GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

# Identification of the transcriptionally active cytochrome P450 repertoire in *Coffea arabica*

**S.T. Ivamoto[1,2], D.S. Domingues[2], L.G.E. Vieira[3] and L.F.P. Pereira[2,4]**

[1]Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Estadual de Londrina, Londrina, PR, Brasil
[2]Laboratório de Biotecnologia Vegetal, Instituto Agronômico do Paraná, Londrina, PR, Brasil
[3]Universidade do Oeste Paulista, Presidente Prudente, SP, Brasil
[4] Laboratório de Biotecnologia, Embrapa Café/IAPAR, Londrina, PR, Brasil

Corresponding author: L.F.P. Pereira
E-mail: filipe.pereira@embrapa.br

Genet. Mol. Res. 14 (1): 2399-2412 (2015)
Received June 16, 2014
Accepted October 9, 2014
Published March 27, 2015
DOI http://dx.doi.org/10.4238/2015.March.27.25

**ABSTRACT.** Cytochrome P450s (P450s) comprise a gene superfamily encoding enzymes that are involved in diverse plant metabolic pathways that produce primary and secondary metabolites such as phenylpropanoids, terpenoids, nitrogen-containing compounds, and plant hormones. They comprise one of the most diverse gene families in plant evolution. Although there are many studies that aim to characterize P450s in plants, there is no report on the characterization of this superfamily in *Coffea arabica*, where they might be related to plant tolerance to biotic and abiotic stresses, as well as aroma-related compounds. In this study, we report the characterization and annotation of 87 putative P450s from *C. arabica* obtained from the Brazilian Coffee Genome Project and describe their transcriptional pattern in different tissues and coffee organs. To validate our approach, we measured the transcriptional profile of the *CaCYP81D8_1* gene by quantitative polymerase chain reaction in leaves, flowers, and fruits. This study is the first effort to present and analyze the P450 superfamily

 

in *C. arabica*, which may assist in understanding the chemical diversity of coffee secondary metabolites.

**Key words:** Cytochrome P450; Coffee; Expressed sequence tags (ESTs); Transcriptome; Candidate genes

## INTRODUCTION

Cytochrome P450s (P450s) comprise a superfamily of genes present in all living organisms, from bacteria to plants and animals. They constitute a large and diverse group of heme proteins and are named for the spectral absorbance peak of their carbon-monoxide-bound species at 450 nm (Nelson, 2009). Currently, 5100 P450s sequences are annotated in plants, 1461 in vertebrate animals, 2137 in insects, 2960 in fungi, 1042 in bacteria, 27 in Archaea, and 2 in virus (Nelson, 2011). A total of 3651 P450s are identified in 11 plant genomes: *Arabidopsis thaliana*, *Carica papaya*, *Vitis vinifera*, *Glycine max*, *Solanum lycopersicum esculentum*, *Oryza sativa*, *Brachypodium distachyon*, *Selaginella moellendorffii*, *Physcomitrella patens*, *Chlamydomonas reinhardtii*, and *Volvox carteri*. Besides these, 1449 sequences from 255 plant species derived from incomplete sequenced genomes are available at the cytochromes (P450s) homepage (http://drnelson.uthsc.edu/CytochromeP450.html).

P450s correspond to one of the largest protein families in higher plants and include many events of gene duplication and conversion. This fact probably arises from the high plasticity required for plants to adapt to several environmental conditions (e.g., protection from pathogens and predators). Moreover, P450s also participate in a wide range of biochemical pathways to produce primary and secondary metabolites that include precursors of membrane sterols and structural macromolecules such as lignin, cutin, carotenoids, and suberins and in the biosynthesis of pigments, antioxidants, and defense compounds including flavonoids, phenolic esters, coumarins, glucosinolates, cyanogenic glucosides, isoprenoids, and alkaloids (Hamberger and Bak, 2013). P450s also contribute to the homeostasis of phytohormones and signaling molecules by controlling their biosynthesis (gibberellins, auxin, brassinosteroids, and jasmonate) and catabolism (brassinosteroids and abscisic acid). In addition to their physiological substrates, P450s metabolize and usually detoxify exogenous molecules such as pesticides and pollutants (Bak et al., 2011). Despite the increasing knowledge of the biological functions of plant P450s in recent years, most functions remain completely unknown (Bak et al., 2011).

The range of reactions catalyzed by P450 is extremely diverse, but the most common reaction is a monooxygenase reaction, which is consistently linked with a gain in the bioactivity of plant metabolites. P450s have the ability to catalyze region- and stereospecific hydroxylation reactions usually based on oxygen molecule activation with the insertion of its atoms into the substrate (S) and the reduction of the other oxygen molecule to form water ($S + O_2 + NADPH + 2H^+ \rightarrow SOH + H_2O + NADP^+$) (Bak et al., 2011). In order to be active, CYPs need to be coupled with a protein partner to deliver one or more electrons (Bak et al., 2011).

Plant P450s were originally grouped as A-type or non-A-type based on clustering clades in phylogenetic trees (Durst and Nelson, 1995). The A-type P450s are involved in specialized plant metabolism (synthesis of lignin, alkaloids, flavonoids, and cyanogenic glucosides), and non-A-types P450s are involved in sterol and lipid oxygenation and hormone metabolism. Furthermore, based on the available sequences, plant P450s can be classified in

11 phylogenetically distinct clans named according to the lowest-numbered family member and in 127 different families (Hamberger and Bak, 2013).

Coffee is one of the world's most consumed beverages and is a very complex mixture of substances, including biologically active classes of biochemical compounds such as flavonoids, phenolics, alkaloids, and terpenoids (Leroy et al., 2006). It is well known that P450s control key steps in the biosynthetic routes of plant specialized metabolites; consequently, P450s are involved in secondary compound synthesis in coffee. Despite their importance for establishing the chemical diversity of coffee beans associated with good cup qualities, there are no reports in the literature about putative functions of P450s and transcriptional activity patterns between different tissues in *Coffea arabica*. In this study, we used a public transcriptome database (Mondego et al., 2011) to identify, annotate, and study the transcriptional pattern of the available P450 coffee genes that are still unexplored.

