

RESEARCH PAPER

# Molecular characterization and functional analysis of the $\beta$ -galactosidase gene during *Coffea arabica* (L.) fruit development

Sérgio Araujo Figueiredo<sup>1,3</sup>, Philippe Lashermes<sup>2</sup> and Francisco José Lima Aragão<sup>3,\*</sup>

<sup>1</sup> Universidade Católica de Brasília, Campus II, SGAN 916, 70790-160, Brasília, DF, Brazil

<sup>2</sup> IRD-Institut de Recherche pour le Développement, UMR RPB (CIRAD, IRD, Université Montpellier II), BP 64501, 34394 Montpellier Cedex 5, France

<sup>3</sup> Embrapa Recursos Genéticos e Biotecnologia, Laboratório de Introdução e Expressão de Genes, PqEB W5 Norte, 70770-900, Brasília, DF, Brazil

\* To whom correspondence should be addressed. E-mail: aragao@cenargen.embrapa.br

The nucleotide sequence data reported in this paper is deposited in the GenBank Nucleotide Database under accession number HQ283330 (Ca $\beta$ Gal).

Received 8 October 2010; Revised 6 November 2010; Accepted 7 November 2010

## Abstract

$\beta$ -D-Galactosidase (EC 3.2.1.23) has been detected in several plant species, and is characterized in different organs and tissues by its ability to hydrolyse terminal non-reducing  $\beta$ -D-galactosyl residues from  $\beta$ -D-galactoside polymers. In the present paper the cloning and the biochemical and molecular characterization of *Coffea arabica*  $\beta$ -galactosidase expressed in the pericarp and the endosperm of coffee fruits in all phases of ripeness are described. It was found that coffee  $\beta$ -galactosidase is not evenly transcribed throughout fruit ripening, oscillating with two distinct peaks: the first peak when immature fruits are at the active growing stage and the second when fully developed coffee fruits are completely ripe. Both *in vitro* enzymatic activity of coffee fruit protein extracts and *in vivo* histochemical assay of freshly harvested coffee fruits confirmed the uneven transcription of  $\beta$ -galactosidase as fruit maturation advanced. Partial genomic DNA sequencing indicated a complex arrangement of nine putative exons. *In silico* translation of the cloned coding sequences clearly revealed the cloned gene as  $\beta$ -galactosidase, with the presence of a signal peptide directing the enzyme to the apoplast. Two isoforms were distinguished by sequencing reverse transcription-PCR transcripts, one expressed in young and adult leaves and another in stems, petals, and coffee fruit endosperm and pericarp. Southern blot analysis indicates that there are at least two copies of this gene in the *C. arabica* genome that could explain the presence of two  $\beta$ -galactosidase isoforms.

**Key words:** Cell wall, *Coffea arabica*, coffee, fruit ripening,  $\beta$ -galactosidase.

## Introduction

Tissue softening during fruit ripening is a consequence of cell wall structural weakening due to the activity of several cell wall-degrading enzymes (Fischer and Bennett, 1991). The application of analytical biochemistry and molecular techniques revealed that this physiological process is more complex than previously expected. The cell wall structure is a matrix composed of several different polysaccharides, and an increase has been reported in the solubility of both hemicellulosic and pectic fractions, with the release of

carbohydrate residues from the side chains of the polysaccharide matrix during fruit maturation (Sakurai and Nevins, 1997; Brummell and Harpster, 2001).

Despite the well characterized activity of polygalacturonases on the depolymerization of the cell wall pectic polymers, it appears that polygalacturonase activity alone is not sufficient to induce fruit softening (Giovannoni *et al.*, 1989). The release of arabinosyl and galactosyl residues from pectic side chains is also an important event involved

in fruit softening (Gross, 1984; Gross and Sams, 1984; Redgwell *et al.*, 1992), which is related to the activity of  $\beta$ -galactosidases (EC 3.2.1.23). This class of cell wall-degrading enzymes belongs to the glycosyl hydrolase enzymatic group, which is characterized by its ability to hydrolyse  $\beta$ -D-galactosyl residues from the non-reducing terminals of  $\beta$ -D-galactosides (Esteban *et al.*, 2003). In higher plants,  $\beta$ -galactosidase is known to be the only enzyme able to hydrolyse galactosyl residues from cell wall polysaccharides by an *endo*-cleavage of  $\beta$ -1,4-galactan (Smith *et al.*, 1998). Its ability to release galactose from carbohydrates, galactolipids, and glycoproteins, and from cell wall components during cell expansion, senescence, and fruit ripening has been identified in the literature (De Veau *et al.*, 1993; Buckeridge and Reid, 1994; Ross *et al.*, 1994).

$\beta$ -Galactosidase enzymatic activity and gene expression have been measured in different plant tissues undergoing developmental changes, such as pollen (Rogers *et al.*, 2001), seeds (Sekimata *et al.*, 1989; Buckeridge and Reid, 1994), and seedlings (Li *et al.*, 2001; Esteban *et al.*, 2005). In fruits,  $\beta$ -galactosidase activity was reported during fruit development and ripening for several species including tomato (Smith *et al.*, 1998, 2002; Smith and Gross, 2000), carambola (Balasubramaniam *et al.*, 2005), chick-pea (Esteban *et al.*, 2005), pear (Tateishi *et al.*, 2001), pepper (Biles *et al.*, 1997; Ogasawara *et al.*, 2007), mango (Prasanna *et al.*, 2005), apple (Ross *et al.*, 1994), rice (Chantarangsee *et al.*, 2007), *Arabidopsis* (Dean *et al.*, 2007), strawberry (Trainotti *et al.*, 2001), and orange (Wu and Burns, 2004), among others. In coffee species,  $\beta$ -galactosidase enzymatic activity was first described by Golden *et al.* (1993). These authors reported that *Coffea*  $\beta$ -galactosidase was not able to release galactose from pectins, but did so from arabinogalactans and galactans, which are two distinct side chains on the pectin backbone. They evaluated  $\beta$ -galactosidase enzymatic activity during coffee fruit ripening and verified that immature fruits presented low activity; however, as coffee fruits went into the ripening stage,  $\beta$ -galactosidase presented a peak with a 6-fold activity increase, followed by a slight decrease in completely ripe coffee fruits. These authors also compared  $\beta$ -galactosidase activity with four other glycosidases and verified that during coffee fruit ripening,  $\beta$ -galactosidase presented the highest enzymatic activity among all four cell wall-degrading enzymes analysed. However, they did not mention whether  $\beta$ -galactosidase was expressed in the pericarp or in the endosperm.

Although coffee is an important commodity, few studies related to *Coffea* fruit development have been published (De Castro and Marraccini, 2006; Geromel *et al.*, 2006, 2008; Maluf *et al.*, 2009). However, it should be noted that efforts have been directed towards understanding the complex integrated transcription network involved in coffee seed development (Salmona *et al.*, 2008; Joët *et al.*, 2009). Apart from the biochemical characterization of coffee cell wall components (Bradbury and Halliday, 1990; Rogers *et al.*, 1999; Redgwell and Fischer, 2006), only a few genes related to coffee fruit cell wall-degrading enzymes have been cloned and biochemically characterized, such as

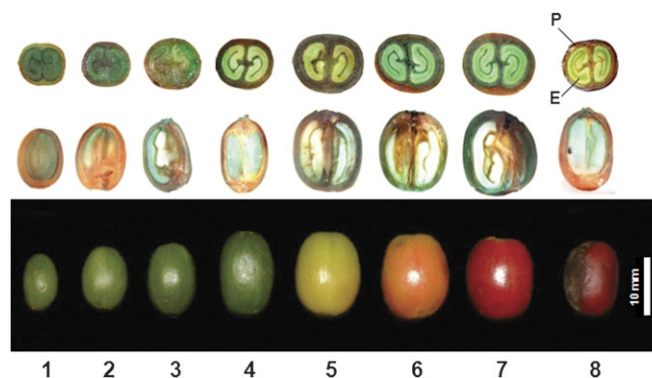
$\alpha$ -galactosidase (Marraccini *et al.*, 2005) and endo- $\beta$ -mannanases (Marraccini *et al.*, 2001). Molecular approaches are also being used to expand the knowledge related to coffee bean grain composition and biochemical aspects involved in coffee beverage quality. In this context, Salmona *et al.* (2008) investigated transcription networks specific to *C. arabica* seed development, combining cDNA arrays and real-time reverse transcription-PCR (RT-PCR). Even though these authors identified the presence of a  $\beta$ -galactosidase transcript (gathered with several other genes in cluster-IV1), it was only mentioned without any further exploration. As far as is known, there are no reports specifically on the molecular analysis of coffee  $\beta$ -galactosidases.

In this work, the isolation, molecular characterization, and functional analysis of the  $\beta$ -galactosidase gene of *C. arabica* (*Ca $\beta$ Gal*) during coffee fruit development are presented, distinguishing both temporal and spatial  $\beta$ -galactosidase expression.

## Materials and methods

### Plant material

Field-grown *C. arabica* cv Icatu Vermelho fruits were manually harvested from October to February at Centro de Pesquisa Agropecuária dos Cerrados/EMBRAPA, Planaltina, DF, Brazil (15°35'35"S, 47°42'30"W, 1007 m asl), with a lowest and a highest mean annual temperature of 12.6 °C and 30.1 °C, respectively. Coffee fruits at different ripening phases were classified based on their phenological stages from 1 to 8, according to fruit size and pericarp colouration (Fig. 1). In the first four stages, the immature fruits were visually classified by the increment in both volume and length and, from stage 5 to 8, the fully grown coffee fruits were classified by pericarp colour changes. Once fruits were harvested and classified, the pericarps and endosperms were separated, identified, independently weighed, frozen in liquid nitrogen, and stored at -80 °C.



