Analysis of expressed sequence tags from oil palm (*Elaeis guineensis*)☆

Stefan Jouannic[a,*], Xavier Argout[a,1], Frédéric Lechauve[a], Cécile Fizames[b], Alain Borgel[a], Fabienne Morcillo[a], Frédérique Aberlenc-Bertossi[a], Yves Duval[a], James Tregear[a]

[a] *IRD/CIRAD Oil Palm Laboratory, Centre IRD Montpellier, UMR 1098, 911 avenue Agropolis, 34394 Montpellier cedex 5, France*
[b] *Biochimie et Physiologie Moleculaires des Plantes, UMR 5004 CNRS/ENSA-M/INRA/UM2, Place Viala, 34060 Montpellier Cedex 2, France*

Received 10 January 2005; revised 7 March 2005; accepted 8 March 2005

Available online 18 April 2005

Edited by Takashi Gojobori

Abstract **This is the first report of a systematic study of genes expressed by means of expressed sequence tag (EST) analysis in oil palm, a species of the Arecales order, a phylogenetically key clade of monocotyledons that is not widely represented in the sequence databases. Five different cDNA libraries were generated from male and female inflorescences, shoot apices and zygotic embryos and unidirectional systematic sequencing was performed. A total of 2411 valid EST sequences were thus obtained. Cluster analysis enabled the identification of 209 groups of related sequences and 1874 singletons. Putative functions were assigned to 1252 of the set of 2083 non-redundant ESTs obtained. The EST database described here is a first step towards gene discovery and cDNA array-based expression analysis in oil palm.**

Keywords: Expressed sequence tag; Oil palm; Flower; Shoot apex; Embryo; *Mantled*

1. Introduction

Oil palm (*Elaeis guineensis* Jacq.) is a monocotyledonous plant of the palm family (Arecaceae), originating from West Africa. It is a single-stemmed palm (i.e., it possesses a single shoot apical meristem) and produces male and female inflorescences successively on the same plant. Oil palm has an estimated haploid genome size about 1000 or 1700 Mbp, according to reference DNA used (see the Plant DNA *C*-values Database, http://www.rbgkew.org.uk/cval/homepage.html and [1]), divided among 16 pairs of chromosomes. The Arecaceae, which form the only family in the order Arecales, is a monophyletic group, comprising 190 genera and approximately 2400 species [2]. The Arecales are phylogenetically related to the orders Poales (grasses), Zingiberales (bananas) and Commelinades (water hyacinth), belonging to the Commelinid subclass [3]. Several species of the Arecaceae family are of economic interest for food production (oil palm, coconut palm, date palm, and peach palm) and for ornamental use (*Butia*, *Washingtonia* and *Phoenix* species). Some species are also of medicinal interest such as saw palmetto (*Serenoa repens*) used in benign prostate hyperplasia treatment.

Oil palm is a species of particular economic importance as it is the second largest source of edible oil in the world after soybean. Oil palm is cultivated in the inter-tropical regions of Asia, Latin America and Africa. Planting material consists of *tenera* hybrids (bearing fruits with shells of intermediate thickness) originating from crosses between *dura* (thick shell) and *pisifera* (thin shell) types. Since each selection cycle lasts for around 10 years, genetic improvement is very slow. As a consequence of the biological characteristics of oil palm (long life cycle and no natural vegetative reproduction) and a high heterogeneity prevalent among hybrids, breeding strategies are labour intensive and time consuming. Clonal propagation of elite material through tissue culture has thus been developed for mass propagation. Cloning of oil palm is performed by inducing somatic embryogenesis on calli derived from various tissue sources. Although this approach has been used with success in a number of laboratories, a proportion of the regenerants show an epigenetic homeotic flowering abnormality known as *mantled* [4], which is observed only on palms produced by tissue culture. This somaclonal variant involves an alteration in the identity of third whorl organs in flowers of both sexes and is similar to the class B floral mutants identified in *Arabidopsis thaliana* and *Antirrhinum majus* [5]. The abnormality may result in partial or complete flower sterility, thus directly affecting oil production. Data from various studies suggest that the *mantled* phenotype is associated with a global hypomethylation of the genome and changes in its methylation pattern, but not with major rearrangements of transposable elements [6–8]. In order to understand the molecular mechanisms involved in the control of early development and flower formation of oil palm, a characterisation of the oil palm gene transcriptome was undertaken. Partial sequencing of cDNA libraries to generate expressed sequence tags (ESTs) is an effective first step towards gene discovery, the characterisation of transcription patterns, notably by means of cDNA array-based expression analyses, and also the generation of molecular and genetic markers [9].

