362 unigenes that were assigned with GO terms (second level GO terms). The three GO categories, biological process (BP), molecular function (MF), and cellular component (CC) are presented.

Other genes were involved in glycolysis/gluconeogenesis (45 unigenes), pyrimidine metabolism (36 unigenes), oxidative phosphorylation (32 unigenes), amino sugar and nucleotide sugar metabolism (31 unigenes), the pentose phosphate pathway (33 unigenes), and the citric cycle (28 unigenes). A representative KEGG map for the citric cycle is shown in Supplementary Fig. 3. Our data serve as an additional source of genes related to the basic metabolic processes during date palm development.

The unigenes related transcription factors, defense/stress and fruit development

Blastx searches revealed some proteins that could be identified as putative transcription factors with matches in the Plant TFDB. The most abundant TF families included ARF (8), followed by PHD (8), bHLH (11), MYB (10), CAMTA (4), FAR1 (8), CPP (6), NAC (4), DOF (3), HB-others (8), M-type (5), SBPs (4), GATA (3), B3 (3), bZIP (4), YABBY (5) and C3H (4). The transcription factors CO-like (2), LSD (2) GRAS, and EIL contained 2 unigenes each, while HSF, TCP, SAP, Trihelix, WRKY, C2H2, and E2F contained one sequence each (Supplementary Table 6). ESTs that are related to defense response were also identified in this study. Among the ESTs that matched genes with known or putative functions, approximately 102 unigenes are involved in stress and defense, accounting for 2.3% (107/4647) of all unigenes. Supplementary Table 7 shows the non-redundant ESTs that share similarities with genes related to defense and stress response, according to GO classifications and previously published data. Four unigenes related to the wound-response family were found in the EST library. These were classified as "response to wound". In another class, the most abundant sequences encoded Glutathione. In total, 11 unigenes containing the glutathione enzyme were identified. Another class of genes involved in "response to heat," was also represented in the library. There were 13 unigenes that were identified, including heat shock protein, heat shock transcription factors and 8 unigenes that represented chaperone proteins. Of the stress- and defense-related unigenes, 4 were possibly related to disease resistance proteins. In addition, twenty three unigenes that encode E3 ubiquitin-protein ligase were also identified. Some other defense-related unigenes were also found, including two unigenes belonging to the drought-responsive protein family, 3 unigenes belonging to the metal tolerance protein family, 2 related to superoxide dismutase and 5 related to the alcohol dehydrogenase family. Based on the Blastx results, the

comprehensive annotation for the detected genes revealed some classes of genes that are directly or indirectly related to fruit development (Supplementary Table 8). Some of these genes are related to the cell cycle, zinc finger family, starch metabolism, and sugar metabolism. KEGG metabolic pathways supported the involvement of these classes in a number of important developmental and metabolic pathways, including growth/death, sugar metabolism (fructose/mannose, amino/nucleotide sugar metabolisms), replication/repair, translation, cell cycle, starch/sucrose/galactose metabolisms and glycolysis/gluconeogenesis. A group of proteins involved in primary metabolism were also identified, including fructose-bisphosphate aldolase, pectin methylesterase, invertase, myo-inositol-1-phosphate synthase, omega-3 fatty acid desaturase, and aspartate aminotransferase (Supplementary Table 8). In the secondary metabolism category, the chalcone-flavanone isomerase family protein is identified, which is related to fruit development. In relation to growth and water, aquaporin and expansin b1 are two genes that were identified and are involved in the loosening of the cell wall during fruit development. In the ethylene metabolism category, s-adenosylmethionine synthetase was present. In addition to this, we identified unigenes related to fruit ripening, which encode most of the genes involved in sugar metabolism and other pathways (Supplementary Table 8). The unigenes identified that are involved in sugar metabolism include sucrose synthase, fructokinase 1, soluble starch synthase 2-3, phosphoglucomutase, alpha-glucan phosphorylase, phosphoglycerate kinase, ribulose-bisphosphate carboxylase oxygenase small subunit, fructose-bisphosphatase, triosephosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase, malate dehydrogenase, glucose-6-phosphate isomerase, phosphoenolpyruvate carboxylase, fructose-bisphosphate aldolase, pyruvate kinase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate mutase, phosphoenolpyruvate carboxykinase, 6-phosphofructokinase 3, aspartate aminotransferase, alpha-amylase, sucrose phosphate synthase ii, and neutral invertase (Supplementary Table 8).