**Fig. 1.** Classification of field-grown *C. arabica* fruits at eight developmental stages based on fruit size (immature coffee fruits, stages 1–4) and on pericarp colouration (ripening coffee fruits, stages 5–8) is presented in the lower panel. Transversal and longitudinal coffee fruit sections are presented, respectively, in the upper panel, presenting blue coloured regions that correspond to *in vivo*  $\beta$ -galactosidase histochemical activity. P, pericarp; E, endosperm.

*RNA extraction, cDNA synthesis, and RT-PCR*

Total RNAs were extracted from ~100 mg of both pericarps and endosperms of *C. arabica* fruits, previously stored at  $-80^{\circ}\text{C}$ , using Micro-to-midi Total RNA Purification System (Invitrogen, USA). cDNAs were prepared from these RNAs with SuperScript™ III Reverse Transcriptase (Invitrogen, USA) using oligo(dT)<sub>20</sub> according to the manufacturer's instructions.

Semi-quantitative  $\beta$ -galactosidase transcription expression was carried out by RT-PCR, which was optimized in order to allow this analysis from coffee fruit pericarp and endosperm at different ripening stages, using the elongation factor 1 $\alpha$  gene (*EF-1 $\alpha$* ) as an internal control. Three biological replicates of coffee fruit endosperm and pericarp at each developmental stage were used in this analysis. For  $\beta$ -galactosidase transcription analysis,  $\beta$ GALPLF (AGAAGCACTCCCGAGATGTGG) and  $\beta$ GALPLR (TGGCA-CATAACCCATGG) were used as primers, amplifying a single fragment of 504 bp. The PCRs were carried out using 1  $\mu\text{l}$  of cDNA, Invitrogen 10 $\times$  PCR buffer, 1.5 mM MgCl<sub>2</sub>, 10 mM of each dNTP, 10  $\mu\text{M}$  of each gene-specific primer, and 1 U of *Taq* polymerase (Invitrogen, USA) in a final volume of 25  $\mu\text{l}$ . The PCR amplification program was optimized, and consisted of one cycle at  $94^{\circ}\text{C}$  for 2 min, followed by 30 cycles of  $94^{\circ}\text{C}$  for 30 s,  $50^{\circ}\text{C}$  for 1 min, and  $72^{\circ}\text{C}$  for 1 min, after which the reactions were kept at  $72^{\circ}\text{C}$  for another 10 min on PTC-100<sup>®</sup> Peltier Thermal Cycler equipment (MJ Research, USA). For *EF-1 $\alpha$*  gene expression, identical reactions were prepared in parallel, except that EF-1F (TGTTGCTGTTAAGGATTTGAAGCG) and EF-1R (AACAGTTTGACGCATGTCCCTAAC) were used as primers, amplifying a fragment of 358 bp. The PCR amplification cycle was the same as that used for  $\beta$ -galactosidase.

The RT-PCRs were run on a 1% (w/v) agarose gel by electrophoresis, and the relative intensity of the amplified fragments was determined by analysis of the gel images using the Quantity One<sup>®</sup> program v4.6.3. Relative quantification was done by comparing the intensities of *Ca $\beta$ Gal* and *EF-1 $\alpha$*  bands. Three biological replicates were analysed per coffee fruit at each developmental stage.

*Genomic DNA extraction and Southern analysis*

Genomic DNA was extracted from 200 mg of healthy adult leaves of both *C. arabica* cv Icatu Vermelho and *C. arabica* cv Catuai Vermelho using the cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987) with a modified extraction buffer containing 2% (w/v) CTAB, 2% (w/v) PVP (polyvinylpyrrolidone), 100 mM Tris-HCl pH 8, 25 mM EDTA, and 2 mM sodium chloride. At the moment of extraction, 2% (v/v)  $\beta$ -mercaptoethanol and 2% (w/v) sodium metabisulphite were added.

For Southern blot analysis, 50  $\mu\text{g}$  of genomic DNA was individually digested with *Dra*I, *Eco*RI, and *Hind*III, followed by electrophoresis in a 0.8% (w/v) agarose gel at  $25\text{ V h}^{-1}$ . The digested samples were transferred to a nylon membrane Hybond™-N<sup>+</sup> (Amersham Biosciences, UK) by the alkaline capillary method (Sambrook and Russell, 2001), followed by pre-hybridization with salmon sperm at  $65^{\circ}\text{C}$  for 2 h. A 426 bp probe from the original coffee  $\beta$ -galactosidase expressed sequence tag (EST) was PCR amplified with the  $\beta$ GALPLF (AGAAGCACTCCCGAGATGTGG) and  $\beta$ RACE (TTCTCGTGAAAAGCTCTCTCCG) primers, using the amplification cycle previously described for gene expression analysis. The probe was radioactively labelled using Ready-To-Go™ DNA Labeled Beads (dCTP) (GE Healthcare, UK), according to the manufacturer's instructions. The membrane was hybridized at  $55^{\circ}\text{C}$  for 16 h, followed by two sequential washes using  $2\times\text{SSC}$ , 0.1% (w/v) SDS at room temperature, and two washes with  $1\times\text{SSC}$ , 0.1% (w/v) SDS at  $55^{\circ}\text{C}$ . The washed membrane was placed in an imaging plate cassette from Bio Imaging Analyzer (Fuji Photo Film Co. Ltd, Japan) for 48 h, and analysed using FUJIFILM Fluorescent Image Analyzer FLA-3000 (FUJIFILM Life Science, USA).

*Extraction of  $\beta$ -galactosidase from coffee fruits*

$\beta$ -Galactosidase was extracted independently from coffee fruit pericarp and endosperm at all eight ripening stages. Approximately 1 g of each sample was ground in liquid nitrogen, and extracted in ice-cold extraction buffer containing 0.1 M sodium citrate pH 4.6, 1 M sodium chloride, 13 mM EDTA, 1% (w/v) PVP-10, and 100  $\mu\text{g ml}^{-1}$  PMSF (phenylmethylsulphonyl fluoride), which was added at the ratio of 5 ml of extraction buffer per gram of fresh weight. Samples were kept on ice for 5 min with periodic mixing, followed by centrifugation at 8000 rpm for 20 min at  $4^{\circ}\text{C}$ . The supernatants containing  $\beta$ -galactosidase were harvested, and stored at  $-20^{\circ}\text{C}$ .

*Protein estimation*

Protein was estimated from the supernatant extracted for  $\beta$ -galactosidase enzymatic analysis by Bradford's method using bovine serum albumin as standard (Bradford, 1976).

 *$\beta$ -Galactosidase enzymatic assay*

$\beta$ -Galactosidase enzymatic assay was performed according to Golden *et al.* (1993). Fruit extracts from coffee pericarps and endosperms containing  $\beta$ -galactosidase were thawed on ice, and 20  $\mu\text{g}$  of total soluble protein was used per sample. Extracts were incubated with 13 mM PNPG (4-nitrophenyl- $\beta$ -D-galactopyranoside; Sigma-Aldrich, USA), incubated at  $37^{\circ}\text{C}$  for 30 min, and the absorbance was measured at 405 nm using iMark™ Microplate Reader (Bio-Rad, USA). Each sample was measured in triplicate. The standard curve was prepared using a 2-fold serial dilution from 50 mM of *p*-nitrophenol (Sigma-Aldrich, USA) in triplicate.

 *$\beta$ -Galactosidase histochemical assay*

$\beta$ -Galactosidase *in vivo* histochemical assay was performed according to Chantarangsee *et al.* (2007). Briefly, freshly harvested coffee fruits at all ripening stages were longitudinally and transversally sectioned, and incubated in staining solution containing X-Gal [50 mM sodium phosphate pH 7; 10 mM EDTA; 0.1% (w/v) sarcosyl; 20% (v/v) methanol; 1 mM X-Gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside)] for 4 h at room temperature. After this period, the staining solution was discarded, the reaction was stopped, and the samples were decolourized in methanol, and stored at room temperature for visual analysis (Fig. 1).

*Cloning  $\beta$ -galactosidase*

The partial genomic  $\beta$ -galactosidase sequence was obtained using Genome Walking™ Universal Kit (Clontech, USA) with *Dra*I, *Fsp*I, *Sca*I, *Ssp*I, *Sma*I, *Nru*I, and *Stu*I to digest the genomic DNA according to the manufacturer's instructions.

*BAC library screening analysis*

A bacterial artificial chromosome (BAC) library constructed from the genomic DNA of *C. arabica* cv IAPAR 59 was used (Noir *et al.*, 2004). High-density colony filters were prepared using a Genetix Q-Bot. BAC clones were double-spotted onto Hybond-N<sup>+</sup> filters (Amersham-Pharmacia) in a  $4\times 4$  pattern as described by Tomkins *et al.* (1999). In total, 27 648 clones of the BAC library were spotted onto a  $22.5\times 22.5$  cm filter representing, approximately,  $4\times$  equivalent-arabica genomes (equal to eight times basic coffee genome equivalents since two diploid subgenomes, C<sup>a</sup> and E<sup>a</sup>, are associated in *C. arabica*).

