Mol Biol Rep (2011) 38:1329–1340
DOI 10.1007/s11033-010-0234-8

Unraveling new genes associated with seed development and metabolism in *Bixa orellana* L. by expressed sequence tag (EST) analysis

Virgínia L. F. Soares · Simone M. Rodrigues · Tahise M. de Oliveira · Talisson O. de Queiroz · Lívia S. Lima · Braz T. Hora-Júnior · Karina P. Gramacho · Fabienne Micheli · Júlio C. M. Cascardo · Wagner C. Otoni · Abelmon S. Gesteira · Marcio G. C. Costa

Received: 6 May 2010 / Accepted: 11 June 2010 / Published online: 19 June 2010

Abstract The tropical tree *Bixa orellana* L. produces a range of secondary metabolites which biochemical and molecular biosynthesis basis are not well understood. In this work we have characterized a set of ESTs from a non-normalized cDNA library of *B. orellana* seeds to obtain information about the main developmental and metabolic processes taking place in developing seeds and their associated genes. After sequencing a set of randomly selected clones, most of the sequences were assigned with putative functions based on similarity, GO annotations and protein domains. The most abundant transcripts encoded proteins associated with cell wall (prolyl 4-hydroxylase), fatty acid (acyl carrier protein), and hormone/flavonoid (2OG-Fe oxygenase) synthesis, germination (MADS FLC-like protein) and embryo development (AP2/ERF transcription factor) regulation, photosynthesis (chlorophyll a–b binding protein), cell elongation (MAP65-1a), and stress responses (metallothionein- and thaumatin-like proteins). Enzymes were assigned to 16 different metabolic pathways related to both primary and secondary metabolisms. Characterization of two candidate genes of the bixin biosynthetic pathway, *BoCCD* and *BoOMT*, showed that they belong, respectively, to the carotenoid-cleavage dioxygenase 4 (CCD4) and caffeic acid O-methyltransferase (COMT) families, and are up-regulated during seed development. It indicates their involvement in the synthesis of this commercially important carotenoid pigment in seeds of *B. orellana*. Most of the genes identified here are the first representatives of their gene families in *B. orellana*.

Keywords Apocarotenoids · Bixin · Carotenoid-cleavage dioxygenase · O-methyltransferase

Electronic supplementary material The online version of this article (doi:10.1007/s11033-010-0234-8) contains supplementary material, which is available to authorized users.

V. L. F. Soares · T. M. de Oliveira · T. O. de Queiroz · B. T. Hora-Júnior · F. Micheli · J. C. M. Cascardo · M. G. C. Costa (✉)
Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, Ilhéus, BA, Brazil
e-mail: mcosta@labbi.uesc.br

S. M. Rodrigues · W. C. Otoni
Departamento de Biologia Vegetal, Universidade Federal de Viçosa, Viçosa, MG, Brazil

L. S. Lima · K. P. Gramacho
Centro de Pesquisas do Cacau, CEPLAC, Ilhéus, BA, Brazil

F. Micheli
CIRAD-BIOS, UMR DAP, Montpellier, France

A. S. Gesteira
Embrapa Mandioca e Fruticultura Tropical, Cruz das Almas, BA, Brazil

Introduction

Bixa orellana L. (known as annatto or achiote) is a perennial tree species, native to the tropical Americas, which produces a wide variety of secondary metabolites through its tissues, including several carotenoid derivatives, terpenoids, tocotrienols, arenes, and flavonoids [1–6]. *B. orellana* is economically important as it is the sole source of the natural pigment bixin, an apocarotenoid extracted from the pericarp of the seeds and widely used as a soluble color additive in foods and cosmetics since the pre-Colombian times [7]. Bixin ranks second after caramel as the most used natural dye in the industry [8].

Large-scale partial sequencing of anonymous cDNA clones and the subsequent identification of their functions through homology searches in public databases is a powerful tool to clarify biosynthetic pathways and their associated genes. This approach, usually referred to as Expressed Sequence Tags (EST) analysis, has been extensively applied to understanding the biochemistry of natural products, such as fatty acids [9], terpenes [10], phenylpropanoids [11] and flavonoids [12], in organs of different plant species. These works showed that partial cDNA sequences, i.e. ESTs, can be successfully used to identify putative clones for a wide range of gene products related to both primary and secondary metabolisms.

Notwithstanding its economic importance, the transcriptome of *B. orellana* remains poorly characterized, with only 870 ESTs reported from a subtracted cDNA library of developing seeds [13]. Further characterization of ESTs would provide a quick, low-cost, and efficient way to expand the pool of genes associated to seed development and metabolism in *B. orellana*. In this work we report the analysis of 792 anonymous clones from a non-normalized cDNA library from developing seeds of the commercial cv. 'Bico-de-Pato'.

Materials and methods

Plant material

Samples of seeds at different developmental stages (14, 28, 42 and 63 days after anthesis) (Fig. 1) were collected from 14-year-old annatto plants (*Bixa orellana* L. cv. 'Bico-de-Pato') grown at the Federal University of Viçosa, Brazil. After harvesting, the samples were immediately immersed in RNA*later*™ RNA stabilization reagent (Qiagen, USA), incubated overnight at 4°C, and then maintained at −20°C for future use.

cDNA library construction and sequencing

RNA was isolated by using the method described by Rodrigues et al. [14]. RNA samples were then treated with DNase I (Amersham Biosciences, EUA) and later purified using the QiaGen Rneasy Mini Protocol RNA Cleanup Kit (Qiagen, USA). The purity and concentration of the isolated RNA were checked by electrophoresis in 1% (w/v) agarose gel and spectrophotometry at 260 and 280 nm.

RNA samples from the different developmental stages were pooled and used for cDNA library construction using the Creator™ SMART™ cDNA Library Construction Kit (Clontech, USA), according to the manufacturer's instructions. The cDNAs were cloned into ρGEM-T Easy vector (Promega Co., USA) under random orientation and the recombinant plasmids were subsequently introduced into *Escherichia coli* ElectroMAX DH10B competent cells (Invitrogen, USA) by electroporation. Successful recombinants were obtained by screening white colonies of electroporated *E. coli* grown on Luria–Bertani (LB) medium [15] supplemented with 100 μg/ml ampicillin, 20 μg/μl IPTG and 2 μg/μl Xgal. Quality control of cDNA insertion was performed by PCR amplifications with M13 forward (5-GT AAAACGACGGCCAGT-3′) and reverse (5′-CAGGAAAC AGCTATGAC-3′) primers of randomly selected clones. The size of amplified DNA fragments was checked on 1% (w/v) agarose gel stained with ethidium bromide (0.5 μg/ml).

Individual white colonies were selected and incubated overnight at 37°C and 260 rpm orbital agitation in 96 wells plates containing LB supplemented with ampicillin (100 μg/ml). Plasmid DNA was extracted by using the alkaline lysis procedure [15]. The supernatant was cleared off from protein and cellular debris by filtration (0.22 μm) prior to alcohol precipitation.