## MATERIAL AND METHODS

### *In silico* mining of putative *C. arabica* P450s

P450 contig sequences were searched on the CafESTs database (http://www.lge.ibi.unicamp.br/coffea) that comprises 187,142 expressed sequence tags (ESTs) of *C. arabica* produced by the Brazilian Coffee Genome (CafESTs) and resulted in 32,007 contigs (Mondego et al., 2011). *Arabidopsis* P450 sequences already described in the literature (Nelson, 2009) were used for a local similarity search on the CafESTs database. In addition, all contigs were annotated based on the translated nucleotide basic local alignment search tool (BLASTX) match against the National Center for Biotechnology Information (NCBI) non-redundant (nr) proteins and UniProt (http://www.uniprot.org) results using a minimum E-value cutoff of $1e^{-20}$. We checked the presence of specific P450 domains in all contigs using the NCBI conserved domain platform (NCBI-CDD) and Blast2GO tools.

### Functional annotation and subfamily classification

*C. arabica* P450 functional analysis was carried out by analysis in Blast2GO (version 2.7.0; Conesa and Götz, 2008). We classified *Coffea* P450-related contigs according to their molecular function, cellular components, and biological process using default parameters. Contigs were also mapped in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using the GO-Enzyme Code Mapping tool and protein domains obtained from InterProScan (Quevillon et al., 2005). Coffee P450 subfamilies were classified using the criteria established on the *Arabidopsis* P450 website (http://www.p450.kvl.dk; Paquette et al., 2009).

### Phylogenetic analysis

Phylogenetic analyses were done using MEGA6 (Tamura et al., 2013). *Coffea* P450 amino acid sequences were aligned using MUSCLE and then used to produce a neighbor-joining tree using the Jones-Taylor-Thornton substitution evolution model. Partial deletion alignment gaps were considered for analysis. The phylogeny was tested with the bootstrap method using 1000 replications. Only full-length CYP sequences were used in the phylogenetic analysis, and bootstrap values below 50 are not shown.

## *In silico* expression pattern analysis

The raw digital gene expression counts were normalized using the reads per kilobase of transcript per million mapped reads (Mortazavi et al., 2008) method to correct the digital gene expression counts for bias caused by sequence size and total EST numbers in each cDNA library described in a previous report (Vieira et al., 2006). These values were used to generate a heatmap with the Genesis software, version 1.7.6 (Sturn et al., 2002).

## RNA isolation, purification, and cDNA synthesis

Total RNA of *C. arabica* cv. IAPAR59 mature leaves, flowers, and fruit perisperm at 3 development stages [90, 120, and 150 days after flowering (DAF)] was isolated as described by Chang et al. (1993). Total RNA was purified using the Pure Link Micro to Midi Total RNA Purification System (Invitrogen, Life Technologies, Carlsbad, CA, USA) and treated with DNase (Invitrogen). The RNA integrity was verified by 1% agarose gel electrophoresis, and its concentration and purity were determined using a NanoDrop ND-100 spectrophotometer. The absence of genomic DNA contamination was confirmed by polymerase chain reaction (PCR) using glyceraldehyde-3-phosphate-dehydrogenase gene (*GAPDH*) primers (Cruz et al., 2009) with 100 ng RNA (data not shown). Complementary DNA (cDNA) was synthesized using SuperScript III Reverse Transcriptase (Invitrogen) according to manufacturer instructions with a final volume of 20 µL using 5 µg total RNA. The final cDNA products were diluted tenfold prior to use in quantitative real-time PCR (qPCR).

## Primer design and amplification efficiency

*CaCYP81D8_1* primers were designed using the Primer Express software version 3.0 (**Table S1**). Primer specificity was verified using dissociation curve analysis, and the amplicon length was verified by 1% agarose gel electrophoresis. Primer efficiency (99%) was calculated by the LinRegPCR software (Ramakers et al., 2003).

## qPCR and transcriptional activity data analysis

Total RNAs were extracted from bud flowers, leaves, and fruits of 9 *C. arabica* cv. IAPAR59 full-grown plants maintained at Instituto Agronômico do Paraná. The transcript abundance for *CaCYP81D8_1* was analyzed by qPCR (7500 Fast Real-Time PCR System, Applied Biosystems) using the SYBR Green PCR Master Mix (Applied Biosystems). The reaction mixture contained 12.5 µL SYBR Green master mix, 1 µL of each primer (10 µM), 1 µL of cDNA diluted 1:10, and Milli-Q water to a total volume of 25 µL. Thermal conditions were 95°C for 10 min, followed by 40 cycles of 95°C for 30 s and 60°C for 60 s. Melting curves were analyzed to verify the presence of a single product including a negative control. All reactions were performed with 3 technical and biological replicates. Relative expression was calculated as $(1 + E)^{-\Delta\Delta Ct}$, where $\Delta Ct_{target} = Ct_{target\ gene}\text{-}Ct_{GAPDH}$ and $\Delta\Delta Ct = \Delta Ct_{target}\text{-}\Delta Ct_{reference\ sample}$. Perisperm tissue at 150 DAF was used as the calibrator sample. Gene expression levels were normalized using the *GAPDH* gene as recommended by Cruz et al. (2009) (**Table S1**).

 

## RESULTS AND DISCUSSION

### P450 identification

A total of 87 coffee sequences similar to plant CYP P450 monooxygenases were selected from the CafESTs database. They were classified in 28 different subfamilies (Tables 1 and 2). All *C. arabica* CYPs displayed a domain related to the P450 monooxygenase superfamily (CL12078, CypX superfamily, Pfam database). Most of them (86 of 87) had their domain confirmed in other databases, such as InterPro (IPR001128 and IPR001433), Panther (PTHR19384, PTHR24286, PTHR24298, PTHR24300, and PTHR25943), and COG (COG2124) (Table 2). The strict criteria used to search coffee P450s may explain the low number of contigs identified. The CYPs were further classified in subfamilies based on the domain search and annotation process in the *Arabidopsis* P450 website.