The filters were hybridized with three different PCR-generated probes corresponding to different regions of the 504 bp *C. arabica*  $\beta$ -galactosidase EST. Single amplified fragments were agarose purified and used as probes. Probes were labelled with [<sup>32</sup>P]dCTP according to the manufacturer's recommendations (Megaprime DNA labelling systems kit, Amersham). Filter hybridization was carried out as described by Sambrook and Russell (2001). Hit

BAC clones were grown for 48 h at 37 °C in 1.5 ml of 2YT medium containing chloramphenicol (12.5 µg ml<sup>-1</sup>), and BAC DNA was extracted using an alkaline lysis procedure (Sambrook and Russell, 2001).

#### DNA sequencing and sequence analysis

Sequencing was carried out by MacroGen Inc. (Maryland, USA) using BigDye ver. 3.1 sequencing chemistry (Applied Biosystems, Inc., USA) with the instrument 3730x1 DNA Analyzer from Life Technologies (Applied Biosystems, Inc., USA). The BLAST program was used for searching homology against nucleotide and amino acid sequence databases at GenBank. The motifs of the deduced amino acid sequence were searched using the PROSITE algorithm (<http://ca.expasy.org/prosite/>).

#### Molecular phylogenetic analysis

In order to perform a phylogenetic analysis, partially cloned *CaβGal* translated coding sequence was compared with 36 plant β-galactosidase proteins available at the National Center for Biotechnology Information (NCBI) GenBank. The following accessions were selected: *Arabidopsis thaliana* BGAL6 (CAB64742), *Arabidopsis lyrata* subsp. *lyrata* BGAL6 (XP\_002864867), *Asparagus officinalis* (CAA54525), *Brassica oleracea* (CAA59162), *Capsicum annuum* (AAK40304), *Carica papaya* (ACPI8875), *Cicer arietinum* (CAA06309), *Citrus sinensis* (AAK31801), *Fragaria×ananassa* βGal1 (CAC44500), *Fragaria×ananassa* βGal2 (CAC44501), *Fragaria×ananassa* βGal3 (CAC44502), *Glycine max* (ACF22882), *Gossypium hirsutum* (AAS76480), *Solanum lycopersicum* (formerly *Lycopersicon esculentum*) TBG2 (AAF70821), *S. lycopersicum* TBG3 (AAF70822), *S. lycopersicum* TBG4 (AAC25984), *S. lycopersicum* TBG6 (AAF70825), *S. lycopersicum* TBG7 (AAF70823), *Malus×domestica* (P48981), *Mangifera indica* SP26 (AAB61470), *Nicotiana tabacum* TP5 (CAC13966), *Oryza sativa* Indica group (ACC64531), *Oryza sativa* Japonica group (AAM34271), *Petunia×hybrida* BGAL2 (ACC60982), *Persea americana* PaGAL4 (BAF31234), *Pyrus communis* (CAH18936), *Populus trichocarpa* (EEE91791), *Prunus persica* (ABV32545), *Pyrus pyrifolia* PpGal5 (BAD91082), *Raphanus sativus* RsBGAL1 (BAD20774), *Ricinus communis* (XP\_002521778), *Sorghum bicolor* (XP\_002437187), *Vigna radiata* (AAF67341), *Vitis vinifera* BG1 (AAK81874), *Zea mays* (ACG42995). *Coffea arabica* β-galactosidase (CAI47559) was used as outgroup to root the phylogenetic tree. The β-galactosidase sequences were aligned using BioEdit software (<http://www.mbio.ncsu.edu/BioEdit>) and the phylogenetic analysis was performed using the ClustalW algorithm through the MEGA4 program (Tamura *et al.*, 2007). A bootstrap method ( $n=1000$ ) was employed to evaluate the reliability of tree topologies (Felsenstein, 1985) and the Maximum Parsimony (MP) algorithm was used to construct a β-galactosidase phylogenetic tree.

## Results and Discussion

### β-Galactosidase gene isolation and characterization

β-Galactosidase (EC 3.2.1.23) participates actively in fruit maturation in a diverse group of plants and, although β-galactosidase presents the highest activity during coffee fruit maturation among other glycosidases (Golden *et al.*, 1993), to our knowledge no further investigation related to this cell wall-degrading enzyme has been carried out.

As no data related to *C. arabica* β-galactosidase were available from any public database, initially two primers were designed based on highly conserved nucleotide regions of several aligned β-galactosidase ESTs obtained from the

NCBI GenBank using the MegAlign program (Lasergene, USA), and a PCR was performed with cDNAs extracted from mature coffee fruits (stage 7; Fig. 1). One single band was amplified (data not presented), sequenced, and blasted against the NCBI translated nucleotide database, identifying the sequence as β-galactosidase, which was named *CaβGAL*.

The nucleotide sequence was partially solved by genome walking (Genome Walking™ Universal Kit, Clontech, USA) which basically consists of PCR amplifications of single-digested genomic DNA with blunt-ended restriction enzymes, followed by the ligation of specific adaptors (provided by the manufacturer) to both ends of these DNA fragments. These adaptor-ligated DNA pools ('libraries') are submitted to two PCR amplifications per library: (i) the first PCR uses the outer adaptor primer (AP1) provided in the kit and an outer, gene-specific primer (GSP1) provided by the researcher. The primary PCR mixture is then diluted and used as a template for a nested PCR with a nested adaptor primer (AP2) and a nested gene-specific primer (GSP2). However, blunt-ended restriction enzymes other than those recommended by the manufacturer were used here, because the partial EST sequence presented restriction sites for the restriction enzymes provided with the kit. Several different PCR cloning approaches published in peer-reviewed articles such as iPCR (Ochman *et al.*, 1988), single-site Sda-PCR (MacGregor and Overbeek, 1991), and single oligonucleotide nested PCR (Antal *et al.*, 2004), as well as RACE (Rapid Amplification of cDNA Ends) using GeneRacer™ (Invitrogen, USA) were also tested without success. In order to sequence the gene of interest, a *C. arabica* BAC library (Noir *et al.*, 2004) was screened with three *CaβGal* probes (data not presented), identifying eight putative *C. arabica* β-galactosidase BAC clones. Among these, four clones hybridized with all three probes, which were used for sequencing the gene of interest. These putative *CaβGal* genes were purified and sequenced by MacroGen Corp. USA ([www.macrogen.com](http://www.macrogen.com)) by a primer walking approach. The data analysis sequence confirmed the previous sequences and expanded both upstream and downstream from the known nucleotide sequence.

A partial genomic nucleotide sequence of 3948 bp was obtained based on the assembly of overlapping genomic DNA sequences derived from *C. arabica* using both genome walking DNA fragments and putative positive BAC clones sequenced by primer walking. The sequence was entered in the NCBI GenBank database with the accession number HQ283330 (*CaβGal*).

*In silico* analysis of the cloned *CaβGal* nucleotide sequence blasted against the NCBI protein database (blastx) clearly indicated the gene of interest to be β-galactosidase. Nine exons were identified, producing a putative 1005bp β-galactosidase-coding sequence (EST), which translated into 334 amino acid residues (mol. wt=37 651 Da) (Figs 2A, 4B). The presence of several exon regions is not unique to the *CaβGal* gene. Ahn *et al.* (2007) reported the highly complex structure of 17 *A. thaliana* β-galactosidase genomic sequences, identifying from 12 to 19 exons among

**A**

```

1   ATGGGTGCTTTTGGTTGAGCTGTTTTGGCTTGTTAATGGTTATGTGGACAACACTACAAGG
1   M G A F W L S C F G L L M V M W T T T R

61  GGTGGTGTGAGGGAGGACAAGTTTCTTACGATGGAAGATCACTGATAATTGAAGGGCAA
21  G G V E G G Q V S Y D G R S L I I E G Q

121 AGAAAGCTCCTGTTTTCTGGTTCAATTCATTATCCTCGAAGCACTCCCAGATGTGGCCA
41  R K L L F S G S I H Y P R S T P Q M W P

181 TCTTTGATATCAAAGGCCAAACATGGTGGACTAGATGTGATTGAAACCTATGTCTTTTGG
61  S L I S K A K H G G L D V I E T Y V F W

241 AATCTTCATGAACCTCGACATGGGCAGCAGTATGATTTTAAGGGAAGACATAATATAGTG
81  N L H E P R H G Q Q Y D F K G R H N I V

301 AGATTCATTAGAGAAATCAAGCTCATGGCCTCTATGCATTCATCCGGATTGGACCCTTC
101 R F I R E I Q A H G L Y A F I R I G P F

361 ATTGAGGCTGAATGGACTTACGGGGTCTACCATTTTGGCTGCATGATGTACCAGGAATT
121 I E A E W T Y G G L P F W L H D V P G I

421 GTTTATCGATCTGATAATGAACCTTTAAGGTGCAGTATCATATGCAAACTTCACAACA
141 V Y R S D N E P F K V Q Y H M Q N F T T

481 AAAATTGTCAACTGTTCAAATCAGAAGGATTATATGCTCCAAGGAGGGCCTATCATT
161 K I V N L F K S E G L Y A P Q G G P I I

541 CTTCAACAGATCGAGAATGAGTACAAAATGCGGAGAGAGCTTTTCACGAGAAAGGACCG
181 L Q Q I E N E Y K N A E R A F H E K G P

601 CCTTATGTTTCAGTGGGCAGCTGCCATGGCTGTCGGACTCCAACGGGTGTCCATGGGTT
201 P Y V Q W A A A M A V G L Q T G V P W V

661 ATGTGCAAGCAAGATGATGCTCCTGATCCTGTGGTAATAAACACATGTAATGGGAGGACG
221 M C K Q D D A P D P V V I N T C N G R T

721 TGTGGAGAAACATTTGTAGGCCTAATTCACCAAACAAGCCAGCAATCTGGACAGACAAT
241 C G E T F V G P N S P N K P A I W T D N

781 TGGACAAGTTTGTTCAGTATCATGGCGGTACAAATTTGGCAGGACAGGTTTCAGCATTT
261 W T S L F Q Y H G G T N F G R T G S A F

841 GTA CTACAAGCTATTATGATGAAGCTCCTATTGATGAATACGGTGGTTTGATTAGGCAA
281 V L T S Y Y D E A P I D E Y G G L I R Q

901 CCAAATGGGGTCATCTGAAGCAGTTACATTAGTAATAAAGTCATGCTCAGACTCTA
301 P K W G H L K Q L H S V I K S C S Q T L

961 CTT CATGGAGTCATATCTGTTTCTCCACTTGGTCAGCAGCAAGAA
321 L H G V I S V S P L G Q Q Q E