Discussion

The EST analysis provides a powerful and rapid means of generating the transcriptome of specific cell types, which can be used for the identification of differentially expressed genes. The ESTs provide an excellent resource for novel gene discovery and confirmation of *in silico* gene predictions. Large scale EST sequencing provides a direct link to transcribed portions of genome, and with the emergence of next generation sequencing technology, there is rapid screening for novel genes based on various transcriptomic profiles. The high quality and high local demand of this Sukary variety in Saudi market were among the reasons for uncovering its molecular nature via EST sequencing. The ESTs identified here and their various categories are comparable with those reported by Yin et al., (2012) and Al-Mssallem et al., (2013). The functional aspects of the unigenes were revealed, including transcription factors, defense, abiotic tolerance and other fruit development genes, which are of particular interest. Among the identified ESTs, a relatively large group of transcription factors were identified (78 unigenes). Transcription factors (TFs) play important roles in the regulation of cellular pathways in the response to biotic and abiotic stimuli and basic developmental processes. The enrichment of TFs may enhance modulation of cellular redox levels and assist with the avoidance or delay of stress-like processes. Zinc finger proteins are generally considered as DNA-binding transcription factors (Laity et al., 2001) and play an important role in the floral organ morphogenesis and gametogenesis (Huang et al., 2006, Yilmaz and Mittler, 2008). Other important transcription factors in our data are part of the MADS-box family that is involved in floral development (Theißen, 2001). Another important category of the identified ESTs was the plant defense genes. Plants respond to diseases and other biotic or abiotic stresses by activating a broad range of defenses, including the activation of pathogenesis-related (PR) genes at both local and systemic sites, crosslinking of cell wall proteins, generation of reactive oxygen species (ROS) and local programmed cell death. Previous studies in Arabidopsis showed that glutathione genes exhibited a diverse range of responses to jasmonates, salicylic acid, ethylene, and oxidative stress that was induced by heavy metals and hypoxic stress in rice roots (Wagner et al., 2002; Moons, 2003). Heat shock genes are crucial for abiotic stress defense mechanisms. They are widely distributed in nature and involved in protein refolding and assembly, which are induced by drought and salinity (Alamillo et al., 1995, Campalans et al., 2001); thus, restoring plant function after abiotic stress. Furthermore, the presence of E3 ubiquitin-protein ligase among the unigenes indicates a regulatory mechanism for controlling various responses to external stimuli. Ubiquitination is associated with proteasome-mediated protein degradation regulates protein function in a proteasome-independent way. In plants, ubiquitination plays an important role in controlling environmental and endogenous signals, including responses to pathogens (Hare et al., 2003). Moreover, the involvement of E3 ligase in the plant pathogen response was previously identified in Arabidopsis RING finger proteins, RPM1-interacting protein 2 (RIN2) and RIN3 (Kawasaki et al., 2005) and in rice (Zeng et al., 2004). Another two genes identified are homologous to aquaporin, stress protecting proteins family that facilitates water uptake by forming pores; thereby, playing an important role in cell growth and photosynthesis activities after dehydration (Oono et al., 2003). The differential accumulation of this protein has been reported in drought tolerant varieties (Montalvo-Hernandez et al., 2008). Some annotated unigenes belong to the vacuolar-type H+-ATPase (VHA2) family, which plays an important role in generating electrochemical gradients to drive solute transport and related water flux. Expression of H+-ATPase is not tissue specific. It induces responses to external environmental stimuli via cell specific signal transduction (Hentzen et al., 1996). The multiple gene family of ATPase has also been reported in *Arabidopsis thaliana* (Pardo and Serrano, 1989, Houlne and Boutry, 1994) and *Vicia faba* (Nakajima et al., 1995). Five unigenes encoded ethylene response factor (ERF1). It is a member of a novel family of plant-specific transcriptional factors in *Arabidopsis thaliana* (Nakano et al., 2006). ERFs affect number of developmental processes and are also differentially adapted to biotic or abiotic stress conditions, such as pathogen attack, wounds, extreme temperatures, and drought (Ecker, 1995; Penninckx et al., 1996, O'Donnell et al., 1996). An ERF is activated by either ethylene (ET) or jasmonate (JA) and may also be activated synergistically by both hormones (Lorenzo et al., 2003). It regulates defense response genes in the necrotrophic fungi *B. cinerea* and *P. cucumerina* by integrating ET and JA defense responses in Arabidopsis (Berrocal-Lobo et al., 2002; Lorenzo et al., 2003). The chitinases are considered PR proteins (Salzman et al., 1998; Pocock et al., 2000). Chitinases constitute a large family of enzymes with hydrolytic activity against a linear polymer of b-1, 4-N-acetylglucosamine, or chitin, which is a major component in

