Characterization and Expression Analysis of Genes Directing Galactomannan Synthesis in Coffee

MARTIAL PRÉ, VICTORIA CAI ЛET, JULIEN SOBILO and JAMES MCCARTHY*
Centre de Recherche Nestlé, 101 Av. Gustave Eiffel, Notre Dame d’Oé, BP 49716 - 37097 Tours, France

Received: 11 January 2008 Returned for revision: 5 March 2008 Accepted: 21 April 2008 Published electronically: 18 June 2008

* For correspondence. E-mail james.mccarthy@rdto.nestle.fr

INTRODUCTION

Plant cell walls are complex and dynamic composites of polysaccharides, proteins plus lignin, and the polysaccharide component can be subdivided into three broad categories: pectins, hemicelluloses and cellulose. The cell wall provides mechanical support to plant cells and acts as a primary barrier against pathogen attack, mechanical injury and other environmental stresses. The biophysical properties of the wall are also responsible for the strength and flexibility of plant tissue during cell expansion (Somerville et al., 2004). In addition, degradation of seed cell wall carbohydrates constitutes a predominant source of energy during germination in various species. As well as being important for the plant, cell wall-derived carbohydrates have many commercial applications, ranging from food, fuel and textiles to building materials. For example, the consistency of jellies and jams are derived from food, fuel and textiles to building materials. For

The seeds of guar (Cyamopsis tetragonoloba) and coffee (Coffea arabica and Coffea canephora). In coffee, the galactomannans can represent up to 25% of the mass of the mature green coffee grain, and they exert a significant influence on the production of different types of coffee products. The objective of the current work was to isolate and characterize cDNA encoding proteins responsible for galactomannan synthesis in coffee and to study the expression of the corresponding transcripts in the developing coffee grain from C. arabica and C. canephora, which potentially exhibit slight galactomannan variations. Comparative gene expression analysis was also carried out for several other tissues of C. arabica and C. canephora.

Methods cDNA banks, RACE-PCR and genome walking were used to generate full-length cDNA for two putative coffee mannan synthases (ManS) and two galactomannan galactosyl transferases (GMGT). Gene-specific probe-primer sets were then generated and used to carry out comparative expression analysis of the corresponding genes in different coffee tissues using quantitative RT-PCR.

Key Results Two of the putative galactomannan biosynthetic genes, ManS1 and GMGT1, were demonstrated to have very high expression in the developing coffee grain of both Coffea species during endosperm development, consistent with our proposal that these two genes are responsible for the production of the majority of the galactomannans found in the grain. In contrast, the expression data presented indicates that the ManS2 gene product is probably involved in the synthesis of the galactomannans found in green tissue.

Conclusions The identification of genes implicated in galactomannan synthesis in coffee are presented. The data obtained will enable more detailed studies on the biosynthesis of this important component of coffee grain and contribute to a better understanding of some functional differences between grain from C. arabica and C. canephora.

Key words: Coffea, galactomannans, mannan synthase, galactomannan galactosyl transferase, coffee grain.

*For correspondence. E-mail james.mccarthy@rdto.nestle.fr

© The Author 2008. Published by Oxford University Press on behalf of the Annals of Botany Company. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org

Coffee grain polysaccharides are primarily the arabinogalactans, plus cellulose and xylglucans, with each representing approximately 17% and 8%, respectively, of the mass in the mature coffee grain (Oosterveld et al., 2003).

Whereas the biochemical composition of the cell wall has been examined in a variety of plant species, only a small proportion of the hundreds of proteins predicted to be involved in cell wall biosynthesis and metabolism have been identified and characterized. Profiling of gene transcript abundance during cotton (Gossypium hirsutum) fibre development resulted in the identification of the first plant cellulose synthase (CesA) gene (Pear et al., 1996). Based on predicted protein sequences, all the putative Arabidopsis glycan synthase genes were discovered and then grouped into nine clearly distinguishable families: the CesA family, which includes the authentic cellulose synthase genes, as well as eight additional families of structurally related genes designated the ‘cellulose synthase-like’ genes (CsiA-CsiH; Richmond and Somerville, 2000). Until recently, the genes responsible for the synthesis of the plant cell mannan were not known. The first gene isolated that encodes a mannan synthase was CtManS from guar seeds, which make very large quantities of galactomannans (Dhugga et al., 2004). The putative guar CtManS cDNA was demonstrated to encode a functional mannan synthase enzyme in transformed soybean somatic embryos. These soybean embryos, which normally have no detectable mannan synthase activity, exhibited significant mannan synthase activity when the CtManS cDNA sequence was over-expressed (Dhugga et al., 2004). Recently, a functional study has been carried out on recombinant proteins generated from a number of CsiA genes from moss (Phycomitrella patens), loblolly pine (Pinus taeda), rice (Oryza sativa) and Arabidopsis thaliana (Liepmann et al., 2005, 2007). The results demonstrated that in all these plant organisms, some members of the CsiA gene family encoded proteins with mannan synthase activity. Similarly, a member of the CsiC family from Arabidopsis was shown to encode for an enzyme involved in xyloglucan biosynthesis (Cocuron et al., 2007), suggesting that the different Csi families participate in the synthesis of distinct polysaccharides.

The enzyme responsible for the transfer of galactose (Gal) residues to a mannose (Man) on the mannan backbone is a member of the glycosyl transferase (GT) family of proteins called galactomannan galactosyltransferase (GMGT). The mannan synthase (ManS) and GMGT enzymes are localized in the membrane of Golgi vesicles and are believed to work together very closely to determine the statistical distribution of galactosyl residues along the mannan chain (Edwards et al., 2004). The Man/Gal ratio varies between the plant species that produce galactomannans (Reid and Meier, 1970; Bailey, 1971), suggesting that the distribution and the degree of galactosyl substitution of the mannan chain synthesized is, in part, determined by the levels and specific activity of the biosynthetic enzymes. For example, the fenugreek (Trigonella foenum-graecum) seed contains a galactomannan that is almost fully Gal substituted (Man/Gal = 1.1), and a gene encoding a GMGT isolated from this plant has been implied to play an important role in this high Man/Gal ratio (Edwards et al., 1999). Consistent with this proposal, over-expression of the fenugreek GMGT in tobacco and Lotus japonicus seed endosperm produced galactomannans with increased Gal substitution compared with non-transformed endosperms (Reid et al., 2003; Edwards et al., 2004).

The precise chemical composition of coffee grain galactomannans has been studied in some detail with several groups reporting degrees of galactosyl substitution ranging from 2:1 to 130:1 (Wolform et al., 1961; Bradbury and Halliday, 1990; Fischer et al., 2001; Redgwell et al., 2003; Oosterveld et al., 2003). Such variable results are likely to be due to the different methodologies used, in addition to differences in the exact developmental stage of the samples used. Indeed, Redgwell et al. (2003) have shown that the Man/Gal ratio changes during coffee grain development, falling from a value of approx. 2:1 to 7:1 during early development to a ratio of 7:1 to 40:1 near maturity. Regardless of these disparities, all the published data indicate that the galactomannans in the mature coffee grain have a rather low level of galactosylation.

Despite the high amount of galactomannan in the coffee grain and the implicit importance of the enzymes that participate in galactomannan synthesis, little information is available on these genes in coffee. Therefore, we have isolated and characterized two full-length cDNAs encoding mannan synthases and identified four cDNAs encoding highly related glycosyl transferases from C. canephora. Expression analysis shows that two of these genes, ManSI and GMGT1, are highly expressed in the developing grain, suggesting that they are involved in most of the galactomannan synthesis seen in the grain. We show that interesting differences exist in the levels of expression of these two genes between arabica and robusta, and this observation is discussed in the context of the differences seen between the extractability of arabica versus robusta coffees (Thaler, 1979; Trugo, 1985).