```

**Fig. 2.** Partial putative *C. arabica*  $\beta$ -galactosidase EST. (A) The nucleotide sequence (upper line) and the predicted amino acid sequence (lower line). Deduced amino acids are shown in one-letter code. The predicted signal peptide is in bold. Underlining indicates the predicted glycosyl hydrolase 35 family putative active site. (B) Alignment of the *C. arabica*  $\beta$ -galactosidase predicted amino acid sequence with the *A. thaliana* (CAB 64742)  $\beta$ -galactosidase amino acid sequence. The box indicates the predicted glycosyl hydrolase 35 family putative active site.

them. Not all intron sequences were shared by all 17 *At $\beta$ Gal* genes, suggesting that, during evolution, indel mechanisms might have been present in order to explain the observed genomic organizational physical pattern.

To investigate whether the deduced  $\beta$ -galactosidase protein is targeted to the cell wall, analysis of the deduced amino acid sequence with the SignalP 3.0 Server program (Emanuelsson *et al.*, 2007) predicted a eukaryotic signal

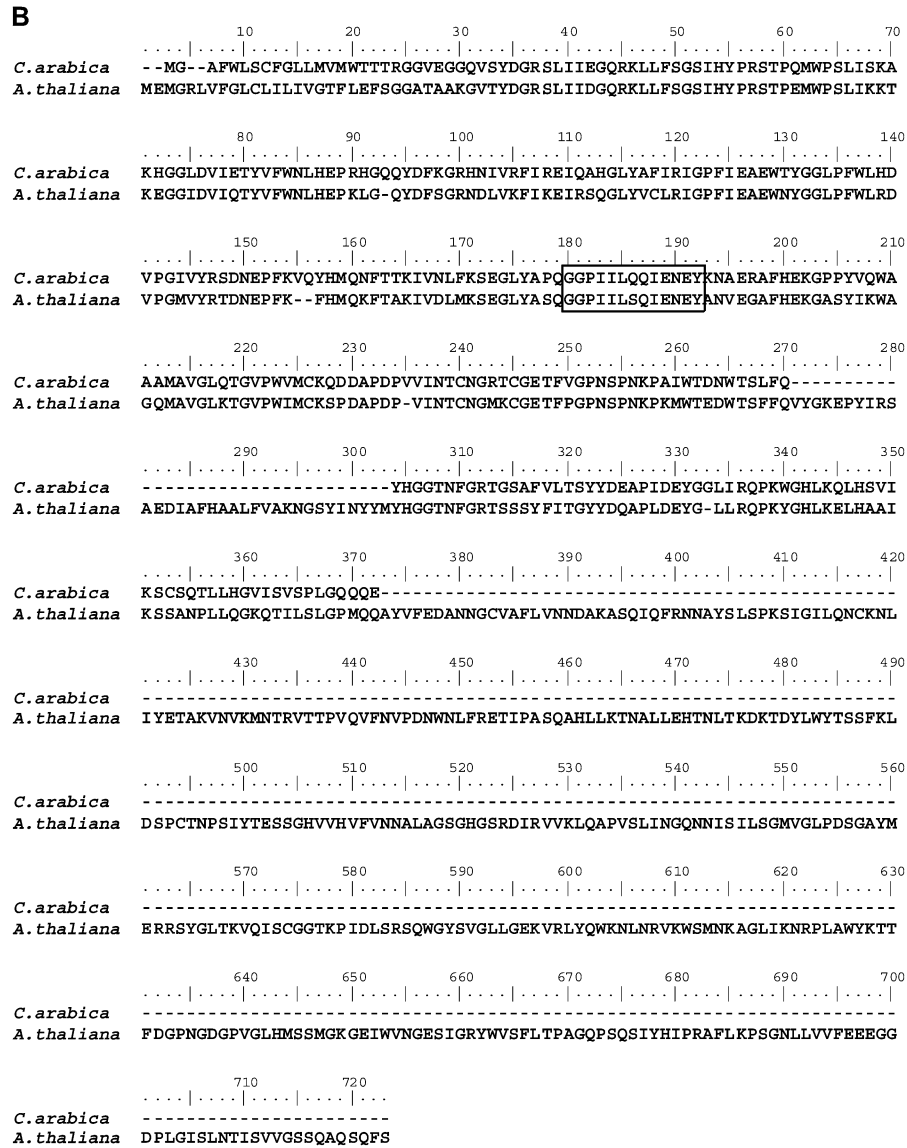


Fig. 2. Continued

peptide at the polypeptide N-terminus (signal peptide probability=1.000) with a maximum cleavage site probability (0.834) between glycine residues at positions 21 and 22 (Fig. 2A). This result was confirmed by the TargetP (Emanuelsson *et al.*, 2000) algorithm, indicating that the mature polypeptide should be targeted to the secretory pathway, where it might be involved in cell wall galactosyl modification. The presence of a signal peptide in the unprocessed form of the CaβGal enzyme indicates that it should be transported into the endoplasmic reticulum, and thence to the cell walls, where it is functionally active. However, immunolocalization and cell fractionation experiments will be required to confirm the cellular localization of CaβGal in coffee fruits.

Detailed analysis of the deduced *CaβGal* translated nucleotide sequence clearly indicated the presence of the glycosyl hydrolase superfamily 42 conserved motif, within

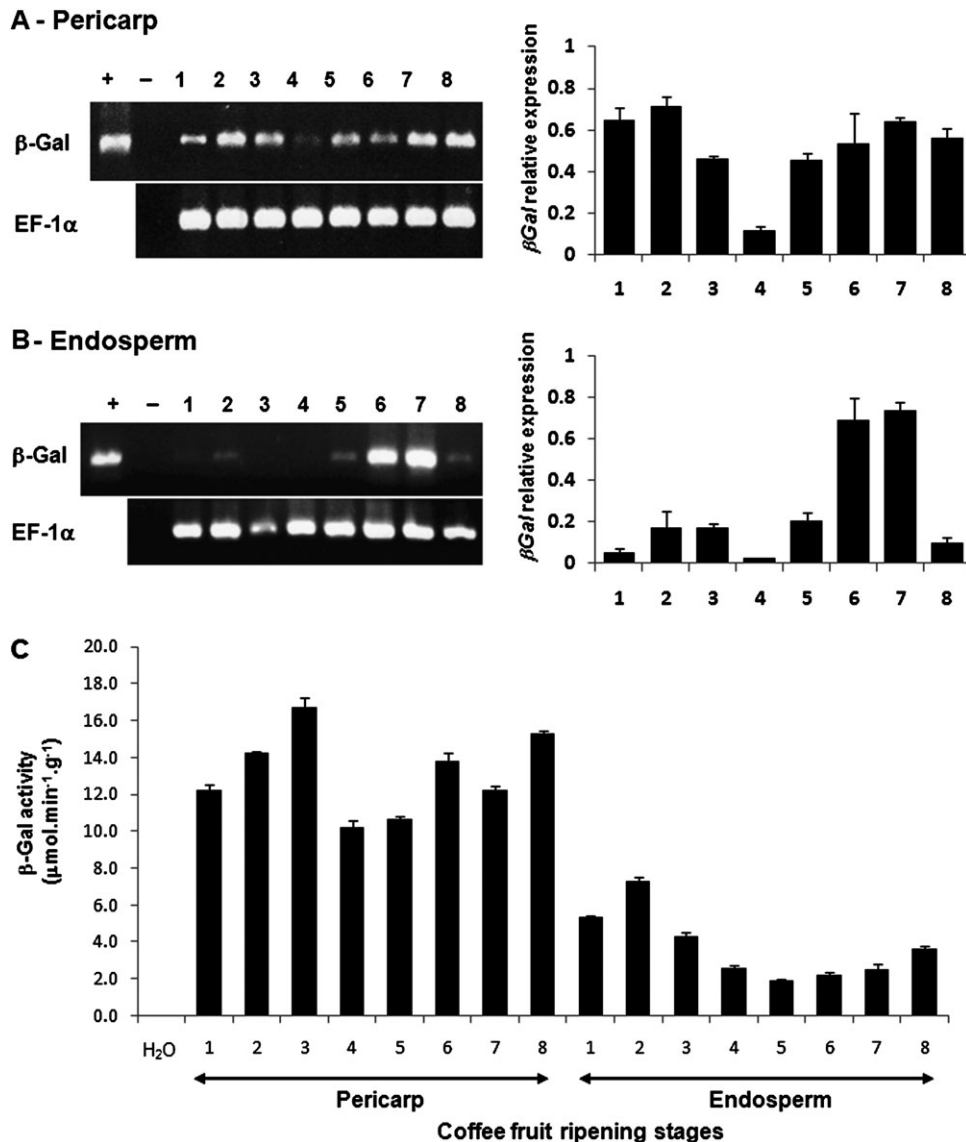
which the glycosyl hydrolase family 35 active site sequence G-G-P-I-I-L-Q-Q-I-E-N-E-Y was identified (<http://ca.expasy.org/prosite/>) (Fig. 2A). This amino acid sequence is a characteristic motif specific to β-galactosidases; however, slight modifications within this region are also observed (Trainotti *et al.*, 2001; Triantafillidou and Georgatsos, 2001; Ahn *et al.*, 2007).

