Available online at www.sciencedirect.com
ScienceDirect
Phytochemistry 68 (2007) 1510–1520
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Functional genomic analysis of *Arabidopsis thaliana* glycoside hydrolase family 35

Young Ock Ahn [a], Meiying Zheng [a], David R. Bevan [b], Asim Esen [a], Shin-Han Shiu [c], Jonas Benson [d], Hsiao-Ping Peng [d], Joseph T. Miller [d,e], Chi-Lien Cheng [d], Jonathan E. Poulton [d], Ming-Che Shih [d,e,*]

[a] *Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA*
[b] *Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA*
[c] *Department of Plant Biology, Michigan State University, S-306 Plant Biology Building, East Lansing, MI 48824, USA*
[d] *Department of Biological Sciences, University of Iowa, Iowa City, Iowa 52242, USA*
[e] *Roy J. Carver Center for Comparative Genomics, 200 Biology Building East, University of Iowa, Iowa City, Iowa 52242, USA*

Received 2 November 2006; received in revised form 20 February 2007
Available online 26 April 2007

Abstract

Catalysing the hydrolysis of terminal β-galactosyl residues from carbohydrates, galactolipids, and glycoproteins, glycoside hydrolase Family 35 (β-galactosidases; BGALs) are widely distributed in plants and believed to play many key roles, including modification of cell wall components. Completion of the *Arabidopsis thaliana* genome sequencing project has, for the first time, allowed an examination of the total number, gene structure, and evolutionary patterns of all Family 35 members in a representative (model) angiosperm. Reiterative database searches established a multigene family of 17 members (designated *BGAL1–BGAL17*). Using these genes as query sequences, BLAST and Hidden Markov Model searches identified *BGAL* genes among 22 other eukaryotes, whose genomic sequences are known. The *Arabidopsis* ($n = 17$) and rice ($n = 15$) *BGAL* families were much larger than those of *Chlamydomonas*, fungi, and animals ($n = 0$–4), and a lineage-specific expansion of *BGAL* genes apparently occurred after divergence of the *Arabidopsis* and rice lineages. All plant *BGAL* genes, with the exception of *Arabidopsis BGAL17* and rice Os 9633.m04334, form a monophyletic group. Arabidopsis *BGAL* expression levels are much higher in mature leaves, roots, flowers, and siliques but are lower in young seedlings. *BGAL8*, *BGAL11*, *BGAL13*, *BGAL14*, and *BGAL16* are expressed only in flowers. Catalytically active BGAL4 was produced in the *E. coli* and baculoviral expression systems, purified to electrophoretic homogeneity, and partially characterized. The purified enzyme hydrolyzed *p*- and *o*-nitrophenyl-β-D-galactosides. It also cleaved β-(1,3)-, β-(1,4)-, and β-(1,6)-linked galactobiosides and galactotriosides, showing a marked preference for β-(1,3)- and β-(1,4)-linkages.

Keywords: *Arabidopsis thaliana*; Cruciferae; Functional genomics; Phylogenetic analysis; Heterologous expression; Glycoside hydrolase family 35; β-Galactosidases

1. Introduction

Glycoside hydrolases (GHs) are widely distributed enzymes that hydrolyse the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. Based on amino acid sequence similarities, these enzymes have been classified into 108 families, whose distinguishing features and representative members are described in the Carbohydrate-Active Enzymes database (Coutinho and Henrissat, 1999; http://www.cazy.org/). GH Family 35 contains β-galactosidases (BGALs; EC 3.2.1.23), which catalyse the hydrolysis of

* Corresponding author. Address: Department of Biological Sciences, University of Iowa, Iowa City, Iowa 52242, USA. Tel.: +1 319 335 2071; fax: +1 319 335 1069.
E-mail address: ming-che-shih@uiowa.edu (M.-C. Shih).

0031-9422/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.phytochem.2007.03.021

terminal β-galactosyl residues from carbohydrates, galactolipids, and glycoproteins. These enzymes are widespread in bacteria and eukaryotes. Found throughout the higher plant body, they are believed to play key roles in modification of cell wall components during fruit ripening (Pressey, 1983; Carey et al., 1995; Smith et al., 1998, 2002), loosening of the cell wall during growth (Dopico et al., 1989; Sekimata et al., 1989), flower senescence (Ragothama et al., 1991), xyloglucan mobilization in cotyledons (de Alicântara et al., 1999), and galactolipid turnover (Bhalla and Dalling, 1984). Additionally, the possibility that certain BGALs may act biosynthetically in vivo has been raised by demonstration of their transglycosylation activities and reverse hydrolysis in vitro (Yoon and Ajisaka, 1996). BGALs have been purified from many plant sources, especially from ripening fruits and germinating seeds, and commonly occur as multiple forms (Dey and del Campillo, 1984; Lee et al., 2003; Li et al., 2001; Triantafillidou and Georgatsos, 2001; Balasubramaniam et al., 2005). The nature and physiological significance of this microheterogeneity are poorly understood, but recent molecular approaches indicate that BGALs are encoded by multigene families in several plant species (Ross et al., 1994; Smith and Gross, 2000). The best understood is the tomato *BGAL* family, which has at least seven members that share 33–79% amino acid identity (Smith and Gross, 2000). Their respective cDNAs possess signal sequences predicted to target individual isoforms to the apoplast, plasma membrane, ER, or chloroplast.

The completion of the *Arabidopsis thaliana* genome sequencing project (The Arabidopsis Genome Initiative, 2000) has provided an ideal opportunity to examine the gene structure and evolutionary patterns for all members of GH Family 35 hydrolases within a representative (model) angiosperm. In this study, 17 *BGAL* genes were identified in the Arabidopsis genome, which were used as query sequences to perform BLAST and Hidden Markov Model (HMM) searches to identify recognize *BGAL* genes in 22 other eukaryotes whose genomic sequences are known. A phylogenetic tree was then constructed to examine evolutionary relationships between these genes. In addition, the expression profiles of Arabidopsis *BGAL* genes were examined at different developmental stages. Finally, the overexpression of Arabidopsis BGAL4 in *Escherichia coli* and baculoviral expression systems and its partial characterization are reported.

2. Results and discussion

2.1. Properties of predicted Arabidopsis BGAL proteins

A search of the Arabidopsis genome identified 17 GH Family 35 *BGAL* genes. cDNAs corresponding to 14 of these genes (*BGAL1* through *BGAL14*) had already been isolated and sequenced by Gy et al. (2000). We have designated the remaining three genes *BGAL15* through *BGAL17* (Table 1). Predicted protein lengths range from 697 (BGAL17) to 988 (BGAL14) amino acids, corresponding to molecular masses ranging from 70 to 112.1 kDa. All 17 predicted Arabidopsis BGAL possess the active-site consensus sequence G-G-P-[LIVM](2)-x(2)-Q-x-E-N-**E**-[FY] typical of GH Family 35 members, within which the second Glu is believed to act as the proton donor in the catalytic mechanism (Henrissat, 1998). In addition, it has been shown that Glu-268 of human lysosomal BGAL functions as the catalytic nucleophile (McCarter et al., 1997). This Glu residue is invariant in the motif P-N-K-x-x-K-P-K-M-W-T-**E**-x-W in plant BGALs as well as several Family 35 enzymes from non-plant species (McCarter et al., 1997; Blanchard et al., 2001). This Glu residue is also conserved in the consensus motifs of 16 of the Arabidopsis BGALs; BGAL17 lacks this motif but still has the Glu residue conserved at the comparable position.

Comparison of available BGAL sequences from other plant species reveals that these hydrolases fall into two distinct groups based on molecular masses and other characteristics. Group 1 members, which include the tomato cDNA clone TBG4 and BGALs from apple, carnation, lupin, and papaya, are 721–731 amino acids in length (Ragothama et al., 1991; Buckeridge and Reid, 1994; Ross et al., 1994; Smith and Gross, 2000; Lazan et al., 2004). Group 2 BGALs, exemplified by asparagus BGAL and the remaining six tomato cDNAs, are far longer (832–888 amino acids); this is mainly due to the existence of a lectin-like domain in the carboxyl termini of these proteins (King and Davies, 1995). It is unknown whether the latter domains are functional, but their existence and conserved nature raise the interesting possibility that they may increase the catalytic efficiency of Group 2 BGALs by anchoring these hydrolases to their polymeric substrates. With the exception of BGAL14 and BGAL17, Arabidopsis BGALs fall into either Group 1 (BGAL2, BGAL4, BGAL5, BGAL6, BGAL10, and BGAL12) or Group 2 (BGAL1, BGAL3, BGAL7, BGAL8, BGAL9, BGAL11, BGAL13, BGAL15, and BGAL16). As predicted by PSORT, all Arabidopsis BGALs, with the exception of BGAL14, have an N-terminal signal peptide, and many appear targeted to the cell exterior (Table 1), suggesting their potential involvement in cell wall modification. Finally, the majority of these hydrolases may be glycoproteins, because they possess multiple ($n = 1$–9) potential *N*-glycosylation sites.