Sequencing reactions were performed on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, USA), using

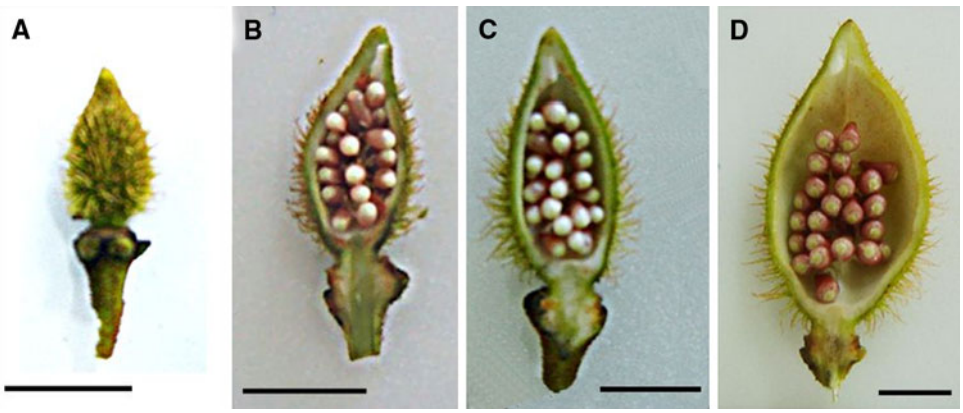


Fig. 1 Fruits and seeds of *B. orellana* at 14 (**a**), 28 (**b**), 42 (**c**), and 63 (**d**) days after anthesis (daa). Bar = 10 mm

the Big Dye Terminator DNA Sequencing Kit (Applied Biosystems, USA) and the M13 reverse primer.

Sequence analysis and annotation

ESTs were clustered and assembled into contigs using the program SEQMAN from the laser-gene package (DNAstar, USA). ESTs were annotated by reference to Gene Ontology (GO) and to the 'nr' section of GenBank using BLAST2GO (http://blast2go.bioinfo.cipf.es/) [16] and BLASTX [17], respectively. Homologies with expected-values $\leq$1E−3 were considered significant. GO annotations were divided into categories: Biological Process (BP), Cellular Component (CC) and Molecular Function (MF). The annotation was improved by comparing the sequences against the InterPro domains database using InterProScan [18] at the EBI Server.

Sequences encoding plant carotenoid cleavage dioxygenases (CCDs) and O-methyltransferases (OMTs) were identified and compared to the 'nr' section of the GenBank. For these two gene families, dendrograms were generated using ClustalW [19] with Neighbor-Joining [20] and a bootstrap of 1000.

Quantitative real-time PCR (qPCR)

The expression pattern of *B. orellana* CCD (BoCCD) and OMT (BoOMT) genes was investigated during different stages of seed development. Total RNA isolation and treatment was carried out as previously described and the first strand cDNA was synthesized by using the First Strand cDNA Synthesis Kit (Fermentas, USA). The relative abundance of the mRNAs was quantified using ABI Prism 7500 and Sequence Detection System 1.6.3 (Applied Biosystems, USA). The *B. orellana* gene encoding for 18S rRNA (Bo 18S rRNA) was selected as endogenous control. Primers of the two target genes and endogenous control were designed according to their expressed sequence tags (ESTs) in our library, using the Primer Express Software, version 2 (Applied Biosystems, USA). These are listed in Electronic supplementary material (ESM) (Table S1).

All the real-time PCR reactions triplicates included: 75–100 ng of cDNA, 200 nM of each primer, 2 µl SYBR Green 1 × (Applied Biosciences, USA), 1 × PCR buffer, 200 mM dNTPs, 3 mM $MgCl_2$, ½ 50 × Rox, 1U Platinum® Taq DNA Polymerase (Invitrogen, USA), and sterile Milli-Q water for a final volume of 20 µl. The following steps were used in the qPCR reactions: (1) activation of Taq DNA polymerase at 50°C for 2 min, (2) initial denaturation at 95°C for 10 min, (3) denaturation at 95°C for 15 s, (4) annealing at 60°C for 30 s and (5) extension at 60°C for 1 min. Steps 3–5 were repeated for 40 cycles. Non-template control (NTC) reactions were also performed in triplicates for each pair of primers, and two biological repeats were used in all experiments. To verify that only a single PCR product was generated for the amplified transcript, the multicomponent data for each sample was subsequently analyzed using Dissociation Curve 1.0 (Applied Biosystems, USA). The relative gene expression levels were obtained by relative quantification (RQ) according to the $2^{-\Delta\Delta Ct}$ method [21]. A non-parametric *t*-test was performed in order to compare the expression level of each gene at each different stage of seed development.

Results

EST library construction

'Bico-de-Pato' is an accession of *B. orellana* that produces a high number of capsules per inflorescence and seeds per capsules. The bixin content per seed is $\geq$2.8% [22]. The standard EST library was designed to represent the main characteristics of interest in this study, namely seed yield and color. The library covered all pivotal developmental stages of the seeds. This strategy was planned to obtain the widest representation of the *B. orellana* seed transcriptome, in order to make the identification of genes of agronomic interest easier. The library contained an estimated insert size distribution ranging from 500 to 2000 bp (data not shown).

EST sequencing and clustering

A total number of 792 clones were sequenced. After low quality and vector trimming of raw sequences, 651 sequences larger than 100 bp were assembled into 82 contigs containing from 2 to 58 sequences each; 42 sequences were singlets. About 87% of contigs had less than 10 ESTs, and 29 of these had only two sequences. Twelve percent of the contigs had 10 to 25 sequences and only 1% had 58 sequences (ESM Fig. S1). The sample size of the final unigene set resulting from the assembling step was 124 sequences. About 68% of the ESTs were between 300 and 599 pb (ESM Fig. S2). Only 3% of the ESTs were smaller than 200 bp. The rate of redundancy calculated as [1-(number of unigenes/number of ESTs)] × 100 was ~81%.

EST annotation and functional classification

Gene Ontology (GO) annotation of ESTs was performed with BLAST2GO software (http://blast2go.bioinfo.cipf.es/) [16], which assigns GO annotations through a three-step procedure: BLAST against protein databases, retrieval of all GO annotations for a specified number of BLAST hits

(Mapping), and GO assignment through an evaluated annotation rule (Annotation). The rate of BLASTX hits was found to be 649 among 651, i.e. 99.7%, with 464 (71.5%) being significant at $E < 0.001$ (ESM Fig. S3) and ranging from 1E−4 to 1E−105. More than 80% of the hits showed E-values higher than 1E−10 (ESM Fig. S4) and half of them (45.5%) were between 1E−19 and 1E−10.

About 70% (326) of the ESTs showing a significant hit were mapped, and 95% (310) of them were later annotated (ESM Fig. S3). The rate of no-hit (Not blasted) was 2 ESTs (0.3%), the absence of GO annotation (No mapping) accounted for 138 ESTs (29.7%), and the sequence that did not fulfill quality parameters of trustable annotation (Not annotated) accounted for 16 ESTs (4.9%). From the sequences with a significant BLAST hit without GO term assignment, 6.4% were hypothetical (1.3%) or predicted proteins (5.1%), indicating the uncertainty in the functional characterization of the hit sequences.

A total of 651 ESTs were annotated with 1167 terms, distributed among the main Gene Ontology categories: Biological Process (47.5%), Molecular Function (79%) and Cellular Component (52.8%). There were 189 (29%) sequences annotated in all three GO categories, and 55 (8.4%) had at least two annotations. The most expressive defined categories in Biological Processes were related to proline hydroxylation (60 ESTs; 9.2%), regulation of transcription (49 ESTs; 7.5%), translation and post-translational modifications (48 ESTs; 7.4%), defense response (43 ESTs, 6.6%), response to biotic and abiotic stimulus (28 ESTs; 4.3%), ribosome biogenesis (22 ESTs, 3.4%), biosynthesis of fatty acids (19 ESTs, 2.9%), and oxidation reduction (14 ESTs, 2.2%) (Table 1). Other processes represented included hydrogen peroxide catabolic process (8 ESTs, 1.2%), photosynthesis and light harvest (5 ESTs, 0.8%), and aerobic respiration (4 ESTs, 0.6%).

The Molecular Function describes the biochemical activity carried out by the gene product. Some of the most expressive categories included iron ion binding (68 ESTs, 10.4%), oxidoreductase activity (61 ESTs, 9.4%), protein binding (46 ESTs, 7.1%), and transcription factor activity (43 ESTs, 6.6%) (Table 2).