The alignment of the CaβGal predicted amino acid sequence with a closely related β-galactosidase protein sequence from *A. thaliana* (CAB 64727) available at the NCBI GenBank database indicated that ~45.64% of the *CaβGal* coding sequence has been cloned (Fig. 2B). In this analysis, it could also be observed that the aligned amino acid sequences presented 74.25% homology, and the glycosyl hydrolase family 35 motif G-G-P-I-I-L-[Q/S]-Q-I-E-N-E-Y, characteristic of β-galactosidases, was clearly conserved between these two species.

$\beta$ -Galactosidase transcription analysis, *in vitro* enzymatic activity and *in vivo* histochemical assay

A semi-quantitative  $\beta$ -galactosidase transcription analysis from coffee fruit pericarp and endosperm was performed by RT-PCR, and *in vitro* enzymatic activity was evaluated based on the enzymatic degradation of the substrate PNPG by  $\beta$ -galactosidase into galactose and *p*-nitrophenol. Transcription analysis revealed that *Ca $\beta$ Gal* was not evenly expressed in both tissues during fruit development (pericarp, Fig. 3A; endosperm, Fig. 3B), and both presented a similar pattern with a transcription peak in active growing

immature fruits (Fig. 1, stage 2) and another peak in near fully ripe coffee fruits (Fig. 1, stages 6 and 7). Considering *Ca $\beta$ Gal* expressed in the pericarp, one peak was clearly observed in fast growing coffee fruit (stage 2), followed by an expression decay (stages 4 and 5) and a second peak when fruits were fully ripe (stage 7) (Fig. 3A). However, despite two peaks of *Ca $\beta$ Gal* expressed in the endosperm also being observed, initially there was a low transcription level in immature fast growing coffee fruit (stage 2), followed by no detectable expression in the next two phases. As fruits went into more advanced ripening stages, the endosperms presented a sharp increase in  $\beta$ -galactosidase



**Fig. 3.** Expression of  $\beta$ -galactosidase in field-grown *C. arabica* fruits at different ripening stages. (A) Pericarp *Ca $\beta$ Gal* RT-PCR products are presented in the left upper lane. Constitutively expressed transcripts of the elongation factor 1 $\alpha$  gene (*EF-1 $\alpha$* ) by RT-PCR were used as control and are presented in the left lower lane. A histogram presenting the relative expression of *Ca $\beta$ Gal* and *EF-1 $\alpha$*  band intensity is presented in the right panel; number of biological replicates=3. (B) Endosperm *Ca $\beta$ Gal* RT-PCR products are presented in the left upper lane. Constitutively expressed transcripts of *EF-1 $\alpha$*  by RT-PCR were used as the control and are presented in the left lower lane. A histogram presenting the relative expression of *Ca $\beta$ Gal* and *EF-1 $\alpha$*  band intensity is presented in the right panel; number of biological replicates=3; (+) positive control, plasmid carrying  $\beta$ -galactosidase EST; (-) negative control, water. (C) *Coffea arabica*  $\beta$ -galactosidase enzymatic activity during fruit development is presented for both pericarp and endosperm; number of biological replicates=3.

expression with a transcription peak at nearly full mature stages (stage 6 and 7). A low *CaβGal* expression in the endosperm was also observed in the last stage when coffee fruit started to dehydrate (Fig. 3B). Considering β-galactosidase enzymatic activity during coffee fruit development, it could be clearly observed that coffee pericarp also presented two peaks (Fig. 3C), with a similar pattern observed during transcriptional analysis (Fig. 3A). On the other hand, although it could be observed that coffee β-galactosidase enzymatic activity also presented peaks (one during the early endosperm developing stage and another when coffee fruit were fully ripe), the enzymatic activity did not reflect the transcription level observed in the endosperm (Fig. 3B). This unexpected result could be related to some post-transcription regulation mechanism, which needs to be further investigated. It should be noted that *in vitro* β-galactosidase enzymatic activity in coffee pericarps presented, approximately twice the activity compared with the endosperm. This result suggests that specific hydrolytic events involving this cell wall-degrading enzyme are more intense in the pericarps than in the endosperm, probably related to the composition of the cell wall in these tissues.

In order to evaluate both the spatial and temporal distribution of β-galactosidase enzymatic activity in *C. arabica* fruits, an *in vivo* histochemical analysis was performed (Fig. 1). The blue areas indicate β-galactosidase activity due to the enzymatic degradation of X-Gal substrate yielding galactose and 5-bromo-4-chloro-3-hydroxyindole, which is further oxidized into 5,5'-dibromo-4,4'-dichloro-indigo, an insoluble blue product (Sambrook and Russell, 2001). Both longitudinal and transversal fresh stained coffee fruit sections revealed areas with similar staining patterns. A strong expression was observed in endosperms at the initial development stages (phases 1–3), decreasing during the transition from fully developed green coffee fruit into the initial ripening stages (phases 4 and 5), followed by another peak activity when fully ripe (phases 6 and 7), and decreasing in the final developmental stage when fruits began dehydration (stage 8) (Fig. 1, upper lanes). In the pericarp, β-galactosidase expression presented a similar pattern with two peaks: the first peak in initial developmental phase 3, and a second peak when coffee fruits were fully ripe (phases 6 and 7), similar to β-galactosidase expression in the endosperm (Fig. 1, upper lanes).

*In vivo* β-galactosidase histochemical analysis presented a similar pattern to that observed by *in vitro* β-galactosidase enzymatic activity, confirming that this gene was not uniformly expressed throughout coffee fruit development. Both presented two distinct peaks during coffee fruit development as observed by RT-PCR, indicating that this gene might be involved in specific cell wall events at different moments during coffee fruit development. The inconsistent *CaβGal* transcription and enzymatic activity observed mainly in the coffee endosperm suggests that post-transcriptional gene expression regulation might be related to *CaβGal* enzymatic activity. These results observed in coffee are not unique. Ross *et al.* (1994) observed similar

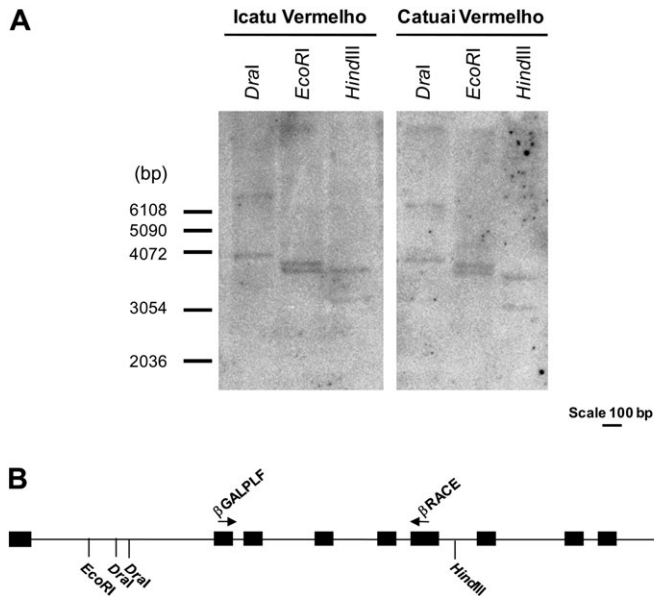
results in apple fruit maturation, detecting strong β-galactosidase activity without detectable expression by using northern blot analysis. In coffee fruit development, Marraccini *et al.* (2005) studied the expression of another galactosidase (α-galactosidase) and observed that even though transcription was concomitant with enzymatic activity, the enzymatic activity persisted despite the fact that no transcription was detected in the following phase analysed. On the other hand, Golden *et al.* (1993) reported an increased β-galactosidase enzymatic activity as coffee fruits matured and ripened. Immature fruits which presented the lowest activity levels and the highest β-galactosidase activities were observed as fruits were at the semi-ripe stage (equivalent to phase 6, Fig. 1), with a 4-fold increase in specific activity compared with the initial phase analysed. These results partially confirmed the present data, but a detailed analysis cannot be performed since their assayed extracts were derived from entire berries. More recently, Joët *et al.* (2009) analysed several different biochemical pathways during coffee seed development (the pericarps were excluded from these analyses) by real-time RT-PCR assays, including carbohydrate metabolism. Among several enzymes involved in this complex pathway, β-galactosidase expression was evaluated. These authors also reported the presence of two *βGal* transcription peaks during coffee seed (endosperm) development with a very similar pattern to that observed in the present RT-PCR analysis, confirming the results.