**MATERIAL AND METHODS**

**Plant material**

Roots, young leaves, branches, flowers and fruit at different stages of development [small green fruit (SG), large green fruit (LG), yellow fruit (Y) and red fruit (R)] were collected from Coffea arabica L. ‘Caturra T2308’ grown under greenhouse conditions (25°C, 70% relative humidity) in Tours, France, and from C. canephora (robusta) ‘BP409’ grown in the field at the Indonesian Coffee and Cacao Research Center (ICCRI), Indonesia. Leaf samples at different stages of development [very young (VYL), young (YL), mature (ML) and old (OL)] were harvested from C. arabica ‘T2308’ and C. canephora ‘BP409’ plants grown under the greenhouse conditions described above. Fresh tissues were frozen immediately in liquid nitrogen and then stored at −80°C until used for RNA extraction. Coffee cherries from an additional six field-grown varieties, at four different developmental stages, were first separated into pericarp and grain tissues before RNA preparation. The construction and analysis of the EST libraries generated from leaves,
Extraction of total RNA and RT-PCR

Samples were ground into a powder and total RNA was extracted using the method described previously (Rogers et al., 1999). Alternatively, for low amounts of tissue (<0.2 g) total RNA was purified using the RNeasy Plant Mini Kit (Qiagen) with an additional on-column RNase-free Dnase I treatment (Qiagen) following the manufacturer’s instructions.

Two different methods were used to obtain the cDNAs. In Method 1, cDNAs were prepared using the protocol described in the Superscript III Reverse Transcriptase kit (Invitrogen) as follows: 1 µg total RNA sample plus 100 ng oligo dT(18) (Sigma) was made up to 12 µL final volume with DEPC-treated water. This mixture was subsequently incubated at 70 °C for 10 min and then rapidly cooled on ice. Next, 4 µL of 5× first strand buffer, 1 µL of 0-1 M DTT, 1 µL of RNase Out Ribonuclease Inhibitor (Invitrogen) and 1 µL of dNTP mix (10 mm each) were added. This reaction mix was pre-incubated at 42 °C for 2 min before adding 1 µL SuperScript III Rhnase H-Reverse transcriptase (200 U µL⁻¹). Subsequently, the reaction mix was incubated at 25 °C for 10 min and then at 42 °C for 50 min, followed with enzyme inactivation by heating at 70 °C for 10 min. Alternatively, cDNAs were prepared following Method 2, identically to the protocol described in the Transcripter Reverse Transcriptase kit from Roche: 2 µg total RNA plus 270 ng poly dT(20) (Sigma) was made up to 12 µL final volume with DEPC-treated water. This mixture was subsequently incubated at 65 °C for 10 min and then rapidly cooled on ice. Next, 4 µL 5× Transcriptor RT buffer, 20 U RNase OUT Ribonuclease inhibitor, 1 mm of each dNTP and 10 U Transcriptor Reverse Transcriptase (Roche) were added to a final volume of 20 µL. This reaction was incubated at 55 °C for 30 min, followed by enzyme inactivation at 85 °C for 5 min. The cDNAs obtained from either method were treated with 2 U RNase H (Invitrogen) at 37 °C for 20 min to degrade remaining mRNAs and were then diluted ten-fold in sterile water and stored at −20 °C for later use in 5’ RACE and QRT-PCR experiments.

5’ RACE reactions (rapid amplification of cDNA ends)

5’ RACE experiments were performed to recover the missing 5’ coding sequence of the putative coffee mannan synthases CcManS1 and CcManS2, the galactosyl transferase CcGMGT1, and the CcXT1 genes. The synthesis of cDNA for the 5’ RACE experiments was carried out using RNA from C. canephora (‘BP409’) at the yellow (Y) stage. The experimental approach is described by Lepelley et al. (2007) and closely follows the methods described in the kit for the 5’ RACE system for Rapid Amplification of cDNA Ends kit (Invitrogen). cDNA was purified using SNAP columns (Invitrogen) according to the manufacturer’s instructions, recovered in 50 µL of sterilized water and then stored at −20 °C. The 5’ RACE experiments all began with a TdT tailing of the SNAP-purified cDNA: 25 µL of the purified cDNA, 11.5 µL of DEPC-treated water, 5 µL of 5× TdT tailing buffer (Invitrogen) and 2.5 µL of 2 mm dCTP. The reactions were then incubated at 94 °C for 3 min, followed by chilling on ice. One µL of TdT was then added and the reaction was incubated for 10 min at 37 °C. The reactions were terminated by heating 10 min at 65 °C and again placed on ice.

The first round of 5’ RACE was performed in a final 50 µL volume, as follows: 5 µL of each tailed cDNA, 5 µL of 10× PCR buffer (ThermoPol buffer), 400 nm of both Gene Specific Primer 1 and AAP primers (see Table 1 for specific primers used), 200 µm of each dNTP, and 2.5 U of Taq DNA polymerase (BioLabs). The first round PCR cycling conditions were: 94 °C for 2 min; then 40 cycles of 94 °C for 1 min, annealing temperature (noted in Table 1) for 1 min, and 72 °C for 2 min. An additional final step of elongation was done at 72 °C for 7 min. The second round PCR reactions were performed in a final 50 µL volume, as follows: 5 µL of 100-fold diluted PCR product from the first round, 5 µL of 10× PCR buffer (LA buffer II Mg⁺⁺ plus), 200 nm of both Gene Specific Primer 2 and AAAP primers (Table 1), 200 µm of each dNTP, and 0.5 U of DNA polymerase Takara LA Taq (Cambrex Bio Science). The cycling protocol was: 94 °C for 2 min; then 40 cycles of 94 °C for 1 min, annealing temperature (Table 1) for 1 min, and 72 °C for 1 min 30 s. An additional final step of elongation was done at 72 °C for 7 min. PCR products from each round were then analysed by agarose gel electrophoresis and ethidium bromide staining.

Genome Walker

Genome Walker libraries were constructed by using the Universal Genome Walker kit according to the manufacturer’s instruction (BD Biosciences Clontech). Briefly, coffee genomic DNA from C. arabica ’T2308’ and C. canephora ’BP409’ plants (2.5–5 µg) was digested at 37 °C overnight with four different restriction enzymes (DraI, EcoRV, PvuII and StuI) in separate reactions. After purification with phenol/chloroform extraction and ethanol precipitation, DNA restriction fragments were ligated to Genome Walker adapters (5'-GTAATACGACTCACTATA GGCCACCGTGGTGCAGGGCACCACCGCTTGTGCTCGTG-3') at 16 °C overnight. In a 50 µL PCR mix, 1 µL of each Genome Walker DNA library was used as a template in the primary PCR, and 1 µL of 100-fold diluted primary PCR products was used as a template in secondary PCR. Primary and secondary PCR were performed in ThermoPol Reaction Buffer (Biolabs) with 200 µm dNTP (each), 2.5 U of Taq DNA polymerase (Biolabs) and with appropriate templates and primer sets (0-2 µm each) as indicated in Table 1 using primer sequences described in Table 2. Primary PCR was started with seven cycles consisting of 94 °C for 25 s and 72 °C for 3 min, followed by 32 cycles consisting of 94 °C for 25 s and 67 °C for 3 min, and a final extension at 67 °C for 7 min. Secondary PCR was started with five cycles consisting of 94 °C for 25 s and 72 °C for 3 min, followed by 20
cycles consisting of 94 °C for 25 s and 67 °C for 3 min, and a final extension at 67 °C for 7 min.