Table 1 Functional classification according to the category of Biological Process in GO

Biological Process	Number of ESTs	Fraction of ESTs (%)
Peptidyl-proline hydroxylation to 4-hydroxy-L-proline	60	9.2
Regulation of transcription	49	7.5
Translation and post-translational protein modification	48	7.4
Defense response	43	6.6
Response to biotic and abiotic stimulus	28	4.3
Ribosome biogenesis	22	3.4
Fatty acid biosynthetic process	19	2.9
Oxidation reduction	14	2.2
Hydrogen peroxide catabolic process	8	1.2
tRNA aminoacylation for protein translation	6	0.9
Xenobiotic metabolic process	6	0.9
Photosynthesis, ligth harvesting	5	0.8
Aerobic respiration	4	0.6
Copper ion transport	4	0.6
Mitocondrial electron transport, cytochrome c to oxygen	4	0.6
Protein-chromophore linkage	4	0.6
Cytolysis	3	0.5
Transcription	3	0.5
DNA replication and repair	2	0.3
Intracellular protein transport	2	0.3
Superoxide metabolic process	2	0.3
ATP synthesis coupled proton transport	1	0.2
Glutamine biosynthetic process	1	0.2
Lignin biosynthetic process	1	0.2
Lipid metabolic process	1	0.2
Nitrogen fixation	1	0.2
Pathogenesis	1	0.2
Proteolysis	1	0.2
Regulation of protein metabolic process	1	0.2
Transport	1	0.2
Tricarboxylic acid cycle	1	0.2

Abundant genes in the EST library

The most frequent transcripts of the cDNA library are shown in Table 3. They are about 43% of the total number of annotated ESTs. Transcripts frequently found in the library were associated to genes from (i) cell wall (prolyl 4-hydroxylase), fatty acid (acyl carrier protein), and hormone/flavonoid (2OG-Fe oxygenase family) synthesis, (ii) germination (MADS FLC-like protein 3) and embryo development (AP2/ERF transcription factor) regulation, (iii) photosynthesis (chlorophyll a-b binding protein), (iv) cell elongation (MAP65-1a) and (v) stress and defense responses (metallothionein- and thaumatin-like proteins). For example, prolyl 4-hydroxylase displayed a redundancy of 60 ESTs (9.2%). The highest redundancy was 97 ESTs (14.9%) and was observed for the function "protein", which in most cases did not match any known protein motif by InterProScan analysis. The functions of these proteins still need clarification.

Table 2 Functional classification according to the category Molecular Function in GO

Molecular Function	Number of ESTs	Fraction of ESTs (%)
Iron ion binding	68	10.4
Oxidoreductase activity	61	9.4
Protein binding	46	7.1
Transcription factor activity	43	6.6
Structural constituent of ribosome	26	4.0
Cofactor binding	19	2.9
Phosphopantetheine binding	19	2.9
Acyl carrier activity	19	2.9
ATP binding	10	1.5
DNA binding	10	1.5
Catalase activity	8	1.2
Metal ion binding	8	1.2
Aminoacyl-tRNA ligase activity	6	0.9
Translation initiation factor activity	6	0.9
Sequence-specific DNA binding	4	0.6
Chlorophyll binding	4	0.6
Copper ion transmembrane transporter activity	4	0.6
Cytochrome-c oxidase activity	4	0.6
Magnesium ion binding	4	0.6
RNA binding	3	0.5
Zinc ion binding	3	0.5
Helicase acitivity	2	0.3
Nucleic acid binding	2	0.3
Superoxide dismutase activity	2	0.3
Acyltransferase activity	1	0.2
Caffeate O-methyltransferase activity	1	0.2
Carotenoid-cleavage dioxigenase 4b activity	1	0.2
DNA ligase (ATP) activity	1	0.2
Glutamate-ammonia ligase activity	1	0.2
GTP binding	1	0.2
Heat shock protein binding	1	0.2
Hydrogen ion transporting ATP synthase activity	1	0.2
Hydrolase activity, acting on ester bonds	1	0.2
Isocitrate dehydrogenase (NAD+) activity	1	0.2
Peptide binding	1	0.2
Peptidyl-prolyl cis-trans isomerase activity	1	0.2
Protein dimerization activity	1	0.2
Protein serine/threonine kinase activity	1	0.2
Protein transmembrane transporter activity	1	0.2
Proton-transporting ATPase activity	1	0.2
RNA polymerase II transcription factor activity	1	0.2
Serine-type peptidase activity	1	0.2
Signal peptidase activity	1	0.2
Small conjugating protein ligase activity	1	0.2
Sulfotransferase activity	1	0.2
Transporter activity	1	0.2
tRNA binding	1	0.2

Table 3 Protein functions and EST frequencies in the cDNA library of seeds from *Bixa orellana*

Putative function	EST sequences (%)	E-value
Protein	97 (14.9%)	2.56E−34
Prolyl 4-hydroxylase	60 (9.2%)	6.74E−47
60 s ribosomal protein l38	25 (3.8%)	5.44E−16
Acyl carrier protein	18 (2.8%)	5.25E−08
MADS FLC-like protein 3	10 (1.5%)	7.92E−07
2OG-Fe oxygenase family protein	9 (1.4%)	2.12E−45
Chloroplast initiation factor 3	6 (0.9%)	8.36E−11
Microtubule-associated protein MAP65-1a	6 (0.9%)	2.53E−55
AP2/ERF domain-containing transcription factor	5 (0.8%)	1.91E−07
Chlorophyll a-b binding protein	5 (0.8%)	5.02E−09
Conserved hypothetical protein [Ricinus communis]	5 (0.8%)	8.43E−07
Metallothionein-like protein	5 (0.8%)	2.89E−17
Pathogenesis-related thaumatin-like protein	5 (0.8%)	2.27E−84
Thaumatin-like protein	5 (0.8%)	2.81E−85
Zf-hd protein dimerisation region containing protein	5 (0.8%)	9.95E−19
Copper transporter	4 (0.6%)	3.85E−38
Cytochrome c oxidase subunit vb	4 (0.6%)	9.31E−49
Predicted protein [Populus trichocarpa]	4 (0.6%)	6.64E−20
Total	278 (42.8%)	

Genes related to metabolic pathways

To identify the *B. orellana* sequences that putatively encode for structural enzymes of metabolic pathways, the Enzyme Commission (EC) numbers of the predicted proteins were determined. This analysis indicated the occurrence of three main classes of enzymes: oxidoreductases (EC:1) (74 ESTs), transferases (EC:2) (101 ESTs) and

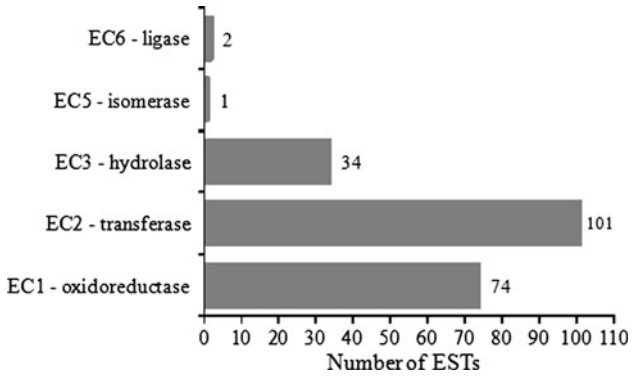


Fig. 2 EC classes detected among the ESTs from the cDNA library

hydrolases (EC:3) (34 ESTs) (Fig. 2). Based on the EC number, it was possible to characterize each putative enzyme coded by an EST sequence, defining to which metabolic pathway it belongs. EC numbers allowed the assignment of 16 metabolic pathways, including those related to seed yield (nitrogen metabolism) and synthesis of secondary metabolites (flavonoids, phenylpropanoids, alkaloids, and terpenoids) (Table 4).