#### *Southern blot and isoform analysis*

The genomic organization of the *CaβGal* gene was investigated by Southern blot analysis under medium stringency (Fig. 4A). Genomic DNA was single-digested with *DraI*, *EcoRI*, and *HindIII*, none of which presented restriction sites within the 426 bp β-galactosidase cDNA probe used for hybridization (Fig. 4B). Two slight bands were observed with *DraI*, *EcoRI*, and *HindIII* digestions. The same hybridization pattern was observed in both Icatu Vermelho and Catuai Vermelho *C. arabica* cultivars, suggesting that this gene pattern is conserved in these cultivars. These results indicated that there were at least two copies of β-galactosidase genes in the *C. arabica* genome. However, gene copy number is not a trivial issue in this species, because it is an amphidiploid resulting from a natural cross between diploid species *C. eugenoides* and *C. canephora* (Lashermes *et al.*, 1999). Hence it could be interpreted as either the result of two homeologous loci (one in each subgenome) or the presence of two duplicate/paralogous genes rather than two isoforms of the same gene.

Even though the focus here was on β-galactosidase expressed in coffee fruits at different developmental stages, β-galactosidase expression was also observed by PCR amplification of cDNAs derived from young and adult coffee leaves, petals, and stems (data not presented). All tissues amplified a single 501 bp DNA fragment that was sequenced and aligned. The sequence alignment indicated the presence of two distinct groups of PCR-amplified





**Fig. 4.** Genomic DNA analysis. (A) Southern blot of *C. arabica* Icatu Vermelho and *C. arabica* Catuai Vermelho genomic DNA digested with *DraI*, *EcoRI*, and *HindIII* restriction enzymes and hybridized with a  $\beta$ -galactosidase gene probe. Molecular markers are indicated. (B) Schematic diagram of the partially cloned *Ca $\beta$ Gal* genomic DNA indicating the exons (boxes) and the introns (lines). A restriction map is also presented, as well as the position of the primers used to amplify the probe from *Ca $\beta$ Gal* EST.

*Ca $\beta$ Gal*: one group composed of coffee fruit pericarps, endosperms, petals, and stems, and the other composed of young and adult coffee leaves. These two groups presented 71.06% homology considering the nucleotide sequence (Fig. 5A), suggesting that there are, at least, two  $\beta$ -galactosidase isoforms present in the *C. arabica* genome and confirming the Southern blot results (Fig. 4A). Analysis of the predicted amino acid sequence of both isoforms presented 74.25% homology (Fig. 5B). It should be noted that, despite presenting amino acid divergence, both isoforms conserved the glycosyl hydrolase family 35 motif G-G-P-I-I-(L/M)-Q-Q-I-E-N-E-Y which is characteristic of  $\beta$ -galactosidase (<http://ca.expasy.org/prosite/>) (Trainotti *et al.*, 2001; Triantafyllidou and Georgatsos, 2001; Ahn *et al.*, 2007).

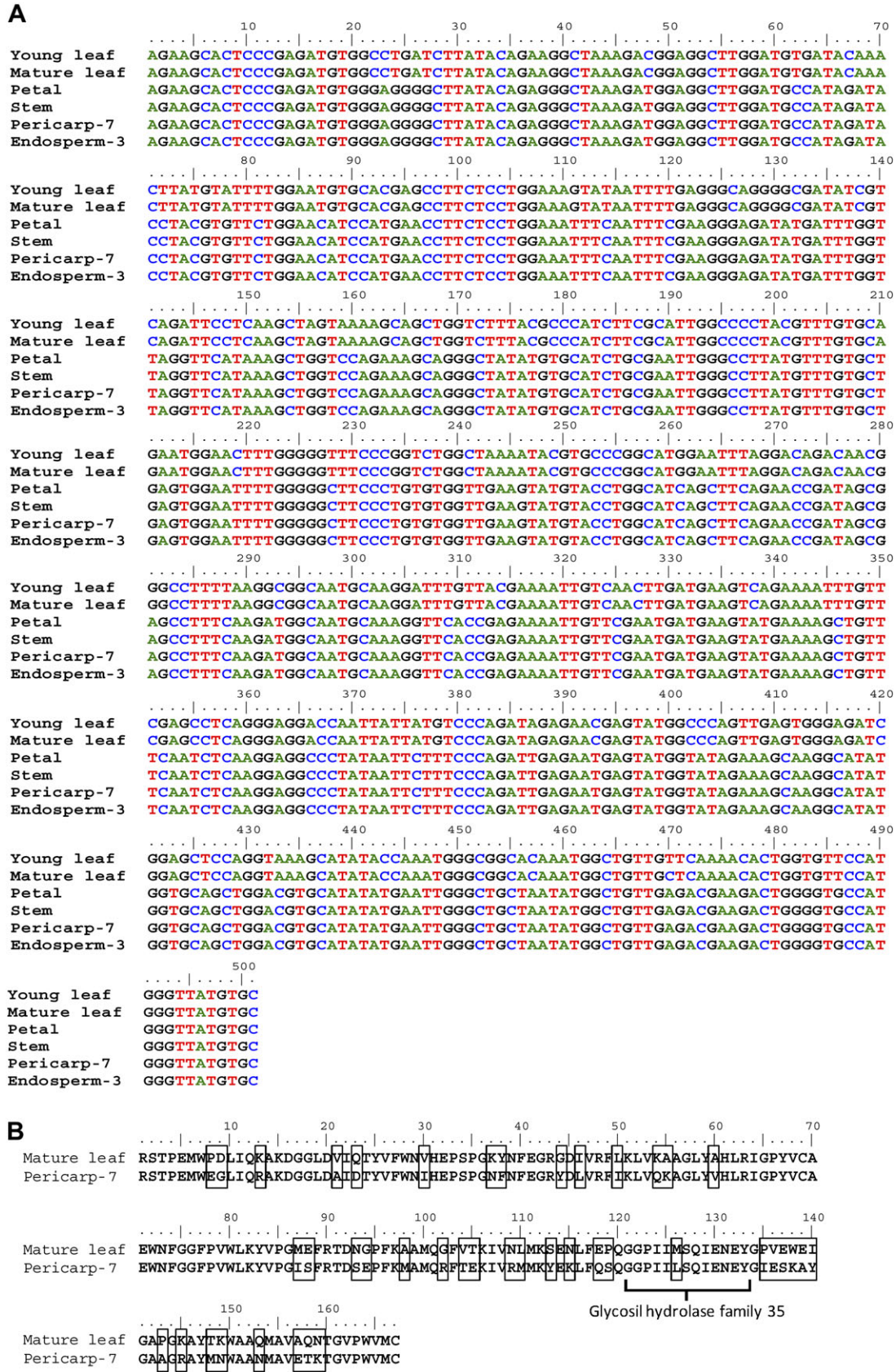
The presence of *Ca $\beta$ Gal* isoforms suggests that this gene might be involved in specific metabolic events which are taking place in different tissues and organs of coffee plants, mainly related to the composition of the cell wall matrix. This appears to be the case for *Ca $\beta$ Gal*, since *in silico* analysis predicted the presence of a signal peptide sequence. However, it should be noted that not all  $\beta$ -Gals present metabolic pathway signaling (Ahn *et al.*, 2007). In coffee fruits, the sequenced *Ca $\beta$ Gal* genes expressed in coffee pericarp and endosperm were identical (Fig. 5A), despite showing different transcription levels and enzymatic activity throughout coffee fruit development and ripening (Fig. 3A, B). This preliminary result suggests that gene expression might be regulated in the promoter region, either by tissue-

specific promoters or by control by specific transcription factors. These hypotheses should be further investigated by cloning the 5'-untranslated region (UTR) from both tissues. Another point to be considered is regarding *Ca $\beta$ Gal* RT-PCR analysis: RT-PCR fragments amplified at early developmental stages (endosperm at stage 3, Fig. 1) and at late stages (pericarp at stage 7, Fig. 1) were sequenced and both DNA sequences were identical (Fig. 5A), indicating that *Ca $\beta$ Gal* transcription analysis was mainly related to the isoform expressed in coffee fruits. Nevertheless, the possibility of the coffee leaf isoform being expressed cannot be definitely excluded, since the set of primers used was not specific to the coffee fruit isoform and every single amplified fragment was not sequenced. The activity of *Ca $\beta$ Gal* in coffee leaves, stems, and petals is beyond the scope of this study and should be addressed in the future in order to understand its involvement in these organs.  $\beta$ -Galactosidase isoforms have been identified in several plant species, including by Prasanna *et al.* (2005) who identified three  $\beta$ -galactosidase isoforms in mango. In tomato, Smith and Gross (2000) identified seven  $\beta$ -galactosidase isoforms, but only isoform TGB4 was active during fruit ripening; and, in avocado (*Persea americana*), Tateishi *et al.* (2001) identified not only three  $\beta$ -galactosidase isoforms, but also their unique expression pattern during fruit ripening. In strawberry fruit maturation, Trainotti *et al.* (2001) identified three  $\beta$ -galactosidase isoforms, but only isoform *Fa $\beta$ Gal1* increased expression during fruit maturation. These results suggest that  $\beta$ -galactosidase isoforms are not unique to *C. arabica*. They appear to be ubiquitously present in several plant species under the control of a multigene family, but not all isoforms appear to be related to ripening.

