Elucidation of the Putative Glycosides Hydrolases Responsible for the Smoky Aroma in Tomato

MSc Thesis

Orianne Gudenschwager

Registration number: 800328288110
MSc Plant Biotechnology

Supervisors: Dr. Yury Tikunov
Dr. Arnaud Bovy

ABSTRACT

Some tomato genotypes possess a characteristic flavor that is called “smoky” and it is clearly determined by the presence of three volatiles: guaiacol, methyl salicylate and eugenol. Their production is regulated by a ripening-induced modification of the corresponding diglycosides precursors. Smoky genotypes release guaiacol, methyl salicylate and eugenol upon tissue disruption through the action of glycosides hydrolases. The nature and the substrate specificity of these enzymes remain unknown. In this thesis project, the hydrolytic activity towards diglycoside precursors of smoky volatiles was studied through the expression of recombinant proteins from candidate tomato genes in *N. benthamiana* and *E. coli*. Only one candidate gene encoding for α-L-arabinofuranosidase/β-D-xylosidase showed hydrolytic activity towards the diglycosides precursors. However, its action did not involve the cleavage of smoky volatiles from their glycoconjugates, but the cleavage of a pentose moiety, which function might play a role in cell wall modification. A tea primeverosidase was expressed in *N. benthamiana* showing a high hydrolytic activity to many of the substrates analyzed. This enzyme could be used in further studies as an effective positive control. *N. benthamiana* showed endogenous hydrolytic activity which was revealed by the large amount of guaiacol and eugenol released by the wild type leaves. Therefore, *E. coli* protein expression under an optimized experimental setup might be a better alternative for the study of hydrolytic activities of putative glycosides hydrolases. Further studies are needed to reveal the mechanisms involved in the emission of smoky aroma that could contribute to the development of new strategies with the ultimate aim of producing tomatoes with qualities that are consistent with consumer preferences.

TABLE OF CONTENTS

1. INTRODUCTION

The quality of fresh tomato is related to many attributes, such as appearance, texture, colour and flavour, being this last one relying on a combination of taste and aroma (Beckles, 2012). In addition to its biological importance to fruit species survival, aroma is a key contributor to fruit flavour perception and plays a primary role in consumer acceptability. However, the efforts to provide fresh fruits of high quality over a long period of time are made primarily in terms of appearance and texture. This often result in fruits with poor flavour quality, which is currently the main complaint of consumers of fresh tomatoes (Ortiz-Serrano and Gil, 2007).

A complex mixture of a large number of odour-active volatile compounds determines the aroma characteristics (Dixon and Hewett, 2000), where differences in the concentrations and thresholds of key volatiles ultimately determine the distinctive aroma of a particular fruit species or cultivar (Defilippi et al., 2009). Over 400 aroma volatiles have been identified in fresh tomato and upon tissue disruption (Díaz de León-Sánchez et al., 2009), but it is known that only a small number of volatile compounds have an impact on the perception of tomato flavour by consumers (Tikunov et al., 2010). It has been proposed that a combination of cis-3-hexenal, cis-3-hexenol, hexanal, 1-penten-3-one, 3-ethylbutanal, trans-2-hexenal, 6-methyl-5-hepten-2-one, methyl salicylate, 2-isobutylthiazole, and β-ionone, at appropriate concentrations, provides the characteristic fresh ripe aroma of tomato (Baldwin et al., 2000). However, other volatile compounds may also impart background notes to tomato overall flavour. Tikunov et al. (2005) suggested that phenylpropanoid volatiles as guaiacol, methyl salicylate (MeSA), and eugenol are the main contributors to the “smoky” aroma found in some tomato genotypes. The aroma of guaiacol has been described as “pharmaceutical” or “smoky”, while eugenol and MeSA provides the distinctive aroma to cloves (*Syzygium aromaticum*) and wintergreen (*Gaultheria procumbens*), respectively (Tikunov et al., 2010).

Phenylpropanoid volatiles mainly derive from the aromatic amino acid phenylalanine through a series of complex pathways (Osorio et al., 2010). As a result of glycosylation, by which plants make secondary metabolites more stable and facilitate their transport and storage, a large number of phenylpropanoids are found as non-volatile β-glucoside and β-diglycoside conjugates in tomato fruit (Ortiz-Serrano and Gil, 2007; Tikunov et al., 2010). Nevertheless, upon tissue disruption there is a loss of cell compartmentation that leads to the release of volatiles by the action of glycosides hydrolases (GH) (Günata et al., 1998). Tikunov et al. (2010) found that the release of phenylpropanoid volatiles upon tissue disruption of smoky tomato genotypes is regulated by a ripening-induced modification of their glycosylated precursors. Upon tissue disruption, smoky genotypes released methyl salicylate, guaiacol, and eugenol volatiles from mature green to red ripening stages, and the non-volatile precursors of these volatiles were identified as the corresponding diglycosides consisting of a hexose-pentose

moiety. In contrast, non-smoky genotypes from breaker stage onwards did not release smoky volatiles due to a third sugar moiety attached to the precursor, blocking the action of glycoside hydrolases to cleave the volatile glycoconjugate.

Among the widespread group of GH that have been identified in plants, β-glucosidases are the most studied because of their role in several biological processes. This enzyme catalyses the hydrolysis of β-glycosidic bonds of β-glucosides, which leads to the release of a glucose unit and the aglycone moiety (Saino et al., 2008). On the other hand, β-diglycosides can be cleaved either by one or two steps reactions. In a sequential reaction, an exoglycosidase cleaves the O-glycosidic bond connecting the two sugars, and a second β-glucosidase hydrolyses the resultant β-glucoside releasing the volatile aglycone (Günata et al., 1998; Ortiz-Serrano and Gil, 2007). In contrast, the one step reaction involves diglycosidases that hydrolyse disaccharide-glycosides into a disaccharide unit and an aglycone moiety. Diglycosidases have the unique characteristic to specifically cleave the β-glycosidic bond between disaccharide and aglycone, but not cleaving the inter-sugar linkage (Saino et al., 2008).

Among the diglycosidases identified in plants, β-primeverosidase from tea plants (*Camellia sinensis*) has been characterized in detail (Ma et al., 2001; Mizutani et al., 2002). It was found that this enzyme is able to hydrolyse selectively the disaccharide-aglycon bond of β-primeverosides (6-O- β-D-xylopyranosyl-β-D-glucopyranosides) to release a disaccharide primeverose unit and an aglycone, contributing to the floral aroma formation in tea during its manufacturing process (Ijima et al., 1998). Moreover, β-primeverosidase, classified as glycoside hydrolase family 1 (GH1), has shown to possess a high substrate specificity to β-primeverosides (Saino et al., 2008), and even a strong enantioselectivity (Zhou et al., 2014). The mechanism of the glycosidic bond hydrolysis of β-primeverosides is shown in Figure 1.

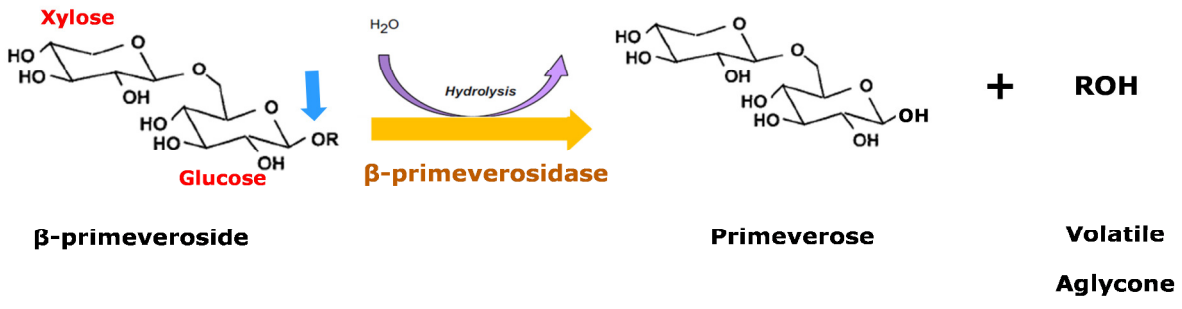


Figure 1. Schematic representation of β-primeverosides cleavage by the action of β-primeverosidase glycoside hydrolase.

In smoky tomato genotypes, the diglycosides identified as the corresponding precursors of smoky aroma-associated phenylpropanoids were later identified as putatives β-primeverosides, according to the retention time and the mass fragmentation patterns observed for these diglycosides precursors (Tikunov et al., 2013). This suggests that a β-primeverosidase-like enzyme could be involved in the cleavage of diglycosides in smoky tomatoes. Further attempts by Tikunov's team (data non published) to identify a β-primeverosidase-like enzyme in tomato, led to the identification of two genes by in silico analysis, Solyc10g045240.1 and Solyc01g074030.2 encoding for putative β-glucosidases. Their protein sequences were found to cluster together with β-primeverosidase from *Camellia sinensis*. An expression analysis of these genes was carried out in smoky tomatoes at different ripening stages, revealing that these genes were highly expressed in fruits at green maturity stage but low expressed from breaker stage onwards. This suggested that these genes may not be involved in diglycosides cleavage in fruits at ripe/red stage. Therefore, other strategies were attempted to identify new candidate genes encoding for GH that could be involved in the production of smoky phenylpropanoid volatiles in tomato. Two new candidate genes have been identified by two different strategies (data non published):

(1) A QTL analysis was performed in an F6 population from a cross of two "non-smoky" cherry type tomato parents (C074 x C085) that showed different hydrolytic activities to guaiacol, MeSA and eugenol. A gene located in chromosome 2, **Solyc02g091680.2** encoding for α-L-arabinofuranosidase/β-D-xylosidase, was found to be the best candidate gene that could explain the differences observed in hydrolytic activities among the F6 population and the parents. Further expression analysis (qPCR) of this gene in both parents showed that gene expression correlates relatively well with the amount of smoky volatiles released (hydrolytic activity).

(2) A microarray analysis was performed using those F6 lines (same population as before) that showed the highest and the lowest hydrolytic activity. A gene located in chromosome 1, **Solyc01g008710.2** encoding for mannan endo-1,4-β-mannosidase, was found as candidate gene based in the highest fold change observed between the two groups.