Analysis of genes related to secondary metabolism

Two candidate genes of the bixin biosynthetic pathway included in the EST collection were selected and further analyzed: (i) a carotenoid-cleavage dioxygenase (CCD) that belongs to a rapidly growing family that cleaves carotenes and xanthophylls to apocarotenoids and (ii) a caffeate O-methyltransferase, which is a member of the distinct family of O-methyltransferases (OMTs) involved in the transfer of methyl groups from S-adenosyl-L-methionine to hydroxyl or carboxyl groups of a wide range of acceptor molecules to generate methyl ether derivatives and that includes lignin, flavonoids, caffeine, bixin and floral scent components.

Multiple alignments of plant CCD and OMT sequences were performed with ClustalW, and dendrograms were constructed with the Neighbor Joining method. The dendrogram of 19 plant CCDs (Fig. 3a) shows that the BoCCD analyzed is closely related to the members of the CCD subclass 4 (CCD4), which includes enzymes involved in the cleavage of β-carotene (CmCCD4b and MdCCD4), lycopene (BoLCD), and zeaxanthin (CsZCD). Considering the dendrogram of nine plant OMTs (Fig. 3b), that of *B. orellana* (BoOMT) was grouped together with the caffeic acid 3-O-methyltransferase (EgCOMT) from *Eucalyptus gunnii*, as a sister clade of 3′ flavonoid O-methyltransferase (CaFOMT) from *Chrysosplenium americanum* and caffeic acid O-methyltransferases 1 (RcOMT1) and 2 (RcOMT2) from *Rosa chinensis*. This shows that BoOMT is a member of the caffeic acid O-methyltransferase family.

Table 4 Metabolic pathways assigned by the Enzyme Comission (EC) number

Pathway	Enzyme	EC number
Nitrogen metabolism	Glutamine synthetase	6.3.1.2
Peptidoglycan biosynthesis	Glutamine synthetase	6.3.1.2
Arginine and proline metabolism	Glutamine synthetase	6.3.1.2
Alanine, aspartate and glutamate metabolism	Glutamine synthetase	6.3.1.2
Flavonoid biosynthesis	Caffeate O-methyltransferase	2.1.1.68
Biosynthesis of phenylpropanoids	Caffeate O-methyltransferase	2.1.1.68
	Isocitrate dehydrogenase (NAD+)	1.1.1.41
Biosynthesis of alkaloids derived from shikimate pathway	Caffeate O-methyltransferase	2.1.1.68
	Isocitrate dehydrogenase (NAD+)	1.1.1.41
Biosynthesis of alkaloids derived from terpenoid and polyketide	Isocitrate dehydrogenase (NAD+)	1.1.1.41
Biosynthesis of alkaloids derived from histidine and purine	Isocitrate dehydrogenase (NAD+)	1.1.1.41
Biosynthesis of alkaloids derived from ornithine, lysine and nicotinic acid	Isocitrate dehydrogenase (NAD+)	1.1.1.41
Biosynthesis of terpenoids and steroids	Isocitrate dehydrogenase (NAD+)	1.1.1.41
Tryptophan metabolism	Catalase	1.11.1.6
Photosynthesis	H^+-transporting two-sector ATPase	3.6.3.14
Oxidative phosphorylation	H^+-transporting two-sector ATPase	3.6.3.14
	Cytochrome-c oxidase	1.9.3.1
Biosynthesis of plant hormones	Isocitrate dehydrogenase (NAD+)	1.1.1.41
Citrate cycle	Isocitrate dehydrogenase (NAD+)	1.1.1.41

To ascertain the importance of both genes in the seed metabolism of *B. orellana*, their expression profiles were also analyzed during the distinct stages of seed development by quantitative real-time PCR. We found that *BoCCD* and *BoOMT* were strongly up-regulated during seed development (Fig. 4). There was a sharp increase in the mRNA levels of both genes at 42 days after anthesis (daa), reaching maximum levels at 63 daa. This induction of expression correlates with the concomitant increase in seed pigmentation and fruit diameter and length. The expression level of *BoOMT* was higher than *BoCCD* in all the different stages of seed development.

Discussion

EST sequencing has been used as a valuable tool in genomic analysis. A number of EST collections from different plant tissues and species are available in public databases. Searches in these databases are a quick way to identify, by similarity, genes of potential interest. However, in the case of *B. orellana*, there are a limited number of deposited sequences in public databases. To get a genome-wide overview of the repertoire of *B. orellana* genes, we have constructed a standard non-normalized cDNA library from developing seeds, covering all their pivotal developmental stages. A total of 792 independent clones were randomly selected and sequenced. These cDNA clones are publicly available from the corresponding author upon request. Overall, 651 ESTs were obtained, yielding 124 putative unique transcripts. Besides of the high redundancy (81%) of the library, it was possible to identify several sequences of interest.

The Gene Ontology annotation of the *B. orellana* unigenes was performed with BLAST2GO (B2GO) [16]. It uses multiple BLAST hits to search for functional annotations and assigns GO terms to the query sequence by applying an annotation algorithm that considers HSP length, e-value, percentage of similarity, Evidence Code of the source annotations and the topology of the Gene Ontology. This represents an important contribution to the EST projects, which in the past years used to perform annotation solely by direct assignment of the GO terms to the best hit of BLAST searches. The B2G method has shown to have a high annotation recall and has been used in other EST projects related to plants [23, 24].

The homology searches performed against different databases allowed annotation of most of the ESTs, with 72% of them displaying some similarity degree. These results are in agreement with those obtained in many EST projects [24–26]. It was observed that 28% of the sequences did not produce significant hits in the BLASTX searches. These sequences were in average 414-bp long. Additional efforts to characterize these sequences did not yield any success, suggesting that the 185 ESTs of *B. orellana* might correspond to *Bixa* exclusive genes since there were no significant hits with more than 8.5 million EST sequences derived from plant species deposited at the GenBank.

The functional classification according to the GO vocabulary showed that the genes were distributed among the three main categories: Biological Process (BP), Molecular Function (MF) and Cellular Component (CC). The fact that 79% of the sequences annotated were distributed in the MF category reveals a major role of this class of genes in the cellular and biochemical events occurring during seed development in *B. orellana*. Transcription factor activity (6.6%), binding (31.5%) and oxidoreductase activity contributed together with 47.5% of the ESTs assigned to this category.