**Table 1.** Cytochrome P450 (P450) gene family from *Coffea arabica.*

| Classification | *Coffea** | Soybean | Rice | Grape | Poplar | *Medicago* | *Arabidopsis* | Moss |
|---|---|---|---|---|---|---|---|---|
| CYP71 Clan | | | | | | | | |
| CYP71 | 21 | 55 | 84 | 24 | 25 | 37 | 52 | 0 |
| CYP73 | 1 | 3 | 3 | 3 | 3 | 1 | 1 | 4 |
| CYP75 | 1 | 7 | 3 | 11 | 3 | 0 | 1 | 0 |
| CYP76 | 8 | 14 | 29 | 24 | 13 | 6 | 8 | 0 |
| CYP77 | 2 | 4 | 2 | 2 | 3 | 2 | 5 | 0 |
| CYP81 | 8 | 12 | 12 | 21 | 28 | 5 | 18 | 0 |
| CYP82 | 7 | 24 | 0 | 34 | 10 | 10 | 5 | 0 |
| CYP83 | 1 | 12 | 0 | 0 | 5 | 9 | 1 | 0 |
| CYP89 | 3 | 8 | 14 | 14 | 10 | 9 | 7 | 0 |
| CYP98 | 2 | 2 | 2 | 1 | 5 | 1 | 3 | 1 |
| CYP701 | 1 | 2 | 5 | 1 | 1 | 1 | 1 | 1 |
| CYP703 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 |
| CYP706 | 1 | 3 | 4 | 9 | 5 | 1 | 7 | 0 |
| CYP72 Clan | | | | | | | | |
| CYP72 | 5 | 6 | 13 | 12 | 22 | 7 | 9 | 0 |
| CYP734 | 1 | 3 | 4 | 2 | 2 | 1 | 1 | 0 |
| CYP714 | 3 | 6 | 5 | 6 | 6 | 3 | 2 | 0 |
| CYP721 | 1 | 6 | 2 | 2 | 5 | 1 | 1 | 0 |
| CYP74 Clan | | | | | | | | |
| CYP74 | 1 | 6 | 4 | 6 | 7 | 4 | 2 | 3 |
| CYP85 Clan | | | | | | | | |
| CYP87 | 1 | 12 | 11 | 2 | 7 | 2 | 1 | 0 |
| CYP90 | 1 | 7 | 5 | 12 | 4 | 4 | 4 | 0 |
| CYP707 | 2 | 7 | 3 | 10 | 5 | 3 | 4 | 0 |
| CYP716 | 3 | 17 | 0 | 7 | 15 | 3 | 2 | 1 |
| CYP720 | 1 | 1 | 0 | 2 | 1 | 1 | 1 | 0 |
| CYP86 Clan | | | | | | | | |
| CYP86 | 1 | 8 | 5 | 9 | 6 | 3 | 11 | 2 |
| CYP94 | 2 | 13 | 18 | 14 | 9 | 4 | 6 | 2 |
| CYP96 | 2 | 9 | 12 | 7 | 5 | 5 | 13 | 0 |
| CYP704 | 2 | 6 | 7 | 5 | 6 | 14 | 3 | 6 |
| CYP97 Clan | | | | | | | | |
| CYP97 | 4 | 3 | 3 | 5 | 3 | 4 | 3 | 3 |

*Cytochrome P450s identified in this study. Table based on Guttikonda et al. (2010).

*C. arabica* CYP genes were compared to 6 selected plant species (Table 1): soybean, *Medicago*, *Arabidopsis*, rice, poplar, grape, and moss based on the study of Guttikonda et al. (2010). The subfamily CYP71 was the highest represented with 21 identified contigs, which was similar to the results reported for other plant species (Guttikonda et al., 2010).

**Table 2.** *Coffea arabica* P450 characterization.

| CafEST database | | Sequence | Conserved domains | | |
|---|---|---|---|---|---|
| Gene name | Name ID | length (bp) | PFAM | InterProScan | Panther |
| CaCYP701A3 | Contig11456 | 1460 | CL12078 | IPR001128 | PTHR24279 |
| CaCYP703A2 | Contig4905 | 128 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP704A2_1 | Contig3043 | 1089 | CL06868 | IPR001128 | PTHR24296 |
| CaCYP704A2_2 | Contig9730 | 676 | CL12078 | IPR001128 | PTHR24296 |
| CaCYP706A7 | Contig15619 | 1070 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP707A1_1 | Contig11624 | 1304 | CL12078 | IPR001128 | PTHR24286 |
| CaCYP707A1_2 | Contig17418 | 1050 | CL12078 | IPR001128 | PTHR24286 |
| CaCYP714A1_1 | GT706425 | 651 | CL12078 | IPR001128 | PTHR24282 |
| CaCYP714A1_2 | GT714829 | 607 | CL12078 | IPR001128 | PTHR24282 |
| CaCYP714A1_3 | Contig12354 | 1057 | CL12078 | IPR001128 | PTHR24282 |
| CaCYP716A1_1 | Contig8089 | 1329 | CL12078 | IPR001128 | PTHR24286 |
| CaCYP716A1_2 | Contig11052 | 793 | CL12078 | IPR001128 | PTHR24286 |
| CaCYP716A2 | Contig4121 | 934 | CL12078 | IPR001128 | PTHR24286 |
| CaCYP71A12 | Contig2624 | 811 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP71A15 | Contig14803 | 684 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP71A21 | Contig3364 | 2019 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP71A24 | GW451309 | 636 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP71A25 | Contig15753 | 646 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP71A25_1 | Contig14459 | 822 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP71A25_2 | Contig8797 | 1254 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP71B1 | Contig4398 | 928 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP71B10 | Contig3941 | 748 | CL12078 | IPR001128 | PTHR24300 |
| CaCYP71B12 | Contig9150 | 628 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP71B13_1 | Contig7947 | 1271 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP71B2 | Contig3451 | 671 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP71B26 | Contig12379 | 765 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP71B34_1 | Contig6233 | 1454 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP71B34_2 | Contig15335 | 681 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP71B34_3 | Contig12264 | 1864 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP71B34_4 | Contig11431 | 657 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP71B37_1 | Contig11890 | 1134 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP71B37_2 | Contig7157 | 914 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP71B37_2 | Contig2728 | 971 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP71B4 | Contig13053 | 1730 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP720A1 | Contig8798 | 785 | CL12078 | IPR001128 | PTHR24286 |
| CaCYP721A1 | Contig687 | 880 | CL12078 | IPR001128 | PTHR24282 |
| CaCYP72A15_1 | Contig6702 | 1477 | CL12078 | IPR001128 | PTHR24282 |
| CaCYP72A15_2 | Contig16992 | 1564 | CL12078 | IPR001128 | PTHR24282 |
| CaCYP72A8_2 | Contig14015 | 880 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP72A8_1 | Contig262 | 775 | CL12078 | IPR001128 | PTHR24282 |
| CaCYP72A9 | Contig6520 | 892 | CL12078 | IPR001128 | PTHR24282 |
| CaCYP734A1 | Contig14807 | 560 | CL12078 | IPR001128 | PTHR24282 |
| CaCYP73A5 | Contig1623 | 1744 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP74A1 | Contig7490 | 761 | CL12078 | NA | NA |
| CaCYP75B1 | Contig9089 | 1627 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP76C2 | Contg3190 | 1020 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP76C3_1 | Contig3858 | 754 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP76C3_2 | Contig15055 | 1612 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP76C4_1 | Contig146 | 813 | CL12078 | IPR001128 | PTHR24300 |
| CaCYP76C4_2 | Contig9076 | 579 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP76C4_2 | Contig12897 | 488 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP76G1_1 | Contig9396 | 757 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP76G1_2 | Contig2124 | 731 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP77B1_1 | Contig4801 | 925 | CL12078 | IPR001128 | PTHR24300 |
| CaCYP77B1_2 | Contig6212 | 882 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP81D2 | Contig14185 | 1122 | CL12078 | IPR001128 | NA |
| CaCYP81D2 | Contig2465 | 798 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP81D8_1 | Contig6304 | 525 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP81D8_2 | Contig16287 | 1828 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP81F1 | Contig14809 | 785 | CL12078 | IPR001128 | PTHR25943 |