Until now, considerable progress has been made in the elucidation of the mechanisms underlying the release of phenylpropanoid volatiles in smoky tomatoes; however, the nature and the substrate specificity of the enzymes involved in the cleavage of their glycosylated precursors remain unknown. Therefore, the main purpose of this project is the identification and functional characterization of the enzyme(s) involved in the cleavage of glycosylated precursors of smoky-aroma volatiles released in tomato. To attain this purpose, the following sub-objectives were pursued:

- Use GH candidate genes identified as most likely related to the emission of smoky phenylpropanoid volatiles in tomato to express mature recombinant proteins in *N. benthamiana* and *E. coli*.
- Study the hydrolytic activity and substrate specificity of the recombinant GH proteins by determining volatiles released and changes in the diglycosides precursors during the enzymatic hydrolysis.

2. MATERIALS AND METHODS

2.1. Tomato extract enriched with diglycosides

360g frozen powder of fruits from a tomato genotype rich in diglycosides were extracted with 3 volumes of 100% methanol by agitation for 3 h at room temperature. Following centrifugation, supernatant was collected and filtered. The methanolic extract was then concentrated in a vacuum rotary evaporator at 30 °C and 120 mbar for 6 h. The remaining extract was passed through a glass column (35 x 1 cm i.d.) packed with pretreated Amberlite XAD-2 resin (Supelco) at a flow rate of 2 mL min^{-1}. Following 2 washing steps with Milli-Q water, bound compounds were eluted by adding 100 ml 100% methanol. The eluted split in 24 vials, each containing around 2.6 ml, was then subjected to evaporation under vacuum at 30 °C until methanol was completely removed. The residues were stored at -20 °C until further analysis.

2.2. Recombinant protein transient expression of candidate glycoside hydrolases in *Nicotiana benthamiana* and enzyme activity assay

2.2.1. Binary vector construction and transformation into *Agrobacterium tumefaciens*

Five constructs were made for further expression in *N. benthamiana,* corresponding to **GH01** (Solyc01g008710.2) and **GH02** (Solyc02g091680.2) genes from non-smoky tomato parents C074 and C085, and β-primeverosidase gene from tea (tea-pvd) as a positive control (GenBank accession number AB088027.1). GH02 ORF possess some SNPs between 74 and 85 parents, therefore different strategies were applied for GH02 cloning. GH02(85) ORF was subjected to site directed mutagenesis (point mutation) in order to generate the restriction sites necessary for vector construction. The same strategy was used for tea-pvd. The general cloning procedure was as follows.

cDNAs previously synthesized from non-smoky tomato parents 74 and 85 at turning stage were amplified by PCR using Q5 HF DNA polymerase kit (New England Biolabs) and primers (A to E) containing restrictions sites (RS), listed in Table 1. For tea-pvd amplification, a tea library previously generated was used as a template with primers (E, F) listed in Table 1. The PCR conditions were setup with an initial denaturation at 98°C for 1 min, followed by 35 cycles of 98°C for 10 s, 60°C for 30 s, 72°C for 2 min, and a final extension at 72°C for 2 min. The resulting PCR products were checked by agarose gel electrophoresis and the corresponding bands were purified with Zymoclean Gel DNA recovery kit (Zymo Research). GH02(85) and GH02(85)' fragments were digested with BsaI and simultaneously ligated according to protocol described in Appendix 1. The ligated product was then amplified by PCR with primers D forward and E reverse (Table 1). The band obtained (2383 bp) together with the fragment GH02(74) amplified in the previous step were then digested with BamHI and XhoI, and cloned into pFLAP50 vector by ligation to the BamHI and SalI sites of a previously digested pFLAP50 DNA. For tea-pvd, PCR fragments were digested with Esp3I and simultaneously ligated according to protocol described in Appendix 1. The ligated product was then amplified by PCR with primers F forward and G reverse (Table 1). The band obtained (1769 bp) together with the GH01(74) and GH01(85) amplified products were digested with BglII and SalI, and cloned into pFLAP50 vector by ligation to the BamHI and SalI sites of a previously digested pFLAP50 DNA.

Table 1. Primers for PCR amplification

	Gene	Primer	Sequence	RS
A.	GH01(74), 1211 bp	Forward	5'-GAGAGAGATCTATGAATAACTCAATCATCTTAATTTTTGTTGC-3'	Bgl II
		Reverse	5'-GAGAGGTCGACCTATGATAGCTTAGAGAGCCTAAGAGA-3'	Sal I
B.	GH01(85), 1211 bp	Forward	5'-GAGAGAGATCTATGAATAACTCAATCATCTTAATTTTTGTTGC-3'	Bgl II
		Reverse	5'-GAGAGGTCGACCTATGATAGCTTAGAGAGCCTAAGAGA-3'	Sal I
C.	GH02(74), 2382 bp	Forward	5'-GAGAGGGATCCATGACCAAAAATATCCATTTCTTGATTCT-3'	BamHI
		Reverse	5'-GAGAGCTCGAGTCACATTTCAATGGATACAACATGTTC-3'	XhoI
D.	GH02(85), 1820 bp	Forward	5'-GAGAGGGATCCATGACCAAAAATATCCATTTCTTGATTCT-3'	BamHI
		Reverse	5'-GAGAGGGTCTCGCTAAGGTACCAAGTCATAGGCAATTTTC-3'	BsaI
E.	GH02(85)', 563 bp	Forward	5'-GAGAGGGTCTCCTTGAGTCATACTCCAAAGTGCCAATG-3'	BsaI
		Reverse	5'-GAGAGCTCGAGTCACATTTCAATGGATACAACATGTTC-3'	XhoI
F.	Tea-pvd, 1218 bp	Forward	5'-GAGAGAGATCTATGATGGCAGCGAAAGGGT-3'	Bgl II
		Reverse	5'-GAGAGGCAGAGAGGTCTTTCAATCCTTTCGGGTACATGAAA-3'	Esp3I
G.	Tea-pvd, 356 bp	Forward	5'-GAGAGCGTCTCAGACCTATTGGTCTACACAAAGGAGAAGT-3'	Esp3I
		Reverse	5'-GAGAGGTCGACCTACTTGAGGAGGAATTTCTTGAACC-3'	Sal I

25 µL DH5α *E. coli* competent cells were transformed with each construct by heat shock (42°C for 45 s) followed incubation in SOC medium for 1h at 37°C. 150 µL of cells for each

transformation reaction, concentrated by centrifugation at low speed, were plated on LB-agar plates containing 100ug/ml ampicillin. Plates were incubated overnight at 37°C, and then three colonies were selected per transformation reaction for following plasmid purification with Qiaprep Spin Miniprep kit (Qiagen). After ORF verification by sequencing, plasmids DNA (GHs and tea-pvd in pFLAP50) digested with AscI and PacI were ligated to the backbone of the expression vector pBBC50 previously digested in AscI/PacI sites. Transformation of DH5α *E. coli* competent cells with an empty vector (negative control) and the expression vector containing the GHs and tea-pvd ORF was performed as described before, following plasmid purification and restriction analysis.

After plasmid verification, constructs were transformed into AGL1 *A. tumefaciens* competent cells. For this, 1µL plasmids were mixed with 50 µL AGL1 cells, incubated on ice for 30 min and transferred to a prechilled electroporation cuvette. Electroporation was carried out by applying 2.5 V at 4.8-5 time constant. 1mL LB was added to the electroporated cells, transferred to Eppendorf tubes and incubated in water bath at 28°C for 1 h. Then, cells were centrifuged at 4000 rpm for 4 min and 850 µL supernatant were removed. The cells were resuspended in the remaining LB volume and then plated on LB-agar plates containing 20 µg/mL rifampicin and 50 µg/ml kanamycin. Plates were incubated at 28°C for 2 days. AGL1 transformants were confirmed by AscI/PacI restriction analysis.

2.2.2. *Agrobacterium tumefaciens*-mediated transformation

Starter cultures of cells harbouring the empty vector, GHs and tea-pvd constructs were prepared by growing single colonies in 4 mL LB containing 34 µg/mL rifampicin and 25 µg/mL kanamycin at 28°C, 220 rpm for 2 days. In addition, a starter culture from a glycerol stock of *A. tumefaciens* harbouring p19/pBIN construct was prepared for further leave infiltration to maximize protein production by suppression of gene silencing. After 2 days, the starter cultures were diluted 1:100 in LB media supplemented with 34 µg/mL rifampicin and 25 µg/ml kanamycin, and 2x50 ml Falcon tubes containing 5ml diluted culture for each construct were grown at 28°C, 220 rpm for 24 h. For p19 cultures, 6x50ml Falcon tubes were incubated in order to have enough volume to be mixed with other cultures prior to infiltration. After 24 h, cells were centrifuged at 4000g for 20 min at room temperature, and after removing the supernatant, cells were resuspended in 10ml infiltration media (10mM MES buffer, 10mM $MgCl_2$ and 100µM acetosyringone) by rolling for 3 h at room temperature. Resuspended cells were then diluted with infiltration media to a final OD_{600}=0.5, and 30ml of each construct culture were mixed with 15mL p19 culture. For infiltration, three leaves in the middle/upper part of three weeks-old *N. benthamiana* plants were injected with the *A. tumefaciens* suspension by pressing a 5mL syringe without metal needle towards the underside of the leaf.

Three plants were infiltrated per construct. Additionally, three plants non-infiltrated (wild type, WT) were included for further analysis. Plants were grown under greenhouse conditions for 5 days to accumulate recombinant proteins. Following this period, infiltrated leaves were harvested and stored at -80 °C for further analysis.

2.2.3. Recombinant protein activity assay

The hydrolytic activity of proteins expressed in *N. benthamiana* leaves was determined by measuring the amount of diglycosides present and volatiles released after feeding recombinant proteins in leaves with the diglycosides enriched extract. The determination of diglycosides was performed in a different enzyme assay from the volatile analyses. First, diglycosides enriched extract was resuspended in sodium citrate pH 4.5 buffer to a final volume that each ml contained diglycosides from 1.5g fresh weight of original non-smoky tomato material. A second batch of diglycosides was resuspended in Milli-Q water to the same final concentration. For the enzyme assay, 0.5g frozen powder of transiently transformed and WT leaves were mixed with 1ml diglycosides enriched solutions (one reaction for each buffer) in 10 mL glass tubes. Tubes were incubated at overnight 37 °C. After the incubation, samples were extracted with 4.5 mL formic acid:methanol (1:1000, v/v) solution by agitation for 30 min. Samples were centrifuged for 15 min and 1mL supernatant was filtered through a 0.2-mm inorganic membrane filter and transferred to 1.5mL vials. These samples were subjected to LC-QTOF-MS analysis for determination of diglycosides as described by Tikunov et al. (2013). Since no authentic standards of diglycosides bound to smoky volatiles exist, the retention time and the mass (m/z) observed in previous studies (Tikunov et al., 2010) were used for the identification of these compounds in the current experiment. Results were expressed as the signal intensity (peak height), which is proportional to the concentration of the ion detected [M-H]-.