Isolation of full-length cDNA

A partial cDNA clone encoding CcManS1 (pcccs46w24c19) was identified in the Cornell C. canephora EST database. Two rounds of 5' RACE were carried out to recover the missing 5'-end of the cDNA clone pcccs46w24c19. However, this RACE1 fragment was shown to be missing the 5'-end of this gene. Therefore, a new 5' RACE PCR was carried out as described (RACE2). This produced a 340-bp long fragment that overlapped the 5'-end of the CcManS1 RACE1 fragment. From the newly obtained 5'-end CcManS1 sequence and the 3'-end sequence information from the cDNA clone pcccs46w24c19, two flanking primers (ManS-Am3 and ManS-Am2, Table 2) were designed to amplify the complete CcManS1 open reading frame (ORF). The PCR reactions were performed in 50 µL reactions as follows: 5 µL of cDNA made from RNA of C. canephora ('BP409') grain at the yellow developmental stage, 5 µL of 10× PCR buffer (LA PCR Buffer II Mg2+ plus), 800 µM of each gene-specific primer, 200 µM of each dNTP, and 0.5 U of DNA polymerase Takara LA Taq (Cambrex Bio Science). After denaturing at 94 °C for 2 min, the amplification consisted of 35 cycles of 1 min at 94 °C, 1 min 30 s at 47 °C, and 3 min at 72 °C. An additional final step of elongation was done at 72 °C for 7 min. The PCR products were then analysed by agarose gel electrophoresis and ethidium bromide staining. Fragments of the expected

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Tissue</th>
<th>C</th>
<th>Primer sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>CcManS1 RACE1</td>
<td>'BP409' grain yellow</td>
<td>55</td>
<td>AAP / RNAi-Pr2</td>
</tr>
<tr>
<td>Round 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Round 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CcManS1 RACE2</td>
<td>'BP409' grain yellow</td>
<td>55</td>
<td>AAP / ManSynGWR249</td>
</tr>
<tr>
<td>Round 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Round 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CcManS2 GW1</td>
<td>'T2308' genomic DNA</td>
<td>67</td>
<td>API / ManS2GP3</td>
</tr>
<tr>
<td>Round 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Round 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CcManS2 GW2</td>
<td>'T2308' genomic DNA</td>
<td>67</td>
<td>API / ManS2GP6</td>
</tr>
<tr>
<td>Round 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Round 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CcManS2 RACE3</td>
<td>'BP409' pericarp yellow</td>
<td>55</td>
<td>AAP / ManS2R2GP3</td>
</tr>
<tr>
<td>Round 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Round 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CoGMGT1 RACE</td>
<td>'T2308' grain yellow</td>
<td>60</td>
<td>AAP / GMGT_30w15m14_Race4</td>
</tr>
<tr>
<td>Round 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Round 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CcXT1 GW</td>
<td>'BP409' genomic DNA</td>
<td>67</td>
<td>API / GMGT3 GSP3</td>
</tr>
<tr>
<td>Round 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Round 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The primers, the plant tissue from which the template cDNA originated and the annealing temperatures are given for the various 5' RACE and GW PCR reactions. The DNA sequences of the primers are given in Table 3.

**Table 2. DNA sequences of the primers used for the different 5' RACE and Genome Walker experiments**

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence 5’–3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAP</td>
<td>GGCCACCGCTCGACTAGTACGGGIIGGGIIGGGIIGGGIIGGGIIGGG</td>
</tr>
<tr>
<td>AUAP</td>
<td>GGCCACCGCTCGACTAGTAC</td>
</tr>
<tr>
<td>RNAi-Pr2</td>
<td>GAACATGTTGACGAGCCT</td>
</tr>
<tr>
<td>ManSynGWR249</td>
<td>GCCCGCAGGACTTCATTCGTGGAG</td>
</tr>
<tr>
<td>ManSRace2</td>
<td>ATACTTGGTATATCGTTTCCTTCC</td>
</tr>
<tr>
<td>ManSRace1</td>
<td>TGGACATCCAATCACATCGC</td>
</tr>
<tr>
<td>ManS-Am3</td>
<td>CAGGTCGCCCCAGGATCACAAATTTCC</td>
</tr>
<tr>
<td>ManS-Am2</td>
<td>GCCGCCCTCCTTGTGACCAATTTACTTTC</td>
</tr>
<tr>
<td>ManS2-GSP4</td>
<td>ACATTTAGAGTGAGAGTGAGAT</td>
</tr>
<tr>
<td>ManS2-GSP5</td>
<td>TCTCCATACCCAGGCTCCTTAAG</td>
</tr>
<tr>
<td>ManS2-GSP6</td>
<td>AGACAGCCAGCCACATGCC</td>
</tr>
<tr>
<td>ManS2-R2-GSP2</td>
<td>CCCCCACGTCAACTCACTCACCAG</td>
</tr>
<tr>
<td>ManS2-R2-GSP3</td>
<td>CTATTAGATGAGATGAGAT</td>
</tr>
<tr>
<td>ManS2-Fw</td>
<td>TGTTGCCCCCAGATTAT</td>
</tr>
<tr>
<td>GMGT_30w15m14_race2</td>
<td>CAGTGGAGGCAATGAAATGACAAATCTG</td>
</tr>
<tr>
<td>GMGT_30w15m14_Race4</td>
<td>TTCTCCAAGCGTCCCCCAG</td>
</tr>
<tr>
<td>GMGT-Fwd1</td>
<td>TTCTCCATACCCAGGCTCCTTAAG</td>
</tr>
<tr>
<td>GMGT-Rev</td>
<td>CCCCCACGTCAACTCACTCACCAG</td>
</tr>
<tr>
<td>GMGT3-GSP1</td>
<td>GGTCTTTGGATATTCTTGATACTTGG</td>
</tr>
<tr>
<td>GMGT3-GSP2</td>
<td>CAGTGGAGGCAATGAAATGACAAATCTG</td>
</tr>
<tr>
<td>GMGT3-GSP3</td>
<td>TCGAGGCTCGACTAGTAC</td>
</tr>
<tr>
<td>GMGT3-GSP4</td>
<td>CTTCAAGCGTCCCCCAG</td>
</tr>
<tr>
<td>GMGT3-GSP5</td>
<td>GCCGCCCTCCTTGTGACCAATTTACTTTC</td>
</tr>
<tr>
<td>GMGT3-GSP6</td>
<td>ACATTTAGAGTGAGAGTGAGAT</td>
</tr>
<tr>
<td>GMGT-Fwd1</td>
<td>TCTCCATACCCAGGCTCCTTAAG</td>
</tr>
<tr>
<td>GMGT-Rev</td>
<td>AGACAGCCAGCCACATGCC</td>
</tr>
<tr>
<td>GMGT3-GSP1</td>
<td>CCCCCACGTCAACTCACTCACCAG</td>
</tr>
<tr>
<td>GMGT3-GSP2</td>
<td>TTCTCCATACCCAGGCTCCTTAAG</td>
</tr>
<tr>
<td>GMGT3-GSP3</td>
<td>CAGTGGAGGCAATGAAATGACAAATCTG</td>
</tr>
<tr>
<td>GMGT3-GSP4</td>
<td>TTCTCCATACCCAGGCTCCTTAAG</td>
</tr>
<tr>
<td>GMGT3-GSP5</td>
<td>AGACAGCCAGCCACATGCC</td>
</tr>
<tr>
<td>GMGT3-GSP6</td>
<td>CCCCCACGTCAACTCACTCACCAG</td>
</tr>
</tbody>
</table>

Pre´ et al. — Galactomannan Biosynthesis in Coffee Grain

210

Downloaded from https://academic.oup.com/aob/article-abstract/102/2/207/184418 by guest on 18 January 2019
size were then cloned in vector pCR4-TOPO using the TOPO TA Cloning Kit for Sequencing (Invitrogen) according to the instructions given by the manufacturer.