Continued on next page

**Table 2.** Continued.

| CafEST database | | Sequence | Conserved domains | | |
|---|---|---|---|---|---|
| Gene name | Name ID | length (bp) | PFAM | InterProScan | Panther |
| CaCYP81F2 | Contig11655 | 645 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP81F4 | Contig9152 | 589 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP81K1 | Contig15338 | 765 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP82C2 | Contig17481 | 757 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP82C4_1 | Contig14863 | 1805 | CL12078 | IPR001128 | NA |
| CaCYP82C4_2 | Contig14986 | 1625 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP82C4_3 | Contig1552 | 1114 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP82C4_4 | Contig7644 | 694 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP82C4_5 | Contig14863 | 1805 | CL12078 | IPR001128 | NA |
| CaCYP82G1 | GT672395 | 560 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP83B1 | Contig15696 | 1082 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP86A8 | Contig10063 | 1188 | CL12078 | IPR001128 | PTHR24296 |
| CaCYP87A2 | Contig5381 | 1618 | CL12078 | IPR001128 | PTHR24286 |
| CaCYP89A5 | Contig4244 | 1350 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP89A9_1 | Contig728 | 749 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP89A9_2 | Contig2896 | 583 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP90A | GW448593 | 766 | CL12078 | IPR001128 | PTHR24286 |
| CaCYP94B1 | GW461079 | 483 | CL12078 | IPR001128 | PTHR24296 |
| CaCYP94D1 | GW436444 | 669 | CL12078 | IPR001128 | PTHR24296 |
| CaCYP96A10 | Contig6171 | 634 | CL12078 | IPR001128 | PTHR24296 |
| CaCYP96A9 | Contig10936 | 1336 | CL12078 | IPR001128 | PTHR24296 |
| CaCYP97A3 | Contig15748 | 1398 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP97B3_1 | GW447951 | 592 | CL12078 | IPR001128 | PTHR24305 |
| CaCYP97B3_2 | GW483987 | 752 | CL12078 | IPR001128 | PTHR25943 |
| CaCYP97C1 | Contig11227 | 1843 | CL12078 | IPR001128 | PTHR24305 |
| CaCYP98A3_1 | Contig10703 | 1868 | CL12078 | IPR001128 | PTHR24298 |
| CaCYP98A3_2 | Contig14347 | 683 | CL12078 | IPR001128 | PTHR24298 |

CaCYP = *Coffea arabica* cytochrome P450; NA = not available.

## Functional annotation analysis of putative *C. arabica* P450s

The CYP71 subfamily, the most represented among coffee unigenes (Tables 1 and 2), is related to alkaloid biosynthesis (Schröder et al., 1999), herbicide detoxification (Siminszky et al., 1999), camalexin biosynthesis (Nafisi et al., 2007) and hydroxylation/oxidation reactions in the tyrosine hydroxylase synthesis and cyanogenic glucosides (Bak et al., 2011). These CYPs may be related to the amount of caffeine and trigonelline, traditionally alkaloids, observed in coffee beans.

Five *Coffea* CYPs are related to phytohormone biosynthesis: 1 CYP90 member for brassinosterols (Ohnishi et al., 2006), 2 CYP94 members for jasmonoyl L-isoleucine (Koo et al., 2011), and 2 CYP707 members for abscisic acid and gibberellin (Kushiro et al., 2004). Twenty-two contigs belong to the subfamilies CYP72 (5), CYP74 (1), CYP76 (8), and CYP81 (8) (Cabello-Hurtado et al., 1998; Swaminathan et al., 2009; Guttikonda et al., 2010; Zhu et al., 2012), which are related to plant protection against herbivores and herbicide detoxification.

Among the several CYP subfamilies related to secondary metabolism, we identified 3 contigs similar to CYP714, which is involved in alkaloid production (Zhu et al., 2006). CYP703 (Morant et al., 2007) and CYP86 (Höffer et al., 2008), represented by 2 contigs in our annotation, are related to fatty acid metabolism. Seven contigs were identified as CYP82, which is possibly involved in the biosynthesis of volatile compounds in flowers (Tholl et al., 2011).

We also identified CYPs from the subfamilies related to chlorogenic acid biosynthesis. One *CaCYP73* gene and 2 *CaCYP98* genes were classified as putative cinnamate 4-hydroxylase

and putative *p*-coumaroyl-3-hydroxylase, respectively (Fraser and Chapple, 2011). Their enzyme activities were already described in a previous study with *Coffea* species (Joët et al., 2009).

*CaCYP97A1* is a P450 with a putative function related to carotenoid biosynthesis in *C. arabica* and *Coffea canephora* (Simkin et al., 2008). It encodes a ε-hydroxylase that catalyzes the lutein formation from β-carotene. Carotenoids are essential to the photosynthetic apparatus, detoxifying reactive oxygen species, and they participate in plastidial adaptation to changes in environmental light conditions (Simkim et al., 2008).