A second enzyme assay was performed for the determination of volatiles released after enzymatic hydrolysis. Leave tissue samples and diglycosides extract resuspended in sodium citrate pH 4.5 buffer (same amounts as described in previous enzyme assay) were mixed in a 10mL crimp cap vial, immediately capped and incubated overnight at 37°C. Next day, enzymes were deactivated by heating at 70°C for 1 h and then samples were subjected to SPME-GC-MS analysis under conditions described by Tikunov et al. (2013). Smoky volatiles were identified by matching compound mass spectra to the NIST mass library. Results were expressed as the signal intensity (peak height), which is proportional to the concentration of the ion detected [M-H]-.

2.3. Recombinant protein expression of candidate glycoside hydrolases in *Escherichia coli* and enzyme activity assay

2.3.1. Recombinant protein expression in *Escherichia coli*

Coding sequences without signal peptide of GH01 and GH02 genes from parents 74 and 85 were previously cloned by Jos Molthoff in pACYCDUET expression vector to generate constructs containing a His-tag sequence and another set without it. For constructs including His-tag, the mature sequences of GH01(74) and GH01(85) were cloned in the expression vector using SacI and AscI restriction sites, while GH02(74) and GH02(85) mature sequences were cloned using BamHI and AscI. For constructs without His-tag, the mature sequences of GH01(74) and GH01(85) were cloned in the expression vector using NcoI and PstI restriction sites, while GH02(74) and GH02(85) mature sequences were cloned using NcoI and Bam HI. pACYCDUET-GH constructs and an empty vector (negative control) were transformed into E. coli BL21 (DE3) cells (New England Biolabs) according to manufacturer's instructions. Colonies harbouring the constructs were picked from the selection plates and grown overnight in LB media supplemented with 34 µg/mL chloramphenicol at 37°C. Plasmid purification following restriction analysis were performed for plasmid verification. After confirmation, a starter culture was grown overnight at 37°C in 5 mL LB medium containing 50 µg/mL chloramphenicol and 1% glucose. The starter culture was diluted 1:100 in 20mL 2xYT medium containing 50 µg/mL chloramphenicol and incubated for 2 h at 37°C to reach OD_{600} 0.6-0.8. At this point, protein expression was induced by adding 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG), and the culture was incubated overnight at 18°C with agitation 250rpm. The cultures were then centrifuged for 15 min at 3400rpm and the supernatant removed. Immediately, cells pellet was resuspended on ice with 1mL sodium citrate pH 4.5 buffer and transferred to 2mL Eppendorf tubes containing 0.2g zirconium/silica beads 0.1 mm (Biospec). Cells were then disrupted by shaking 10 s in a fastprep machine (speed 6.5), incubated 5 min on ice and shaken again for 10 s. Cells were centrifuged at 13000g, 4°C for 10 min and the supernatant, crude protein extract, was collected for further analysis.

2.3.2 His-tag purification

Crude extract of GH01(74) protein containing His-tag was subjected to purification for further SDS-PAGE analysis in order to confirm the expression of the recombinant protein. Previous to purification, the protein crude extract buffer was exchanged to a protein binding buffer (50mM NaH_2PO_4, 300Mm NaCl, 20mM imidazole pH 8.0) using Amicon Ultra-0.5 filter 10K according to manufacturer's instructions. Then, 500 µL protein crude extract in binding buffer was mixed with 50 µL of Ni-NTA beads suspension and incubated on an end-over-end shaker for 1 h at

room temperature. With the help of a magnetic separator to keep beads attached to the tube walls, the supernatant corresponding to the unbound protein was removed and stored for further analysis. Ni-NTA beads were washed three times with 500 µL binding buffer and supernatants were collected and stored. After washing steps, 25 µL of elution buffer (50mM NaH_2PO_4, 300Mm NaCl, 250mM imidazole pH 8.0) were added to beads, mixed, incubated for 1 min and eluted proteins were collected for further analysis.

2.3.3 SDS-PAGE

The crude extract proteins and the GH01(74) protein His-tag purified were analyzed by SDS polyacrylamide gel electrophoresis. 15 µL of crude protein extract was mixed with 5 µL of 4x loading buffer containing β-mercaptoethanol, while 25 µL of his-tag purified enzyme were mixed with 8 µL of loading buffer. Samples were denatured at 95°C for 10 min and then 20 and 25 µL of crude extract proteins and purified enzyme, respectively, were loaded to a 12% polyacrylamide gel. After electrophoresis, the gel was stained with Coomassie Brilliant Blue for 1 h and washed with acetic acid:methanol:water (10:25:65, v/v/v) de-staining solution. Precision Plus Protein™ dual color (Bio-Rad) was used as the marker for protein molecular weight determination.

2.3.4 Recombinant protein activity assay

The hydrolytic activity of proteins expressed in *E. coli* was determined by measuring the amount of diglycosides present and volatiles released after feeding the crude extract proteins with the diglycosides enriched extract. Two enzyme assay reactions were setup for determination of diglycosides as products. One reaction contained 118 µL crude protein, 1mL diglycosides enriched extract in sodium citrate pH 4.5 buffer, and 58.8 µL sodium citrate pH 4.5 buffer. The second reaction included 118 µL crude protein, 1mL diglycosides enriched extract in sodium citrate pH 4.5 buffer, and 0.5g frozen powder of tomato fruit from a genotype with low hydrolytic activity (20-105). Reactions were incubated overnight at 37°C. After incubation, samples were extracted with 4.8 mL formic acid:methanol (1:1000, v/v) solution by agitation for 30 min. Samples were centrifuged for 15 min and 1mL supernatant was filtered through a 0.2-mm inorganic membrane filter and transferred to 1.5mL vials. These samples were subjected to LC-QTOF-MS analysis and signals identified as described for *N. benthamiana* assay.

Same enzyme assay reactions were performed for determination of volatiles released as hydrolysis products. The reactions were setup in capped 10mL crimp cap vial and incubated overnight at 37°C. This time proteins were not denatured. Samples were subjected to SPME-GC-MS analysis and signals identified as described for *N. benthamiana* assay.

2.4. Statistical analysis

Independent samples t-test analyses were performed to test for significant differences between means for the empty vector and transformed plants in *N. benthamiana* assay.

3. RESULTS

3.1 *N. benthamiana* expressed protein assays

LC-MS analysis of diglycosides precursors of smoky volatiles was carried out in order to determine whether the putative GHs expressed in *N. benthamiana* possess hydrolytic activity towards diglycosides substrate. **Figure 2** shows the results of this analysis for enzyme activity assays performed in water and sodium citrate pH 4.5 buffers. Retention times (RT) and masses of diglycosides precursors identified are shown in Appendix 2.

In **Figure 2**, the positive control, tea-pvd, showed a significantly high hydrolytic activity towards diglycosides precursors of guaiacol, eugenol and MeSA in both buffers with respect to the control plants. GH02 also showed some hydrolytic activity towards eugenol diglicosides precursors, which significantly decreased under the action of GH02(74) and GH02(85) in sodium citrate pH 4.5 buffer. In water, only GH02(85) cleaved eugenol precursors to a significantly lower levels. MeSA precursors were not affected by GHs action, only a small but still significant increase was observed in diglycosides levels under the action of GH02(85) in water. In general, large variability in MeSA diglycosides was observed in enzymatic assays performed with GH01(74), GH01(85) and GH02(74), which could have masked the effect of these GHs on MeSA substrates. GH01 from both parents did not show a significant hydrolytic activity towards any smoky volatile diglycoside precursor in any buffer used.

Through an untargeted analysis of LC-MS data, an increase of some metabolites was found in assays performed with GH02 in both buffers. These metabolites were then identified as guaiacol and eugenol attached to a hexose moiety according to the RT and masses observed. **Figure 3** shows the amount of these glycosides after hydrolysis assays performed in sodium citrate pH 4.5 buffer, where there is a clear significant increase of guaiacol-hexose and eugenol-hexose under the action of GH02 proteins with respect to the control. No effect was observed for MeSA glycoside. Interestingly, the increase in eugenol-hexose matched with the decreased levels observed for the corresponding diglycosides specially in sodium citrate pH 4.5 buffer (**Figure 2**), suggesting that the action of GH02 towards diglycosides is mainly the cleavage of

their pentose moiety. This effect was also observed in water (data not shown for eugenol-hexose) but to a less extent, indicating that these enzymes perform best under tomato matrix conditions. On the other hand, tea-pvd showed a highly significant hydrolytic activity towards smoky volatiles glycosides conjugates, suggesting that this protein is also able to cleave glycosides with one sugar moiety. GH01 activity showed not significant differences with respect to the control (**Figure 3)**.