A number of publicly available EST libraries (those containing more than 1000 ESTs) of monocotyledon species have been produced for members of the Poales (wheat, maize, barley, rice, sugarcane, sorghum, etc.), the Asparagales (onion) and

☆ *Accession numbers*. The nucleotide sequence data reported in this paper are available in the DDBJ/EMBL/GenBank databases under the Accession Nos. CN599371–CN601781.

*Corresponding author. Fax: +33 4 67416181.
E-mail address: jouannic@mpl.ird.fr (S. Jouannic).

[1] Present address: CIRAD-AMIS, UMR 1096, TA 40/03, Avenue Agropolis, F-34398 Montpellier cedex 5, France.

Abbreviations: EST, expressed sequence tag; BLAST, basic local alignment search tool; UTR, untranslated region

0014-5793/$30.00
doi:10.1016/j.febslet.2005.03.093

the Acorales (sweetflag) orders, which do not illustrate the whole diversity of the monocotyledons. Oil palm is a representative of the Arecaceae family and the Arecales order, which is a phylogenetically key clade of monocotyledons. This is the first paper describing a large set of expressed genes through EST analysis from a species of the Arecales order, which is poorly represented in the sequence databases. We have summarised and classified 2411 ESTs using a local sequence analysis pipeline and compared data with those of other plant species. The availability of EST data for this important plant group creates the possibility of carrying out comparative studies at both the genome level and for individual accessions of interest, between the Arecales and other monocotyledon and dicotyledon groups.

2. Materials and methods

2.1. Plant material

Male and female inflorescences of 20 and 15 cm length, respectively, were collected from oil palm plants (C1001 seed-derived material and LMC17 clonal line for male and female inflorescences, respectively), growing at the La Mé Experimental Station, Centre National de Recherche Agronomique, Côte d'Ivoire. Spikelets from the median zone of the inflorescences were used for RNA extraction. One cm long leaf segments containing the shoot apical meristem (hereafter referred to as "apices") were collected from young in vitro cultivated oil palm plantlets regenerated from leaf-derived calli as previously described [10]. Two different clonal lines were used. Apices were harvested from plants regenerated from a clonal line obtained from LMC3 line palms carrying the *mantled* abnormality (referred later as "abnormal apex") and from a clonal line obtained from normal LMC249 palms (referred later as "normal apex"). Immature zygotic embryos (3–5.5 months old) were collected from seed-derived oil palms (C1001 genotype) growing at the Pobé Experimental Station, Benin. The seed- and in vitro-derived plant material all originated from *Deli* (*tenera*) × *La Mé* (*dura*) crosses.

2.2. cDNA library construction and sequencing

Five cDNA libraries were constructed from distinct oil palm tissues: shoot apices from normal in vitro cultivated oil palm plantlets ("normal apex" library); shoot apices from in vitro cultivated oil palm plantlets carrying the *mantled* abnormality ("abnormal apex" library); rachillae from a male inflorescence ("male inflorescence" library); rachillae from a female inflorescence ("female inflorescence" library), and immature zygotic embryos ("zygotic embryo" library) (see Section 2.1 for plant material details).

Total RNA extraction and poly(A)+ RNA purification were carried out as described previously [11]. The cDNA libraries were constructed from each poly(A)+ RNA sample using the ZAP-cDNA synthesis kit and the ZAP-cDNA Gigapack® III Gold packaging extract (Stratagene). Individual cloned cDNAs were obtained by in vivo mass-excision according to the manufacturer's instructions, and were randomly isolated and cultivated. Glycerol stocks of individual cDNA clones were arrayed in 96-well microtitre plates for storage of bacteria and isolation of plasmid DNA. Plasmid DNA extraction was carried out using the NucleoSpin Robot-96 Plasmid extraction kit (Machery–Nagel) with an automated DNA extraction robot (Biorobot 9600 Qiagen). cDNA inserts were sequenced with M13 Reverse primer using the Big Dye terminator kit (Perkin–Elmer) with an automated DNA capillary sequencer ABI 3700 (Perkin–Elmer).