2.2. Structure of the Arabidopsis BGAL gene family

To assess relationships between Arabidopsis GH Family 35 members, we plotted the intron–exon organization patterns of these genes according to their phylogenetic distances (Fig. 1). The number of exons ranges from 12 to 19 for *BGAL* genes. The intron–exon junctions are mostly conserved among *BGAL* genes with the exception of *BGAL17*. *BGAL1*, *BGAL8*, *BGAL9*, and *BGAL16* possess 19 exons separated by 18 introns, with exons 2 through

Table 1
Predicted properties of Arabidopsis BGALs

Gene	Gene ID	ORF (# aa)	Mass (kDa)	pI[a]	Signal peptide[b]	Likely destination[c]	Glycosyl. Sites #[d]
BGAL1	At3g13750	847	93,658	8.66	+	O	1
BGAL2	At3g52840	727	81,999	8.61	+	O	1
BGAL3	At4g36360	856	95,193	7.23	+	ERM	–
BGAL4	At5g56870	724	80,591	8.88	+	O	1
BGAL5	At1g45130	733	81,433	8.37	+	ERM	–
BGAL6	At5g63800	718	79,759	9.11	+	O	4
BGAL7	At5g20710	826	92,720	7.75	+	O	–
BGAL8	At2g28470	838	91,866	8.21	+	O	5
BGAL9	At2g32810	887	99,198	8.40	+	O	7
BGAL10	At5g63810	741	83,101	8.57	+	O	6
BGAL11	At4g35010	845	95,569	9.90	+	ERM	–
BGAL12	At4g26140	728	81,661	8.04	+	O	3
BGAL13	At2g16730	848	95,889	8.92	+	O	9
BGAL14	At4g38590	988	112,101	6.56	–	Cyt	4
BGAL15	At1g31740	786	86,350	5.22	*	O	9
BGAL16	At1g77410	815	91,648	9.50	+	ERM	–
BGAL17	At1g72990	697	78,639	8.58	*	O	5

+, cleavable signal peptide; *, uncleaved signal peptide; –, lacks signal peptide.
[a] Determined using EditSEQ in the Lasergene Sequence Analysis software.
[b] Predicted by SignalP v3.0 (HMM analysis).
[c] Predicted by PSORT. ERM: endoplasmic reticulum; O: outside; Cyt: cytosol.
[d] Predicted using NetNGlyc.

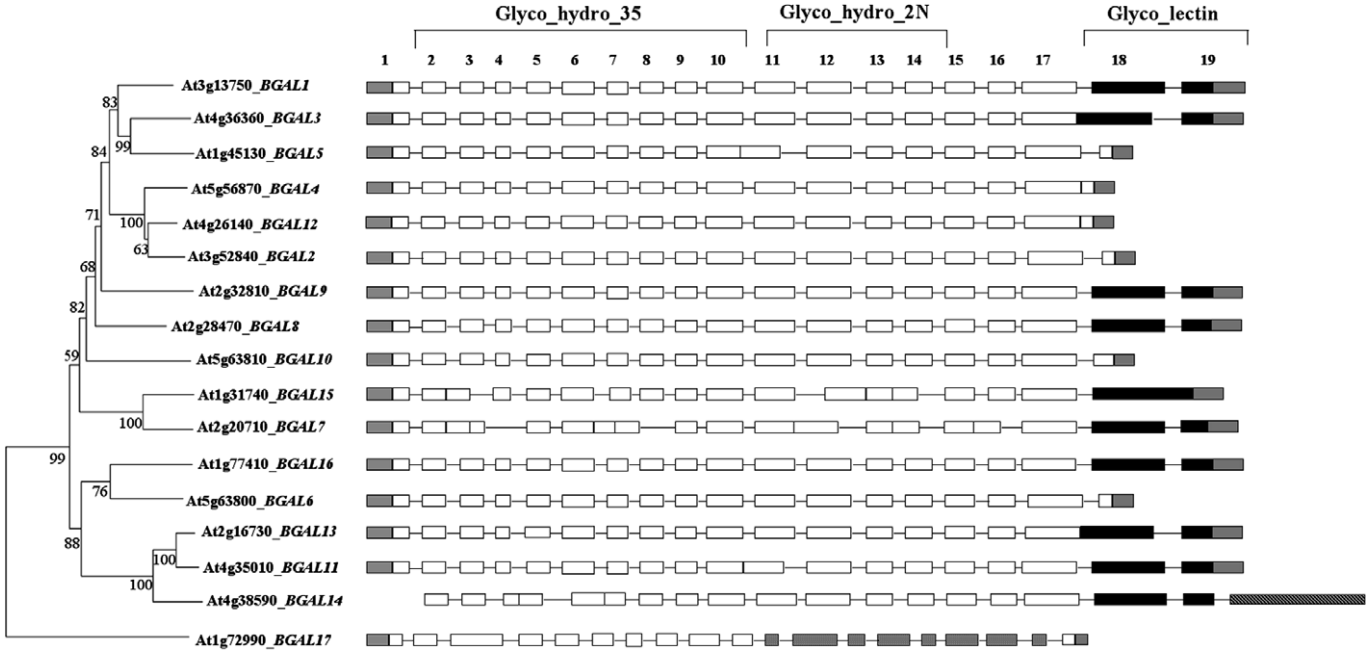

Fig. 1. Intron–exon organization and domain structures of Arabidopsis *BGAL* genes. Exons are represented by rectangles, whereas introns are represented by solid lines between exons. Exon and intron lengths are drawn to scale. Black rectangles: Gal_lectin domains, gray rectangles, 5′-UTR and 3′-UTR; hatched rectangle at the BGAL14 C-terminus, PRP1_N domain.

10 encoding the Glyco_hydro_35 domain, exons 11 through 14 encoding the Glyco_hydro_2_N domain, and exons 18 and 19 encoding the Gal_lectin domain. Based on the parsimony principle, we propose that the gene structure shown by these four genes represents the ancestral intron–exon organization for all Arabidopsis *BGAL* genes excluding *BGAL17*. Differential intron loss in the regions containing the Glyco_hydro_35 and Glyco_hydro_2N domains likely accounts for the lower number of introns possessed by other Arabidopsis *BGAL* genes. For example, the intron between exon 2 and 3 was probably present in the *BGAL* common ancestor but was lost in the common ancestor of *BGAL15* and *BGAL7*. Regarding the lectin-like domain exhibited by Group 2 BGALs, we propose that the ancestor of all Arabidopsis *BGAL* genes excluding *BGAL17* possessed Glyco_hydro_35, Glyco_hydro_2N,

and Gal_lectin domains, but that the lectin-like domain was lost in Group 1 *BGAL* genes (*BGAL2*, *BGAL4*, *BGAL5*, *BGAL6*, *BGAL10*, and *BGAL12*). Based on the tree topology, three independent intron-loss events can account for the *BGAL* gene structure. In contrast, a minimum of 7 intron-insertion events would be needed to explain the current intron distribution patterns among Arabidopsis *BGAL* genes.

The intron–exon organization patterns of *BGAL17* and *BGAL14* are particularly noteworthy (Fig. 1). First, the exon-intron junctions of *BGAL17* are similar to those of animal *BGAL* genes but differ from those of the other Arabidopsis *BGAL* genes, even for the exons that encode the respective Glyco_hydro_35 domains. Secondly, with the exception of *BGAL17*, all Arabidopsis *BGALs* share strong sequence identities along their entire coding sequences. In contrast, *BGAL17* shows sequence similarity with other Arabidopsis *BGALs* only within its N-terminal Glyco_hydro_35 domain. However, this domain shares higher sequence identity with those of animal *BGALs* than with those of the other Arabidopsis and plant *BGALs*. *BGAL14*, which contains an extra exon encoding a PRP1_N domain, probably results from the fusion of *BGAL* with a PRN1_N domain-containing gene.