A partial cDNA clone encoding CcManS2 (pcccs30w34j20) was identified in the Cornell C. canephora EST database. Two rounds of Genome Walker followed by one round of 5’ RACE were carried out successively to recover the missing 5’ sequence of CcManS2 (Table 1) using the primers listed in Table 2. From the newly obtained 5’-end CcManS2 sequence and the 3’-end information from the cDNA clone pcccs30w34j20, two flanking primers (ManS2-Fw and ManS2-Rev, Table 2) were designed to amplify the complete CcManS2 open reading frame (ORF). PCR cloning reactions were performed as described above for CcManS1.

A partial cDNA clone encoding GMGT1 (pcccs46w8o23) was identified in the Cornell C. canephora EST database. To recover the missing 5’ sequence, one round of 5’ RACE was carried out on cDNA from C. arabica ‘T2308’ grain tissue (Tables 1 and 2). The fragment obtained was 1120 bp long and overlapped the 5’-end of the cDNA clone pcccs46w8o23. Amplification of the complete GMGT1 ORF was performed as before using the above-mentioned cDNA collection and the primers GMGT-Fwd1 and GMGT-Rev (Table 2).

The full-length sequence of CcXT1 was obtained by generating an in silico contig between the in silico sequence of unigene SGN-U352025 identified in the Cornell EST database and a genomic sequence encoding the 3’ coding region of this gene, which was recovered by the Genome Walker technique using robusta ‘BP409’ genomic DNA and the primers listed in Table 2. The genomic PCR amplicon was 1651 bp long and overlapped the 3’-end of unigene SGN-U352025. The complete in silico contig sequence obtained was 2951 bp long and contains a continuous ORF.

Phylogenetic analyses

Full-length amino acid sequences encoding glycan synthases and glycosyl transferases from Arabidopsis thaliana (At), potato (St), Ipomoea trifida (It), Coffea arabica (Ca), Coffea canephora (Cc), Cyamopsis tetragonolobus (Ct), Senna occidentalis (So), Lotus japonicus (Lj), and fenugreek (Tf) were aligned using ClustalW (Thompson et al., 1994) with the Gonnet protein weight matrix, pairwise gap opening/extension penalties of 10/0.1, and multiple alignment gap opening/extension penalties of 10/0.2. The alignments were performed using the Laser Gene software package (DNASTAR). Rooted trees were constructed from the aligned sequences using the default parameters of the software and bootstrapped (n = 1000 trials) to create the final trees.

Quantitative RT-PCR

The cDNAs used for gene expression analysis by Q-PCR were prepared according to Method 1 for the complete tissue sets of arabica ‘T2308’ and robusta ‘BP409’, and according to Method 2 for the other grain samples plus the leaf developmental samples. Quantitative PCR using TaqMan probes was carried out as described by Lepelley et al. (2007). All reactions (25 μL final volume) were performed in 1 × TaqMan Universal PCR master mix (Applied Biosystems), 800 nM of each gene-specific primer and 200 nM of the corresponding TaqMan probe. Four μL of 100-fold diluted cDNA was used as a template, which corresponded to approximately 0.25 ng of original RNA. The Q-PCR primers and TaqMan probes used were designed with the PRIMER EXPRESS software (Applied Biosystems) and are listed in Table 3. The reaction mixture was incubated at 50 °C for 2 min, then at 95 °C for 10 min, followed by 40 amplification cycles of 15 s at 95 °C and 1 min at 60 °C. Samples were quantified in the GeneAmp 7500 Sequence Detection System (Applied Biosystems). Transcript levels were normalized to the levels of the constitutively expressed control gene RPL39 that codes for a ribosomal protein. The amplification efficiencies of the primers/probes sets were tested on various dilutions of the plasmid containing the corresponding gene and were all found to be near 100%.

RESULTS

Galactomannan-synthesizing enzymes in coffee

Because galactomannans represent the major carbohydrate reserve of the coffee endosperm, we decided to identify and study the genes involved in the synthesis of this polymer from Coffea. Furthermore, considering the possibility that some measurable differences exist between the galactomannans found in the mature grain of canephora (robusta) and arabica Coffea species (Redgwell et al., 2003), we were also

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward</th>
<th>Reverse</th>
<th>MGB Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>MonS1</td>
<td>AATGTCATGTCCTCGCGGCTATGCA</td>
<td>AACTCCGCGTCTGCTTCAAAAGTC</td>
<td>CAAAGCACGAAATAT</td>
</tr>
<tr>
<td>MonS2</td>
<td>CTTGATTAGAAGATTTGGATGCTATTAC</td>
<td>GGAATGAGAATCTGCTGTTGCCATTA</td>
<td>ACATGGAGATAAAGACAGA</td>
</tr>
<tr>
<td>GMGT1</td>
<td>CGCCCTCGCCGCGCCTGA</td>
<td>ATTCGAGAAGCAGGCTCCCA</td>
<td>CCAAGCTCAGGCTCTT</td>
</tr>
<tr>
<td>GMGT2</td>
<td>CCATCGGGAAGTCTGCGTAAAA</td>
<td>ACGAGGGGATTCATTCACTTGTAAA</td>
<td>AAAGAAATCCCGGTAAGAAGA</td>
</tr>
<tr>
<td>XT1</td>
<td>CCTGTTTCTTACGGTCATCGTTT</td>
<td>CTGCGGTTGTCGCAAATTTTTC</td>
<td>AAGAGCTAATCTGCGTGGCCT</td>
</tr>
<tr>
<td>GT1</td>
<td>CACCAAAAGCCCGGACATT</td>
<td>CGACTAATACAGCGAGGATGA</td>
<td>TCAGAGATTTGCAGGCTCC</td>
</tr>
<tr>
<td>RPL39</td>
<td>GAAACAGCCCACATCCCTTATG</td>
<td>CGAGGCTTGCAGGCTTTA</td>
<td>ATGAGCAGGTGAGGCA</td>
</tr>
</tbody>
</table>

MGB probes were labelled with a quencher dye 6-carboxy-tetramethyl-rhodamine (TAMARA) at the 3’ end and with a fluorescent reporter dye 6-carboxyfluorescein (FAM) at the 5’ end, except for RPL39 which was labelled with a fluorescent reporter dye VIC at the 5’ end.
interested in examining the temporal and spatial expression of the two genes thought to be involved, mannan synthase (ManS) and galactomannan galactosyltransferase (GMGT), in both species. As there is currently no significant information on the genes for galactomannan synthesis in coffee, we first needed to isolate and characterize coffee cDNAs encoding the two key enzymes. To begin this gene search, we used a recently identified gene sequence that encodes a biochemically characterized mannan synthase (ManS) gene from guar (Cyamopsis tetragonoloba; Dhugga et al., 2004) and a characterized galactomannan galactosyltransferase (GMGT) gene sequence from fenugreek (Trigonella foenum-graecum; Edwards et al., 1999). Using this DNA sequence information as a starting point, we set out to isolate and characterize full-length cDNA for related genes of coffee, using a Nestlé/Cornell coffee EST database containing about 13 175 unigenes generated from C. canephora tissue samples, including grain at different developmental stages (Lin et al., 2005; http://www.sgn.cornell.edu).