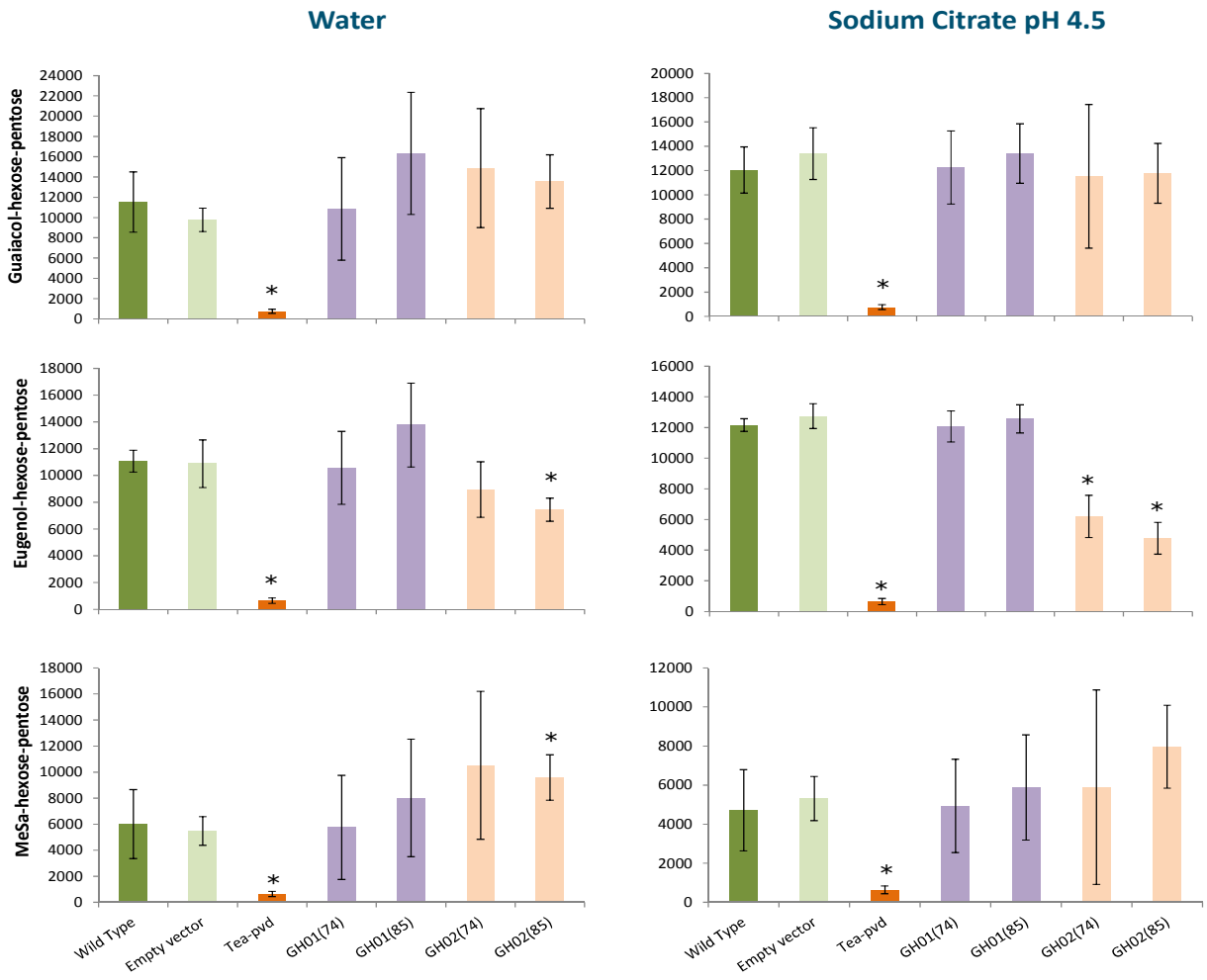


Figure 2. Levels of smoky volatiles-diglycosides precursors found in *N. benthamiana* expressed proteins assays performed in water (left) and sodium citrate pH 4.5 (right) buffers by LC-MS analysis. Samples included wild type and leaves transformed with empty pBBC50 vector (negative controls), tea-pvd (positive control) and GH01 and GH02 genes from parents C074 and C085. Values are means ± standard deviation, n=3. * Mean is significantly different with respect to the empty vector at p-value < 0.05.

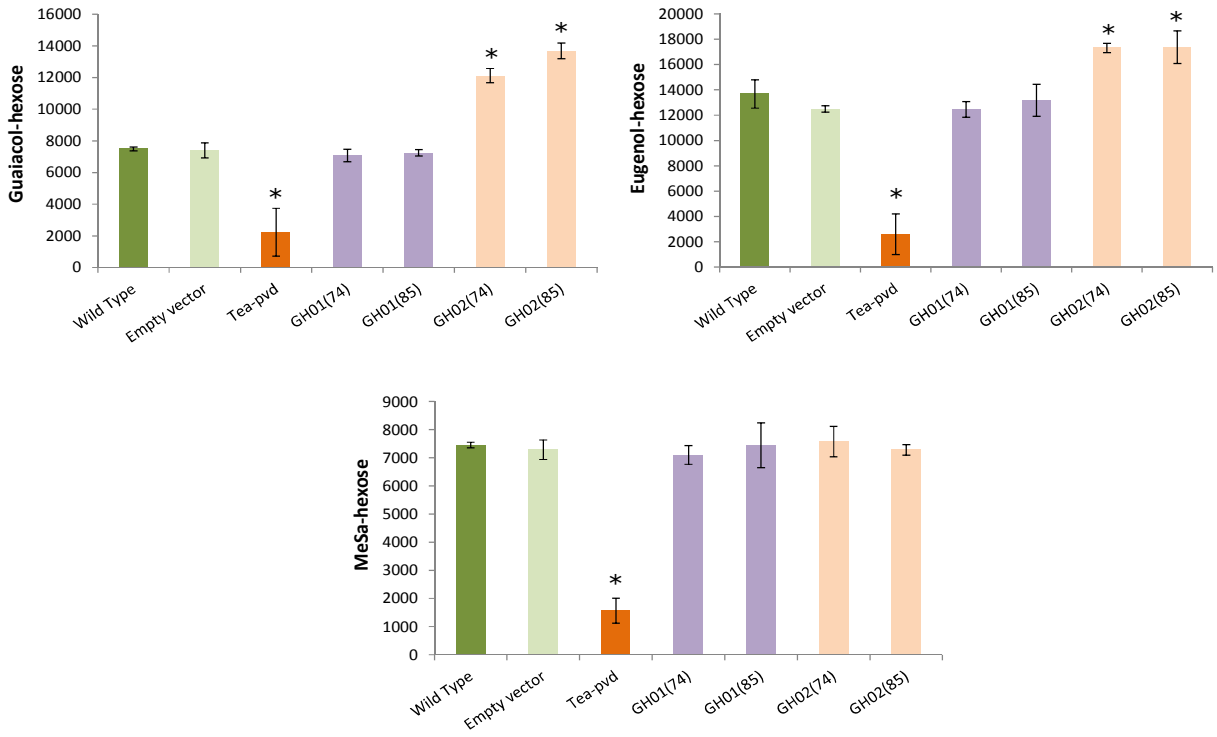


Figure 3. Levels of smoky volatiles-glycosides found in *N. benthamiana* expressed proteins assays performed in sodium citrate pH 4.5 buffer by LC-MS analysis. Samples included wild type and leaves transformed with empty pBBC50 vector (negative controls), tea-pvd (positive control) and GH01 and GH02 genes from parents C074 and C085. Values are means ± standard deviation, n=3. * Mean is significantly different with respect to the empty vector at p-value < 0.05.

GC-MS analysis of smoky volatiles released from enzymatic assays showed no significant differences in levels of guaiacol, eugenol and MeSA volatiles in reactions using GHs proteins with respect to the control (empty vector) (**Figure 4**). Release of volatiles by tea-pvd was also not significant. However, tea-pvd showed a trend to produce more eugenol than the empty vector and wild type, but its effect compared to the empty vector resulted not significant mainly due to the large variability showed by this control. An extra control was added to this experiment, where the volatiles released by hydrolysis were determined in WT leaves not adding tomato diglycosides as substrates (Wild Type-DG). No production of guaiacol and eugenol was observed for this control. This contrast with the large release of guaiacol and MeSA observed in WT fed with substrate, suggesting that N. benthamiana leaves possess

endogenous GH able to cleave tomato diglycosides, and especially those attached to guaiacol and MeSA. In adittion, MeSA levels were largely produced by WT leaves compared to the leaves transformed, suggesting a suppressive effect of Agrobacterium-mediated transformation on the release of MeSA in *N. benthamiana* assays.

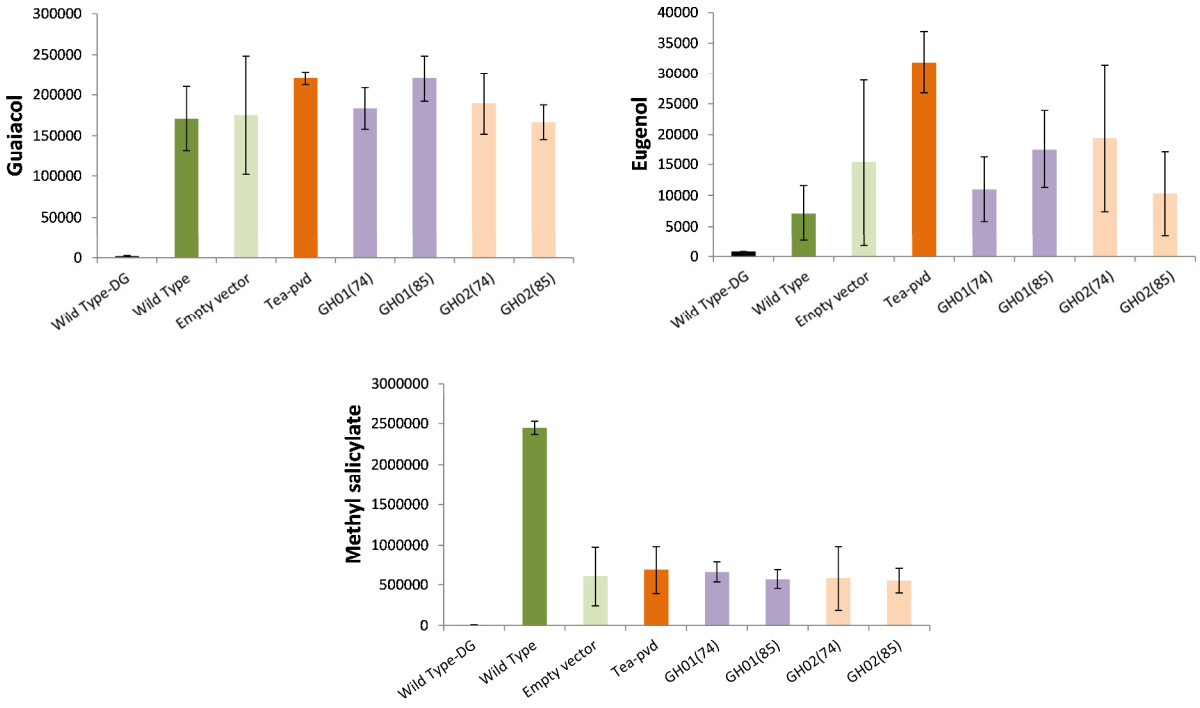


Figure 4. Levels of smoky volatiles found in *N. benthamiana* expressed proteins assays performed in sodium citrate pH 4.5 buffer by GC-MS analysis. Samples included: wild type leaves fed with diglycosides and without diglycosides (Wild Type-DG); leaves transformed with empty pBBC50 vector (negative controls), tea-pvd (positive control) and GH01 and GH02 genes from parents C074 and C085, fed with diglycosides. Values are means ± standard deviation, n=3.