2.3. Sequence processing and analysis

Sequence data were analysed using a biprocessor AMD 1.2 GHz Transtec 2500 computer with Linux Debian. The ABI formatted chromatogram sequences were processed automatically using a custom pipeline. The processing was carried out individually for the five sets of ESTs. This pipeline successively linked sequence backup, base calling by PHRED, elimination of sequences shorter than 50 bp and low quality sequences, and vector trimming by Vecscreen (NCBI, ftp://ftp.ncbi.nlm.nih.gov) and Matcher (Emboss, http://www.hgmp.mrc.ac.uk/Software/EMBOSS/Apps/matcher.html). The EST sequences were clustered and assembled using the Stackpack application (provided by SANBI, http://www.sanbi.ac.za/Dbases.html), which incorporates the d2 cluster [12], Phrap and Crawview [13] programs. Groups that contained only one sequence were classified as singletons. To assign functions, the valid ESTs and the assembled consensus sequences were locally aligned by BLASTALL (NCBI, ftp://ftp.ncbi.nlm.nih.gov/blast) to accessions in a local non-redundant protein sequence database with entries from GenPept, Swissprot, PIR, PDF, PDB and NCBI RefSeq, using the BLASTX algorithm with an *E*-value cut off at 10^{-5}. If the EST sequences did not match any database sequences, the BLASTN algorithm was used in conjunction with a nucleotide sequence database, with entries from all traditional divisions of GenBank, EMBL, and DDBJ and with an *E*-value cut off at 10^{-5}. Finally, the sequences, clustering results and basic local alignment search tool (BLAST) data were automatically integrated in a relational database, searchable via a local web browser-based interface. For comparative analyses between the five EST sets, clustering and assembling were performed using the Stackpack facility and integrated into the database with all EST sequences.

The functional classification previously applied to *Medicago truncatula* genes [14] was adapted to oil palm ESTs. Translated or untranslated oil palm ESTs were manually sorted into 16 functional groups and an unclassified group on the basis of sequence comparison to non-redundant GenBank entries.

Statistical analyses of sequence length, G + C content and functional distributions were performed using the STATISTICA program suite (StatSoft, USA).

3. Results and discussion

3.1. Sequencing and clustering of oil palm ESTs

Five cDNA libraries were constructed from normal apex-, abnormal apex-, female inflorescence-, male inflorescence- and zygotic embryo-derived poly(A)+ RNA samples, respectively (see Section 2). Sequencing of the clones from each oil palm cDNA library followed by sequence processing produced a total of 2411 high quality ESTs (Table 1). The oil palm dbEST represents up to 2083 different genes in the form of 209 clusters assembled from two or more ESTs and 1874 singletons. The distribution among the five EST sets derived from the five cDNA libraries used is shown in Table 1. We found a low redundancy within and between the five EST sets, which may be explained in part by the relatively low number of sequences in each set.

A large proportion of ESTs shared no significant similarity with sequences from GenBank (~40%; Table 1 and Fig. 1). Their average length (294 bp) is significantly lower than that of the ESTs sharing similarities with GenBank entries (418 bp) ($F(1.2409) = 563$, $P = 0.00000$; Table 1). This significant difference is also observed individually within each of the five sets of ESTs ($F(4.2401) = 4.42$, $P = 0.001$).

An analysis of G + C content revealed a lower average G + C value for the ESTs displaying no significant similarity (Table 1). The average G + C content of the EST population is 49.6% and the difference between ESTs with hits (51.2%) and ESTs with no hits (47.0%) is highly significant at a threshold of 0.05 ($F(1.2409) = 135$, $P = 0.00000$). The G + C content of the ESTs with hits is homogenous between the different libraries (50.6–52.2%), in contrast to the ESTs with no significant similarity, which are more variable between the 5 libraries (46.0–49.2%). It is interesting to note that in the dicotyledon *Arabidopsis*, and the monocotyledons rice and onion, G + C contents were found to be higher in coding sequences

Table 1
Analysis of oil palm EST libraries

Library	Female inflo	Male inflo	Normal apex	Abnormal apex	Zygotic embryo	All
Total valid ESTs	349	625	313	998	126	2411
EST assigned to clusters	30	77	20	192	32	537
Clusters	13	38	11	83	11	209
Singletons	319	548	293	806	94	1874
Total non redundant ESTs	332	586	304	889	105	2083
Redundancy (%)	8.6	12.3	6.4	19.1	25.4	9.0
Non redundant ESTs with no similarity	109	220	120	371	40	832
Non redundant ESTs unique to organ	283	515	261	807	91	–
Average length (bp)						
No similarity	301^{cd}	289^{de}	328^{c}	289^{d}	258^{e}	294
Hits	381^{b}	441^{a}	444^{a}	415^{ab}	382^{b}	418
GC% content						
No similarity	48.1^{abc}	49.2^{ab}	45.5^{bc}	46.0^{bc}	46.4^{bc}	47.0
Hits	52.0^{a}	51.1^{a}	50.6^{a}	51.1^{a}	52.2^{a}	51.2

Redundancy, ESTs assembled in clusters/total ESTs; All, total non-redundant ESTs between the five sets of ESTs. Different letters mean significant difference of GC% or sequence length (Newman–Keuls test 5% threshold) revealed by two separate analyses.