An additional gene, At2g04060, was listed as a putative *BGAL* in the Cell Wall Navigator (http://bioweb.ucr.edu/Cellwall/) and Carbohydrate-Active Enzymes databases. This gene encodes an ORF of 469 amino acids. Exons 2–9 share >90% sequence identity with exons 8–14 of *BGAL15* (At1g31740), while its exon 1 corresponds to the last half of exon 4 of *BGAL15* (Fig. 2). This suggests that At2g04060 is a truncated duplicate of *BGAL15*. However, At2g04060 contains an extra exon at its 3′-end that does not have any sequence identity with the last exon of *BGAL15*. It should be noted that, although the At2g04060 ORF has an intact sugar-binding domain (Glyco_hydro_2_N), it lacks those exons of the Glyco_hydro_35 domain that contain the two catalytic Glu residues. Because we, therefore, doubt that At2g04060 encodes a protein with BGAL activity, we decided not to annotate it as a *BGAL* gene.

2.3. Evolution of BGAL gene families

We used the Arabidopsis *BGAL* sequences to perform BLAST and HMM searches to identify related *BGAL* genes in *Oryza* (rice), *Chlamydomonas*, and 20 other eukaryotic genomes, whose genome sequences are available. It was found that the Arabidopsis and rice *BGAL* families are much larger than those of *Chlamydomonas*, fungi, and animals (Fig. 3). Whereas *Arabidopsis* and rice have 17 and 15 *BGAL* genes, respectively, only 1–4 *BGAL* genes were found in the fungal and animal genomes examined here, indicative of an expansion of *BGAL* genes in plants. To define relationships between eukaryotic *BGAL* genes, a phylogenetic tree was constructed based on the amino acid sequences of the Glyco_hydro_35 domains belonging to the genes shown in Fig. 3 and the 7 tomato *BGAL* genes (Smith and Gross, 2000). This tree revealed that all plant *BGAL* genes form a monophyletic group with the exception of one gene each from *Arabidopsis* (*BGAL17*; At1g72990) and rice (Os 9633.m04334) that cluster with the animal *BGAL* genes and *Dictyostelium discoideum BGAL* gene DD ng2214 (Fig. 4). Interestingly, a second *Dictyostelium BGAL* gene (DD ng2314) groups closer to the plant-specific *BGAL* genes. It is conceivable that *BGAL17* and Os 9633.m04334 might be derived by horizontal transfer of an animal *BGAL* gene before the divergence of monocots and dicots. The occurrence of such an event would be supported by the observed similarity in exon-intron organization of *BGAL17* and animal *BGAL* genes and by the fact that the *BGAL17* Glyco_hydro_35 domain shares higher sequence identities with those of animal *BGALs* than with those of the other Arabidopsis *BGALs*. Alternatively, because of *Dictyostelium's* position in both plant and animal groups of GH Family 35 genes, it seems more likely that two copies of ancestral *BGAL*s existed in the last eukaryotic common ancestor. According to the latter hypothesis, one copy of ancestral *BGAL* expanded greatly in plants, was retained in some fungal species, but was lost in metazoan animals. In contrast, the other copy was retained in metazoan animals, did not expand in plants, and was lost in *Chlamydomonas*. Additional *BGAL* genomic sequences, especially from lower plants, are needed to distinguish between these two alternatives for the origin of *BGAL17* and its rice ortholog.

The numbers and distribution of rice and Arabidopsis *BGAL* genes shown in Fig. 4 suggest that there is a lineage-specific expansion of plant *BGAL* genes. Based on the parsimony principle for reconciliation of gene trees and species trees (Page and Charleston, 1997), the "speciation" nodes are labeled in Fig. 4 with filled circles indicating the points of divergence between the Arabidopsis and rice lineages. Among 12 speciation nodes, six have experienced various degrees of lineage-specific expansion (i.e. duplication after the split between Arabidopsis and rice lineages) in either the Arabidopsis or the rice lineage (see arrows).

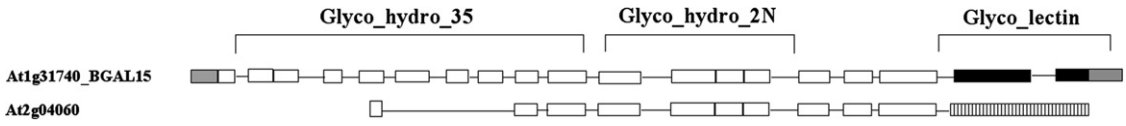

Fig. 2. Alignment of *BGAL15* and At2g04060. The symbols used are the same as those in Fig. 1, except for the last exon of At2g04060, which lacks any sequence identity with BGALs.

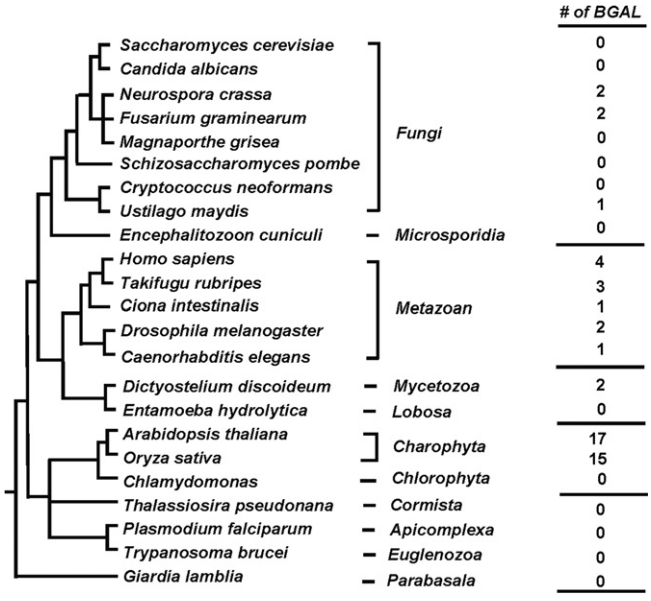

Fig. 3. Distribution of *BGAL* genes among eukaryotes. The Arabidopsis BGAL1 polypeptide sequence was used to search against the protein sequences from 22 other sequenced eukaryote genomes to identify the number of *BGAL* genes in each genome. The tree on the left indicates the phylogenetic relationships between the eukaryotes analyzed, whereas the number of *BGAL* genes in each genome is shown on the right. The tree is a modification of an eukaryote phylogeny published by Baldauf et al. (2000).

2.4. Expression profiles of BGAL genes

Relative RT-PCR was used to determine expression levels of all 17 Arabidopsis *BGAL* genes in different organs and during several developmental stages. Diverse expression profiles were observed (Fig. 5). *BGAL1*, *BGAL2*, and *BGAL3* are expressed constitutively across the different organs and developmental stages, whereas the remaining *BGAL* genes display differential expression patterns. The mRNA levels of *BGAL1*, *BGAL2*, and *BGAL3* are higher in mature leaves, roots, flowers, and siliques but are lower in young seedlings. Five genes (*BGAL8*, *BGAL11*, *BGAL13*, *BGAL14*, and *BGAL16*) are expressed mainly in flowers, whereas *BGAL6* is expressed mainly in roots. However, we were not able to detect the expression of *BGAL12*, *BGAL14*, and *BGAL15* under conditions examined. Results of microarray analyses from various databases and our laboratory indicate that several *BGAL* genes are regulated by abiotic and biotic stresses. For example, *BGAL*1 is induced under hypoxia (Shih, unpublished data), salt stress and pathogen attack (AtGen Express; http://jsp.weigelworld.org/expviz/expviz.jsp; Schmid et al., 2005). The differential tissue-specific expression patterns and stress responsiveness suggest that Arabidopsis *BGALs* function in diverse biological processes.