Other volatiles that contribute to the characteristic tomato aroma were found to be largely released by leaves transformed with tea-pvd (**Figure 5**). Release of phenylethanol and benzyl alcohol was significantly higher by the action of tea-pvd with respect to the control. Tea-pvd also showed an increase in the production of methylbutanol. However, this was not significant with respect to the control mainly due to the large variability shown by the empty vector, and also by the low levels of methylbutanol released that were close to the detection threshold. GHs did not show any activity towards the release of these volatiles.

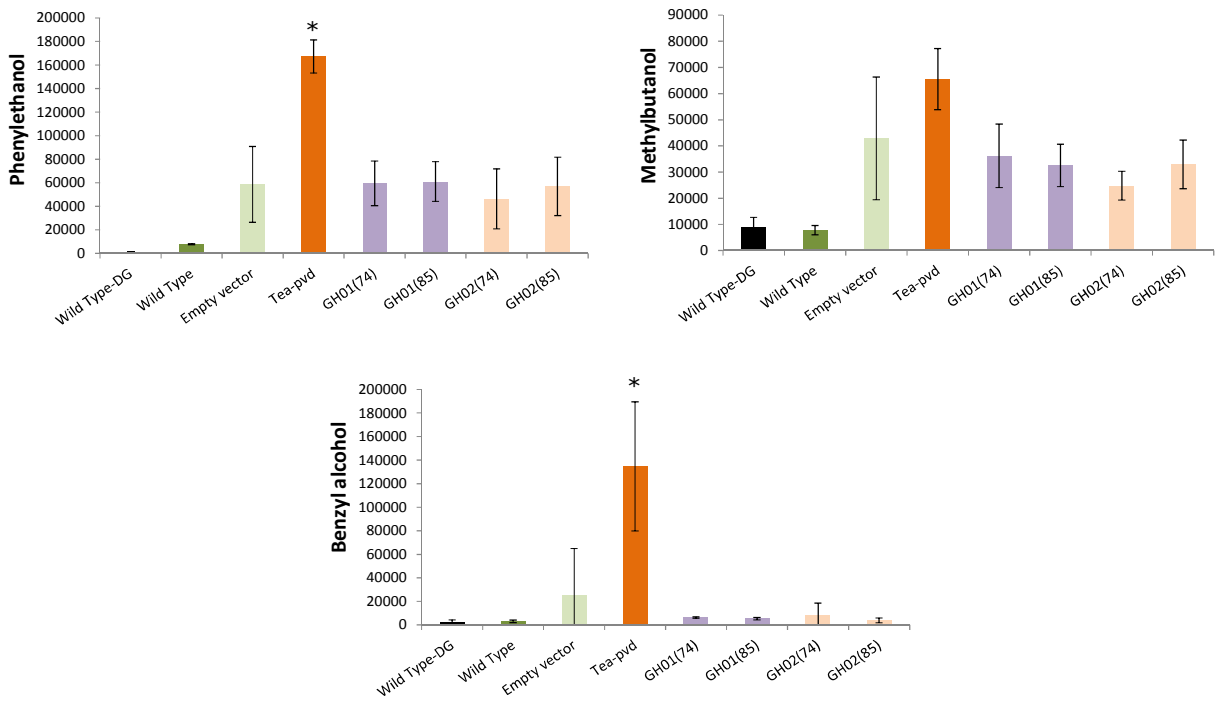


Figure 5. Levels of characteristic tomato volatiles found in *N. benthamiana* expressed proteins assays performed in sodium citrate pH 4.5 buffer by GC-MS analysis. Samples included: wild type leaves fed with diglycosides and without diglycosides (Wild Type-DG); leaves transformed with empty pBBC50 vector (negative controls), tea-pvd (positive control) and GH01 and GH02 genes from parents C074 and C085, fed with diglycosides. Values are means ± standard deviation, n=3. * Mean is significantly different with respect to the empty vector at p-value < 0.05.

Looking at diglycosides precursors among the GC-MS data obtained from hydrolysis assays, a significant decrease of phenylethanol, benzyl alcohol and methylbutanol attached to a hexose and pentose moiety by the hydrolytic action of tea-pvd with respect to the empty vector was observed (**Figure 6**). Thus, these precursors might be the main source for the production of phenylethanol, benzyl alcohol and methylbutanol under these experimental conditions. For phenylethanol, it was possible to identify the nature of the hexose-pentose moiety attached to this volatile according to the RT observed in LC-MS data, and this was identified as glucose-arabinose.

A small but significant decrease of methylbutanol diglycoside was observed in assays using GH02(74) and GH02(85) proteins. However, the cleavage of this diglycoside did not show

significantly higher levels of the corresponding volatile with respect to the control (**Figure 5**). As previously observed for eugenol diglycoside (**Figure 2**), GH02 could have cleaved the pentose moiety of methylbutanol diglycoside and therefore decreasing the amount of this volatile conjugate. This could have been confirmed by looking at methylbutanol-hexose levels; however it was not possible to find this compound among the LC-MS data obtained.

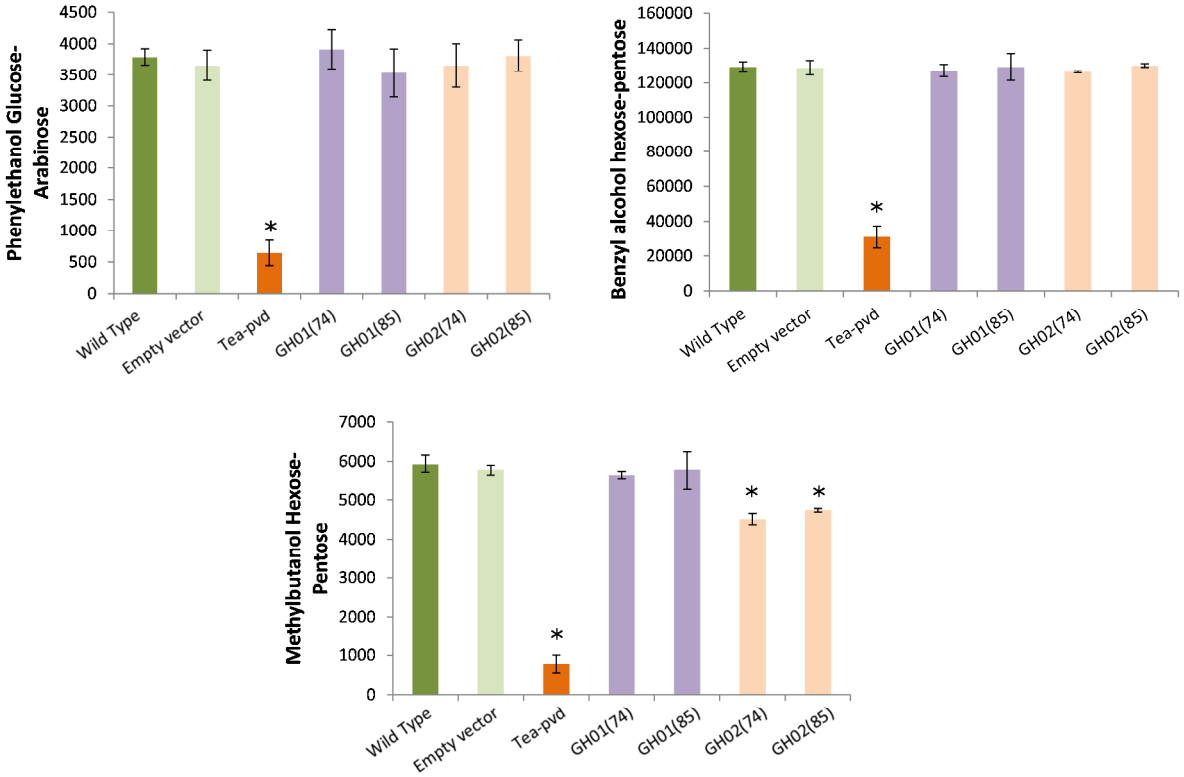


Figure 6. Levels of diglycosides precursors for other characteristic tomato volatiles found in *N. benthamiana* expressed proteins assays performed in sodium citrate pH 4.5 buffer by LC-MS analysis. Samples included wild type and leaves transformed with empty pBBC50 vector (negative controls), tea-pvd (positive control) and GH01 and GH02 genes from parents C074 and C085. Values are means ± standard deviation, n=3. * Mean is significantly different with respect to the empty vector at p-value < 0.05.

3.2 *E. coli* expressed protein assays

Constructs containing GH candidate genes without the signal sequence were used for protein expression in *E. coli.* Moreover, a second set of proteins containing a His-tag sequence were expressed. Crude extracts of expressed proteins were subjected to SDS-PAGE analysis prior to enzymatic hydrolysis assays, and the result is shown in **Figure 7A**. The cell lysis to release the recombinant proteins was performed in sodium citrate pH 4.5 buffer with and without β-mercaptoethanol. However, β-mercaptoethanol interfered in further purification of a protein containing His-tag and also could have potentially interfered GC-MS analysis, therefore proteins in buffer containing this agent were not considered for further analysis. These proteins were included only in SDS-PAGE analysis.