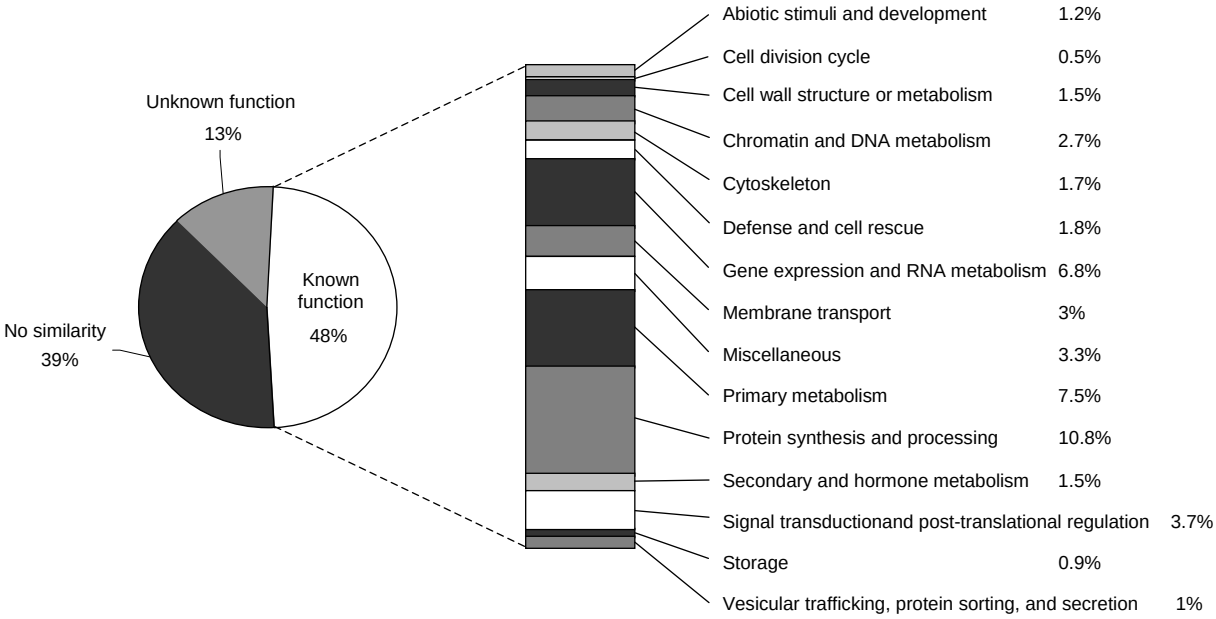

Fig. 1. Functional classification of the oil palm ESTs. All non-redundant ESTs were assigned to a functional category based on highest scoring BLASTX or BLASTN results. Percentages are with respect to the total set of non-redundant ESTs.

than in non-coding sequences [15]. The difference in G + C composition observed between oil palm ESTs with or without similarities suggests that a substantial portion of the sequences displaying no database similarities may contain non-coding regions such as 5′ untranslated regions (UTRs). This hypothesis is compatible with the shorter average length observed for the sequences displaying no significant similarities, which are thus less likely to contain coding sequences. Indeed a significant number of cDNAs isolated and characterised in our lab have been found to possess a 5′ UTR longer than 300 bp (unpublished data). Following comparison with *Arabidopsis*, rice and onion data [15], we found that the G + C content of the oil palm ESTs reported here was more similar to that of rice sequences, which are characterised by G + C contents of about 50% and 35% for coding and non coding sequences, respectively; compared to that of the other two species, which are both characterised by corresponding G + C contents.

A detailed analysis of G + C content distribution (Fig. 2) revealed a unimodal distribution of the ESTs with hits ($N = 1505$, mean = 0.5122, S.D. = 0.0740). In contrast the ESTs with no significant similarities have a bimodal distribution with a higher variance ($N = 906$, S.D. = 0.1032), one with a low G + C content (0.40) and a second one with a high G + C content (0.50) similar with the average G + C content of ESTs sharing similarities with GenBank entries. These results suggest the presence of two populations within the group of ESTs which display no significant similarities. One possible explanation is that the population of ESTs with high G + C content may correspond to coding sequences for proteins with no significant similarities to existing accessions, and that the population with low G + C content corresponds to 5′ UTR sequences. Longer sequences would tend to minimize this phenomenon by decreasing the proportion of 5′ UTRs in sequences obtained, thereby increasing the overall G + C content.