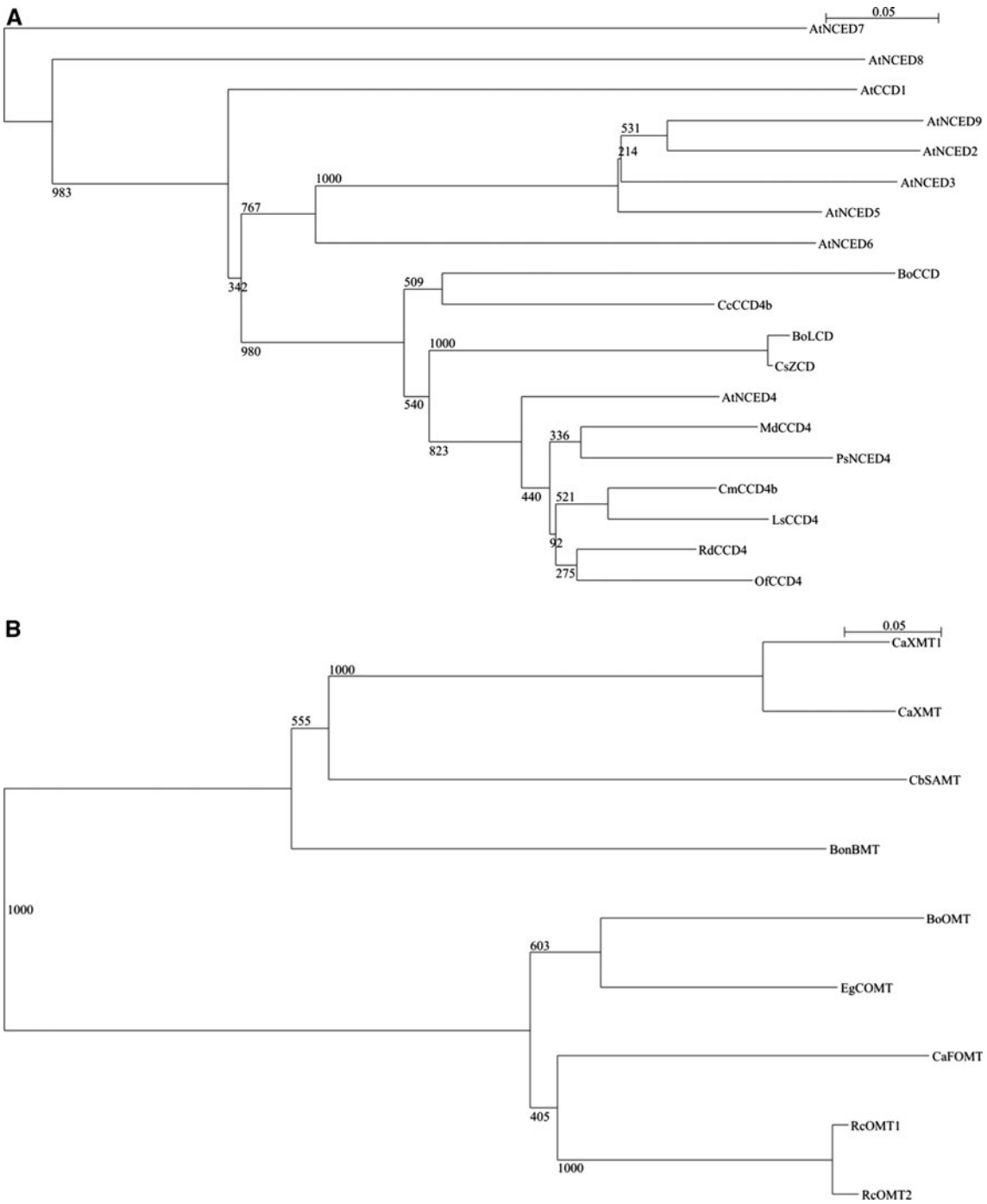
Analysis of the transcripts most frequently found in our non-normalized cDNA library provided insights about the main developmental and metabolic processes occurring during seed development in *B. orellana*. These transcripts included those encoding prolyl 4-hydroxylase, acyl carrier, MADS FLC-like 3, 2OG-Fe oxygenase, AP2/ERF transcription factor, chlorophyll a-b binding, MAP65-1a, metallothionein-like, and thaumatin-like proteins. The majority of these proteins are the first representatives of their gene families in *B. orellana*. Prolyl 4-hydroxylases (P4Hs) are 2-oxoglutarate dioxygenases that catalyze the hydroxylation of peptidyl prolines, playing an important role in the synthesis of cell-wall glycoproteins, such as extensins, hydroxyproline rich glycoproteins (HRGPs), lectins, and arabinogalactan proteins [27, 28]. These glycoproteins are involved in various aspects of plant growth and development, ranging from wall architecture and assembly to cell proliferation, cell-to-cell recognition, and cell expansion [29]. Acyl carrier protein (ACP) is a small acidic protein involved as an essential co-factor in fatty acid synthesis, acyl-ACP desaturation and plastidial acyltransferase reactions [30]. This protein contains a phosphopantetheine prosthetic group that forms a thioester bond with fatty acids, resulting in activation of the carboxyl carbon of the acyl group. Fatty acids are widely found as major carbon and chemical energy reserve in seeds [31]. As a MADS-box transcription factor (so named for *MCM1*, *AGAMOUS*, *DEFICIENS*, and serum-response factor), *FLOWERING LOCUS C* (FLC) is a major regulator of flowering and seed germination responses to seasonal environmental factors [32]. The increase of MADS FLC expression may prevent both flowering and dormancy, in a temperature-dependent process. It has been recently proposed that a high level of MADS FLC expression during seed maturation would allow seeds to overcome primary dormancy during imbibition, most likely following a sequence involving ABA decrease

followed by GA-promoted germination [32]. 2OG-Fe oxygenase is an oxidoreductase of the 2-oxoglutarate (2OG)-Fe(II) oxygenase superfamily, which is widespread in bacteria and eukaryotes [33]. In plants, these enzymes catalyze different hydroxylation and desaturation steps such as those involved in the biosynthesis of flavonoids, cathechins, anthocyanidins [34], gibberellins [35] and ethylene [36]. AP2/ERF proteins are defined by the presence of the AP2/ERF domain, which consists of about 60–70 amino acids involved in DNA binding [37]. It has been demonstrated that the AP2/ERF proteins have important functions in the transcriptional regulation of a variety of biological processes, including embryo development [38]. The functional analysis of *BABY BOOM* (*BBM*), an AP2/ERF domain-containing protein of Arabidopsis, showed that it activates signal transduction pathways leading to the induction of embryo development from differentiated somatic cells, thereby acting as a key regulator of embryo development in plants [38]. MAP65 is a multigene family of 65 kDa microtubule-associated proteins specific to plants [39]. These proteins are involved in the maintenance of the spindle midzone, a structure that is formed between daughter chromatids during anaphase B and is essential to the mitosis completion [40]. MAP65s were reported to be involved in cell elongation of bean epicotyls [41] and carrot suspension cells [39]. Metallothioneins (MTs) are a superfamily of ubiquitously expressed low molecular mass (6–7 kDa) cysteine-rich proteins found in a wide range of organisms, from plants to fungi, with high capacity to form metal-thiolate clusters with copper, cadmium, zinc, mercury and silver [42, 43]. Several functional roles have been attributed to MTs, including heavy metal detoxification, zinc and copper homeostasis [44, 45], scavenging of reactive oxygen species [46], regulation of metalloenzymes and transcription factors [44], and response to stress conditions [47, 48]. Thaumatin-like proteins (TLPs) belong to the Group 5 of the pathogenesis related (PR-5) family commonly found both in monocotyledonous and dicotyledonous species [49]. All members of this PR group are homologous to thaumatin, a sweet-tasting protein isolated from *Thaumatococcus daniellii* fruits [50]. They exhibit antifungal activity, which is correlated with plasma membrane permeabilization [51, 52] and can also act as protease and reverse transcriptase inhibitors [53]. TLPs may be part of the complex defensive system that protects the nutrient-rich seeds of *B. orellana* against microbial growth, as reported in chestnut seeds [54].