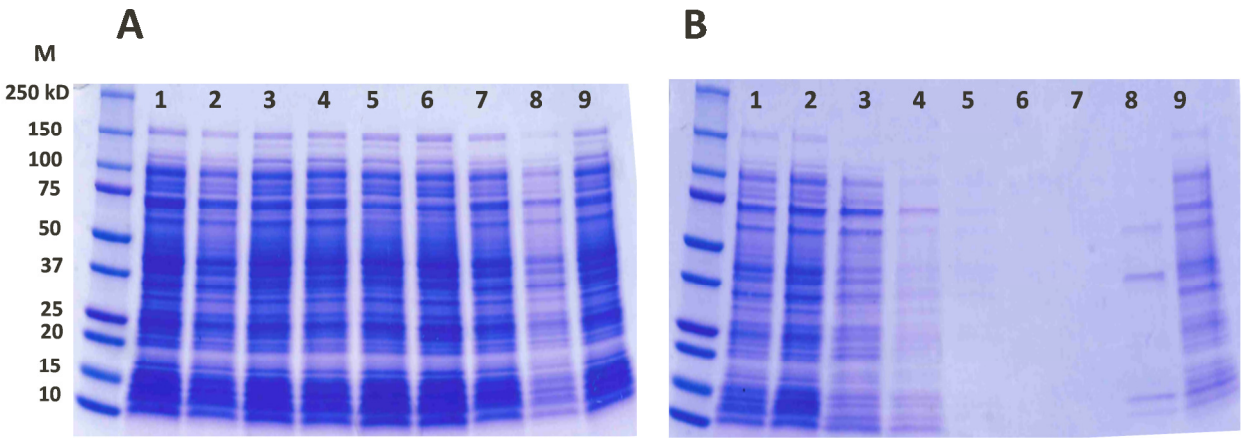


Figure 7. SDS-PAGE analysis of recombinants proteins expressed in *E. coli*. (A) Crude protein extracts in buffer without β-mercaptoethanol (1-5) and buffer including β-mercaptoethanol (7-9). Lanes: 1&7, GH01(74); 2&8, GH01(85); 3&9, GH02(74); 4, GH02(85); 5&6, empty vector. (B) His-tag purified GH01(74) protein in buffer without β-mercaptoethanol (1,3,8) and buffer including β-mercaptoethanol (2, 4, 5-7, 9). Lanes: 1-2, crude extract; 3-4, unbound protein; 5-7, protein from washing steps; 8-9, eluted protein.

According to the nucleotide sequences, the size of GH01 and GH02 proteins should be 45 and 88 kDa, respectively. Crude GH proteins (**Figure 7A**, lanes 1-4 and 7-9) showed some thin bands that were not present in the empty vector (lanes 5-6); however, it was difficult to confirm for sure that these bands corresponded to GH proteins. Therefore, one of the GH proteins, GH01(74), containing a His-tag sequence was purified with Ni-NTA method and a second SDS-PAGE analysis was run (**Figure 7B**). In this gel, it is clear the interference of β-mercaptoethanol in His-tag purification, in which total proteins remained bound to the Ni-NTA beads and were eluted at the final step (**Figure 7B,** lanes 4 and 9). In lane 8, there is a band between 37 and 50 kDa that could correspond to GH01(74) protein (45 kDa). However, a SDS-PAGE analysis of the empty vector is needed to confirm its presence. This was not performed in this project due to limitations in Ni-NTA beads availability.

Crude protein extracts were used to determine their hydrolytic activity towards diglycosides substrate. Only one batch of proteins was expressed per culture and per construct, therefore GC-MS and LC-MS data obtained consisted of one assay/construct. In addition, the assays were performed in two different matrixes, one of them in sodium citrate pH 4.5 buffer and a second batch in a matrix consisting of fresh tomato powder with low hydrolytic activity. Levels of smoky volatiles attached to one and two sugar moieties found in these assays are shown in **Figure 8**. There were not clear differences observed in diglycosides amount between the empty vector and the GH proteins in assays performed in sodium citrate pH 4.5 buffer. While in tomato matrix all diglycosides were dramatically reduced and therefore any potential hydrolytic activity of GH proteins could have been masked by this large background effect. Eugenol diglycosides and eugenol-hexose levels did not show the effect of the pentose cleavage activity observed with *N. benthamiana* expressed GH02 under these experimental conditions. A slight decrease in eugenol diglycosides and minor increase in eugenol-hexose compared to the empty vector can be observed for the assay performed in tomato matrix, but this could have been effect of the hydrolytic activity present in the tomato matrix. Clear differences were only observed for MeSa-hexose in tomato matrix; however, the observed decrease in the amount of this conjugate occurs close to the LC-MS detection threshold and therefore is not a reliable result.

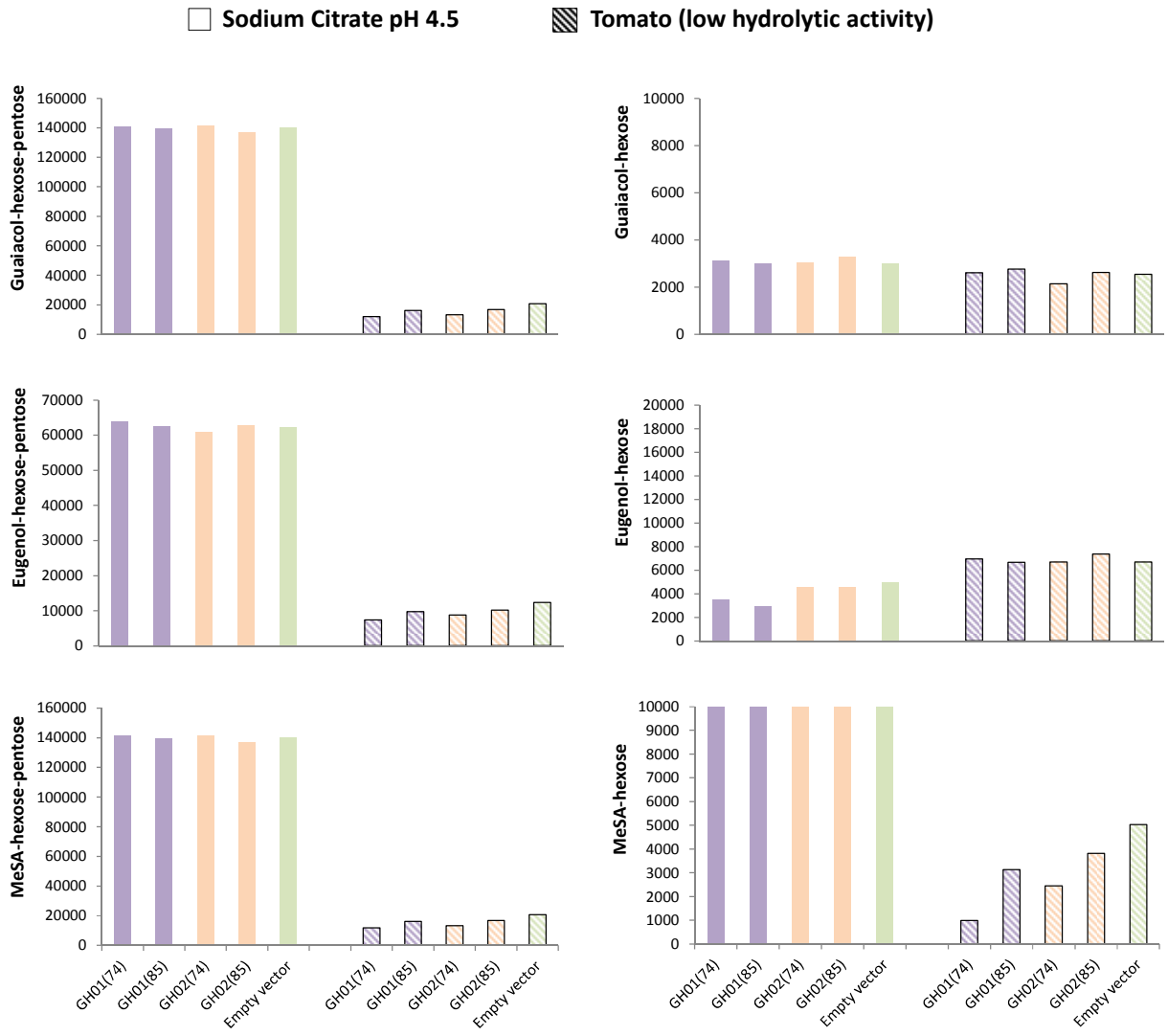


Figure 8. Levels of smoky volatiles diglycosides (left) and glycosides (right) precursors found in *E. coli* expressed proteins assays performed in sodium citrate pH 4.5 buffer and in tomato (low hydrolytic activity) matrix by LC-MS analysis. Samples included empty pACYCDUET vector (negative control), and GH01 and GH02 genes from parents C074 and C085. Values correspond to LC-MS analysis of one sample.

Smoky volatiles were also determined in recombinant protein assays (**Figure 9**). These volatiles were only detected in assays performed in tomato matrix. Only eugenol slightly increased when substrate was used to feed GH02(74), but this could have been effect of the hydrolytic activity of tomato matrix and not from the recombinant protein. MeSA was released in GH01(85) and GH02(85) assays in very low amounts.

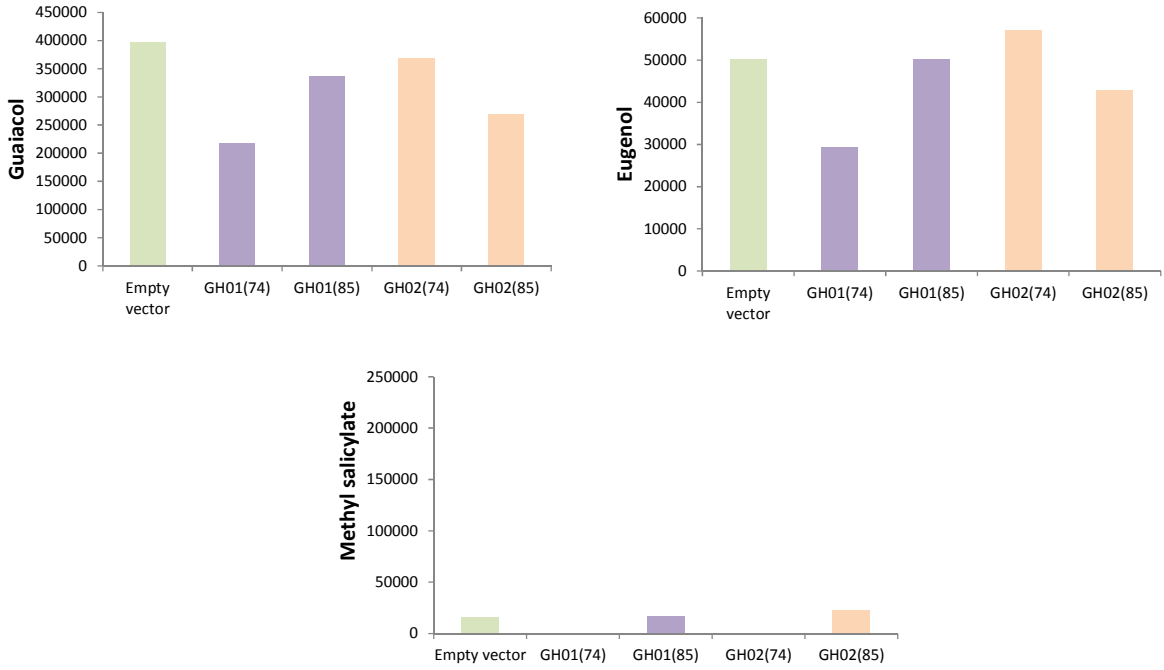


Figure 9. Levels of smoky volatiles released by *E. coli* expressed proteins in tomato (low hydrolytic activity) matrix by GC-MS analysis. Samples included empty pACYCDUET vector (negative control), and GH01 and GH02 genes from parents C074 and C085. Values correspond to GC-MS analysis of one sample.