2.5. Heterologous expression in E. coli and purification of BGAL4

Our primary strategy to determine the biological functions of the Arabidopsis *BGAL* genes is to express each mature hydrolase in a suitable expression system and to identify its natural substrates through in vitro assays. Encoded by At5g56870, BGAL4 was selected as the first target hydrolase for expression in both prokaryotic and eukaryotic expression systems followed by extensive purification and characterization. Recombinant BGAL4 was

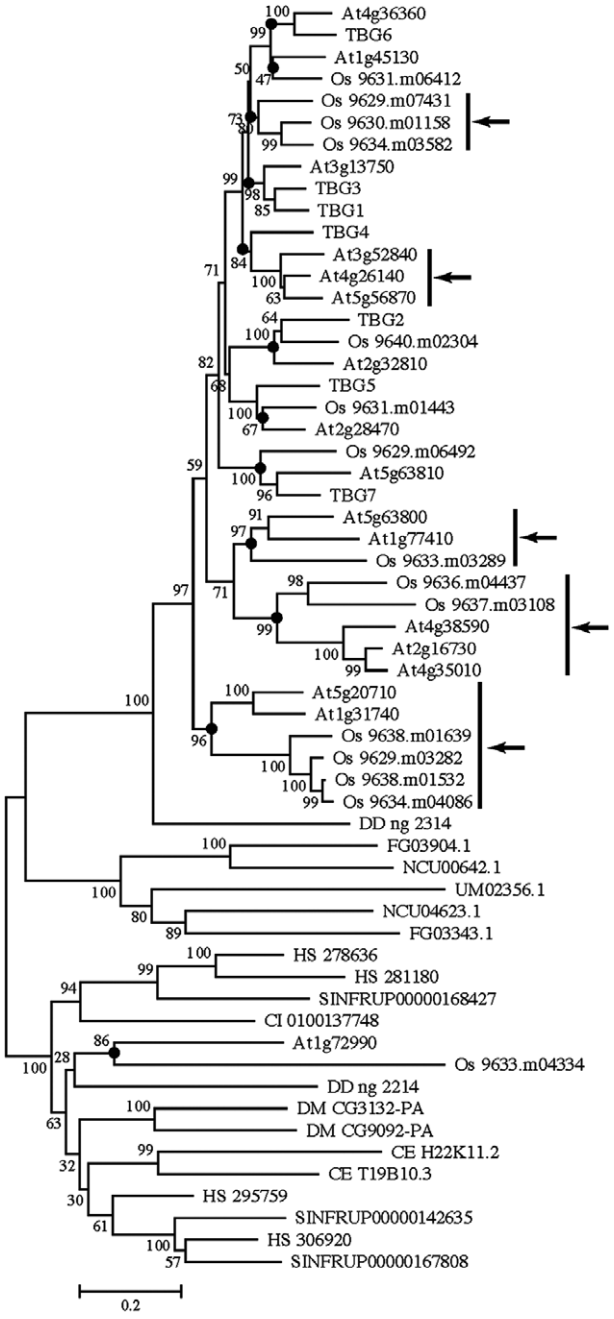

Fig. 4. Phylogenetic relationships between eukaryotic *BGAL* genes. A phylogenetic tree depicting the relationships among *BGAL* genes in the 23 eukaryotic genomes listed in Fig. 3. The seven known tomato *BGAL* genes (*TBG1* through *TBG7*) are also included. The abbreviated species names are as follows: At, *Arabidopsis thaliana*; CE, *Caenorhabditis elegans*; CI, *Clona intestinalis*; DD, *Dictyostelium discoideum*; DM, *Drosophila melanogaster*; FG, *Fusarium graminearum*; HS, *Homo sapiens*; NC, *Neurospora crassa*; Os, *Oryza sativa*; SINFRU, *Tagifugu rubripes*; TBG, *Lycopersicon esculentum*; UM, *Ustilago maydis*.

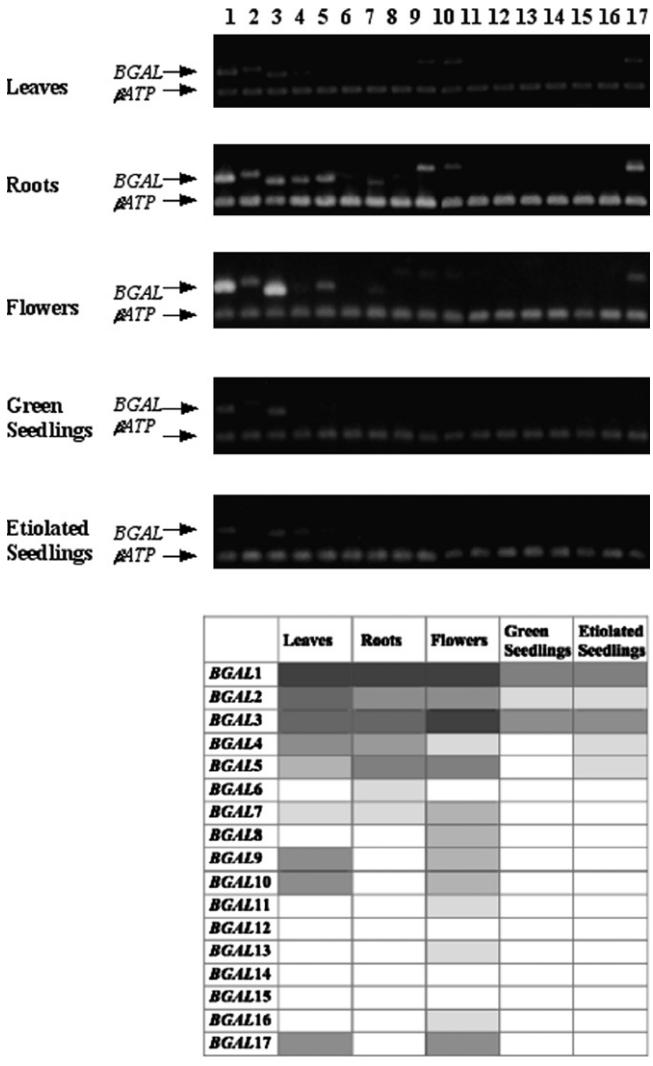

Fig. 5. Expression profiles of *Arabidopsis BGAL* genes at different developmental stages. Top panel: total RNAs from leaves, roots, flowers, and light-grown and etiolated seedlings were isolated and analysed by relative RT-PCR to determine steady-state transcript levels of the 17 Arabidopsis *BGAL* genes. RT-PCR products were analysed by agarose gel electrophoresis. The nuclear gene β*ATP*, which encodes the β-subunit of mitochondrial ATP synthase and expresses at a constant level during these developmental stages, served as internal standard. Bottom panel: The relative intensity of the *BGAL* and β*ATP* bands from each lane was used as an index of the relative expression level of each *BGAL* gene. The level for each *BGAL* at a particular development stage is shown as a shaded box. The darker gray represents a higher expression level, whereas a white box indicates no detectable RT-PCR product.

initially expressed in *E. coli* strain BL21 Codon Plus (DE3) cells but failed to hydrolyse the chromogenic substrates *p*-nitrophenyl-β-D-galactoside (*p*NPGal) and *o*-nitrophenyl-β-D-galactoside (*o*NPGal), even when isolated from the soluble fraction. Because BGAL4 has 10 Cys residues, we postulated that this lack of enzymatic activity might be due to the absence of one or more disulfide bonds critical for correct folding and/or activity. To test this hypothesis, the pET-BGAL4 construct was transformed into *E. coli* Rosetta-gami cells, which are devoid of thioredoxin reductase and therefore provide an oxidizing cytosolic environment for formation of disulfide bonds (Prinz et al., 1997). Indeed, using this strain, we were able to produce mature BGAL4 protein having an N-terminal His-tag in soluble and catalytically active form. The recombinant hydrolase was purified by Ni–Sepharose chromatography followed by hydrophobic interaction chromatography and ion-exchange chromatography, resulting in over 31-fold purification and a 14% yield (Table 2). The Ni-affinity chromatography was relatively ineffective for purification, because BGAL4 did not bind well to this matrix. It is possible that the enzyme's N-terminal His-tag is not sufficiently exposed on the surface for binding. After SE-cellulose chromatography and subsequent removal of the N-terminal His-tag by thrombin cleavage, BGAL4 appeared on SDS–PAGE gels as a single band (Fig. 6a, lane 4), whose estimated molecular mass of 74 kDa is close to that predicted (77.6 kDa) for the mature BGAL4 polypeptide (697 amino acids) plus an additional N-terminal 19 amino acids contributed by the vector. The 74-kDa polypeptide showed specific immunoreactivity with anti-BGAL4 serum on western blots (Fig. 6c). Most importantly, recombinant BGAL4 exhibited BGAL activity toward both *p*NPGal and *o*NPGal, as seen by increasing amounts of free aglycone produced with increasing incubation time or enzyme amount.