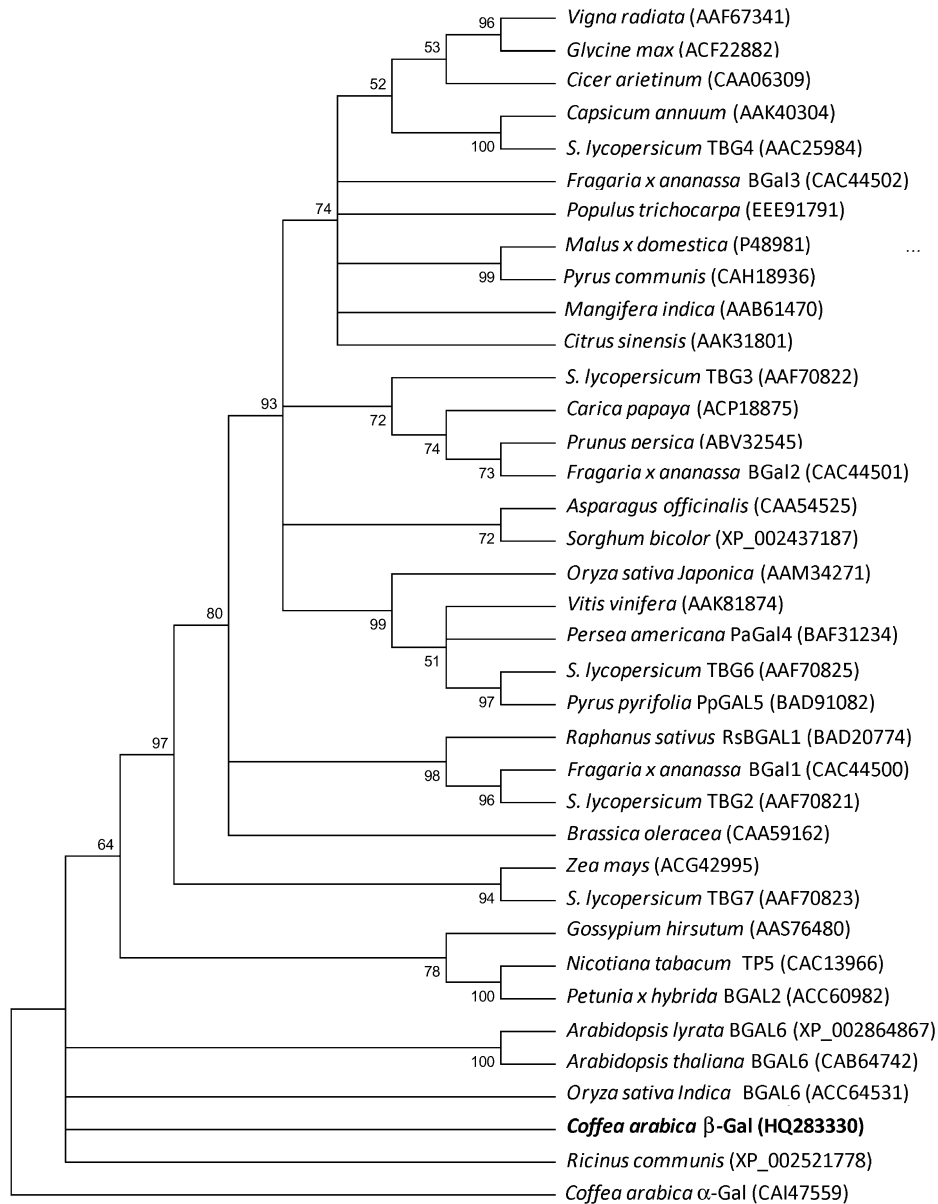
#### Molecular phylogenetic analysis

A  $\beta$ -galactosidase hypothetical evolutionary cladogram was constructed based on the comparison of *Ca $\beta$ Gal* translated coding sequences with another 36  $\beta$ -galactosidase amino acid sequences from different plant species obtained by Blastx search using the NCBI database. Sequences were aligned, manually edited, and phylogenetic analysis was performed using the MP algorithm with a 50% cut-off limit (Fig. 6).

The aligned amino acid sequences were 90.80% variable and 63.22% were parsimonious informative. Analysis of the  $\beta$ -galactosidase phylogenetic tree clearly indicated a similarity between coffee (*C. arabica*), rice (*O. sativa* subspecies Japonica), castor bean (*R. communis*), and *A. thaliana*, placing them genetically apart from other species. However, the high level of heterogeneity observed in the database analysis separated isoforms from the same species, such as in strawberry (*Fa $\beta$ Gal1*, *Fa $\beta$ Gal2*, and *Fa $\beta$ Gal3*), and tomato (several TBGs) into different phylogenetic groups. A clear organizational pattern could not be identified, because it was not possible to correlate this hypothetical evolutionary tree either with species or with family. Similar results were also observed by Ahn *et al.* (2007), where  $\beta$ -galactosidase isoforms from *A. thaliana*, *S. lycopersicon*,



**Fig. 5.**  $\beta$ -Galactosidase isoforms from young and adult *C. arabica* leaves, stems, petals, endosperms, and pericarps. (A) Coffee  $\beta$ -galactosidase EST isoform alignment. (B) Predicted amino acid sequence alignment from the two *C. arabica* isoforms. Open boxes highlight the differences in the amino acid sequences.



**Fig. 6.** Consensus  $\beta$ -galactosidase phylogenetic hypothesis based on the *C. arabica*  $\beta$ -galactosidase deduced amino acid sequence with 36 other plant  $\beta$ -galactosidase proteins using the MP algorithm. The phylogenetic tree was drawn using MEGA4 with a 50% cut-off limit. Bootstrap percentage appears on each branch.

and *O. sativa* were spread throughout their proposed phylogenetic  $\beta$ -galactosidase tree.  $\beta$ -Galactosidases are classified based on their similar enzymatic activities; however, because the composition of plant cell walls is highly variable and dynamic throughout plant growth and development (Campbell and Braam, 1999), the results suggest that  $\beta$ -galactosidases might have evolved in distinct patterns in order to cope with specific biochemical cell wall modification requirements. Despite the observed high variability in amino acid sequence, the biochemically active motifs were conserved, allowing them to perform as  $\beta$ -galactosidases.

This preliminary evolutionary hypothesis should be considered with caution since it was built on a partially sequenced *Ca $\beta$ Gal* translated coding nucleotide sequence.

A more robust analysis must be carried out when *C. arabica*  $\beta$ -galactosidase sequencing is completed, to obtain a better understanding of  $\beta$ -galactosidase gene evolution within plants.

The knowledge acquired with the current molecular and biochemical characterization of *Ca $\beta$ Gal* revealed a more complex expression pattern than previously expected, which now should be put into context with other coffee cell wall-degrading enzymes expressed during fruit development in order to fully understand coffee ripening. The participation of the *Ca $\beta$ Gal* gene and its impact on coffee fruit ripening should be further investigated by gene silencing through genetic engineering, producing biotech coffee plants with reduced *Ca $\beta$ Gal* expression, followed by analysis of coffee fruit ripening. Due to the fact that *Ca $\beta$ Gal* gene expression

was observed in both vegetative and reproductive tissues, it would be interesting to investigate further its influence on plant growth and development. Furthermore, biochemical characterization of carbohydrate metabolism during coffee fruit ripening should also be exploited, correlating the polysaccharide profiles with all known coffee cell wall-degrading enzymes. This information could lead to important biotechnological inputs not only for plant breeding, but also for the coffee industry.

## Acknowledgements

We would like to acknowledge the collaboration of Marie-Christine Combes (IRD, Montpellier, France) for her work on the identification of  $\beta$ -galactosidase clones in *Coffea* BAC libraries.