*CaCYP701A3* is presumably an ent-kaurene oxidase enzyme that catalyzes successive oxidations of ent-kaurene to ent-kaurenoic acid and is required for gibberellin phytohormone biosynthesis (Morrone et al., 2010). Wang et al. (2012) demonstrated that one of the rice CYP701A subfamily members does not catalyze the prototypical conversion of the ent-kaurene C4α-methyl to a carboxylic acid; instead, it carries out hydroxylation at the C3α position in a number of related diterpenes, supporting the hypothesis that the biosynthetic routes for phytohormone production provide a reserve that is frequently recruited in the evolution of secondary metabolism. Cafestol and kahweol, known members of the ent-kaurene family that are exclusively found in *Coffea* spp, are diterpenes of special interest because of their biological activities (anti-inflammatory and anticarcinogenic activities); further functional studies on this CYP subfamily could help elucidate the biosynthetic pathways involved in the formation of these diterpenes.

The top 10 hits against *Coffea* P450 sequences were found in *S. lycopersicum*, *V. vinifera*, *Theobroma cacao*, *Nicotiana tabacum*, *Populus trichocarpa*, *Prunus persica*, *Ricinus communis*, *Cicer arietinum*, and *C. arabica* (NCBI nr database).

We used GO assignments to classify the predicted protein of *C. arabica* contigs (Figure 1).

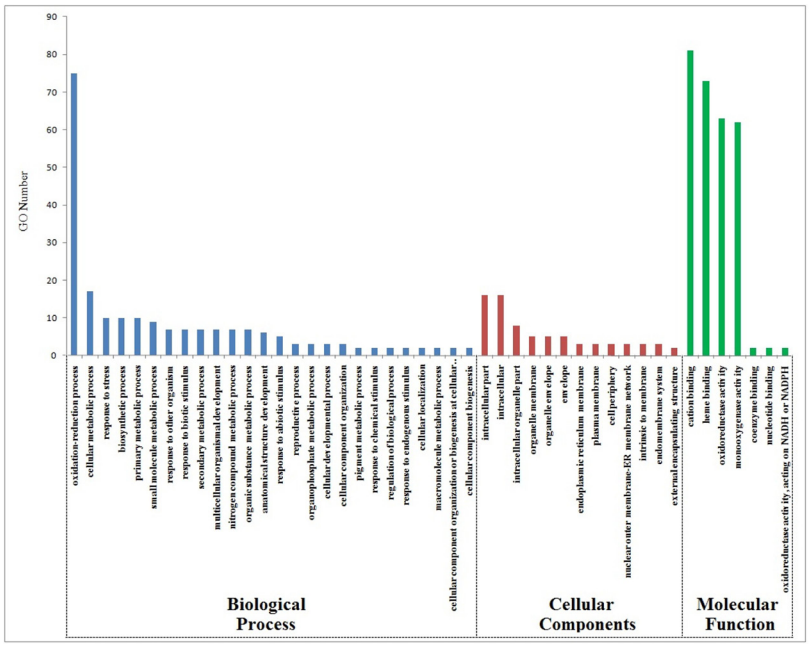


**Figure 1.** Blast2GO results for *Coffea arabica* cytochromes (CYPs). Results are shown for all 3 Gene Ontology (GO) categories: cellular component localization, putative molecular function, and biological process. Numbers represent the quantity of each GO term for each coffee CYP.

Biological process was the ontology with the highest representation, with 303 terms annotated for 10 different levels, followed by cellular component with 51 terms in 8 different levels and molecular function with 49 terms in 7 different levels. In the biological process category, there was a predominance of oxidation-reduction terms, which is a general function of P450 genes. For cellular component, most P450s were related to intracellular components. In the molecular function category, all P450s presented GO terms for cation binding, and the majority presented heme-binding, oxidoreductase, and monooxygenase activities, which corroborates our *in silico* approach to select and identify a catalog of P450s in coffee plants (Figure 1).

GO-enzyme code mapping based on the KEGG database categorized *C. arabica* CYPs in several metabolic pathways including monoterpenoid, flavonoid, flavone, flavonol, diterpenoid, carotenoid, stilbenoid, gingerol, and phenylpropanoid biosynthesis; fatty acid, limonene, and pinene degradation process; and arachidonic acid, alpha-linoleic acid, and phenylalanine metabolism (Figure 1).

## Phylogenetic analysis of predicted P450 families and subfamilies

A phylogenetic tree was generated to confirm the classification of CYP subfamilies using annotated *A. thaliana* CYPs. We observed that *A. thaliana* subfamilies grouped with *C. arabica* CYP sequences. This is an important step to confirm the accuracy of the information obtained by bioinformatic analyses (Figure 2).