2.6. Properties of recombinant BGAL4 enzyme

Recombinant BGAL4 was stable in the pH range 4.5–8.5 and displayed highest activity at pH 4.5 (data not shown). Several other plant β-galactosidases have pH optima ranging from 3.5 to 5.0 (Edwards et al., 1988; Ross et al., 1994; Smith and Gross, 2000; Li et al., 2001; Kotake et al., 2005). Furthermore, Arabidopsis BGAL4 was stable up to 45 °C but lost activity precipitously at higher temperatures; over 90% activity was lost when incubated at 55 °C for 30 min prior to assay. The optimum temperature for enzyme activity was 40 °C (data not shown).

Apparent K_m and V_{max} values were determined for artificial substrates from Lineweaver-Burk plots (Table 3). The K_m values for the *E. coli*-expressed BGAL4 with *p*NPGal and *o*NPGal were 3.4 mM and 2.2 mM, respectively. The V_{max} values for the same substrates were 467 and 1630 nkat/mg protein, respectively. Recombinant BGAL4 also hydrolysed the artificial substrates 4-methylumbelliferyl-β-D-galactoside (4-MUGal) and 6-bromo-2-naphthyl-β-D-galactoside (6-BNGal) in zymogram assays (Fig. 7). D-Galactonic acid-γ-lactone, a transition state analog of galactose, strongly inhibited recombinant BGAL4 ($K_i = 35$ μM), whereas galactose was 170 times less inhibitory ($K_i = 6$ mM) (data not shown). In contrast, lactose exhibited no detectable inhibition.

To probe the glycone specificity of the *E. coli*-expressed BGAL4, we tested its activity toward several *p*NP-glycosides was tested. As Table 4 indicates, this hydrolase showed a pronounced preference for *p*NPGal (100%) and *p*NP-fucoside (18%), with less than 2% activity being

Table 2
Purification of BGAL4 expressed in *E. coli*

Step	Total protein (mg)	Total activity[a] (nkat)	Specific activity (nkat/mg)	Purification (fold)	Recovery (%)
Crude	80.3	822	10.2	1	100
Ni–sepharose	18.9	241	12.7	1.2	29
Toyopearl ether	4.4	153	34.7	3.4	18
SE-cellulose	0.36	117	325	31.8	14

[a] Substrate used: *p*NPGal.

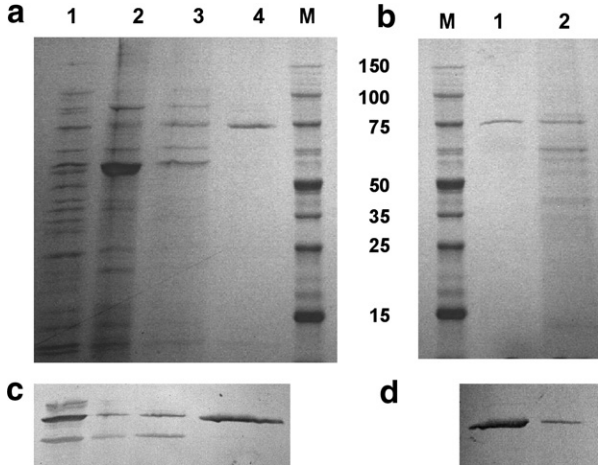

Fig. 6. SDS–PAGE (a and b) and western blotting (c and d) profiles of various BGAL4 preparations. (a) BGAL4 preparations from expression in *E. coli*. Lane 1, crude cell lysate; Lane 2, Ni–Sepharose column eluate; Lane 3, Toyopearl Ether-650 column eluate; Lane 4, SE-cellulose column eluate; Lane M, molecular mass markers. (b) BGAL4 preparations from expression in insect cells. Lane M, molecular mass markers; Lane 1, CM-Sepharose column eluate; Lane 2, crude cell lysate. (c) and (d) western blots of gels identical to those shown in (a) and (b), respectively, but immunostained with anti-BGAL4 serum.

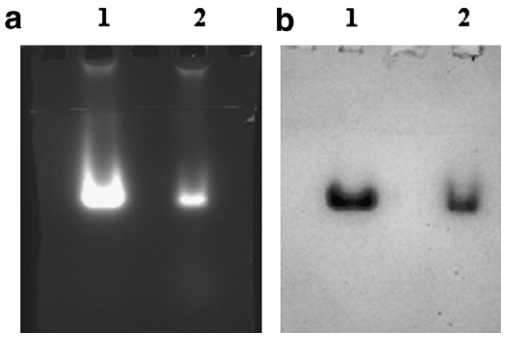

Fig. 7. Hydrolysis of 4-MUGal and 6-BNGal by purified BGAL4 derived from expression in *E. coli* and insect cells. Purified BGAL4 preparations were subjected to native PAGE (8% acid gels), before staining gels with 4-MUGal (a) or 6-BNGal (b). Lane 1, BGAL4 expressed in *E. coli*; Lane 2, BGAL4 expressed in insect cells.

shown toward all other *p*NP-glycosides tested. Given that fucose is 6-deoxy-D-galactose, *p*NP-fucoside hydrolysis is not surprising. It should be noted that β-galactosidases purified from mung bean seedlings also hydrolysed this substrate (Li et al., 2001), but, in contrast to these enzymes, BGAL4 exhibited insignificant activity toward *p*NP-arabinoside. Taken together, our data indicate that both the C_4 and C_6 positions of the galactose moiety are important for substrate binding and catalysis by Arabidopsis BGAL4.

To determine the activity of BGAL4 toward oligosaccharides, the purified enzyme from *E. coli* expression was incubated with β-(1,3)-, β-(1,4)-, and β-(1,6)-linked galactobiosides and galactotrioses. Reaction products were analysed by thin-layer chromatography. As Fig. 8 shows, BGAL4 hydrolysed all three linkage types with a marked preference for β-(1,3)- and β-(1,4)-linkages. However, BGAL4 failed to hydrolyse the galactopolysaccharides (+)-arabinogalactan, gum arabic galactan, and gum guar (data not shown).

2.7. Heterologous expression of BGAL4 in eukaryotic expression systems

Because BGAL4 has a putative *N*-glycosylation site, and because such glycosylation would not occur in *E. coli*, BGAL4 was also expressed in a baculoviral expression system. In brief, the full-length cDNA encoding the BGAL4 precursor protein was cloned into the plasmid pFastBac for expression in insect cells driven by the constitutive polyhedrin promoter. For this, we used the Hi-five cell line was used, which is more efficient for production of recombinant proteins (Invitrogen). Lack of any detectable endogenous

Table 3
BGAL4 kinetic parameters

Recombinant BGAL4 source	Substrate	K_m (mM)	V_{max} (nkat/mg)	k_{cat} (s^{-1})	Catalytic efficiency (s^{-1} mM^{-1})
E. coli	*p*NPGal	3.4 (±0.1)	467 (± 52)	65	19.1
	*o*NPGal	2.2 (±0.3)	1630 (±150)	226	103
Insect	*p*NPGal	3.0 (±0.3)	277 (± 60)	33	11
	*o*NPGal	2.1 (±0.2)	1315 (± 35)	157	75

Enzyme activity was measured under standard assay conditions by varying substrate concentrations from 0 to 40 mM. K_m and V_{max} values were calculated from a Lineweaver–Burk plot using Enzyme Kinetics Pro software (Trinity Software, Inc.).

Table 4
Glycone specificity of purified recombinant BGAL4 expressed in *E. coli*

Substrate	Relative activity (%)
PNP-β-D-galactoside	100
PNP-β-D-glucoside	0.2
PNP-β-D-fucoside	18.8
PNP-β-L-fucoside	1.6
PNP-α-L-arabinoside	0.7
PNP-β-L-arabinoside	0
PNP-β-D-xyloside	0
PNP-β-D-mannoside	0
PNP-β-D-lactoside	0
PNP-β-D-*N*-acetyl-glucosaminide	0
PNP-β-D-glucuronide	1.9
PNP-β-D-cellobioside	0

Substrates were provided at 2 mM final concentration. Reaction rates are expressed as a percentage of that observed with *p*NPGal (100% = 155 nkat/mg protein).