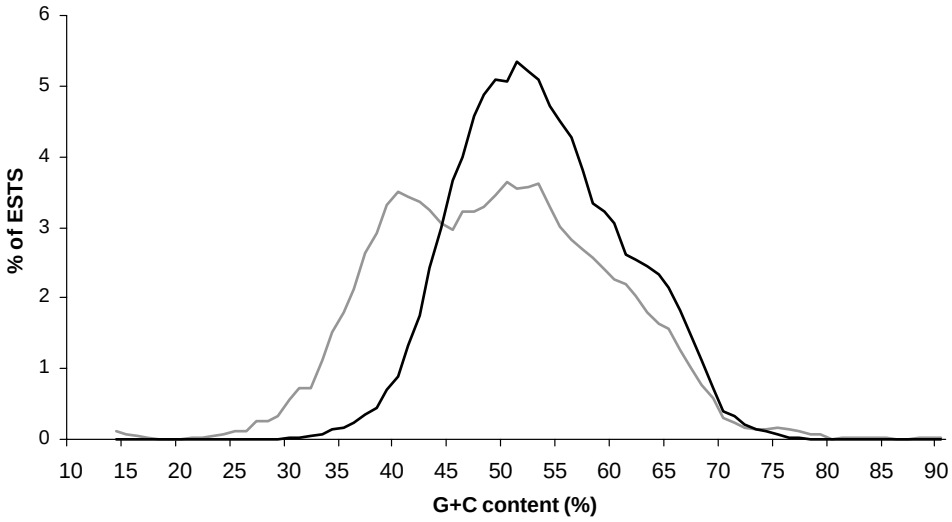

Fig. 2. Differential distribution in the G + C content of all oil palm ESTs with or without significant similarities. Grey curve: G + C content of sequences that do not match with GenBank entries. Black curve: G + C content of sequences that match with GenBank entries with an *E*-value of 10^{-5} or lower.

3.2. Function of expressed oil palm genes

Approximately 61% of the total non-redundant ESTs share significant similarities with GenBank entries (Fig. 1). Putative functions were assigned for 48% of them by comparison to sequences from other organisms. The other 13% matched with sequences of unknown function from the dicotyledon *Arabidopsis*, or the monocotyledons rice and maize. Oil palm ESTs with significant similarities were found to match at similar frequencies to monocotyledonous (rice, 35%) and dicotyledonous (*Arabidopsis*, 37%) sequences. However, this result may be misleading, due to the less advanced state of annotation of the rice genome compared to that of *Arabidopsis*. The ESTs with significant similarities are unequally distributed between the different functional categories. As shown in Fig. 3, in all EST sets, apart from that of the zygotic embryo, the largest functional group of ESTs was that assigned to Protein synthesis and processing (22% of non-redundant ESTs with assigned function), followed by Primary metabolism (15%) and Gene expression and RNA metabolism (14%). This finding is similar to the results obtained from other organisms. It can be noticed that a significantly higher portion of ESTs (α risk = 5%) is classified in the Gene expression and RNA metabolism category for the female and male inflorescence-derived EST sets than for the others (~19% of non-redundant ESTs with assigned function and ~10%, respectively) and in Miscellaneous category for the normal and abnormal apex-derived ESTs than for the others (~11% of non-redundant ESTs with assigned function and ~4%, respectively). This last finding is consistent with the fact that photosynthesis-related proteins have been classified in this category. It can also be noticed that a significantly lower portion of ESTs (α risk = 5%) is classified in the Miscellaneous category for the male inflorescence ESTs. The Storage category is specifically composed of immature zygotic embryo-derived ESTs. This is to be expected as this functional category contains ESTs corresponding to embryo-specific maturation-related proteins. For this specific cDNA library, the non-normalisation of the cDNA is a major limitation to obtaining good gene representation in the library population, as the storage protein mRNAs are highly abundant in the original tissue.

About 39% of the oil palm ESTs share no significant similarities with GenBank entries. The proportion of sequences in this category varied from 33% in the female inflorescence group to 42% in the abnormal apex-derived EST set. This is to be expected because only a small number of sequences from *E. guineensis* and other species of the Arecaceae family are currently available in the sequence databases (374 and 1753 nucleotide sequence entries, respectively; 56 and 818 protein sequence entries, respectively; based on NCBI entries on 08/12/2004), most of them corresponding to microsatellite-containing sequences, rDNA sequences, organelle DNA sequences and relatively few protein coding sequences. A portion of these ESTs seems to consist of 5′ UTRs, on the basis of their lower G + C content (Fig. 2), which are unlikely to be conserved significantly to match with sequences from other species present in the databases. As previously mentioned, a proportion of the oil palm ESTs which share no significant similarities with GenBank entries is likely to consist of translated sequences, on the basis of their G + C content averaging at about 50% (Fig. 2) even though no relationship was found with any accessions in the public databases. Sequences of this type may represent coding sequences specific to oil palm or with low similarity to other species represented in GenBank entries.