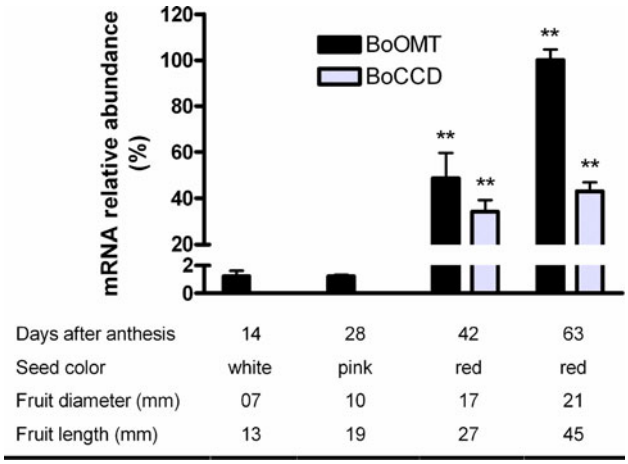
The investigation of the EC number associated to predicted proteins revealed that the functions transferases, oxidoreductases and hydrolases are the three main classes of enzymes present in *B. orellana* seeds. These enzymes were associated to 16 different metabolic pathways related to both primary and secondary metabolisms (Table 4). Metabolic pathways involved in seed yield and synthesis of secondary metabolites, the targets of this study, were present. For instance, glutamine synthase (GS) catalyzes the ATP-dependent assimilation of ammonium into glutamine using glutamate as a substrate. This reaction is now considered to be the major route facilitating the incorporation of inorganic nitrogen (ammonium) into organic molecules (e.g., glutamine and glutamate) and, thereby, controlling plant growth and productivity [55]. In fact, recent findings demonstrated that GS can control seed number and seed size in a number of different plant species [55–58]. This enzyme could play a role in the control of seed number in 'Bico-de-Pato'.

The genes related to the synthesis of secondary metabolites, CCD (*BoCCD*) and caffeate O-methyltransferase (*BoOMT*), were object of a deeper analysis in the present study. The dendrogram analysis showed that BoCCD has a strong similarity to many other CCD4 proteins. CCD4 is a member of the CCD family, which in *A. thaliana* has nine members (CCD1, 4, 7, 8 and NCED2, 3, 5, 6, and 9) that are used to name orthologous proteins in other plant species. The biological function of CCD4 in plants has been elucidated only recently. It has been demonstrated that there are at least two isoforms encoded by CCD4 genes (*CCD4a* and *CCD4b*) [59, 60]. CCD4 of *A. thaliana* has been observed to be located in plastoglobules, where it could play a role in the dark-induced breakdown of carotenoids [61]. RNA interference (RNAi) showed that *CCD4a* from *Chrysanthemum* × *morifolium* is involved in

Fig. 3 Dendrogram analysis of amino acid sequences of plant CCDs (**a**) and OMTs (**b**). Amino acid sequences were aligned using Clustal W and a neighbor-joining tree was constructed using a bootstrap of 1000. The GenBank accession numbers for the sequences are as follows: AtCCD1 (*Arabidopsis thaliana* CCD1, NP_191911), AtNCED2 (*Arabidopsis thaliana* NCED2, NP_193569), AtNCED3 (*Arabidopsis thaliana* NCED3, NP_188062), AtNCED4 (*Arabidopsis thaliana* NCED4, NP_193652), AtNCED5 (*Arabidopsis thaliana* NCED5, NP_174302), AtNCED6 (*Arabidopsis thaliana* NCED6, NP_189064), AtCCD7 (*Arabidopsis thaliana* NCED7, NP_182026.1), AtNCED8 (*Arabidopsis thaliana* NCED8, NP_195007.2), AtNCED9 (*Arabidopsis thaliana* NCED9, NP_177960), BoLCD (*B. orellana* lycopene cleavage dioxygenase, AJ489277), BonBMT (*B. orellana* norbixin methyltransferase, AJ548847), CaFOMT (*Chrysosplenium americanum* 3′ flavonoid O-methyltransferase, AAA80579.1), CaXMT1 (*Coffea arabica* xanthosine methyltransferase 1, BAC43756), CaXMT (*Coffea arabica* xanthosine methyltransferase, BAB39215), CbSAMT (*Clarkia breweri* salicylic acid methyltransferase, AAF00108), CcCCD4b (*Citrus clementina* CCD4b, ABC26012), CmCCD4b (*Chrysanthemum* × *morifolium* CCD4b, BAF36656), CsZCD (*Crocus sativus* zeaxanthin cleavage dioxygenase, AJ489276), EgCOMT (*Eucalyptus gunnii* caffeic acid O-methyltransferase, P46484.1), LsCCD4 (*Lactuca sativa* CCD4, BAE 72094), MdCCD4 (*Malus* × *domestica* CCD4, ABY47995), OfCCD4 (*Osmanthus fragans* CCD4, ABY60887), PsNCED4 (*Pisum sativum* NCED4, BAC10552), RcOMT1 (*Rosa chinensis* caffeic acid O-methyltransferase 1, Q8GU25), RcOMT2 (*Rosa chinensis* caffeic acid O-methyltransferase 2, BAC78827.1), and RdCCD4 (*Rosa* × *damascena* CCD4, ABY60886)



the white color formation in petals by cleaving carotenoids into colorless compounds [59]. *Crocus sativus* CCD4 has been reported to cleave β-carotene at the 9,10 (9′,10′) positions to yield β-ionone [62]. In a comprehensive analysis involving CCD4 from different plant species, it has been recently shown that the recombinant proteins oxidatively cleave their substrates at the same position (9,10 and 9′,10′) but they might have different biochemical and biological functions as they accept different (apo)-carotenoid substrates and show various expression patterns [60]. *Chrysanthemum* × *morifolium* (*CmCCD4a*) and *Malus* × *domestica* (*MdCCD4*) CCD4 accepted the C_{40}-carotenoid β-carotene as substrate, while *A. thaliana* (AtCCD4) and *Rosa* × *damascena* (RdCCD4) CCD4 preferentially cleaved the C_{30}-apocarotenoid substrate 8′-apo-β-caroten-8′-al. *CmCCD4a, MdCCD4* and *RdCCD4* have flower specific expression patterns, while *AtCCD4* is expressed during petal differentiation and anthesis. Besides its



Days after anthesis	14	28	42	63
Seed color	white	pink	red	red
Fruit diameter (mm)	07	10	17	21
Fruit length (mm)	13	19	27	45

Fig. 4 Quantitative RT-PCR analysis of *BoCCD* and *BoOMT* expression in seeds of *Bixa orellana* at 14, 28, 42, and 63 days after anthesis (daa). The levels of expression were normalized to the corresponding value of 18S rRNA. The mRNA accumulation was normalized according to the expression of *BoOMT* at 63 daa. Data are means ± SE of three experimental replicates. ** Significantly different from 14 daa according to Student's *t*-test ($P < 0.01$)

similarity with *B. orellana* LCD (BoLCD) [63] and *Crocus sativa* ZCD (CsZCD) [64], as showed in our dendogram, CCD4 does not cleave linear carotenoids such as lycopene, or carotenoids containing hydroxyl group such as zeaxanthin [60]. Our data on expression analysis revealed that *BoCCD* was significantly induced during seed development, which correlated with the massive accumulation of apocarotenoids accompanying seed maturation [65]. These results strongly suggest that BoCCD plays a key role in the accumulation of apocarotenoids by cleaving carotenes and/or xanthophylls in *B. orellana* seeds.

BoOMT sequence was found closer to EgCOMT, CaFOMT, RcOMT1 and RcOMT2 than to the *B. orellana* norbixin methyltransferase (BonBMT) [63]. The OMTs form a distinct family in which the caffeic acid OMT (COMT) and caffeoyl CoA OMT (CCOMT) are the most studied. COMT methylates caffeic acid/5-hydroxyferulic acid, whereas CCOMT methylates CoA ester [66]. Both enzymes are involved in the biosynthesis of lignin, but it has been reported that a COMT from *R. chinensis* (RcOMT1) is highly expressed in floral organs and efficiently methylates eugenol and isoeugenol to yield the volatile floral scents methyleugenol and isomethyleugenol, respectively, while RcOMT2 does not show any particular tissue specificity and exhibits the highest activity with caffeic acid [67]. Flavonoid and isoflavonoid OMTs involved in the biosynthesis of phytoalexins [68, 69], salicylic acid OMT involved in the biosynthesis of floral scent and methyl salicylate (CbSAMT) [70], norbixin OMT involved in the biosynthesis of norbixin (BonBMT) [63] and several OMTs involved in the biosynthesis of caffeine (CaXMT) [71, 72] have also been identified and characterized. Due to the high variability in function with minimal sequence divergence, the function of the BoOMT identified in our EST library remains to be unveiled. Nevertheless, its up-regulation during seed maturation provides strong evidence of its involvement in the biosynthesis of phenylpropanoid- and/or flavonoid-methylated derivatives, which are present in mature seeds of *B. orellana* [2, 5].