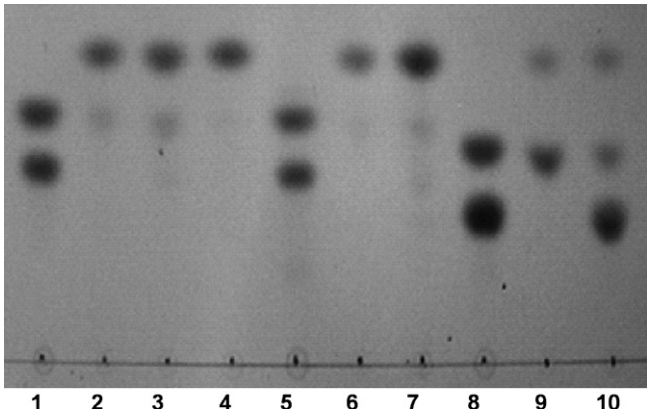

Fig. 8. Thin-layer chromatogram of hydrolysis products of various β-linked galactobioses and galactotrioses incubated with purified BGAL4 expressed in *E. coli*. Lane 1, β-1,3-galactobiose (upper) and β-1,3-galactotriose (lower) standards; Lanes 2 and 3, BGAL4 incubated with β-1,3-galactobiose and β-1,3-galactotriose, respectively; Lane 4, galactose standard; Lane 5, β-1,4-galactobiose (upper) and β-1,4-galactotriose (lower) standards; Lanes 6 and 7, BGAL4 incubated with β-1,4-galactobiose and β-1,3-galactotriose, respectively; Lane 8, β-1,6-galactobiose (upper) and β-1,6-galactotriose (lower) standards; Lanes 9 and 10, BGAL4 incubated with β-1,6-galactobiose and β-1,6-galactotriose, respectively.

BGAL activity in insect cells allowed us to monitor BGAL4 expression and purification unequivocally. BGAL activity was detected 72 h post-infection, predominantly in the growth medium of infected cells. Surprisingly, BGAL4 did not bind to Concanavalin A-Sepharose, suggesting that the recombinant enzyme remained unglycosylated or was glycosylated in an unanticipated manner. Therefore, BGAL4 was purified by cation exchange chromatography on CM-Sepharose, yielding a near homogenous preparation (Fig. 6b) that reacted with anti-BGAL4 serum (Fig. 6d). The recombinant enzymes from expression in *E. coli* and insect cells shared virtually identical pH and temperature profiles for stability and activity (data not shown). Furthermore, insect cell-expressed BGAL4 also hydrolyzed 4-MUGal and 6-BNGal in zymogram assays (Fig. 7), exhibited somewhat lower K_m and V_{max} values with *p*NPGal and *o*NPGal (Table 3), and gave identical hydrolysis patterns with the same galactooligosaccharides (data not shown).

The cDNA encoding the mature BGAL4 protein was also cloned into *Pichia pastoris*, expressing it under the control of the inducible alcohol oxidase promoter. Soluble and catalytically active enzyme was produced and secreted into growth medium, but, because the recombinant protein lacked an affinity tag, we did not attempt to undertake its large-scale purification and characterization.

3. Concluding remarks

3.1. Proposed physiological role of Arabidopsis BGAL4

The most well investigated *BGAL* genes to date are the tomato genes *TBG1* through *TBG7* (Smith and Gross, 2000; Smith et al., 2002; Ishimaru et al., 2005). Our study has now shown that Arabidopsis *BGAL4* belongs to the same orthologous group in the *BGAL* phylogenetic tree as *TBG4* (Fig. 4). In addition to 72% sequence identity, these two hydrolases share similar pH and temperature optima. Recently, it was shown that yeast-expressed TBG4 hydrolyses plant-derived cell wall substrates, with highest activity being exhibited toward the alkali-soluble and chelator-soluble pectin fractions from ripening tomato fruit at the turning stage (Ishimaru et al., 2005). With BGAL4 also predicted to be targeted to the cell exterior, it is tempting to propose that this hydrolase is likewise involved in pectin degradation. Efforts are underway to test its activity toward Arabidopsis cell wall fractions (Zablackis et al., 1995; Willats et al., 2001).

4. Experimental

4.1. Sequence retrieval and computer-based analyses

A. thaliana BGALs were identified by a BLAST search of the TIGR Arabidopsis Genome Annotation Database annotation v.5 (http://www.tigr.org/tdb/e2k1/ath1/ath1.shtml), and their protein domains were obtained by searching against the SMART (Schultz et al., 2000) and Pfam (Sonnhammer et al., 1998) databases. To find BGALs from other eukaryotes, the HMMs of BGALs were retrieved from Pfam (Pfam:Glyco_hydro_35) to search against the protein sequences from 22 other eukaryote genomes as described in Shiu et al. (2004, 2005). The Arabidopsis genome was analysed in the same way to check for potential changes in annotation. These BGAL protein sequences were aligned using ClustalW (Higgins et al., 1996; http://www.ebi.ac.uk/clustalw/index.html) with the default setting. Phylogenetic trees were constructed using the neighbor-joining method (Saitou and Nei, 1987) as implemented in MEGA software package (Kumar et al.,

2004) with the pairwise deletion option for treating gaps/missing data, Poisson correction for multiple substitutions, and 1,000 bootstrap replicates. The percent bootstrap supports are shown if >50%. The tree is rooted at midpoint. Microarray data for all Arabidopsis *BGAL* genes except *BGAL15* were retrieved from AtGen Express (http://www.weigelworld.org/resources/microarray/AtGenExpress/). GeneSpring 7.0 software (Silicon Genetics) was used for cluster analyses.

4.2. RNA isolation and RT-PCR analysis

Total RNA was isolated by an acidic phenol protocol adapted from the procedures described by Chomczynski and Sacchi (1987). Prior to RT-PCR reactions, RNA samples were treated twice with DNase I to deplete contaminating genomic DNA. First-strand cDNA synthesis was performed in a 20 µl reaction mixture containing 1 µg RNA, 0.5 µg of random hexamer, 40 µM each of the four dNTPs, 2 µl 10× buffer (500 mM Tris–HCl, pH 8.3, room temperature, 750 mM KCl, and 30 mM $MgCl_2$), 200 units of Moloney Murine Leukemia Virus reverse transcriptase, and 20 units of RNase inhibitor. The reaction mixture was incubated at 42 °C for 60 min and stopped by heating at 65 °C for 10 min. To determine the temporal and tissue-specific expression patterns of *BGAL* genes, relative PCR was performed using the level of β*ATP* (gene encoding ATP synthase β-subunit) as the internal standard (Sellner and Turbett, 1998; Schmittgen et al., 2000). Reactions contained 4 µl 10× Mg-free buffer, 4 µl 25 mM $MgCl_2$, 1 µl 10 mM dNTPs, 1 µl 5 pmol/µl each of forward and backward primers, and 1 µl single-stranded cDNA. The solution was overlaid with 40 µl mineral oil. The reaction mixture was maintained at 94 °C for 5 min, during which time 2 units of *Taq* polymerase in 10 µl 1× buffer were added. The subsequent thermal cycling conditions were 94 °C for 30 s, 54 °C for 30 s, and 72 °C for 90 s for 25 cycles, before products were analysed by agarose gel electrophoresis. After electrophoresis, the resolved PCR products were imaged by UV illuminator and digitally photographed (DC120 digital camera; Eastman Kodak, Rochester, NY). The intensities of the cDNA bands were quantified by computerized image analysis using the NIH Image Analysis Program 1.62f.