the cell walls of most pathogenic fungi and the exoskeleton of insects. Four unigenes related to chitinases were identified in this study. Heat shock proteins (HSPs) were also identified. Most HSPs function as molecular chaperones that promote folding, perform structural maintenance, regulate of a subset of proteins involved in signal transduction, regulate cell cycle control, and aid in adaptation to a range of internal and external stresses (Sabehat et al., 1998; Iba, 2002; Wang et al., 2003). In addition to this, they are also involved in developmental regulation. In tomato (*Lycopersicon esculentum*) and strawberry (*Fragaria* spp.) plants, the level of plastid-localized Hsp (pTOM111) increased several fold in ripening fruit and in response to heat stress. Metallothioneins are small, Cys-rich proteins and are implicated in the detoxification of metal ions and reactive oxygen species, as well as in control of cellular redox potential. Oxidative stress promotes the enhanced accumulation of metallothionein transcripts (Navabpour et al., 2003). The antioxidant properties of metallothioneins have been shown in previous research (Akashi et al., 2004). Some of the identified genes in this study may have a role in enhancing Sukary adaptability to harsh environmental conditions. These results suggest the probable existence of different tolerance mechanisms to biotic and abiotic stresses within this variety. Fruit development is characterized by changes in various biological processes, including cell division and enlargement, primary and secondary metabolism, and resistance/ susceptibility to abiotic/biotic stresses. We identified a few classes of genes that are related to fruit development comparable with those previously described for date palm (Yin et al., 2012; Al-Mssallem, et al., 2013). Some of these genes are related to the cell cycle, the zinc finger family, starch metabolism, and sugar metabolism. The KEGG metabolic pathways also indicated the involvement of these genes in a number of important pathways, such as growth/death, sugar metabolism (fructose/mannose, amino/ nucleotide sugar metabolisms), replication/repair, translation, cell cycle, the starch/sucrose/galactose metabolisms and glycolysis/gluconeogenesis. A group of proteins involved in the primary metabolism and secondary metabolism categories was also identified. In addition, unigenes relating sugar metabolism and other pathways were detected. Their role in fruit development in the date palm has been investigated recently (Yin et al., 2012, Al-Mssallem, et al., 2013). This study also revealed new unigenes that may be used for plant development and quality parameters. Further studies based on bioinformatics and functional genomic tools are needed for the characterization of their possible functions. These genes may be of particular interest for crop improvement and manipulation.

Materials and Methods

Plant materials and cDNA library construction

Samples for RNA isolation were taken from female flowers, developed fruits at different stages, and leaves from a single Sukary female tree. Total RNA was isolated using Invitrogen's Concert™ Plant RNA Reagent (Invitrogen Ltd, Paisley PA4 9RF, UK). RNAs from different tissues were pooled together and sent to Evrogen (Miklukho-Maklaya Str, Moscow, Russia) for cDNA synthesis and library normalization (CS011-2B LEVEL 2B). Normalized cDNA species were cloned into a pAL17.3 cloning vector. Positive bacterial colonies were shipped back to our lab where titration was carried out according to Evrogen recommendations. Bacterial colonies were cultured overnight in LB liquid media supplemented with kanamycin (50 µg/ml, final concentration) as a selectable marker, followed by plasmid minipreparation using the QIAprep Spin MiniPrep Kit (QIAGEN, Hilden, Germany). Sequencing reactions were performed in a thermal cycler using the Applied Biosystems BigDye® Terminator v3.1 kit according to the protocols supplied by the manufacturer. The total volume, 20 µl, contained 2.5 µl of mini prep plasmid DNA, 0.5 µl of T7 primer (10 µM stock), 3 µl of 5x Sequencing Buffer, 2 µl of ABI Big Dye terminator v3.1 and 12 of µl Milli-Q water. The PCR program was set to 30 cycles for 20 sec at 96^{o}C, 30 sec at 50^{o}C, and 4 min at 60^{o}C. The sequence reactions were EtOH precipitated and analyzed in an ABI 3130*xl* -16 capillaries sequencer (Applied Biosystems).