In summary, we have characterized a set of ESTs from a non-normalized cDNA library of *B. orellana* that provided information on the genes involved in the main developmental and metabolic processes taking place during seed maturation. *BoCCD* and *BoOMT*, two genes candidate genes of the bixin biosynthetic pathway included in the EST collection, were showed to belong to the carotenoid-cleavage dioxygenase 4 (CCD4) and caffeic acid O-methyltransferase (COMT) families, respectively. The expression analysis showed that both genes are highly induced during seed development, which indicates their involvement in the synthesis of this important carotenoid pigment in seeds of *B. orellana*. The functional analysis of these genes is now underway in order to clarify their precise roles in the synthesis of bixin in *B. orellana*.

Acknowledgments This research was supported by BNB (Research grant no. 2004-1-22 to MGCC) and CNPq (Research grant no. 473619/04-04 to WCO). We gratefully acknowledge the PhD scholarships to V.L.F. Soares and S.M. Rodrigues by FAPESB (Fundação de Amparo à Pesquisa do Estado da Bahia) and FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais), respectively. T.M de Oliveira and T.O. de Queiroz are recipients of PIBIC/CNPq and IC/FAPESB fellowships, respectively. We thank Dr. Claudia Fortes Ferreira (Embrapa, Brazil) and Dr. Nicolas Carels (Fiocruz, Brazil) for helpful advice and critical reading of the manuscript.

References

1. Lawrence BM, Hogg JW (1973) Ishwarane in *Bixa orellana* leaf oil. Phytochemistry 12:2995
2. Harbone JB (1975) Flavonoid bisulfates and their co-occurrences with ellagic acid in the Bixaceae, Frankeniaceae, and related families. Phytochemistry 14:1331–1337
3. Jondiko IJO, Pattenden G (1989) Terpenoids and an apocarotenoid from seeds of *Bixa orellana*. Phytochemistry 28:3159–3162
4. Frega N, Mozzon M, Bocci F (1998) Identification and estimation of tocotrienols in the annatto lipid fraction by gas chromatography mass spectrometry. J Am Oil Chem Soc 75:1723–1727
5. Pino JA, Correa MT (2003) Chemical composition of the essential oil from annatto (*Bixa orellana* L.) seeds. J Essent Oil Res 15:66–67
6. Shilpi JA, Taufiq-Ur-Rahman Md, Uddin SJ, Alam Md S, Sadhu SK, Seidel V (2006) Preliminary pharmacological screening of *Bixa orellana* L. leaves. J Ethnopharmacol 108:264–271
7. Giuliano G, Al-Babili S, Von Lintig J (2003) Carotenoid oxygenases: cleave it or leave it. Trends Plant Sci 8:145–149

8. Mercadante AZ, Pfander H (1998) Carotenoids from annatto: a review. Recent Res Devel Agric Food Chem 2:79–91
9. Lu C, Wallis JG, Browse J (2007) An analysis of expressed sequence tags of developing castor endosperm using a full-length cDNA library. BMC Plant Biol 7:42
10. Crowhurst RN, Gleave AP, MacRae EA et al (2008) Analysis of expressed sequence tags from Actinidia: applications of a cross species EST database for gene discovery in the areas of flavor, health, color and ripening. BMC Genomics 9:351
11. Yoshida K, Nishiguchi M, Futamura N, Nanjo T (2007) Expressed sequence tags from *Cryptomeria japonica* sapwood during the drying process. Tree Physiol 27:1–9
12. Hocher V, Auguy F, Argout X, Laplaze L, Franche C, Bogusz D (2006) Expressed sequence-tag analysis in *Casuarina glauca* actinorhizal nodule and root. New Phytol 169:681–688
13. Jako C, Coutu C, Roewer I, Reed DW, Pelcher LE, Covello PS (2002) Probing carotenoid biosynthesis in developing seed coats of *Bixa orellana* (Bixaceae) through expressed sequence tag analysis. Plant Sci 163:141–145
14. Rodrigues SM, Soares VLF, Oliveira TM, Gesteira AS, Otoni WC, Costa MGC (2007) Isolation and purification of RNA from tissues rich in polyphenols, polysaccharides, and pigments of annato (*Bixa orellana* L.). Mol Biotechnol 37:220–224
15. Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Plainview, New York
16. Conesa A, Götz S, García-Gomez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
18. Labarga A, Valentin F, Anderson M, Lopez R (2007) Web services at the European Bioinformatics Institute. Nucleic Acids Res 35:w6–w11
19. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weigh matrix choice. Nucleic Acids Res 22:4673–4680
20. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
21. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. Methods 25:402–408
22. Cruz AC, Costa MGC, Otoni WC (2008) Aspectos gerais da cultura e beneficiamento da semente de urucum. In: Stringheta PC, Silva PI (eds) Pigmentos de urucum: extração, reações químicas, usos e aplicações. Suprema, Viçosa, pp 11–23
23. Ma J, Morrow D, Fernandes J, Walbot V (2006) Comparative profiling of the sense and antisense transcriptome of maize lines. Genome Biol 7:R22
24. Terol J, Conesa A, Colmenero JM et al (2007) Analysis of 13000 unique *Citrus* clusters associated with fruit quality, production and salinity tolerance. BMC Genomics 8:31
25. Horn R, Lecouls AC, Callahan A et al (2005) Candidate gene database and transcript map for peach, a model species for fruit trees. Theor Appl Genet 110:1419–1428
26. Vecchietti A, Lazzari B, Ortugno C, Bianchi F, Malinverni R, Caprera A, Mignani I, Pozzi C (2009) Comparative analysis of expressed sequence tags from tissues in ripening stages of peach (*Prunus persica* L. Batsch). Tree Genet Genome 5:377–391
27. Keskiaho K, Hieta R, Sormunen R, Myllyharju J (2007) *Chlamydomonas reinhardtii* has multiple prolyl 4-hydroxylases, one of which is essential for proper cell wall assembly. Plant Cell 19:256–269
28. Koski MK, Hieta R, Hirsilã M, Könka A, Myllyharju J, Wierenga RK (2009) The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. J Biol Chem 2484:25290–25301
29. Asif MH, Trivedi PK, Misra P, Nath P (2009) Prolyl-4-hydroxylase (5AtP4H1) mediates and mimics low oxygen response in *Arabidopsis thaliana*. Funct Integr Genomics 9:525–535
30. Suh MC, Schultz DJ, Ohlrogge JB (1999) Isoforms of acyl carrier protein involved in seed-specific fatty acid synthesis. Plant J 17:679–688
31. Buchanan BB, Gruissem W, Jones RL (2002) Biochemistry and molecular biology of plants. American Society of Plant Physiologists, Rockville
32. Chiang GCK, Barua D, Kramera EM, Amasino RM, Donohue K (2009) Major flowering time gene, FLOWERING LOCUS C, regulates seed germination in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 106:11661–11666
33. Aravind L, Koonin E (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. Genome Biol 2:1–8
34. Lukacin R, Britsch L (1997) Identification of strictly conserved histidine and arginine residues as part of the active site in *Petunia hybrida* flavanone 3b-hydroxylase. Eur J Biochem 249:748–757
35. Hedden P, Phillips AL (2000) Gibberellin metabolism: new insights revealed by the genes. Trends Plant Sci 5:523–530
36. Wang KL, Li H, Ecker JR (2002) Ethylene biosynthesis and signaling networks. Plant Cell 14:S131–S151
37. 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
38. Boutilier K, Offringa R, Sharma VK et al (2002) Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. Plant Cell 14:1737–1749
39. Chan MG, Smertenko A, Hussey PJ, Naldrett M, Bottrill A, Lloyd CW (2003) Identification of a MAP65 isoform involved in directional expansion of plant cells. FEBS Lett 534:161–163
40. Smertenko AP, Kaloriti D, Chang HY, Fiserova J, Opatrny Z, Hussey PJ (2008) The C-terminal variable region specifies the dynamic properties of *Arabidopsis* microtubule-associated protein MAP65 isotypes. Plant Cell 20:3346–3358
41. Sawano M, Shimmen T, Sonobe S (2000) Possible involvement of 65 kDa MAP in elongation growth of azuki bean epicotyls. Plant Cell Physiol 41:968–976
42. Vasák M, Hasler DW (2000) Metallothioneins: new functional and structural insights. Curr Opin Chem Biol 4:177–183
43. Bilecen K, Ozturk UH, Duru AD et al (2005) *Triticum durum* metallothionein: isolation of the gene and structural characterization of the protein using solution scattering and molecular modeling. J Biol Chem 280:13701–13711
44. Andrews GK (2000) Regulation of metallothionein gene expression by oxidative stress and metal ions. Biochem Pharmacol 59:95–104
45. Brouwer M, Syring R, Brouwer TH (2002) Role of a copper-specific metallothionein of the blue crab, *Callinectes sapidus*, in copper metabolism associated with degradation and synthesis of hemocyanin. J Inorg Biochem 88:228–239
46. Ebadi M, Leuschen MP, El Refaey H, Hamada FM, Rojas P (1996) The antioxidant properties of zinc and metallothionein. Neurochem Int 29:159–166
47. Dunaeva M, Adamska I (2001) Identification of genes expressed in response to light stress in leaves of *Arabidopsis thaliana* using RNA differential display. Eur J Biochem 268:5521–5529
48. Oztur ZN, Talame V, Deyholos M, Michalowski CB, Galbraith DW, Gozukirmizi N, Tuberosa R, Bohnert HJ (2002) Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley. Plant Mol Biol 48:551–573