4.3. Heterologous expression and purification of BGAL4 in E. coli

A cDNA fragment encoding the mature BGAL4 hydrolase was amplified from the RAFL11-07-J10 clone (GenBank Accession No. AY054589; At5g56870) by PCR using sense primer BGAL4-1 (5′-GAATTCTCAGTCTCTTATGATCGTAAAGC-3′; *Eco*RI site underlined) and antisense primer BGAL4-2 (5′-GCGGCCGCAAACGAATCAAACAGCTTG-3′; *Not*I site underlined). The amplicon was first cloned into the entry vector pBluescript SK(+). After blue-white selection and confirming the presence of recombinant plasmid by colony PCR, a single colony was used as inoculum to grow a culture for plasmid isolation. The plasmid was digested with *Eco*RI and *Not*I to release the BGAL4 insert, which was then gel purified and cloned into the expression vector pET-28a (Novagen). This procedure adds a hexahistidine sequence to the N-terminus of the recombinant protein, and the resulting construct pET-BGAL4 was transformed into *E. coli* strain Rosetta-gami. After screening for recombinants by colony PCR, a positive clone was selected and used for recombinant protein production. Cultures were grown to log phase ($A_{600} = 0.8$) at 30 °C, before adding isopropyl-β-D-thiogalactopyranoside (0.6 mM final concentration) for induction of BGAL4 expression. Cultures were then incubated at 22 °C for 24 h. Cells were harvested by centrifugation at 5,000 g for 10 min and lysed in a French press, before centrifugation twice at 15,000 g for 30 min. The recombinant protein was first purified from lysates by Ni–Sepharose affinity chromatography following the protocol provided by the vendor (Amersham Pharmacia Biotech). To remove minor protein contaminants, the affinity column eluate was then applied to a Toyopearl Ether-650 column (Toso-Haus), which had been pre-equilibrated with 40% (w/v) $(NH_4)_2SO_4$ in buffer A (50 mM sodium citrate, pH 6.0). After washing the column with the same buffer to remove unbound proteins, BGAL4 was eluted with a linear gradient of 40–0% $(NH_4)_2SO_4$ in buffer A at a flow rate of 1 ml/min. Fractions were assayed for BGAL activity, and the active fractions, which eluted between 10% and 5% $(NH_4)_2SO_4$, were pooled and applied to a sulphoxyethyl (SE)-cellulose (Whatman) column pre-equilibrated with buffer A. After extensive washing with buffer A, bound proteins were eluted with a linear gradient of 0–800 mM NaCl in buffer A at a flow rate of 1 ml/min. BGAL4 eluted from the SE-cellulose column with 150–200 mM NaCl. Prior to SDS–PAGE analysis, the N-terminal His-tag was removed by thrombin cleavage.

4.4. Enzyme and protein assays

BGAL activities were routinely determined by measuring the liberation of aglycone from either *p*NPGal or *o*NPGal. All assays were performed in triplicate in wells of 96-well microtiter plates. The 220-µl reaction mixture contained 5 mM substrate, 20 mM citrate buffer (pH 4.5), and 30 µl of diluted enzyme solution. After adding the enzyme, the reaction mixture was incubated at 37 °C for 15–30 min before terminating the assay with 80 µl of 1 M Na_2CO_3. As control, a mixture without enzyme was included in each assay. The amount of aglycone released was determined spectrophotometrically at 405 nm (*p*-nitrophenol) or 410 nm (*o*-nitrophenol). Calibration curves were constructed using 0–0.5 mM *p*-nitrophenol and 0–2.5 mM *o*-nitrophenol ranges. Protein concentration was determined by the Bradford method (1976) using the Coomassie Blue G-250 protein assay reagent (Pierce) with bovine serum albumin as standard.

4.5. Immunological methods

To express BGAL4 in *E. coli* for antiserum production, a cDNA fragment encoding the mature BGAL4 hydrolase was amplified from the RAFL11-07-J10 clone (GenBank Accession No. AY054589; At5g56870) by PCR and cloned into the expression plasmid pET21a. The resulting construct was transformed into *E. coli* BL21-CodonPlus-RIL competent cells (Stratagene). For induction and expression experiments, a positive clone was grown in 200-ml cultures and induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside at 25 °C for 6 h. After induction, the cells were harvested for protein extraction with an optimized extraction buffer (Cicek and Esen, 1999). The insoluble protein fraction (i.e. inclusion body) containing BGAL4 was solubilized in 6 M urea and electrophoresed on a 12% (w/v) preparative SDS–PAGE gel. The 78-kDa BGAL4 band was visualized by staining the gel briefly with Coomassie Blue R-250. After excision and destaining with 50% MeOH:H_2O (1:1) methanol, the band was rehydrated in 1× PBS, homogenized by grinding to pass through a 20-gauge syringe needle, and mixed with Freund's Complete Adjuvant for the initial immunization and with Freund's Incomplete Adjuvant for subsequent immunizations. Following pre-immune bleeding, the rabbit was immunized 4 times at 2-week intervals, before being bled out under euthanasia. The antiserum was recovered and stored at −20 °C until used.

4.6. Electrophoresis and western blotting

The purity of fractions exhibiting BGAL4 activity and the molecular mass of the BGAL4 polypeptide were determined by SDS–PAGE on 10% (w/v) polyacrylamide minigels according to Laemmli (1970). Protein was detected by staining with Coomassie Brilliant Blue R-250, and molecular masses were estimated using a mixture of calibration standards covering the 15–150 kDa range (Amresco). For native PAGE, proteins were resolved on 8% (w/v) gels (stacking gel: 0.5 M Tris–HCl, pH 6.8; resolving gel: 1.5 M AcOH–KOH, pH 5.0) (Hamer and Rickwood, 1981). After electrophoresis, gels were equilibrated in buffer B (50 mM citrate-sodium phosphate buffer, pH 4.5) for 10 min before incubation with 1 mM 4-MUGal in buffer B at 37 °C for 1 h. The fluorescent aglycone product was visualized under UV light in a MultiImage light cabinet (Alpha Innotech). Alternatively, gels were incubated with 0.05% 6-BNGal and 0.75 mg/ml Fast Blue RR salt in 50 mM acetate buffer (pH 5.0) at 37 °C overnight. For western blotting, proteins were transferred electrophoretically to a PVDF membrane (Millipore) in 1× blotting buffer (25 mM Tris-125 mM glycine/0.25% SDS/20% MeOH). Immunodetection was performed with anti-BGAL4 serum and horseradish peroxidase-conjugated goat anti-rabbit antibodies following standard procedures.

4.7. TLC analysis

The β-(1,3)- and β-(1,6)-linked galactobioses and galactotrioses were prepared from larch wood arabinogalactan, whereas the β-(1,4)-linked galactooligosaccharides were prepared from soybean arabinan galactan, respectively (Kotake et al., 2005). Galactobioses and galactotrioses were incubated with BGAL4 in buffer B at 37 °C overnight. Reaction products were resolved by TLC on silica gel 60 F_{254} plates (Merck 5715, 0.25 mm) using the solvent system *n*-BuOH–HOAc–H_2O (3:1:2) and detected by spraying with 0.2% (w/v) naphthoresorcinol in H_2SO_4–EtOH (1:19) and heating on a heat block at ~80 °C (Anderson and Stoddart, 1966).

Acknowledgements

We are grateful to RIKEN BioResource Center (Ibaraki, Japan) for providing the RAFL11-07-J10 clone (GenBank Accession No. AY054589). We acknowledge the expert technical assistance of the Center for Comparative Genomics (University of Iowa). This work was supported by the Arabidopsis 2010 Program of the National Science Foundation (MCB-0114666 to J.E.P. and MCB-0115937 to A.E.). We also thank Dr. Toshihisa Kotake, Saitama University, Japan, for kindly providing galactooligosaccharide substrates.

References

Anderson, D.M.W., Stoddart, J.F., 1966. Separations of sugars on Chromagrams. Carbohyd. Res. 1, 417–418.

Balasubramaniam, S., Lee, H.C., Lazan, H., Othman, R., Ali, Z.M., 2005. Purification and properties of a β-galactosidase from carambola fruit with significant activity towards cell wall polysaccharides. Phytochemistry 66, 153–163.

Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., Doolittle, W.F., 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290, 972–977.

Bhalla, P.L., Dalling, M.J., 1984. Characteristics of a β-galactosidase associated with the stroma of chloroplasts prepared from mesophyll protoplasts of the primary leaf of wheat. Plant Physiol. 76, 92–95.

Blanchard, J.E., Gal, L., He, S., Foisy, J., Warren, R.A., Withers, S.G., 2001. The identification of the catalytic nucleophiles of two β-galactosidases from glycoside hydrolase family 35. Carbohyd. Res. 333, 7–17.

Bradford, M.M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254.