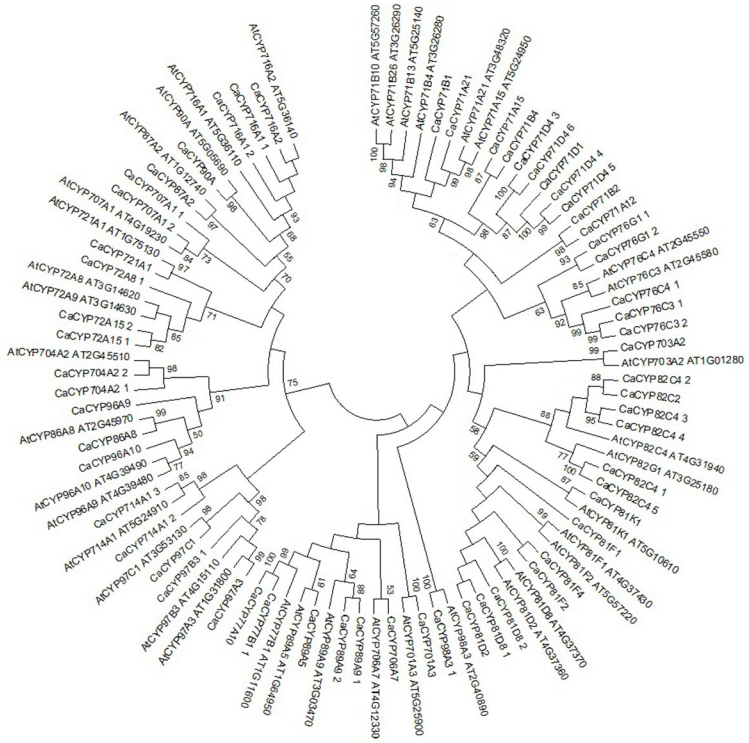


**Figure 2.** Evolutionary relationships of *Coffea arabica* and *Arabidopsis thaliana* CYP proteins.

In Figure 2, we observed that the *CaCYP71* and *CaCYP76* subfamilies grouped in the same clade, which may suggest that they share similar functions (e.g., herbicide detoxification; Siminszky et al., 1999; Swaminathan, et al., 2009). The same occurred for *CaCYP81* and *CaCYP82*, which are putatively related to glucosinolate biosynthesis (Cabello-Hurtado et al., 1998; Tholl et al., 2011). Another 2 subfamilies, *CcCYP86* and *CaCYP704*, also share similar functions that are probably related to fatty acid hydroxylation (Höffer et al., 2008) and fatty acid biosynthesis, respectively (Li et al., 2010).

## *In silico* expression analysis

Specific tissue expression patterns for each ethylene response factors (ERF) gene family (Lima et al., 2011), as well as the identification of candidate genes potentially associated with somatic embryogenesis (Silva et al., 2013), have been reported by *in silico* analysis of *C. arabica* EST libraries. Here, we performed a similar analysis of the transcription patterns of *C. arabica* CYPs using EST libraries derived from different organs/tissues and growth conditions, providing an initial framework to study different aspects of CYP gene expression in *Coffea*.

We observed that *CaYP71B34_3*, *CaCYP71A25_2*, *CaCYP77B1_1*, *CaCYP87A2*, *CaCYP97A3*, *CaCYP97C_1*, and *CaCYP701A3* were highly expressed in EST libraries from more than one organ/tissue, including leaves, fruits, cell suspension, stress treatments, embryogenic calli, and seeds, which suggests that these genes are expressed constitutively and may be involved in a variety of essential processes in the cells (Figure 3).

We also annotated CYPs that showed specialized transcriptional patterns in specific tissues or environmental conditions (Figure 3). *CaCYP81D8_1* was most expressed in mature leaves, while *CaCYP703A2* was highly expressed in flower buds. Similarly, *CaCYP701A3*, *CaCYP71A25_2*, *CaCYP71B2*, *CaCYP76C4_1*, and *CaCYP76C4_1* were most expressed in germinating seeds, while *CaCYP701A3*, *CaCYP71B37_1*, *CaCYP76G1_1*, *CaCYP82C2*, *CaCYP83B1*, and *CaCYP94B1* were highly expressed in different embryogenic calli libraries. Some CYPs showed high transcript levels in flower buds and leaves (*CaCYP716A1_1* and *CaCYP73A5*).

Regarding abiotic stress conditions, *CaCYP71A25_1*, *CaCYP81D8_2*, *CaCYP98A3_2*, and *CaCYP716A1_2* were most expressed in the leaves of plants under water deficit, while *CaCYP706A7* was highly induced in suspension cells stressed with aluminum. For biotic stresses, *CaCYP707A1_1* was most expressed in roots treated with acibenzolar-*S*-methyl, *CaCYP71B34_2* was expressed in stems infected with *Xylella* spp, and *CaCYP74A_1* was expressed in leaves infected with leaf miner and coffee leaf rust.

In order to confirm the transcriptional profile obtained by electronic northern, we selected the *CaCYP81D8_1* gene for further investigation using qPCR.

## *CaCYP81D8_1* transcriptional profile validation

The *CYP81* subfamily mediates the in-chain hydroxylation of several fatty acids. This enzyme is typically found in higher plants and differs from those already isolated from other living organisms (Cabello-Hurtado et al., 1998). In *C. arabica*, *in silico* analysis showed that this enzyme is highly induced in EST libraries from leaves (Figure 3).

The transcriptional pattern of *CaCYP81D8_1* in *C. arabica* plants was measured in 5 tissues by qPCR. We observed high gene expression in leaves and low levels in flowers and in fruits collected at different DAF, corroborating the *in silico* approach (Figure 4).