Sequence processing and assembly

The sequence trace files were base-called using the Phred program, and low-quality bases (<Q20, 99% accuracy) were eliminated from sequence ends, followed by SeqClean (Pertea et al., 2003) to shorten the Poly-A/T to 5 continuous bases. The vector and other contaminating microbial sequences were removed using the VecScreen program (http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html). After trimming, the EST sequences that were shorter than 100 bp were discarded, and those greater than 100 bp were deposited into the dbEST division of GenBank (LIBEST_028537). A total of 6943 high quality sequences were used to assembled into contigs and singletons using Codon Code Aligner software (CodonCode Corporation, http://www.codoncode.com/)

Annotation and functional classification

BLAST searches were carried out after clustering and assembly to identify similarities between the ESTs and other sequences already deposited in public databases. All of the unigene sequences were compared to GenBank non-redundant protein and nucleotide databases using either the Blastx (E-value$\leq 10^{-6}$) or the Blastn (E-value$\leq 10^{-6}$) program (Altschul et al., 1990). Gene Ontology (GO) annotations were performed with BLAST2GO (Conesa et al., 2005; Gotz et al., 2008) based on sequence similarity. Furthermore, Inter ProScan was performed, and the results were merged with GO annotations to improve them. Finally, the analysis of biological pathways was performed using the KEGG (Kyoto Encyclopedia of Genes and Genomes) (Ogata et al., 1999).

Conclusion

In this study, 6671 ESTs from the Sukary variety of the date palm were established and made available to the public domain (LIBEST_028537). Based on the known sequences, a total of 3690 assembled sequences have been successfully annotated. Additionally, various bioinformatics tools revealed that a fraction of these unique sequences may be involved in the developmental pathway. Stress response and fruit development-related genes were also identified, which provides genomic information for future investigations. Therefore, the results from this study open avenues for further in-depth molecular studies of the date palm and improved cultivation.

Acknowledgements

This work was supported by NSTIP strategic technologies program number BIO164-2 in the Kingdom of Saudi Arabia.

References

Akashi K, Nishimura N, Ishida Y, Yokota A (2004) Potent hydroxy radical scavenging activity of drought-induced type-2 metallothionein in wild watermelon. Biochem Biophys Res Commun. 323:72–78.

Alamillo J, Almoguera C, Bartels D, Jordano J (1995) Constitutive expression of small heat shock proteins in vegetative tissues of the resurrection plant *Craterostigma plantagineum*. Plant Mol Biol. 29:1093–1099.

Altschul SF, Gish W, Miller W, Myers EW (1990) Lipman DJ: Basic local alignment search tool. J Mol Biol. 215:403–410.

Al-Mssallem IS, Hu S, Zhang X, Lin Q, Liu W, Tan J, Yu X, Liu J, Pan L, Zhang T, Yin Y, Xin C, Wu H, Zhang G, Ba Abdullahm MM et al (2013) Genome sequence of the date palm *Phoenix dactylifera* L. Nat Commun. 4: 2274.

Al-Mssallem IS (1996) Date palm. Arabian Global Encyclopedia. 7: 182–187.

Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM, Al-Azwani EK, Chaluvadi S, Pontaroli AC, Debarry J et al (2011) *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). Nat Biotechnol. 29:521–527.

Al-Shahib W, Marshall PJ (2003) The fruit of the date palm: its possible use as the best food for the future. Int J Food Sci Nutr. 54(4):247–259.

Berrocal-Lobo M, Molina A, Solano R (2002) Constitutive expression of ethylene- response-factor1 in Arabidopsis confers resistance to several necrotrophic fungi. Plant J. 29:23–32.