Table 2 lists the most highly expressed genes observed in the EST collection as a whole, indicating the number of the corresponding ESTs present in the five libraries. They represent mostly typical ''housekeeping'' genes. The most highly represented transcripts in the EST collection as a whole code for glycine-rich RNA binding proteins. This type of protein has been identified in both angiosperms and gymnosperms and is highly conserved [16]. Several studies have suggested their involvement in RNA processing at several levels [17]. The other abundant transcripts code for DnaJ-like proteins, elongation factor 1-α, α-tubulin, metallothionein-like proteins and ribosomal proteins. It can be noticed that some ESTs,

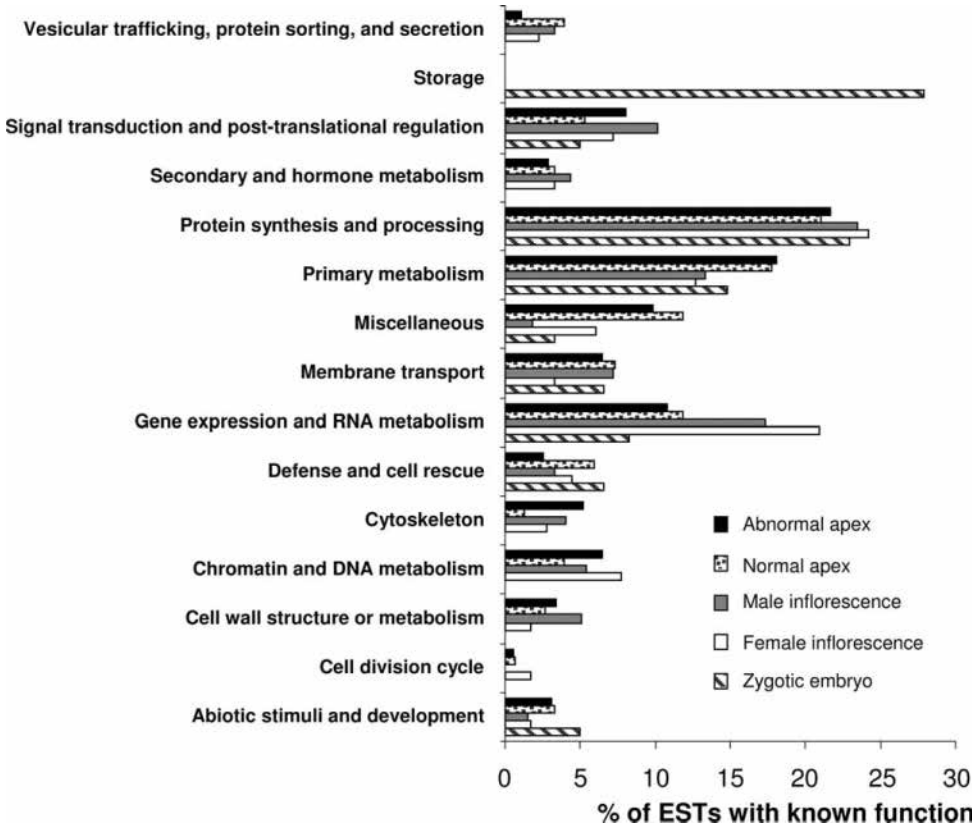

Fig. 3. Distribution amongst the different functional categories of the non-redundant ESTs from the five libraries. ESTs with no significant similarities and ESTs that match to unknown proteins were not included in this analysis.