Isolation and characterization of cDNA encoding mannan synthases from coffee

A BLAST search with the biochemically characterized guar mannan synthase protein sequence (accession number AAR23313) identified two coffee unigenes showing high similarity to the query sequence. The sequences from unigenes SGN-U351137 (CcManS1) and SGN-U350420 (CcManS2) encoded different proteins (Table 4). The 17 ESTs associated with CcManS1 were exclusively detected in cDNA libraries from developing seeds whereas the 3 ESTs from CcManS2 were found in seed and pericarp libraries. The longest cDNA corresponding to each unigene (pcccs46w24c19 for CcManS1 and pcccs30w34j20 for CcManS2) was isolated and sequenced. The inserts for CcManS1 and CcManS2 were found to be 1349 bp and 1180 bp long and encode partial ORF sequences of 362 and 262 amino acids (aa), respectively. Experiments using the 5’RACE and/or the Genome Walker approaches were carried out to obtain the missing sequence information for CcManS1 and CcManS2 (see Material and Methods). Re-amplification of the respective full-length cDNA revealed that CcManS1 (pVC6) and CcManS2 (pVC13) encode two distinct proteins of 530 aa and 537 aa, respectively. The deduced protein sequences of CcManS1 and CcManS2 were aligned using CLUSTALW with the full-length sequences of the biochemically characterized mannan synthase from guar (CtManS, Dhugga et al., 2004) and from Arabidopsis (AtCsmA2 and AtCsmA9, Liepman et al., 2005), as well as with other structurally related glycosyl transferase proteins from family 2, according to the CAZymes classification system (www.cazy.org). The glycosyl transferase family 2 includes the cellulose synthase (CesA) proteins and cellulose synthase-like (Csl) proteins. A rooted phylogenetic tree was constructed from the multiple alignment obtained (Fig. 1). The CcManS1 and CcManS2 sequences grouped in a cluster with the CtManS, AtCsmA2, AtCsmA9 and ItManS proteins. This cluster also contained other Arabidopsis CesA proteins whose functions remain to be clarified (Liepman et al., 2005, 2007) and proteins from the sub-class C, including the xylosyl transferase AtCslC4 (Cocuron et al., 2007). The cellulose synthase CesA proteins and the Csl proteins of subclasses B, D and E clearly grouped in another distinct cluster. A manually optimized alignment (see Supplementary Information, fig. 1, available online) revealed that the CcManS1 sequence exhibits 75 % and 66 % identity with CtManS and AtCsmA9, respectively, whereas CcManS2 was more related to AtCsmA2 and ItManS than to the other mannan synthase proteins, suggesting that coffee CcManS1 and CcManS2 genes are orthologs of Arabidopsis AtCsmA9 and AtCsmA2, respectively.

Expression analysis of the coffee ManS1 and ManS2 genes in different tissues of C. canephora using quantitative RT-PCR

In order to confirm the implied grain specificity of ManS1 and to determine the tissue specificity for ManS2, expression of these two genes was studied by quantitative RT-PCR. Transcript levels were analysed in the grain and fruit (i.e. pericarp) at four different developmental stages, and in roots, branches, leaves and flowers of C. canephora (robusta ‘BP409’). The relative expression profiles of ManS genes in relation to the constitutively expressed gene RPL39 are presented in Fig. 2. The results clearly show that ManS1 is highly and exclusively expressed in the grain of robusta, although large changes in expression levels occur during maturation of the grain (Fig. 2A). No expression was detected at the earliest stage of grain development tested (SG), but then ManS1 expression progressively increased during grain development to reach a peak of expression at the yellow stage (YG) followed by lower transcript levels as the grain matured further (RG). Little or no ManS1 expression was detected in the roots, branches, leaves, pericarp and flower tissues. The same analysis for ManS2 showed a

<table>
<thead>
<tr>
<th>Name</th>
<th>Unigene code</th>
<th>EST number</th>
<th>Tissue*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mannan synthase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CcManS1</td>
<td>SGN-U351137</td>
<td>17</td>
<td>Seed 46w (13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Seed 30w (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Seed 18w (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Seed 30w (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pericarp (1)</td>
</tr>
<tr>
<td>Glycosyl transferase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CcGMGT1</td>
<td>SGN-U350503</td>
<td>6</td>
<td>Seed 30w (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Seed 46w (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Seed 18w (1)</td>
</tr>
<tr>
<td>CcGMGT2</td>
<td>SGN-U347226</td>
<td>7</td>
<td>Leaves (7)</td>
</tr>
<tr>
<td>CcXT1</td>
<td>SGN-U352025</td>
<td>5</td>
<td>Leaves (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Whole cherry (2)</td>
</tr>
<tr>
<td>CcGT1</td>
<td>SGN-U354294</td>
<td>1</td>
<td>Leaves (1)</td>
</tr>
</tbody>
</table>

* 18w, 30w, 46w: seeds collected 18, 30, 46 weeks after pollination, respectively. The numbers in brackets indicate the number of ESTs in each library.
ManS2 was expressed mainly in the pericarp at early (SP) and late (RP) developmental stages, an observation that suggests this gene may have two bursts of expression in the pericarp, one early and one late. More importantly, no ManS2 transcripts were detected in the grain tissue of robusta, except at the earliest stage tested (SG). It is also noted that the level of ManS2 transcripts in the small green grain, and the levels seen for the other tissues expressing ManS2 were much lower than the transcript levels of ManS1 seen in the grain.

Isolation and characterization of cDNA encoding galactomannan galactosyl transferases from coffee

In order to identify genes encoding enzymes involved in the addition of galactosyl residues on the mannan backbone in coffee, the biochemically characterized GMGT protein sequences from *Lotus japonicus* (accession number AJ567668; Edwards et al., 2004) and from fenugreek (accession number CAB52246; Edwards et al., 1999) served as query sequences for a BLAST search against the Nestlé/Cornell coffee EST database. This screen identified four unigenes with high sequence identity to the GMGTs from *Lotus japonicus* and/or fenugreek (Table 4).

All four sequences fell into the glycosyl transferase family 34, according to the CAZymes classification system. This family includes genes encoding both α(1,6)- and α(1,2)galactosyltransferase as well as α(1,6) xyloglucan xylosyltransferases from several plant species. A full-length cDNA clone (pcccs26f9) corresponding to unigene SGN-U347226 was isolated, sequenced and found to contain an insert of 1695 bp. This cDNA
encodes a protein of 447 aa. We subsequently obtained two additional full-length cDNA sequences (see Materials and Methods). The plasmid pVC11, representing unigene SGN-U350503, was cloned from arabica and has an insert of 1626 bp. The encoded protein has 448 aa. A third full-length sequence, representing unigene SGN-U352025, was obtained in silico. The resulting contig sequence was shown to be 2951 bp long and contained a continuous ORF of 460 aa. Only a partial 5' cDNA sequence currently exists for unigene SGN-U354294. An optimized alignment of the three full-length cDNAs with the most homologous biochemically characterized sequences in the protein database (see Supplementary Information, fig. 2, available online) demonstrated that the full-length proteins represented by unigenes SGN-U350503 and SGN-U347226 were the most homologous with the galactomannan galactosyl transferases from fenugreek (TfGMGT) and *Lotus japonicus* (LjGMGT). The protein product of unigene SGN-U350503 showed 59.5% and 58.4% identity with TfGMGT and LjGMGT, respectively. The protein product of unigene SGN-U347226 showed 48% and 48.3% identity with TfGMGT and LjGMGT, respectively. We therefore annotated the protein encoded by unigene SGN-U350503/pVC11 as CaGMGT1 and the protein encoded by unigene SGN-U347226/pcccs26f9 as CcGMGT2. The most homologous biochemically characterized sequence in the protein database for unigene SGN-U352025 by a BLAST search was found to be AxtX2 (77% identity), a protein with xyloglucan xylosyltransferase activity (At4g02500; Cavalier and Keegstra, 2006). Therefore the encoded protein was named CcGT1. Due to absence of over 200 amino acids at the C terminus, the partial protein sequence encoded by unigene SGN-U354294 could only be clearly annotated as a glycosyl transferase, and thus was named CcGT1.

The four coffee sequences available were aligned using CLUSTALW with the biochemically characterized GMGTs from fenugreek (TfGMGT), guar (CtGMGT), *Lotus japonicus* (LjGMGT) and with related proteins from *Senna occidentalis* (SoGMGT), potato (StGTase) and other *Arabidopsis* galactosyl transferase and xylosyl transferase proteins from family 34. A rooted phylogenetic tree was constructed from this multiple alignment (Fig. 3).