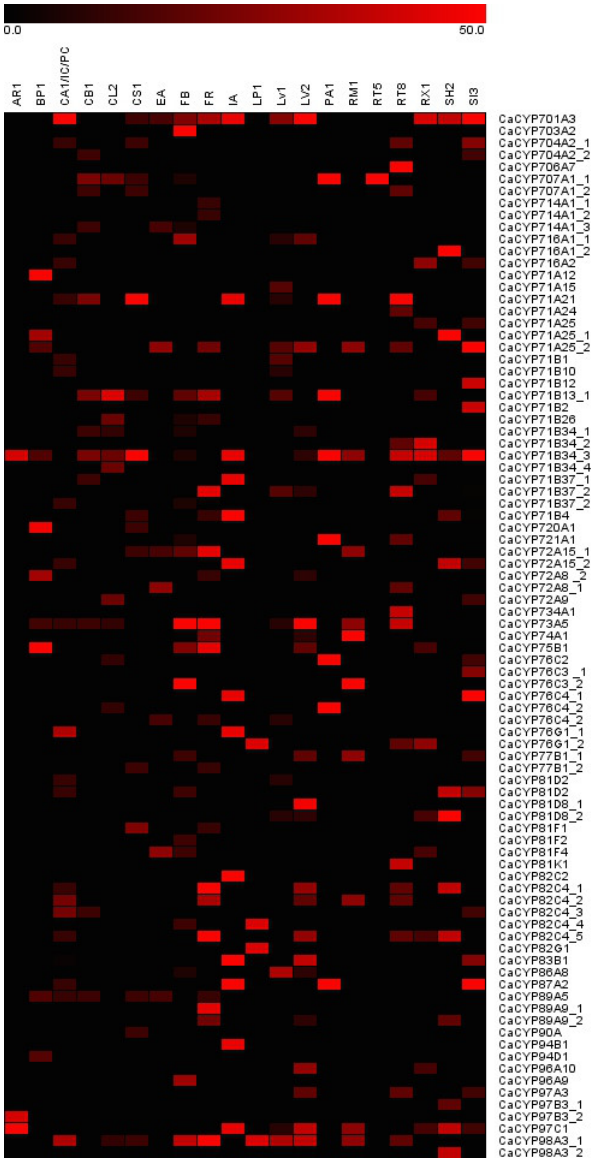


**Figure 3.** Transcriptional CYP patterns by electronic northern. The normalized numbers of reads for the transcripts in each library are represented in a scale from black to red. Coffee libraries are as follows (Vieira et al., 2006; Mondego et al., 2011): seedlings and leaves treated with arachidonic acid (AR1); suspension cells treated with acibenzolar-*S*-methyl (BP1); non-embryogenic calli with and without 2,4 dichlorophenoxyacetic acid (CA1, IC, and PC); suspension cells treated with acibenzolar-S-methyl and brassinosteroids (CB1); hypocotyls treated with acibenzolar-S-methyl (CL2); suspension cells treated with NaCl (CS1); embryogenic calli (EA1 and IAc); flower buds in different developmental stages (FB); flower buds + pinhead fruits + fruits at different stages (FR); seedlings and leaves treated with arachidonic acid (LP1); young leaves from the orthotropic branch (LV1); mature leaves from plagiotropic branches (LV2); primary embryogenic calli (PA1); leaves infected with leaf miner and coffee leaf rust (RM1); roots with acibenzolar-S-methyl (RT5); suspension cells stressed with aluminum (RT8); stems infected with *Xylella* spp (RX1); water deficit-stressed field plants (pool of tissues) (SH2); and germinating seeds (whole seeds and zygotic embryos) (SI3).

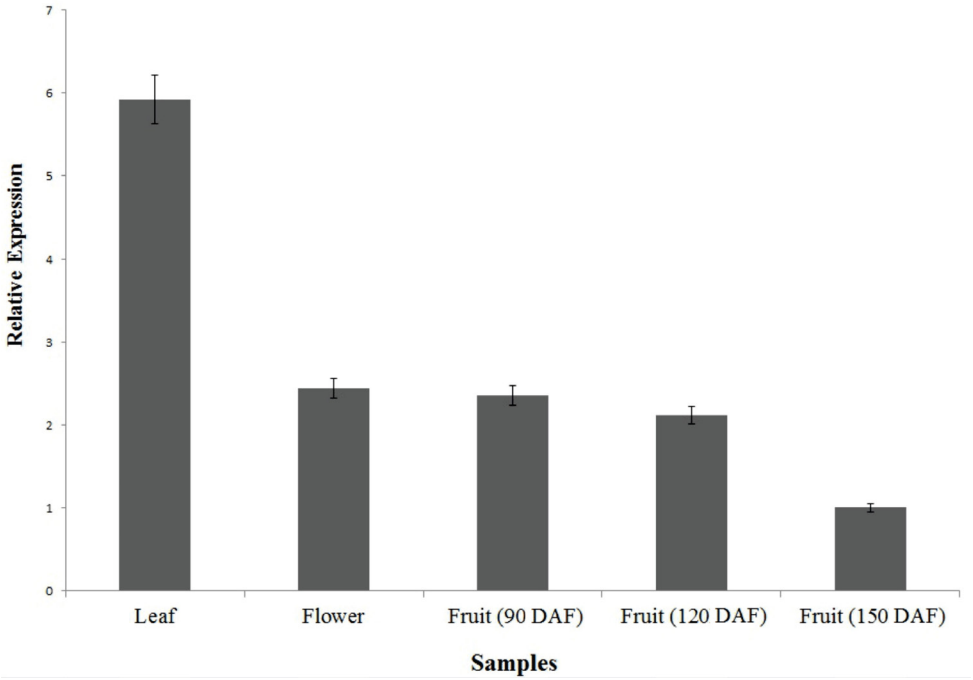


**Figure 4.** Transcriptional activity of *CaCYP81D8_1* in 3 organs (leaves, flowers, and fruit perisperm) at 3 stages: 90, 120 and 150 days after flowering (DAF). For each tissue, 3 independent plants were pooled for RNA extraction. Each tissue represents average data with standard error from 3 technical replicates.

## Future prospects

In this study, we report the identification and classification of *C. arabica* P450s according to families and subfamilies. This is, to our knowledge, the first effort to give an overall view of the actively transcribed P450 enzymes of this important tree crop. Our data offer a starting point for studies of the P450 proteins involved in important metabolic pathways, especially concerning chemical compounds related to cup quality and stress tolerance. In this way, the data presented here open new possibilities to find candidate genes that directly or indirectly are connected to many metabolic pathways that have significant impact on coffee breeding and biotechnology. Further integration of genetics, next-generation sequencing, bioinformatics, and biochemistry profiling tools may also provide further insights in the repertoire of all P450s in *C. arabica*.

## ACKNOWLEDGMENTS

Research supported by the Brazilian Consortium for Coffee Research and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). S.T. Ivamoto received a student fellowship from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). D.S. Domingues received a fellowship from Fundação Araucária, and L.G.E. Vieira and L.F.P. Pereira are CNPq research fellows.

## Supplementary material

## REFERENCES


Bak S, Beisson F, Bishop G and Hamberger B (2011). Cytochromes P450. In: The *Arabidopsis* Book American Society of Plant Biologists.

Cabello-Hurtado F, Batard Y, Salaun JP, Durst F, et al. (1998). Cloning, expression in yeast, and functional characterization of CYP81B1, a plant cytochrome P450 that catalyzes in-chain hydroxylation of fatty acids. *J. Biol. Chem.* 273: 7260-7267.

Chang S, Puryear J and Cairney J (1993). A simple and efficient method for isolating RNA from pine trees. *Plant. Mol. Biol. Rep.* 11: 113-116.

Conesa A and Götz S (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008: 619832.

Cruz F, Kalaoun S, Nobile P and Colombo C (2009). Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. *Mol. Breed.* 23: 607-616.

Durst F and Nelson DR (1995). Diversity and evolution of plant P450 and P450-reductases. *Drug Metabol. Drug Interact.* 12: 189-206.

Fraser CM and Chapple C (2011). The Phenylpropanoid Pathway in *Arabidopsis*. In: The *Arabidopsis* Book American Society of Plant Biologists.