Table 2
Transcripts which predominate in the oil palm EST libraries

Protein	ESTs in cluster	Singletons	cDNA libraries
Glycine-rich RNA-binding protein	20	12	5 (A0, A1, F, M, Z)
DnaJ protein	11	8	5 (A0, A1, F, M, Z)
Oleosin	17	2	1 (Z)
7S globulin	14*	7	1 (Z)
Elongation factor 1-α	7	8	5 (A0, A1, F, M, Z)
Alpha tubulin	9	4	4 (A0, A1, F, M)
Metallothionein-like protein	6	3	4 (A0, A1, F, M)
BURP domain-containing protein	10	1	2 (A0, A1)
Ribosomal protein L23	5	4 (48)	3 (A0, A1, M)
Ribosomal protein L16	5	4 (48)	3 (A0, A1, M)
Lipid transfer protein	7	4	4 (A0, A1, F, M)
PVR3-like protein	7	–	2 (A0, A1)

A0, normal apex; A1, abnormal apex; F, female inflorescence; M, male inflorescence; Z, zygotic embryo; *, 14 ESTs in two distinct clusters; (48), total number of singletons matching to the ribosomal protein class.

such as those encoding a BURP domain-containing protein and a PVR3-like protein, are detected only in apex-derived libraries. The BURP domain is present in a range of plant proteins with diverse patterns of expression, but its function remains to be elucidated [18]. PVR3-like proteins are similar to non specific-lipid transfer proteins. These proteins may be involved in a number of different biological functions, putative roles including the transport of cuticular components, growth inhibition of bacterial and fungal pathogens and adaptation to various environmental conditions [19]. The last class of abundant encoded protein in the ESTs is that which contains the storage proteins (7S globulins and oleosins). This category is specific to zygotic embryo-derived ESTs. One of the goals in the establishment of the oil palm dbEST was to identify organ specific genes or genes affected by the *mantled* abnormality through the comparative analyses of female and male inflorescence on one hand and of the normal and abnormal apices on the other. Due to the very high number of singletons it is very difficult to identify specific genes only on the basis of comparative sequence analysis. This difficulty will need to be resolved by the development of an EST-based macro-array approach.

4. Conclusion

The data presented here represent the first overview of oil palm genes expressed in different organs of the plant and represents an important contribution to the publicly accessible sequence data available for *E. guineensis* and more generally for the Arecaceae family. Most of the currently available sequences for this family correspond to microsatellite, rDNA and organelle DNA sequences with very few protein coding sequences. This resource provides a starting point for cDNA array-based expression analysis in order to reveal changes in gene expression levels during development in oil palm. Future EST work planned as a follow-up to this study will concentrate on subtractive and normalised libraries so as to target specific aspects of development within the plant. It should also provide the basis for identifying genes affected in their expression pattern by the *mantled* abnormality. More generally, this EST resource may contribute in the future to phylogenetic studies and to wider scale efforts to compare plant genomes through comparative genomics. In addition to the data presented here, it should be noted that an oil palm EST resource is available on the Malaysian Palm Oil Board (MPOB) web site (http://palmoilis.mpob.gov.my/palmgenes.html), but that the sequences in question are not currently available in the public databases.

An important effort has been undertaken to establish a genetic map for this species, notably by using RFLP and microsatellite markers [20–22]. In this context, this set of ESTs will be a source of homologous gene targeted markers (GTMs) [23], for the establishment of a reference genetic map used in the development of breeding strategies (QTL identification, marker-assisted selection).

Acknowledgements: We thank Benoit Piégu and Richard Cook for helpful discussions and Xavier Sabau and Conchita Ferraz for technical help in DNA preparation and sequencing, respectively. We also thank Tim Tranbarger for critical reading of the manuscript. The authors also acknowledge the generous support of colleagues at CNRA La Mé Experimental Station in Côte d'Ivoire and INRAB Pobé in Benin for providing plant material. This work was partially funded by the Montpellier-LR Genopole.