Buckeridge, M.S., Reid, J.S., 1994. Purification and properties of a novel β-galactosidase or exo-(1 → 4)-β-D-galactanase from the cotyledons of germinated *Lupinus angustifolius* L. seeds. Planta 192, 502–511.

Carey, A.T., Holt, K., Picard, S., Wilde, R., Tucker, G.A., Bird, C.R., Schuch, W., Seymour, G.B., 1995. Tomato exo-(1 → 4)-β-D-galactanase 1. Isolation, changes during ripening in normal and mutant tomato fruit, and characterization of a related cDNA clone. Plant Physiol. 108, 1099–1107.

Chomczynski, P., Sacchi, N., 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate phenol-chloroform extraction. Anal. Biochem. 162, 156–160.

Cicek, M., Esen, A., 1999. Expression of soluble catalytically active plant (monocot) β-glucosidases in *E. coli*. Biotechnol. Bioeng. 63, 392–400.

Coutinho, P.M., Henrissat, B., 1999. Carbohydrate-active enzymes: an integrated database approach. In: Gilbert, H.J., Davies, G., Henrissat, B., Svensson, B. (Eds.), . In: Recent Advances in Carbohydrate Bioengineering. The Royal Society of Chemistry, Cambridge, pp. 3–12.

de Alicântara, P.H.N., Dietrich, S.M.B., Buckeridge, M.S., 1999. Xyloglucan mobilization and purification of a (XLLG/XLXG) specific β-galactosidase from cotyledons of *Copaifera langsdorfii*. Plant Physiol. Biochem. 37, 653–663.

Dey, P.M., del Campillo, E., 1984. Biochemistry of the multiple forms of glycosidases in plants. In: Meister, A. (Ed.), . In: Adv. Enzymol. Relat. Areas Mol. Biol., Vol. 56. Wiley, New York, pp. 141–249.

Dopico, B., Nicolas, G., Labrador, E., 1989. Partial purification of cell wall β-galactosidases from *Cicer arietinum* epicotyls. Physiol. Plant 75, 458–464.

Edwards, M., Bowman, Y.J., Dea, I.C., Reid, J.S., 1988. A β-D-galactosidase from nasturtium (*Tropaeolum majus* L.) cotyledons. Purification, properties, and demonstration that xyloglucan is the natural substrate. J. Biol. Chem. 263, 4333–4337.

Gy, I., Kreis, M., Lecharny, A., 2000. The β-galactosidases are encoded by a multigene family in *A. thaliana*. GenBank Accessions CAB64773-64750.

Hamer, B.D., Rickwood, D., 1981. Preparation and electrophoresis of polyacrylamide gels. In: Hames, B.D., Rickwood, D. (Eds.), Gel Electrophoresis of Proteins: A Practical Approach. IRL Press Ltd., London and Washington, DC, pp. 23–42.

Henrissat, B., 1998. Glycosidase families. Biochem. Soc. Trans. 26, 153–156.

Higgins, D.G., Thompson, J.D., Gibson, T.J., 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. 266, 382–402.

Ishimaru, M., Smith, D., Gross, K., 2005. Yeast expressed tomato β-galactosidases 1, 4 and 5 have activity against synthetic and plant derived cell wall substrates. Hortscience 40, 1092.

King, G.A., Davies, K.M., 1995. Cloning of a harvest-induced β-galactosidase from tips of harvested asparagus spears. Plant Physiol. 108, 419–420.

Kotake, T., Dina, S., Konishi, T., Kaneko, S., Igarashi, K., Samejima, M., Watanabe, Y., Kimura, K., Tsumuraya, Y., 2005. Molecular cloning of a β-galactosidase from radish that specifically hydrolyzes β-(1 → 3)- and β-(1 → 6)-galactosyl residues of arabinogalactan protein. Plant Physiol. 138, 1563–1576.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings Bioinform. 5, 150–163.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 277, 680–685.

Lazan, H., Ng, S.Y., Goh, L.Y., Ali, Z.M., 2004. Papaya β-galactosidase/galactanase isoforms in differential cell wall hydrolysis and fruit softening during ripening. Plant Physiol. Biochem. 42, 847–853.

Lee, D.H., Kang, S.G., Suh, S.G., Byun, J.K., 2003. Purification and characterization of a β-galactosidase from peach (*Prunus persica*). Mol. Cells 15, 68–74.

Li, S.C., Han, J.W., Chen, K.C., Chen, C.S., 2001. Purification and characterization of isoform of β-galactosidases in mung bean seedlings. Phytochemistry 57, 349–359.

McCarter, J.D., Burgoyne, D.L., Miao, S., Zhang, S., Callahan, J.W., Withers, S.G., 1997. Identification of Glu-268 as the catalytic nucleophile of human lysosomal β-galactosidase precursor by mass spectrometry. J. Biol. Chem. 272, 396–400.

Page, R.D.M., Charleston, M.A., 1997. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. Mol. Phylogen. Evol. 7, 231–240.

Pressey, R., 1983. β-Galactosidases in ripening tomatoes. Plant Physiol. 71, 132–135.

Prinz, W.A., Aslund, F., Holmgren, A., Beckwith, J., 1997. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. J. Biol. Chem. 272, 15661–15667.

Ragothama, K.G., Lawton, K.A., Goldsborough, P.B., Woodson, W.R., 1991. Characterization of an ethylene-regulated flower senescence-related gene from carnation. Plant Mol. Biol. 17, 61–71.

Ross, G.S., Wegrzyn, T., MacRae, E.A., Redgwell, R.J., 1994. Apple β-galactosidase activity against cell wall polysaccharides and characterization of a related cDNA clone. Plant Physiol. 106, 521–528.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., Lohmann, J., 2005. A gene expression map of *Arabidopsis* development. Nature Genet. 37, 501–506.

Schmittgen, T.D., Zakrajsek, B.A., Mills, A.G., Gorn, V., Singer, M.J., Reed, M.W., 2000. Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. Anal. Biochem. 285, 194–204.

Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P., Bork, P., 2000. SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res. 28, 231–234.

Sekimata, M., Ogura, K., Tsumuraya, Y., Hashimoto, Y., 1989. A β-galactosidase from radish (*Raphanus sativus* L.) seeds. Plant Physiol. 90, 567–574.

Sellner, L.N., Turbett, G.R., 1998. Comparison of three RT-PCR methods. Biotechniques 25, 230–234.

Shiu, S.-H., Karlowski, W.M., Pan, R., Tzeng, Y.H., Mayer, K.F., Li, W.-H., 2004. Comparative analysis of the receptor-like kinase family in Arabidopsis and rice. Plant Cell 16, 1220–1234.

Shiu, S.-H., Shih, M.-C., Li, W.-H., 2005. Higher retention rate of transcription factor families in plants than in animals and their preferential retention among duplicated plant genes. Plant Physiol. 139, 18–26.

Smith, D.L., Starrett, D.A., Gross, K.C., 1998. A gene coding for tomato fruit β-galactosidase II is expressed during fruit ripening. Plant Physiol. 117, 417–423.

Smith, D.L., Abbott, J.A., Gross, K.C., 2002. Down-regulation of tomato β-galactosidase 4 results in decreased fruit softening. Plant Physiol. 129, 1755–1762.

Smith, D.L., Gross, K.C., 2000. A family of at least seven β-galactosidase genes is expressed during tomato fruit development. Plant Physiol. 123, 1173–1183.

Sonnhammer, E.L.L., Eddy, S.R., Birney, E., Bateman, A., Durbin, R., 1998. Pfam: multiple sequence alignments and HMM-profiles of protein domains. Nucleic Acids Res. 26, 320–322.

The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815.

Triantafillidou, D., Georgatsos, J.G., 2001. Barley β-galactosidase: structure, function, heterogeneity, and gene origin. J. Protein Chem. 20, 551–562.

Willats, W.G., McCartney, L., Mackie, W., Knox, J.P., 2001. Pectin: cell biology and prospects for functional analysis. Plant Mol. Biol. 47, 9–27.

Yoon, J.H., Ajisaka, K., 1996. The synthesis of galactopyranosyl derivatives with β-galactosidases of different origins. Carbohyd. Res. 292, 153–163.

Zablackis, E., Huang, J., Müller, B., Darvill, A.G., Albersheim, P., 1995. Characterization of the cell-wall polysaccharides of *Arabidopsis thaliana* leaves. Plant Physiol. 107, 1129–1138.