4. DISCUSSION

Putative glycoside hydrolases were expressed in *N. benthamiana* and *E. coli* and their hydrolytic activity towards diglycosides precursors to produce smoky volatiles was studied.

β-primeverosidase has shown a high hydrolytic activity towards diglycosides in *Camellia sinensis* (Ma et al., 2001; Mizutani et al., 2002). This feature made us chooses this enzyme to be expressed in *N. benthamiana* and be used as a positive control in hydrolysis assays with *N. benthamiana* expressed proteins from tomato. Even though the expressed tea-pvd in tobacco showed differences in one amino acid with respect to the *Camellia sinensis* β-primeverosidase (Appendix 3), its functionality was not affected according to the high hydrolytic activity observed towards the substrates. Tea-pvd produced in this experiment cleaved completely the diglycosides precursors of smoky volatiles guaiacol, eugenol and MeSA. Other volatiles

precursors were also largely cleaved, as phenylethanol, benzyl alcohol and methylbutanol diglycosides (**Figure 2** and **6**). Tea-pvd showed also hydrolytic activity towards volatiles attached to one sugar moiety, which is not in agreement with the high specificity towards primeverosides described by other studies (Saino et al., 2008; Zhou et al., 2014). Ma et al. (2001) found that β-primeverosidase possess a strong enantioselectivity, cleaving very efficiently 2-phenylethanol from the diglycoside composed of a xylose-glucose moiety (primeveroside), while a poor hydrolytic activity has been observed towards 2-phenylethanol attached to an arabinose-glucose moiety. Tea-pvd showed a high hydrolytic activity towards this diglycoside (**Figure 6**); however, these differences could be due to the large period of incubation used in this experiment (>12 h), resulting in an important decrease by cleavage of these substrates even though the hydrolytic activity of this enzyme was low.

GH02 was the only protein expressed in *N. benthamiana* that showed some hydrolytic activity. GH02 genes from parents C074 and C085 possess some SNPs; however their activity did not show clear differences towards the different substrates. Even though this protein was presumed to be involved in the cleavage of diglycosides precursors of smoky volatiles, the results found in eugenol diglycosides and monoglycosides suggests that this protein is able to cleave the pentose moiety of these glycoconjugates and not the glycosidic bond of the volatile-glycoside. This feature fits with the function proposed for this enzyme, α-L-arabinofuranosidase/β-D-xylosidase. An α-L-arabinofuranosidase has been cloned and characterized in tomato cv. Ailsa Craig by Itai et al. (2002), who suggested that this enzyme is involved in cell wall modifications through the breakdown of xylans and arabinoxylans during tomato developing and ripening. Perhaps this function, although is not directly related, correlates well with the phenotype observed in smoky tomatoes. In previous experiments, GH02 candidate gene found by QTL analysis showed relatively good correlations between expression levels and the amount of smoky volatiles released (data non published). This could be due to this gene might be close or linked to other gene(s) in the QTL region encoding for proteins involved in the cleavage of the diglycoside precursors of smoky volatiles.

GH01 did not show any clear activity towards diglycosides precursors. GH01 was proposed as a candidate gene due to the highest changes observed in its expression between genotypes with contrasting hydrolytic activities. However, as explained for GH02, this could have resulted from the presence of other genes in the proximity of GH01 that are expressing the smoky phenotype. On the other hand, a phylogenetic analysis (data non published) showed that this particular gene together with other members of family 5 are relatively close to family 1 where β-primeverosidase belongs. GH proteins have been classified into 85 families based in their sequence similarities. However, even though proteins in the same family show high similarities in their amino acid sequences, proteins sometimes show different substrate specificities

although they are in the same family (Dies and Henrissat, 1995). Therefore this proximity to family 1 does not necessarily suggest a similar function.

In general, expression of proteins in *N. benthamiana* showed to be a good strategy to study their hydrolytic activity by monitoring the diglycosides precursors of smoky volatiles. However, the analysis of volatiles released showed that *N. benthamiana* possess a large background of hydrolytic activity towards their precursors, which significantly masked the activity of the positive control (tea-pvd) especially for guaiacol production. Moreover, MeSA releasing seems to be particularly affected by agrobacterium-mediated transformation, being sequestered by some mechanism relying on this technique. Therefore, expression of recombinant proteins in *E. coli* could be a better alternative to study their hydrolytic activity without a background noise, especially if smoky volatiles have to be monitored.

In this experiment, *E. coli* assays did not provide the same results as those found in *N. benthamiana* for GH02. However, it is not possible to make conclusions about the technique, since no positive control was used and no biological and technical replicates were considered for the analysis. β-primeverosidase has been successfully expressed in E. coli and its hydrolytic activity has been largely studied with this system (Saino et al., 2008; Zhou et al., 2014), which provides a precedent that this technique could be suitable to express other candidate proteins that might be involved in the production of smoky volatiles. The expression of β-primeverosidase in *E. coli* has involved the use of His-tag sequence for its further purification, and the hydrolytic activity has been proven with the His-tag purified enzyme. This is something that must be taken into account for further studies.

5. CONCLUSSIONS

The hydrolytic activity of putative tomato glycosides hydrolases towards diglycosides precursors of smoky volatiles was studied. The candidate GH expressed in *N. benthamiana* did not show hydrolytic activity for the release of smoky volatiles. GH02 might cleave their precursors but in a position that do not produce the release of the smoky volatiles.

The effectiveness of tea-pvd as a positive control was proven, although its activity reflected by the smoky volatiles released was masked by the action of endogenous hydrolases in *N. benthamiana* as well as unknown blocking mechanisms. Tea-pvd represents a good alternative for a positive control in experimental conditions where no background activities are found. This could be the case for assays involving proteins expressed in *E. coli*.

An optimized experimental setup for expression in *E. coli* is needed for an effective and reliable study of hydrolytic activities of potential glycosides hydrolases in smoky tomatoes.

Other candidate genes have been found trough QTL and microarray analysis that could be involved in the cleavage of diglycosides precursors of smoky volatiles. Moreover, a β-primeverosidase like enzyme has been identified in silico previously by this team. The expression of this enzyme in E. coli without the signal sequence might also bring an interesting outcome.

Overall, this report has provided some insights in how to proceed and what has to be improved for the expression of new candidate proteins in other organisms and the analysis of their activities.

6. REFERENCES

Baldwin, E. A.; Scott, J. W.; Shewmaker, C. K.; Schuch, W. Flavor Trivia and Tomato Aroma: Biochemistry and Possible Mechanisms for Control of Important Aroma Components. *HortScience* **2000**, 35(6), 1013-1022.

Beckles, D. Factors affecting the postharvest soluble solids and sugar content of tomato (Solanum lycopersicum L.) fruit. *Postharvest Biol. Technol.* **2012**, 63, 129-140.

Davies, G.; Henrissat, B. Structures and mechanisms of glycosyl hydrolases. Structure **1995**, 3(9), 853-859.

Defilippi, B. G.; Manríquez, D.; Luengwilai, K.; González-Agüero, M. Aroma Volatiles: Biosynthesis and Mechanisms of Modulation during Fruit Development. *Adv. Bot. Res.* **2009**, 50, 1-37.

Díaz de León-Sánchez, F.; Pelayo-Zaldívar, C.; Rivera-Cabrera, F.; Ponce-Valadez, M.; Ávila-Alejandre, X.; Fernández, F.J.; Escalona-Buendía, H. B.; Pérez-Flores, L. J. Effect of Refrigerated Storage on Aroma and Alcohol Dehydrogenase Activity in Tomato Fruit. *Postharvest Biol. Technol.* **2009**, 54, 93-100.

Dixon, J.; Hewett, E. W. Factors Affecting Apple Aroma/Flavour Volatile Concentration: A Review. *New Zeal. J. Crop Hort*. **2000**, 28, 155-173.

Günata , Z.; Blondeel, C.; Vallier, M. J.; Lepoutre , J. P.; Sapis, J. C.; Watanabe, N. An Endoglycosidase from Grape Berry Skin of Cv. M. Alexandria Hydrolyzing Potentially Aromatic Disaccharide Glycosides. *J. Agric. Food Chem.* **1998**, 46(7), 2748-2753.

Ijima, Y.; Ogawa, K.; Watanabe, N.; Usui, T.; Ohnishi-Kameyama, M.; Nagata, T.; Sakata, K. Characterization of *β*-Primeverosidase, being Concerned with Alcoholic Aroma Formation in Tea Leaves to be Processed into Black Tea, and Preliminary Observations on its Substrate Specificity. *J. Agric. Food Chem*. **1998**, 46, 1712-1718.

Itai, A.; Ishihara, K.; Bewley, J.D. Characterization of expression, and cloning, of -D-xylosidase and -L-arabinofuranosidase in developing and ripening tomato (*Lycopersicon esculentum* Mill.) fruit. J. Exp. Botany **2003**, 54(393), 2615-2622.

Ma, S. J.; Mizutani, M.; Hiratake, J.; Hayashi, K.; Yagi, K.; Watanabe, N.; Sakata, K. Substrate Specificity of *β*-Primeverosidase, a Key Enzyme in Aroma Formation During Oolong Tea and Black Tea Manufacturing. *Biosci. Biotechnol. Biochem*. **2001**, 65, 2719-2729.