The three functionally characterized galactosyl transferases from guar, fenugreek and *Lotus japonicus* formed a large distinct group that also contained CaGMGT1, CcGMGT2 and the partial sequence CcGT1. This result is consistent with our annotation of CaGMGT1 and CcGMGT2 as galactomannan galactosyl transferases. The fact that CcGT1 also fell into this group suggests that the corresponding protein may also be a GMGT. In contrast, the CcXT1 sequence clearly fell into another distinct group whose closest members include AxtX1 and AxtX2, two *Arabidopsis* proteins that showed xylosyl transferase activity after expression in heterologous systems (Faik et al., 2002; Cavalier and Keegstra, 2006). This latter observation supports our annotation of this protein as a xylosyl transferase. Finally, the observation that all six ESTs found for unigene SGN-U350503 (CcGMGT1) were in the seed, whereas all seven of unigene SGN-U347226 (CcGMGT2) were in leaves, and the single EST in unigene SGN-U354294 (CcGT1) was found in the leaf, suggests that GMGT1 is the key protein for galactomannan synthesis in the coffee grain.

**Expression analysis of the four coffee glycosyl transferase genes in different tissues of C. canephora using quantitative RT-PCR**

To learn more concerning the gene(s) responsible for the transfer of galactose and xylose units in the coffee grain, expression of the four coffee GTases was analysed in the different tissue samples of robusta (Fig. 4). Among the four genes, only GMGT1 was shown to be nearly exclusively expressed in the grain. GMGT1 transcripts were first detected in large green (LG) grain tissue (RQ = 0.06), then rose to reach a maximum transcript level at the yellow stage (YG; RQ = 0–11). A greatly reduced level of transcripts was detected at maturity (RG; RQ = 0–02). No significant GMGT1 expression was observed in other plant tissue. This grain-specific expression profile of GMGT1 was very similar to that observed for ManS1 (Fig. 2A). In contrast, the expression levels of GMGT2, XT1 and GT1 in the different tissue samples were lower in comparison with GMGT1 (Fig. 4B–D). In general, these other GTases were expressed predominantly in the pericarp tissues, with GMGT2 showing higher expression in the mature pericarp (RP; RQ > 0.08; Fig. 4B).

Interestingly, the red pericarp stage also exhibits an increase in ManS2 transcripts, suggesting that, at least at this stage of the pericarp development, these two genes products may work in concert. It will be interesting in the future to look at the galactomannan levels and galactose-to-mannose ratios in different stages of pericarp development. The transcript levels of GMGT2, XT1 and GT1 in the grain were barely detected as their relative quantification levels were close to the detection threshold. From these data, it appears that GMGT1 is the only gene whose expression is relatively strong in the grain and closely follows seed endosperm development. Thus, it is highly plausible that GMGT1 encodes the enzyme responsible for adding galactose units to the bulk of the mannan chains during coffee grain galactomannan biosynthesis.

**ManS1 and GMGT1 transcript accumulation in robusta vs. arabica**

Transcriptional analyses have shown that *CcManS1* and *CcGMGT1* are both expressed specifically in the grain of robusta plants. It was confirmed this was also the case in arabica by examining the expression of these two genes in several tissues of the variety ‘T2308’. Apart from their expression in the grain (see below), we found no significant expression of *ManS1* or *GMGT1* in any of the other arabica tissues examined, although it is possible very low, but detectable, expression of *GMGT1* may occur in the arabica pericarp, branch and leaf tissues (data not shown). These results are consistent with our proposal that the proteins encoded by *ManS1* and *GMGT1* genes participate actively in the biosynthesis of the galactomannans of the coffee grain endosperm. Currently, *C. canephora*
(robusta) and C. arabica (arabica) represent the majority of the coffee cultivated worldwide. These two varieties show differences in agronomical traits, biochemical composition and cup quality. Therefore, it was of interest to determine if these two species may have significant differences in the levels of transcript accumulation, or relative expression, of ManS1 and GMGT1 in the developing grain. To address this question, ManS1 and GMGT1 expression was analysed for several arabica and robusta genotypes at four different grain development stages. The results obtained are presented in Fig. 5. In arabica, except for the ‘T2308’-GH series, the level of GMGT1 transcripts were roughly equivalent, or slightly higher, than the levels seen for the ManS1 transcripts in each sample, suggesting that these transcripts are under a similar transcriptional control. However, there were significant differences in the apparent timing of transcript accumulation between the arabica varieties, with ‘T2308’-GH/‘T2308’-Fld and ‘CCA18’ showing significant levels of ManS1 and GMGT1 transcripts at the ‘small green’ stage, while ‘CCA05’ and ‘CCA07’ only accumulated transcripts to a significant extent at the ‘yellow’ stage. The genotype ‘CCA12’ had an intermediate accumulation pattern. The basis of these differences remains to be determined, but probable factors include variability in the speed of grain versus pericarp development for the cherries of each genotype investigated, as well as differences in the growing conditions. This latter point is exemplified here by the shift to a slightly earlier (in terms of cherry colour) expression of ManS1 and GMGT1 in the grain of greenhouse-grown cherries in France versus cherries harvested from the field in Equador (‘T2308’-GH versus ‘T2308’-Fld, Fig. 5). The transcript accumulation data obtained for ManS1 and GMGT1 during grain development in the robusta samples ‘FRT 07’ and ‘FRT 23’ (Fig. 5), and earlier for ‘BP409’ (Figs 2 and 4), indicate that there is less fluctuation in the time of expression of these genes in robusta grain. As seen for arabica, there is a good correspondence between the accumulation pattern of ManS1 and GMGT1 in all three robusta genotypes examined here. However, although the ManS1 transcript accumulation levels were relatively similar for the three genotypes (for example RQ of 0·98, 2·36 and 2·75 at yellow stage for ‘BP409’, ‘FTR23’ and ‘FRT07’, respectively), regarding the GMGT1 transcripts, two were similar and one was quite low at all stages, for example, RQs of 0·11, 0·73, and 0·07 at the yellow stage for ‘BP409’, ‘FTR23’ and ‘FRT07’, respectively. The reason for the low GMGT1 transcript accumulation seen with ‘FRT07’ is currently not known, but it could be related to a poor detection of the GMGT1 alleles present in ‘FRT07’ using the GMGT1 TaqMan probe set described in the Methods. Testing with other primer-probe sets in the future should help address this possibility.

Expression analysis of ManS2 and the glycosyl transferases in coffee leaves

The data obtained above clearly indicate that expression of the coffee mannan synthases and the glycosyl transferase genes are controlled developmentally. To further investigate...
very low levels of stages of maturity is also presented in Fig. 6. In robusta, script levels of the four detected in the mature arabica leaves. A comparison of transcript accumulation for \( \text{GMGT2}, \text{XT1} \) and \( \text{GT1} \) were seen at the young leaf stage, although the transcript levels were slightly higher at this stage. Again, few \( \text{GMGT1} \) transcripts were detected at this stage. The transcript levels of all four \( \text{GTases} \) genes were very low in the ‘mature’ and old leaf stages. In contrast to robusta, \( \text{GMGT1} \) was found to have the highest level of transcripts of the four genes in arabica leaves during the first three stages examined, with \( \text{GMGT1} \) transcript levels peaking at the young and mature stages. The patterns for \( \text{GMGT2}, \text{XT1} \) and \( \text{GT1} \) transcript accumulation in arabica were relatively similar to those seen for robusta, although the levels of \( \text{GMGT2} \) were slightly higher in the arabica samples. The biggest difference seen between the transcript levels of the four coffee \( \text{GTases} \) in robusta and arabica was the very high level of \( \text{GMGT1} \) expression in arabica. Further experiments are needed to determine if this particular difference is seen between other robusta and arabica varieties, and thus eliminate the possibility that this difference is due to allelic differences between the specific arabica and robusta genotypes studied here.