Bourgis F, Kilaru A, Cao X, Ngando-Ebongue GF, Drira N, Ohlrogge JB, Arondel V (2011) Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. Proc Natl Acad Sci USA. 108(30):12527–12532.

Campalans A, Pages M, Messeguer R (2001) Identification of differentially expressed genes by the cDNA-AFLP technique during dehydration of almond. Tree Physiol. 21: 633-643.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M et al (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21: 3674–3676.

Ecker JR (1995) The ethylene signal transduction pathway in plants. Science. 268: 667–675.

Fang Y, Wu H, Zhang T, Yang M, Yin Y et al (2012) A Complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. PLoS ONE. 7(5): e37164.

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH et al (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36: 3420–3435.

Houlne G, Boutry M (1994) Identification of an *Arabidopsis thaliana* gene encoding a plasma membrane H+-ATPase whose expression is restricted to anther tissues. Plant J. 5: 311-317.

Iba K (2002) Acclimative response to temperature stress in higher plants: approaches of gene engineering for temperature tolerance. Annu. Rev. Plant Biol. 53:225–245.

Hare PD, Seo HS, Yang JY, Chua NH (2003) Modulation of sensitivity and selectivity in plant signaling by proteasomal destabilization. Curr Opin Plant Biol. 6: 453-462.

Hentzen A E, Smart LB, Wimmers LE, Fang HH, Schroeder JI, Bennett AB (1996) Two plasma membrane H+ -ATPase genes expressed in guard cells of vicia faba are also expressed throughout the plant. Plant Cell Physiol. 37: 650–659.

Huang F, Chi YJ, Meng QC, Gai JG, Yu DY (2006) Gm ZFP1 coding a single zinc finger protein is expressed with enhancement in reproductive organs and late seed development in soybean (*Glycine max*). Mol Biol Rep. 33:279–285.

Jain SM, Al-Khayri JM, Johnson DV (2011) Date palm biotechnology. Springer. http://www.springer.com/gp/book/9789400713178

Kawasaki T, Nam J, Boyes DC, Holt BF, Hubert DA, Wiig A, Dangl JL (2005) A duplicated pair of arabidopsis RING-finger E3 ligases contribute to the RPM1-and RPS2-mediated hypersensitive response. Plant J. 44: 258-270.

Laity JH, Lee BM, Wright PE (2001) Zinc finger proteins: new insights into structural and functional diversity. Curr Opin Struct Biol. 11: 39–46.

Lorenzo O, Piqueras R, Sánchez-Serrano JJ, Solano R (2003) Ethylene response factor1 integrates signals from ethylene and jasmonate pathways in plant defense. Plant Cell. 15:165–178.

Malek JA (2010) Next generation DNA sequencing applied to the date palm tree (*Phoenix dactylifera*). Acta Hortic. 882: 249–252.

Mathew et al (2014) A first genetic map of date palm (*Phoenix dactylifera*) reveals long-range genome structure conservation in the palms. BMC Genomics. 15:285.

Montalvo-Hernandez L, Piedra-Ibarra E, Gomez-Silva L, Lira-Carmona R, Acosta-Gallegos JA, Vazquez-Medrano J, Xoconostle-Cazares B, Ruiz-Medrano R (2008) Differential accumulation of mRNAs in drought-tolerant and susceptible common bean cultivars in response to water deficit. New Phytologist. 177:102–113.

Moons A (2003) "Osgstu3 and osgtu4, encoding tau class glutathione s-transferases, are heavy metal- and hypoxic stress induced and differentially salt stress-responsive in rice roots," FEBS Letters. 553(3) 427–432.

Nakajima N, Saji H, Aono M, Kondo N (1995) Isolation of cDNA for a plasma membrane H+-ATPase from guard cells of *Vicia faba* L. Plant Cell Physiol. 36:919-924.

Nakano T, Suzuki K, Fujimura T, Shinshi H (2006) Genome-wide analysis of the ERF gene family in arabidopsis and rice. Plant Physiol. 140:411–432.

Navabpour S, Morris K, Allen R, Harrison E, Mackerness SAH, Buchanan-Wollaston V (2003) Expression of senescence-enhanced genes in response to oxidative stress. J Exp Bot. 54:2285–2292.