49. Vitali A, Pacini E, De Mori P, Pucillo L, Maras B, Botta B, Brancaccio A, Giardina B (2006) Purification and characterization of an antifungal thaumatin-like protein from *Cassia didymobotrya* cell culture. Plant Physiol Biochem 44:604–610
50. van der Wel H, Loeve K (1972) Isolation and characterization of thaumatin I and II, the sweet-tasting proteins from *Thaumatococcus daniellii* Benth. Eur J Biochem 4:221–225
51. Abad LR, D'Urzo MP, Liu D et al (1997) Antifungal activity of tobacco osmotin has specificity and involves membrane permeability. Plant Sci 118:11–23
52. Min K, Ha SC, Hasegawa PM, Bressan RA, Yun DJ, Kim KK (2004) Crystal structure of osmotin, a plant antifungal protein. Proteins 54:170–173
53. Ng TB (2004) Antifungal proteins and peptides of leguminous and non-leguminous origins. Peptides 25:1215–1222
54. Garcia-Casado C, Collada C, Allona I, Soto A, Casado R, Rodriguez-Cerezo E, Gomez L, Aragoncillo C (2000) Characterization of an apoplastic basic thaumatin-like protein from recalcitrant chestnut seeds. Physiol Plant 110:172–180
55. Martin A, Lee J, Kichey T et al (2006) Two cytosolic glutamine synthetase isoforms of maize are specifically involved in the control of grain production. Plant Cell 18:3252–3274
56. Tabuchi M, Sugiyam K, Ishiyama K, Inoue E, Sato T, Takahashi H, Yamaya T (2005) Severe reduction in growth rate and grain filling of rice mutants lacking OsGS1;1, a cytosolic glutamine synthetase1;1. Plant J 42:641–651
57. Hirel B, Le Gouis J, Ney B, Gallais A (2007) The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. J Exp Bot 58:2369–2387
58. Lea PJ, Azevedo RA (2007) Nitrogen use efficiency. 2. Amino acid metabolism. Ann Appl Biol 151:269–275
59. Ohmiya A, Kishimoto S, Aida R, Yoshioka S, Sumitomo K (2006) Carotenoid cleavage dioxygenase (CmCCD4a) contributes to white color formation in chrysanthemum petals. Plant Physiol 142:1193–1201
60. Huang FC, Molnár P, Schwab W (2009) Cloning and functional characterization of carotenoid cleavage dioxygenase 4 genes. J Exp Bot 60:3011–3022
61. Ytterberg AJ, Peltier JB, van WiJk KJ (2006) Protein profiling of plastoglobules in chloroplasts and chromoplasts. A surprising site for differential accumulation of metabolic enzymes. Plant Physiol 140:984–997
62. Rubio A, Rambla JL, Santaella M, Gómez MD, Orzaez D, Granell A, Gómez-Gómez L (2008) Cytosolic and plastoglobule-targeted carotenoid dioxygenases from *Crocus sativus* are both involved in β-ionone-release. J Biol Chem 283:24816–24825
63. Bouvier F, Dogbo O, Camara B (2003) Biosynthesis of the food and cosmetic plant pigment bixin (annatto). Science 300:2089–2091
64. Bouvier F, Suire C, Mutterer J, Camara B (2003) Oxidative remodeling of chromoplast carotenoids: identification of the carotenoid dioxygenase CsCCD and CsZCD genes involved in *Crocus* secondary metabolite biogenesis. Plant Cell 15:47–62
65. Kato OR, Figueirêdo FJC, Belfort AJ, Nogueira OL, Barbosa WC (1992) Época de colheita de sementes de urucu: emergência e teor de corantes. Pesq Agropec Bras 27:1291–1302
66. Han Y, Gasic K, Korban SS (2007) Multiple-copy cluster-type organization and evolution of genes encoding o-methyltransferases in apple. Genetics 176:2625–2635
67. Wu S, Watanabe N, Mita S, Ueda Y, Shibuya M, Ebizuka Y (2003) Two o-methyltransferases isolated from flower petals of *Rosa chinensis* var. spontanea involved in scent biosynthesis. J Biosci Bioeng 96:119–128
68. Christensen AB, Gregersen PL, Olsen CE, Collinge DB (1998) A flavonoid 7-O-methyltransferase is expressed in barley leaves in response to pathogen attack. Plant Mol Biol 36:219–227
69. He X, Reddy JT, Dixon RA (1998) Stress responses in alfalfa (*Medicago sativa* L.). XXII. cDNA cloning and characterization of an elicitor-inducible isoflavone 7-O-methyltransferase. Plant Mol Biol 36:43–54
70. Ross JR, Nam KH, D'Auria JC, Pichersky E (1999) S-Adenosyl-L-methionine: salicylic acid carboxyl methyltransferase, an enzyme involved in floral scent production and plant defense, represents a new class of plant methyltransferases. Arch Biochem Biophys 367:9–16
71. Ogawa M, Herai Y, Koizumi N, Kusano T, Sano H (2001) 7-Methylxanthine methyltransferase of coffee plants. Gene isolation and enzymatic properties. J Biol Chem 276:8213–8218
72. Mizuno K, Okuda A, Kato M, Yoneyama N, Tanaka H, Ashihara H, Fujimura T (2003) Isolation of a new dual-functional caffeine synthase gene encoding an enzyme for the conversion of 7-methylxanthine to caffeine from coffee (*Coffea arabica* L.). FEBS Lett 534:75–81