DISCUSSION

Carbohydrates, mostly in the form of polysaccharides, represent in the region of 50–60\% of the total mass of the mature coffee grain (Fischer et al., 2001; Oosterveld et al., 2003, and references therein) and the majority of these polysaccharides are believed to be localized in the cell wall. In addition to being a major constituent of the green coffee grain, at least some of the coffee polysaccharides are thought to make important contributions to the final coffee beverage (Illy and Viani, 1995), and their solubilization is one of the key issues regarding the yield of soluble coffee (Clifford, 1985). In order to better understand the significance of the polysaccharides to both the coffee grain and the final coffee beverage, it was decided to investigate genes involved in the synthesis of the major coffee grain polysaccharides, i.e. galactomannan. Here, we have described the isolation and characterization of cDNA for a mannan synthase (\( \text{CcManS1} \) and \( \text{CaManS1} \)) and for a galactosyltransferase (\( \text{CtGMGT1} \) and \( \text{CaGMGT1} \)) that most likely encode proteins directly responsible for the synthesis of the galactomannans in the coffee grain of robusta and arabica. We have further described a second mannan synthase cDNA (\( \text{CcManS2} \) and \( \text{CaManS2} \)) that appears to be involved in galactomannan synthesis in green tissues like the leaf, as well as three other cDNAs that probably represent additional GT proteins.

The QRT-PCR expression data obtained for robusta shows that \( \text{ManS1} \) exhibits significant expression only in the coffee grain, and that the level of these transcripts are highest during the period when the endosperm is forming/developing. Consistent with this finding, the \( \text{ManS1 ESTs} \) were also only detected in the robusta cDNA libraries made with RNA extracted from grain (Table 4). Transcripts corresponding to \( \text{ManS2} \) were not significantly detected at the ‘very young’ leaf stage, with \( \text{GMGT2} > \text{XT1} > \text{GT1} \). Only very low levels of \( \text{GMGT1} \) were detected at the very young leaf stage. The same patterns of transcript accumulation for \( \text{GMGT2}, \text{XT1} \) and \( \text{GT1} \) were seen at the young leaf stage, although the transcript levels were slightly higher at this stage. Again, few \( \text{GMGT1} \) transcripts were detected at this stage. The transcript levels of all four \( \text{GTases} \) genes were very low in the ‘mature’ and old leaf stages. In contrast, \( \text{GMGT1} \) was found to have the highest level of transcripts of the four genes in arabica leaves during the first three stages examined, with \( \text{GMGT1} \) transcript levels peaking at the young and mature stages. The patterns for \( \text{GMGT2}, \text{XT1} \) and \( \text{GT1} \) transcript accumulation in arabica were relatively similar to those seen for robusta, although the levels of \( \text{GMGT2} \) were slightly higher in the arabica samples. The biggest difference seen between the transcript levels of the four coffee \( \text{GTases} \) in robusta and arabica was the very high level of \( \text{GMGT1} \) expression in arabica. Further experiments are needed to determine if this particular difference is seen between other robusta and arabica varieties, and thus eliminate the possibility that this difference is due to allelic differences between the specific arabica and robusta genotypes studied here.

the level of control of these genes in green tissue, the accumulation of all six genes was examined in leaf tissue collected at several developmental stages (very young, young, mature, and old leaves) from both a robusta and an arabica variety (Fig. 6). Consistent with the results presented in Fig. 2, transcripts for \( \text{ManS1} \) were not detected in robusta at any leaf stage (data not shown). A similar result was observed for \( \text{ManS1} \) transcripts in arabica leaves, although in this case a trace level of transcripts was detected (\( R_Q < 0.03 \); data not shown). Overall, these results are consistent with \( \text{ManS1} \) being primarily grain specific. In contrast, \( \text{ManS2} \) expression was observed in very young (\( R_Q = 0.1 \)) and young (\( R_Q = 0.2 \)) robusta leaves but were not detected in mature or old robusta leaves (Fig. 6). A similar pattern of \( \text{ManS2} \) expression was observed in arabica leaves, except that \( \text{ManS2} \) transcripts were also detected in the mature arabica leaves. A comparison of transcript levels of the four \( \text{GTases} \) in robusta leaves at different stages of maturity is also presented in Fig. 6. In robusta, very low levels of \( \text{GMGT2}, \text{XT1} \) and \( \text{GT1} \) transcripts were the QRT-PCR expression data obtained for robusta shows that \( \text{ManS1} \) exhibits significant expression only in the coffee grain, and that the level of these transcripts are highest during the period when the endosperm is forming/developing. Consistent with this finding, the \( \text{ManS1 ESTs} \) were also only detected in the robusta cDNA libraries made with RNA extracted from grain (Table 4). Transcripts corresponding to \( \text{ManS2} \) were not significantly

![Fig. 4. Expression of coffee GTases in different tissues of robusta 'BP409'. The expression of each gene was measured in the root, branch and leaf tissues, plus in different stages of developing pericarp and grain of cherries, using quantitative RT-PCR. The value indicates the expression level of the gene of interest relative to that of the constitutively expressed gene RPL39. (A) GMGT1, (B) GMGT2, (C) XT1, (D) GT1. Abbreviations: G, small green stage grain; LG, large green stage grain; YG, yellow stage grain; RG, red stage grain; SP, small green pericarp; LP, large green pericarp; YP, yellow pericarp; RP, red pericarp; Rt, root; Br, branch; Le, leaf. Mean values ± s.e. calculated from triplicate measurements.](https://academic.oup.com/aob/article-abstract/102/2/207/184418)
expressed in the grain, except during the early small green stage (Fig. 2). Expression of ManS2 at this stage, and probably earlier, supports the idea that during this stage at least parts of the developing grain exhibit gene expression patterns seen in green tissues such as leaves. Low levels of ManS2 transcripts were also detected in the pericarp at different stages of development and in other vegetative parts of the coffee plant, suggesting that the ManS2 protein could be responsible for the synthesis of mannans in tissues other than the grain. An exhaustive search of the complete unigene database set for genes encoding glycosyl transferases resulted in the discovery of 13 gene sequences with strong similarities to cellulose synthase or cellulose-like synthase genes (belonging to the glycosyl transferase family 2, according to the CAZY classification system; M. Pré and J. Sobilo, unpubl. data). However, except for ManS1 and ManS2, none of these genes showed sufficient homology with the known mannan synthases to be candidates for involvement in galactomannan synthesis. The fact that no other potential candidate sequences for mannan synthesis, other than the two described here, were found in several available, deeply sequenced, cDNA libraries further underscores that likelihood that the highly expressed ManS1 is the key gene for mannan synthase in the coffee grain. The GMGT1 expression pattern closely resembles that seen for ManS1, with the highest level of transcripts detected at the same grain developmental stages. The observation that ManS1 and GMGT1 are expressed contemporaneously in the grain is consistent with the established model where
galactomannan biosynthesis requires a specific interaction between the ManS and GMGT proteins (Edwards et al., 2002). In addition, because none of the other three glycosyl transferase genes studied here showed significant expression in the grain, it appears very plausible that ManS1 and GMGT1 act in concert to produce most of the galactomannans of the grain. Verification of the roles of these two gene products in both species await new experiments in which expression of one, or both, of these genes are reduced in the grain using techniques such as RNAi.