Guttikonda SK, Trupti J, Bisht NC, Chen H, et al. (2010). Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. *BMC Plant Biol.* 10: 243.

Hamberger B and Bak S (2013). Plant P450s as versatile drivers for evolution of species-specific chemical diversity. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 368: 20120426.

Höffer R, Briesen I, Beck M and Pinot F (2008). The *Arabidopsis* cytochrome P450 *CYP86A1* encodes a fatty acid ω-hydroxylase involved in suberin monomer biosynthesis. *J. Exp. Bot.* 59: 2347-2360.

Joët T, Laffargue A, Salmona J, Doulbeau S, et al. (2009). Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. *New Phytol.* 182: 146-162.

Koo AJ, Cooke TF and Howe GA (2011). Cytochrome P450 CYP94B3 mediates catabolism and inactivation of the plant hormone jasmonoyl-L-isoleucine. *Proc. Natl. Acad. Sci. U. S. A.* 108: 9298-9303.

Kushiro T, Okamoto M, Nakabayashi K, Yamagishi K, et al. (2004). The *Arabidopsis* cytochrome P450 CYP707A encodes ABA 8'-hydroxylases: key enzymes in ABA catabolism. *EMBO J.* 23: 1647-1656.

Leroy T, Ribeyre F, Bertrand B and Charmetant P (2006). Genetics of coffee quality. *Braz. J. Plant Physiol.* 18: 229-242.

Li H, Pinot F, Sauveplane V, Werck-Reichhart D, et al. (2010). Cytochrome P450 family member CYP704B2 catalyzes the {omega}-hydroxylation of fatty acids and is required for anther cutin biosynthesis and pollen exine formation in rice. *Plant Cell* 22: 173-190.

Lima AA, Ságio SA, Chalfun-Junior A and Paiva LV (2011). *In silico* characterization of putative members of the coffee (*Coffea arabica*) ethylene signaling pathway. *Genet. Mol. Res.* 10: 1277-1289.

Mondego JM, Vidal RO, Carazzolle MF, Tokuda EK, et al. (2011). An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. *BMC Plant Biol.* 11: 30.

Morant M, Jorgensen K, Schaller H, Pinot F, et al. (2007). CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. *Plant Cell* 19: 1473-1487.

Morrone D, Chen X, Coates RM and Peters RJ (2010). Characterization of the kaurene oxidase CYP701A3, a multifunctional cytochrome P450 from gibberellin biosynthesis. *Biochem. J.* 431: 337-344.

Mortazavi A, Williams BA, McCue K, Schaeffer L, et al. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5: 621-628.

Nafisi M, Goregaoker S, Botanga CJ, Glawischnig E, et al. (2007). *Arabidopsis* cytochrome P450 monooxygenase 71A13 catalyzes the conversion of indole-3-acetaldoxime in camalexin synthesis. *Plant Cell* 19: 2039-2052.

Nelson DR (2009). The cytochrome p450 homepage. *Hum. Genomics* 4: 59-65.

Nelson DR (2011). Progress in tracing the evolutionary paths of cytochrome P450. *Biochim. Biophys Acta* 1814: 14-18.

Ohnishi T, Szatmari AM, Watanabe B, Fujita S, et al. (2006). C-23 hydroxylation by *Arabidopsis* CYP90C1 and CYP90D1 reveals a novel shortcut in brassinosteroid biosynthesis. *Plant Cell* 18: 3275-3288.

Paquette SM, Jensen K and Bak S (2009). A web-based resource for the *Arabidopsis* P450, cytochromes b5, NADPH-cytochrome P450 reductases, and family 1 glycosyltransferases (http://www.P450.kvl.dk). *Phytochemistry* 70: 1940-1947.

Quevillon E, Silventoinen V, Pillai S, Harte N, et al. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.* 33: W116-W120.

Ramakers C, Ruijter JM, Deprez RH and Moorman AF (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett.* 339: 62-66.

Schröder G, Unterbusch E, Kaltenbach M, Schmidt J, et al. (1999). Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase. *FEBS Lett.* 458: 97-102.

Silva AT, Paiva LV, Andrade AC and Barduche D (2013). Identification of expressed sequences in the coffee genome potentially associated with somatic embryogenesis. *Genet. Mol. Res.* 12: 1698-1709.

Siminszky B, Corbin FT, Ward ER, Fleischmann TJ, et al. (1999). Expression of a soybean cytochrome P450 monooxygenase cDNA in yeast and tobacco enhances the metabolism of phenylurea herbicides. *Proc. Natl. Acad. Sci. U. S. A.* 96: 1750-1755.

Simkin AJ, Moreau H, Kuntz M, Pagny G, et al. (2008). An investigation of carotenoid biosynthesis in *Coffea canephora* and *Coffea arabica*. *J. Plant Physiol.* 165: 1087-1106.

Sturn A, Quackenbush J and Trajanoski Z (2002). Genesis: cluster analysis of microarray data. *Bioinformatics* 18: 207-208.

Swaminathan S, Morrone D, Wang Q, Fulton DB, et al. (2009). CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell* 21: 3315-3325.

Tamura K, Stecher G, Peterson D, Filipski A, et al. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30: 2725-2729.

Tholl D, Sohrabi R, Huh JH and Lee S (2011). The biochemistry of homoterpenes-common constituents of floral and herbivore-induced plant volatile bouquets. *Phytochemistry* 72: 1635-1646.

Vieira LGE, Andrade AC, Colombo CA and Moraes AHA (2006). Brazilian coffee genome project: an EST-based genomic resource. *Braz. J. Plant Physiol.* 18: 95-108.

Wang Q, Hillwig ML, Wu Y and Peters RJ (2012). CYP701A8: a rice ent-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism. *Plant Physiol.* 158: 1418-1425.

Zhu Y, Nomura T, Xu Y, Zhang Y, et al. (2006). Elongated uppermost internode encodes a cytochrome P450 monooxygenase that epoxidizes gibberellins in a novel deactivation reaction in rice. *Plant Cell* 18: 442-456.

Zhu BQ, Xu XQ, Wu YW, Duan CQ, et al. (2012). Isolation and characterization of two hydroperoxide lyase genes from grape berries: HPL isogenes in *Vitis vinifera* grapes. *Mol. Biol. Rep.* 39: 7443-7455.