Oono Y, Seki M, Nanjo T, Narusaka M, Fujita M, Satoh R, Satou M, Sakurai T, Ishida J, Akiyama K (2003) Monitoring expression profiles of arabidopsis gene expression during rehydration process after dehydration using ca 7000 full-length cDNA microarray. Plant J. 34: 868–887.

O'Donnell PJ, Calvert C, Atzorn R, Wasternack C, Leyser HMO, Bowles DJ (1996) Ethylene as a signal mediating the wound response of tomato plants. Science. 274:1914–1917.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H et al (1999) KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 27: 29–34.

Pertea G, Huang X, Liang F, Antonescu V, Sultana R et al (2003) TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics. 19: 651–652.

Pardo JM, Serrano R (1989) Structure of a plasma membrane H+- ATPase gene from the plant *Arabidopsis thaliana*. J Biol Chem. 264: 8551-8562.

Penninckx IA, Eggermont K, Terras FR, Thomma BP, De Samblanx GW, Buchala A, Métraux JP, Manners JM, Broekaer WF (1996) Pathogen induced systemic activation of a plant defens in gene in arabidopsis follows a salicylic acid-independent pathway. Plant Cell. 8:2309–2323.

Pocock KF, Hayasaka Y, McCarthy M, Waters EJ (2000) Thaumatin-like proteins and chitinases, the haze-forming proteins of wine, accumulate during ripening of grape (*Vitis vinifera*) berries and drought stress does not affect the final levels per berry at maturity. J Agric Food Chem. 48:1637–1643.

Salzman RA, Tikhonova I, Bordelon BP, Hasegawa PM, Bressan RA (1998) Coordinate accumulation of antifungal proteins and hexoses constitutes a developmentally controlled defense response during fruit ripening in grape. Plant Physiol. 117:465–472.

Sabehat A, Weiss D, Lurie S (1998) Heat-shock proteins and cross tolerance in plants. Physiol Plant. 103:437–441.

Sghaier-Hammami B, Valledor L, Drira N, Jorrin-Novo JV (2009) Proteomic analysis of the development and germination of date palm (*Phoenix dactylifera* L.) zygotic embryos. Proteomics. 9:2543–2554.

Theiben G (2001) Development of floral organ identity: stories from the MADS house. Curr Opin Plant Biol. 4:75–85.

Wagner U, Edwards R, Dixon DP, Mauch F (2002) Probing the diversity of the arabidopsis glutathione S-transferase gene family. Plant Mol Biol. 49: 515–532.

Wang ZP, Deloire A, Carbonneau A, Federspiel B, Lopez F (2003) An in vivo experimental system to study sugar phloem unloading in ripening grape berries during water deficiency stress. Ann Bot (Lond). 92:523–528.

Wrigley G (1995) Date palm, *Phoenix dactylifera*. In: Smartt J, Simmonds NW (ed), Evolution of crop plants, 2nd ed. Longman, London: pp. 399-403

Yang M, Zhang X, Liu G, Yin Y, Chen K, et al (2010) The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). PLoS ONE. 5(9): e12762. doi:10.1371/journal.pone.0012762

Yilmaz SC, Mittler R (2008) The zinc finger network of plants. Cell Mol Life Sci. 65: 1150–1160.

Yin YX, Zhang XW, Fang YJ, Pan LL, Sun GY, Xin CQ, Abdullah MMB, Yu XG, Hu SN, Al-Mssallem IS et al (2012) High-throughput sequencing-based gene profiling on multi-staged fruit development of date palm (*Phoenix dactylifera*, L.). Plant Mol Biol. 78:617–626.

Zhang H, Jin J, Tang L, Zhao Y, Gu X et al (2011) Plant TFDB 2.0: update and improvement of the comprehensive plant transcription factor database. Nucleic Acids Res. 39: D1114–D1117.

Zeng LR, Qu S, Bordeos A, Yang C, Baraoidan M, Yan H, Xie Q, Nahm BH, Leung H, Wang GL (2004) Spotted leaf11, a negative regulator of plant cell death and defense, encodes a U-box/armadillo repeat protein endowed with E3 ubiquitin ligase activity. The Plant Cell Online. 16: 2795-2808.