Recently, Redgwell et al. (2002) compared the monosaccharide composition of polysaccharides isolated from the mature grain of three robusta and three arabica varieties. Their analysis indicated that the total polysaccharide content of the robusta varieties studied clearly had lower molar percentages of mannose than each of the arabica varieties examined. In addition, their analysis showed the total robusta polysaccharides contained more galactose than that found in any of the arabica varieties. Given that the galactomannans make up around 50% of the coffee grain polysaccharides, such data could suggest that the robusta varieties analysed have less galactomannans than the arabica varieties studied. Furthermore, these data could indicate that robusta galactomannans may have a higher level of galactose substitution. In such a scenario, one might expect a low expression of ManS1 and a high expression of GMGT1 in robusta species, compared with a high expression of ManS1 and a low expression of GMGT1 in arabica species. Therefore, it was of interest to investigate whether the level of transcripts directly linked to galactomannan synthesis in the grain were quantitatively different between robusta and arabica. Analysis of ManS1 and GMGT1 transcript accumulation data obtained for the different grain developmental stages of five arabica and two robusta varieties (Fig. 5) indicates that no general correlation can be made between the mannose content of total polysaccharides measured previously for these species and the robusta and arabica gene expression patterns presented here. Nonetheless, the data obtained show that there is a good correspondence between the relative levels of ManS1 and GMGT1 transcripts at each developmental stage within one variety (which is consistent with our proposal that these genes are regulated in co-ordination). These data also point out clear differences in the levels of these transcripts and the timing of their expression between the different arabica varieties. For example, two varieties (‘CCAO5’ and ‘CCAO7’) showed both late expression and higher levels of GMGT1 transcripts versus ManS1 relative to the other arabica varieties. Some of these differences could be partially due to environmental differences, because there were significant variations in the levels and timing of these transcripts in a single variety when it was grown in the field (‘T2308’-Fld) versus the greenhouse (‘T2308’-GH). However, despite this lack of obvious correspondence between the previously determined composition of the polysaccharides and the ManS1 and GMGT1 transcript levels, more direct comparative work is warranted in the future. In particular, it will be interesting to determine the type and levels of polysaccharides found in varieties ‘CCAO7’ and ‘FRT 23’, which show large differences in transcript ratios for ManS1 and GMGT1. It will also be important to examine the relative levels of ManS1 and GMGT1 proteins in ‘CCAO7’ and ‘FRT 23’ at the different stages of grain development using techniques such as Western blotting.

Examination of the expression of the different genes during leaf development in robusta and arabica confirmed that ManS1 was not expressed in this tissue at any of the stages examined. In contrast, significant expression of ManS2 was seen in the leaves of both robusta and arabica, with highest expression detected in the expanding very young and young leaves, although significant expression was also detected in mature leaves of the arabica variety studied. Transcript levels of GMGT2, XT1 and GT1 were relatively similar for both robusta and arabica (Fig. 6), although as seen for ManS2, the expression of all three genes remained relatively high in the mature leaves of arabica but fell in the mature leaves of robusta. While it is not obvious which of the different GT gene products could be working with ManS2, the transcript levels and expression patterns suggest either GMGT2 and GT1, or both, could partner with ManS2. Future experiments examining where these different proteins are expressed in recombinant cells could resolve this question, as could co-precipitation experiments with protein-specific antibodies. Surprisingly, in contrast to the transcript analysis noted above, GMGT1 transcript analysis in the different leaf samples showed marked differences between robusta and arabica, with the latter species showing a unexpectedly high expression of GMGT1. The earlier GMGT1 expression analysis for robusta variety ‘BP409’ indicated that this gene was primarily grain specific, and this result was confirmed by examining the expression of this gene in the leaves of robusta ‘BP409’ at different stages. The consequence of this gene’s expression in the leaf of arabica variety ‘T2308’ is not currently known, although, if the GMGT1 protein is capable of forming a functional complex with ManS2, one may predict that the high GMGT1 expression of arabica ‘T2308’ may result in leaf galactomannans with a higher galactose/mannose ratio than that seen in robusta leaves. Thus, future experiments should be directed towards examining GMGT1 gene expression in the leaves of other arabica varieties to determine whether high GMGT1 expression is found in all arabica varieties, or is only found in a specific sub-group of arabica coffees. Subsequently, it could also be of interest to directly compare the levels of GMGT protein in both arabica and robusta leaves by Western blotting.

The availability of cDNA for the different coffee ManS and GMGT genes opens new avenues of investigation of the coffee galactomannans, and on the cell wall of the coffee grain generally. For example, it is now possible to couple detailed chemical analysis of the levels and types of mannans present in the grain from different coffee varieties and then explore whether any differences found are correlated with variations in expression of the genes/proteins ManS1 and GMGT1. It is noted that a principal effect of Gal substitution is to increase the solubility of the mannan polymers by preventing the formation of
insoluble crystalline structures between the mannan chains (McCleary et al., 1981). The relatively low galactose content of the coffee galactomannans, especially during late grain development (Fischer et al., 2001), probably has important consequences for the production of soluble coffee, potentially reducing the overall extraction levels and possibly causing precipitation problems in the extracts obtained. The potential to identifying varieties producing either less galactomannans or more heavily galactosylated galactomannans in the grain using gene and protein expression studies opens the possibility of selecting plants that produce more easily extractable coffee. An Arabidopsis ManS protein has been shown to form a beta-linked glucomannan polymer in vitro (Liepman et al., 2005). This observation is of interest because it is not known if there is any significant level of the glucomannan heteropolymer present in the coffee grain. Thus, another new avenue for investigation could be to examine if this polymer exists in significant amounts in the coffee grain and, if so, how important are the ManS1 and ManS2 genes, and their alleles, to the synthesis of this heteropolymer. Finally, several studies have expressed either ManS or GMGT proteins separately in various plant cells (Reid et al., 2003; Dhugga et al., 2004; Edwards et al., 2004) and have shown that this novel expression can alter the amount and/or the level of galactose substitution in the galactomannans produced. However, to date, no studies have been published showing significantly higher galactomannan production due to an over-expression of both recombinant proteins in a single plant cell line. Recently, Dhugga (patent WO/2004/046197) has expressed both genes from guar in soybean seeds. In this case, only a limited increase in galactomannan content was observed, presumably due to an absence of sufficient levels of a necessary precursor(s). Therefore, one future objective of this approach will be to identify plant tissues that can metabolically support substantially increased galactomannan production when ManS and GMGT proteins from plants such as coffee and guar are over-expressed together. An exciting, but much longer-term goal of work on genes involved in polysaccharide synthesis is to assemble different groups of these genes in selected plant tissues to show that novel gene combinations can generate the production of novel polymers. A first step in this direction is work by Reid’s group clearly showing that the structure and quantity of galactomannans can be altered in vivo by either increasing or decreasing the expression of GMGT proteins in the seeds of tobacco and Lotus japonicus (Reid et al., 2003; Edwards et al., 2004).

SUPPLEMENTARY INFORMATION
Additional information on alignment of deduced coffee protein sequences is available online at http://aob.oxfordjournals.org.

ACKNOWLEDGEMENTS
We wish to acknowledge Dr Chenwei Lin and Dr Steve Tanksley for generating the Cornell-Nestlé EST clones used in this work. We thank Valérie Chazalet, Dr Olivier Lerouxel and Dr Christelle Breton for their advice and for many helpful discussions during this work. We also wish to thank Maud Lepelley and Dr Isabelle Privat for supplying some of the cDNA samples used in these experiments and for technical advice.

LITERATURE CITED


**NOTE ADDED IN PROOF**

A full length cDNA called “cccl23g20” representing unigene SGN-U352025 has now been sequenced. This sequence CcXT1b (Accession # EU760961) has 3 base changes from the ‘in silico’ derived sequence CcXT1, which results in three amino acid changes (E to K at position 334, A to V at position 355, and K to M at position 401).