References

[1] Rival, A., Beulé, T., Barre, P., Hamon, S., Duval, Y. and Noirot, M. (1997) Comparative flow cytometric estimation of nuclear DNA content in oil palm (*Elaeis guineensis* Jacq) tissues cultures and seed-derived plants. Plant Cell Rep. 16, 884–887.
[2] Tomlinson, P.B. (1990) The Structural Biology of Palms, Oxford Science Publications.
[3] Chase, M.W. (2004) Monocot relationships: an overview. Am. J. Bot. 91, 1645–1655.
[4] Corley, R.H.V., Lee, C.H., Law, L.H. and Wong, C.Y. (1986) Abnormal development in oil palm clones. The planter, Kuala Lumpur 62, 233–240.
[5] Coen, E.S. and Meyerowitz, E.M. (1991) The war of the whorls: genetic interactions controlling flower development. Nature 353, 31–37.
[6] Jaligot, E., Rival, A., Beulé, T., Dussert, S. and Verdeil, J.L. (2000) Somaclonal variation in oil palm (*Elaeis guineensis* Jacq.): the DNA methylation hypothesis. Plant Cell Rep. 19, 684–690.
[7] Jaligot, E., Beulé, T., Baurens, F.C., Billotte, N. and Rival, A. (2004) Search for methylation-sensitive amplification polymorphisms associated with the *mantled* variant phenotype in oil palm (*Elaeis guineensis* Jacq). Genome 47, 224–228.
[8] Kubis, S.E., Castilho, A.M., Vershinin, A.V. and Heslop-Harrison, J.S. (2003) Retroelements, transposons and methylation status in the genome of oil palm (*Elaeis guineensis*) and the relationship to somaclonal variation. Plant Mol. Biol. 52, 69–79.
[9] Rudd, S. (2003) Expressed sequence tags: alternative or complement to whole genome sequences?. Trends Plant Sci. 8, 321–329.
[10] Pannetier, C., Arthuis, P. and Lievoux, D. (1981) Néoformation de jeunes plantes d'*Elaeis guineensis* à partir de cals primaires obtenus sur fragments foliaires cultivés in vitro. Oléagineux 36, 119–122.
[11] Tregear, J.W., Morcillo, F., Richaud, F., Berger, A., Singh, R., Cheah, S.C., Hartmann, C., Rival, A. and Duval, Y. (2002) Characterization of a defensin gene expressed in oil palm inflorescences: induction during tissue culture and possible association with epigenetic somaclonal variation events. J. Exp. Bot. 53, 1387–1396.
[12] Burke, J., Davison, D. and Hide, W. (1999) d2_cluster: a validated method for clustering EST and full-length cDNA sequences. Genome Res. 9, 1135–1142.
[13] Chou, A. and Burke, J. (1999) CRAWview: for viewing splicing variation, gene families, and polymorphism in clusters of ESTs and full-length sequences. Bioinformatics 15, 376–381.
[14] Journet, E.P., van Tuinen, D., Gouzy, J., Crespeau, H., Carreau, V., Farmer, M.J., Niebel, A., Schiex, T., Jaillon, O., Chatagnier, O., Godiard, L., Micheli, F., Kahn, D., Gianinazzi-Pearson, V. and Gamas, P. (2002) Exploring root symbiotic programs in the model legume *Medicago truncatula* using EST analysis. Nucleic Acids Res. 30, 5579–5592.
[15] Kuhl, J.C., Cheung, F., Yuan, Q., Martin, W., Zewdie, Y., McCallum, J., Catanach, A., Rutherford, P., Sink, K.C., Jenderek, M., Prince, J.P., Town, C.D. and Havey, M.J. (2004) A unique set of 11,008 onion expressed sequence tags reveals expressed sequence and genomic differences between the monocot orders Asparagales and Poales. Plant Cell 16, 114–125.
[16] Veau, B., Oudin, A., Courtois, M., Chenieux, J.-C., Hamdi, S., Rideau, M. and Clastre, M. (2000) Cloning of two cDNAs encoding crGRP2 and crGRP3 (accession nos. AF200323 and AF200322), the first members of the RRM-GRP family in *Catharanthus roseus* (PGR00-049). Plant Physiol. 122, 1459.
[17] Guiltinan, M.J. and Niu, X. (1996) cDNA encoding a wheat (*Triticum aestivum*, cv Chinese spring) glycine-rich nucleic acid-binding protein. Plant Mol. Biol. 30, 1301–1306.
[18] Granger, C., Coryell, V., Khanna, A., Keim, P., Vodkin, L. and Shoemaker, R.C. (2002) Identification, structure, and differential expression of members of a BURP domain containing protein family in soybean. Genome 45, 693–701.
[19] Song, J.Y., Choi, D.W., Lee, J.S., Kwon, Y.M. and Kim, S.G. (1998) Cortical tissue-specific accumulation of the root-specific ns-LTP transcripts in the bean (*Phaseolus vulgaris*) seedlings. Plant Mol. Biol. 38, 735–742.
[20] Billotte, N., Marseillac, N., Risterucci, A.M., Adon, B., Brottier, P., Baurens, F.C., Singh, R., Herran, A., Asmady, H., Billot, C., Amblard, P., Durand-Gasselin, T., Courtois, B., Asmono, D., Cheah, S.C., Rohde, W., Ritter, E. and Charrier, A. (2005) Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). Theor. Appl. Genet. 110, 754–765.
[21] Mayes, S., Jack, P.L., Marshall, D.F. and Corley, R.H.V. (1997) Construction of a RFLP genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). Genome 40, 116–122.
[22] Mayes, S., Jack, P.L. and Corley, R.H. (2000) The use of molecular markers to investigate the genetic structure of an oil palm breeding programme. Heredity 85, 288–293.
[23] Gupta, P.K. and Rustgi, S. (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct. Integr. Genomics 4, 139–162.