## References

- Ahn YO, Zheng M, Bevan DR, et al.** 2007. Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 35. *Phytochemistry* **68**, 1510–1520.
- Antal Z, Rascle C, Fèvre M, Bruel C.** 2004. Single oligonucleotide nested PCR: a rapid method for the isolation of genes and their flanking regions from expressed sequence tags. *Current Genetics* **46**, 240–246.
- Balasubramaniam S, Lee HC, Lazan H, Othman R, Ali ZM.** 2005. Purification and properties of a  $\beta$ -galactosidase from carambola fruit with significant activity towards cell wall polysaccharides. *Phytochemistry* **66**, 153–163.
- Biles CL, Bruton BD, Russo V, Wall MM.** 1997. Characterisation of  $\beta$ -galactosidase isozymes of ripening peppers. *Journal of the Science of Food and Agriculture* **75**, 237–243.
- Bradbury AGW, Halliday DJ.** 1990. Chemical structures of green coffee bean polysaccharides. *Journal of Agricultural and Food Chemistry* **38**, 389–392.
- Bradford MM.** 1976. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein–dye binding. *Analytical Biochemistry* **72**, 248–254.
- Brummell DA, Harpster MH.** 2001. Cell wall metabolism in fruit softening and quality and its manipulation in transgenic plants. *Plant Molecular Biology* **47**, 311–340.
- Buckeridge MS, Reid JS.** 1994. Purification and properties of a novel  $\beta$ -galactosidase or exo- $\beta$ -(1,4)-galactanase from the cotyledons of germinated *Lupinus angustifolius* L. seeds. *Planta* **192**, 502–511.
- Campbell P, Braam J.** 1999. Xyloglucan endotransglycosylases: diversity of genes, enzymes and potential wall-modifying functions. *Trends in Plant Science* **4**, 361–366.
- Chantarangsee M, Tanthanuch W, Fujimura T, Fry SC, Cairns JK.** 2007. Molecular characterization of  $\beta$ -galactosidases from germinating rice (*Oryza sativa*). *Plant Science* **173**, 118–134.
- De Castro RD, Marraccini P.** 2006. Cytology, biochemistry and molecular changes during coffee fruit development. *Brazilian Journal of Plant Physiology* **18**, 175–199.
- De Veau EJI, Gross KC, Huber DJ, Watada AE.** 1993. Degradation and solubilization of pectin by  $\beta$ -galactosidases purified from avocado mesocarp. *Physiologia Plantarum* **87**, 279–285.
- Dean GH, Zheng H, Tewari J, et al.** 2007. The *Arabidopsis* MUM2 gene encode a  $\beta$ -galactosidase required for the production of seed coat mucilage with correct hydration properties. *The Plant Cell* **19**, 4007–4021.
- Doyle JJ, Doyle JL.** 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**, 11–15.
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H.** 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols* **2**, 953–971.
- Emanuelsson O, Nielsen H, Brunak S, von Heijne G.** 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**, 1005–1016.
- Esteban R, Dopico B, Munoz FJ, Romo S, Martín I, Labrador E.** 2003. Cloning of a *Cicer arietinum*  $\beta$ -galactosidase with pectin-degrading function. *Plant and Cell Physiology* **44**, 718–725.
- Esteban R, Labrador E, Dopico B.** 2005. A family of  $\beta$ -galactosidase cDNAs related to development of vegetative tissue in *Cicer arietinum*. *Plant Science* **168**, 457–466.
- Felsenstein J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
- Fischer RL, Bennett AB.** 1991. Role of cell wall hydrolases in fruit ripening. *Annual Review of Plant Physiology and Plant Molecular Biology* **42**, 675–703.
- Geromel C, Ferreira LP, Bottcher A, Pot D, Pereira LFP, Leroy T, Vieira LGE, Mazzafera P, Marraccini P.** 2008. Sucrose metabolism during fruit development in *Coffea racemosa*. *Annals of Applied Biology* **152**, 179–187.
- Geromel C, Ferreira LP, Guerreiro SMC, Cavalari AA, Pot D, Pereira LFP, Leroy T, Vieira LGE, Mazzafera P, Marraccini P.** 2006. Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. *Journal of Experimental Botany* **57**, 3243–3258.
- Giovannoni JJ, DellaPenna D, Bennett AB, Fischer RL.** 1989. Expression of a chimeric polygalacturonase gene in transgenic *rin* (ripening inhibitor) tomato fruit results in polyuronide degradation but not fruit softening. *The Plant Cell* **1**, 53–63.
- Golden KD, John MA, Kean EA.** 1993.  $\beta$ -Galactosidase from *Coffea arabica* and its role in fruit ripening. *Phytochemistry* **34**, 355–360.
- Gross KC.** 1984. Fractionation and partial characterization of cell walls from normal and non-ripening mutant tomato fruit. *Physiologia Plantarum* **62**, 25–32.
- Gross KC, Sams CE.** 1984. Changes in cell wall neutral sugar composition during fruit ripening: a species survey. *Phytochemistry* **23**, 2457–2461.
- Joët T, Laffargue A, Salmona J, Doubeau S, Descroix F, Bertrand B, de Kochko A, Dussert S.** 2009. Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. *New Phytologist* **182**, 146–162.

- Lashermes P, Combes M-C, Robert J, Trouslot P, D'Hont A, Anthony F, Charrier A.** 1999. Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics* **261**, 259–266.
- Li S-C, Han J-W, Chen K-C, Chen C-S.** 2001. Purification and characterization of isoforms of  $\beta$ -galactosidases in mung bean seedlings. *Phytochemistry* **57**, 349–359.
- MacGregor GR, Overbeek PA.** 1991. Use of a simplified single-site PCR to facilitate cloning of genomic DNA sequences flanking a transgene integration site. *PCR Methods and Applications* **1**, 129–135.
- Maluf MP, Silva CC, Oliveira MPA, Tavares AG, Silvarolla MB, Filho OG.** 2009. Altered expression of the caffeine synthase gene in a naturally caffeine-free mutant of *Coffea arabica*. *Genetics and Molecular Biology* **32**, 802–810.
- Marraccini P, Rogers WJ, Allard C, André ML, Caillet V, Lacoste N, Lausanne F, Michaux S.** 2001. Molecular and biochemical characterization of endo- $\beta$ -mannanases from germinating coffee (*Coffea arabica*) grains. *Planta* **213**, 296–308.
- Marraccini P, Rogers WJ, Caillet V, Deshayes A, Granato D, Lausanne F, Lechat S, Pridmore D, Pétiard V.** 2005. Biochemical and molecular characterization of  $\beta$ -D-galactosidase from coffee beans. *Plant Physiology and Biochemistry* **43**, 909–920.
- Noir S, Patheyron S, Combes MC, Lashermes P, Chalhoub B.** 2004. Construction and characterization of a BAC library for genome analysis of the allotetraploid coffee species (*Coffea arabica* L.). *Theoretical and Applied Genetics* **109**, 225–230.
- Ochman H, Gerber AS, Hart DL.** 1988. Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621–623.
- Ogasawara W, Abe K, Nakajima T.** 2007. Pepper  $\beta$ -galactosidase 1 (PBG1) plays a significant role in fruit ripening in bell pepper (*Capsicum annuum*). *Bioscience, Biotechnology, and Biochemistry* **71**, 309–322.
- Prasanna V, Prabha T, Tharanathan R.** 2005. Multiple forms of  $\beta$ -galactosidase from mango (*Mangifera indica* L. Alphonso) fruit pulp. *Journal of the Science of Food and Agriculture* **85**, 797–803.
- Redgwell R, Fischer M.** 2006. Coffee carbohydrates. *Brazilian Journal of Plant Physiology* **18**, 165–174.
- Redgwell RJ, Melton LD, Brasch DJ.** 1992. Cell wall dissolution in ripening kiwifruit (*Actinidia deliciosa*). Solubilization of the pectic polymers. *Plant Physiology* **98**, 71–81.
- Rogers HJ, Maund SL, Johnson LH.** 2001. A  $\beta$ -galactosidase-like gene is expressed during tobacco pollen development. *Journal of Experimental Botany* **52**, 67–75.
- Rogers WJ, Michaux S, Bastin M, Bucheli P.** 1999. Changes to the content of sugars, sugar alcohols, myo-inositol, carboxylic acids and inorganic anions in developing grains from different varieties of Robusta (*Coffea canephora*) and Arabica (*C. arabica*) coffees. *Plant Science* **149**, 115–123.
- Ross GS, Wegrzyn T, MacRae EA, Redgwell RJ.** 1994. Apple  $\beta$ -galactosidase. Activity against cell wall polysaccharides and characterization of a related cDNA clone. *Plant Physiology* **106**, 521–528.
- Sakurai N, Nevins DJ.** 1997. Relationship between fruit softening and wall polysaccharides in avocado (*Persea americana* Mill) mesocarp tissues. *Plant and Cell Physiology* **38**, 603–610.
- Salmona J, Dussert S, Descroix F, Kochko A, Bertrand B, Joët T.** 2008. Deciphering transcriptional networks that govern *Coffea arabica* seed development using combined cDNA array and real-time RT-PCR approaches. *Plant Molecular Biology* **66**, 105–124.
- Sambrook J, Russell DW.** 2001. *Molecular cloning: a laboratory manual*, 3rd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Sekimata M, Ogura K, Tsumuraya Y, Hasimoto Y, Yamamoto S.** 1989. A  $\beta$ -galactosidase from radish (*Raphanus sativus* L.) seeds. *Plant Physiology* **96**, 282–292.
- Smith DL, Abbott JA, Gross KC.** 2002. Down-regulation of tomato  $\beta$ -galactosidase 4 results in decreased fruit softening. *Plant Physiology* **129**, 1755–1762.
- Smith DL, Gross KC.** 2000. A family of at least seven  $\beta$ -galactosidase genes is expressed during tomato fruit development. *Plant Physiology* **123**, 1173–1183.
- Smith DL, Starrett DA, Gross KC.** 1998. A gene coding tomato fruit  $\beta$ -galactosidase II is expressed during fruit ripening. *Plant Physiology* **117**, 417–423.
- Tamura K, Dudley J, Nei M, Kumar S.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596–1599.
- Tateishi A, Inoue H, Shiba H, Yamaki S.** 2001. Molecular cloning of  $\beta$ -galactosidase from Japanese pear (*Pyrus pyrifolia*) and its gene expression with fruit ripening. *Plant and Cell Physiology* **42**, 492–498.
- Tomkins JP, Yu Y, Miller-Smith H, Frisch DA, Woo SS, Wing RA.** 1999. A bacterial artificial chromosome library for sugarcane. *Theoretical and Applied Genetics* **99**, 419–424.
- Trainotti L, Spinello R, Piovano A, Spolaore S, Casadoro G.** 2001.  $\beta$ -Galactosidases with a lectin-like domain are expressed in strawberry. *Journal of Experimental Botany* **52**, 1635–1645.
- Triantafyllidou D, Georgatsos JC.** 2001. Barley  $\beta$ -galactosidase: structure, function, heterogeneity, and gene origin. *Journal of Protein Chemistry* **20**, 551–562.
- Wu Z, Burns JK.** 2004. A  $\beta$ -galactosidase gene is expressed during mature fruit abscission of 'Valencia' orange (*Citrus sinensis*). *Journal of Experimental Botany* **55**, 1483–1490.