Mizutani, M.; Nakanishi, H.; Ema, J.; Ma, S. J.; Noguchi, E.; Inohara-Ochiai, M.; Fukuchi-Mizutani, M.; Nakao, M.; Sakata, K. Cloning of *β*-Primeverosidase from Tea Leaves, a Key Enzyme in Tea Aroma Formation. *Plant Physiol.* **2002**, 130, 2164-2176.

Ortiz-Serrano, P.; Gil, J. V. Quantitation of free and glycosidically bound volatiles in and effect of glycosidase addition on three tomato varieties (*Solanum lycopersicum* L.). *J. Agric. Food. Chem*. **2007**, 55, 9170-9176.

Osorio, S.; Muñoz, C.; Valpuesta, V. Physiology and Biochemistry of Fruit Flavors. In Handbook of Fruit and Vegetable Flavors; Hui, Y.H., Ed.; Wiley: Hoboken, New Jersey, 2010; p 129.

Saino, H.; Mizutani, M.; Hiratake, J.; Sakata, K. Expression and Biochemical Characterization of *β*-Primeverosidase and Application of *β*-Primeverosylamidine to Affinity Purification. *Biosci. Biotechnol. Biochem*. **2008**, 72 (2), 376-383.

Tikunov, Y.; Lommen, A.; De Vos, R.C.H.; Verhoeven, H.A; Bino, R.J.; Hall, R.D.; Bovy, A.G. A novel approach for nontargeted data analysis for metabolomics: large-scale profiling of tomato fruit volatiles. *Plant Physiol.* **2005**, 139, 1125–1137.

Tikunov, Y.M., de Vos, R.C.H., González Paramás, A.M., Hall, R.D., and Bovy, A.G. A Role for Differential Glycoconjugation in the Emission of Phenylpropanoid Volatiles from Tomato Fruit Discovered Using a Metabolic Data Fusion Approach. *Plant Physiol*. **2010**, 152, 55–70.

Tikunov, Y. M.; Molthoff, J. W.; Vos, R. C. H. de; Beekwilder, M. J.; Houwelingen, A. M. M. L. van; Hooft, J. J. J. van der; Nijenhuis-de Vries, M. A.; Labrie, C. W.; Verkerke, W.; Geest, H.C. van de; Viquez Zamora, A. M.; Presa, S.; Rambla Nebot, J. L.; Granell, A.; Hall, R. D.; Bovy, A. G. NON-SMOKY GLYCOSYLTRANSFERASE1 Prevents the Release of Smoky Aroma from Tomato Fruit. *Plant Cell* **2013**, 25(8), 3067-3078.

Zhou, Y.; Dong, F.; Kunimasa, A.; Zhang, Y.; Cheng, S.; Lu, J.; Zhang, L.; Murata, A.; Mayer, F.; Fleischmann, P.; Watanabe, N.; Yang, Z. Occurrence of Glycosidically Conjugated 1‑Phenylethanol and Its Hydrolase *β*‑Primeverosidase in Tea (*Camellia sinensis*) Flowers. *J. Agric. Food Chem*. **2014**, 62(32), 8042-8050.

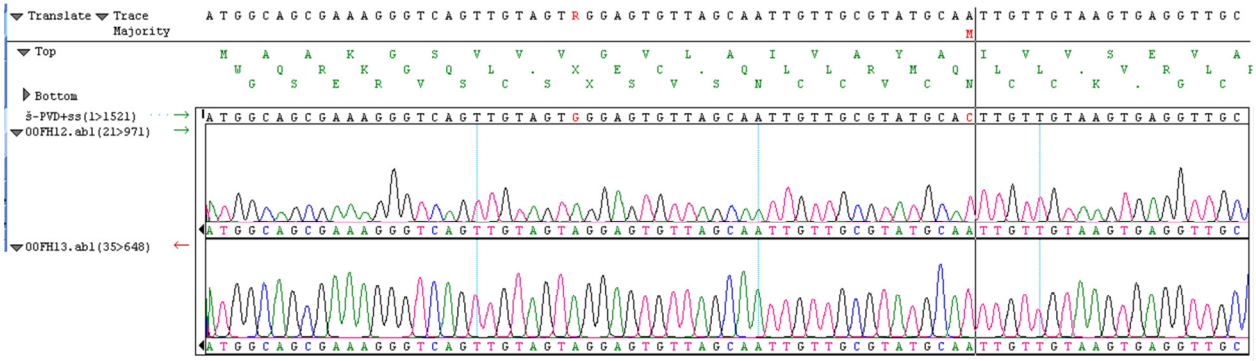
APPENDIX 1. Ligation-restriction reaction.

Mix :75ng PCR fragment 1, 75ng PCR fragment 2, 0.5 µL restriction enzyme, 1 µL T4 ligase buffer, 1 µL T4 ligase enzyme, 2 µL 10mM DTT and X µL H_20 Milli-Q to a final volume 10 µL. Thermocycler program: 25 cycles X (37°C for 2 min, 16°C for 5 min), 1 cycle 80°C for 2 min, 10°C until the end.

APPENDIX 2. RT and masses for volatiles diglycosides conjugates determined by LC-MS analysis.

Compound	mass [M-H]-	RT
Eugenol-hexose-pentose	503.17703	27.23107
Eugenol hexose	371.13467	29.08005
MeSa-hexose-pentose	491.14075	16.81453
MeSa-hexose	491.14063	13.63516
Guaiacol-hexose-pentose	463.14563	13.81555
Guaiacol hexose	331.10333	13.04890
Methylbutanol Hexose-Pentose	427.18195	16.79195
Phenylethanol Glucose-Arabinose	415.16092	17.78407
Benzyl alcohol hexose-pentose	447.15054	13.63516

APPENDIX 3. Change in single amino acid for tea-primeverosidase



APPENDIX 4. LC-MS and GC-MS results for metabolites showed in the report. Tables include means ± standard deviation and p-value obtained by T-test analysis, α 0.05, of transformed samples with respect to the control (empty vector).

Eugenol-hexose-pentose

	SodiumCitrate pH 4.5		Water	
	Mean ± SD	p-value	Mean ± SD	p-value
Wild Type	12155 ± 417		11063 ± 815	
Empty vector	12743 ± 809		10875 ± 1771	
Tea-pvd	642 ± 202	0.000	642 ± 202	0.001
GH01(74)	12064 ± 1010	0.415	10573 ± 2720	0.880
GH01(85)	12564 ± 914	0.832	13760 ± 3128	0.237
GH02(74)	6194 ± 1379	0.002	8944 ± 2072	0.287
GH02(85)	4777 ± 1041	0.000	7441 ± 864	0.039

Eugenol-hexose

	SodiumCitrate pH 4.5		Water	
	Mean ± SD	p-value	Mean ± SD	p-value
Wild Type	13672 ± 1111		12380 ± 1147	
Empty vector	12486 ± 256		9191 ± 354	
Tea-pvd	2594 ± 1597	0.000	700 ± 247	0.000
GH01(74)	12442 ± 621	0.915	10290 ± 1951	0.377
GH01(85)	13163 ± 1262	0.415	9732 ± 402	0.155
GH02(74)	17308 ± 366	0.000	10258 ± 370	0.023
GH02(85)	17369 ± 1283	0.003	14363 ± 118	0.000

MeSa-hexose-pentose

	SodiumCitrate pH 4.5		Water	
	Mean ± SD	p-value	Mean ± SD	p-value
Wild Type	4708 ± 2078		6006 ± 2651	
Empty vector	5315 ± 1131		5481 ± 1100	
Tea-pvd	642 ± 202	0.002	642 ± 202	0.002
GH01(74)	4933 ± 2390	0.815	5758 ± 3998	0.913
GH01(85)	5882 ± 2693	0.754	8021 ± 4514	0.387
GH02(74)	5894 ± 4976	0.854	10522 ± 5681	0.206
GH02(85)	7970 ± 2123	0.129	9585 ± 1747	0.026

Guaiacol-hexose-pentose

	SodiumCitrate pH 4.5		Water	
	Mean ± SD	p-value	Mean ± SD	p-value
Wild Type	12054 ± 1906		11526 ± 2962	
Empty vector	13392 ± 2117		9778 ± 1154	
Tea-pvd	758 ± 202	0.001	758 ± 202	0.000
GH01(74)	12249 ± 3003	0.619	10849 ± 5056	0.739
GH01(85)	13406 ± 2452	0.995	16325 ± 6028	0.138
GH02(74)	11517 ± 5911	0.632	14873 ± 5863	0.214
GH02(85)	11767 ± 2457	0.435	13549 ± 26333	0.085

Methylbutanol Hexose-Pentose

	SodiumCitrate pH 4.5		Water	
	Mean ± SD	p-value	Mean ± SD	p-value
Wild Type	5926 ± 231		5663 ± 250	
Empty vector	5761 ± 139		5263 ± 282	
Tea-pvd	780 ± 224	0.000	989 ± 473	0.000
GH01(74)	5625 ± 94	0.232	5258 ± 196	0.981
GH01(85)	5760 ± 494	0.997	5420 ± 330	0.565
GH02(74)	4516 ± 141	0.000	4656 ± 218	0.042
GH02(85)	4742 ± 45	0.000	4469 ± 281	0.026

Phenylethanol Glucose-Arabinose

	SodiumCitrate pH 4.5		Water	
	Mean ± SD	p-value	Mean ± SD	p-value
Wild Type	3790 ± 132		3554 ± 337	
Empty vector	3656 ± 244		3425 ± 158	
Tea-pvd	642 ± 202	0.000	642 ± 202	0.000
GH01(74)	3911 ± 311	0.327	3404 ± 339	0.928
GH01(85)	3531 ± 384	0.659	3500 ± 220	0.656
GH02(74)	3651 ± 349	0.983	3478 ± 232	0.760
GH02(85)	3812 ± 252	0.485	3818 ± 133	0.030

Benzyl alcohol hexose-pentose

	SodiumCitrate pH 4.5		Water	
	Mean ± SD	p-value	Mean ± SD	p-value
Wild Type	129382 ± 3018		126441 ± 1064	
Empty vector	128811 ± 4205		121884 ± 3357	
Tea-pvd	30888 ± 6445	0.000	49930 ± 8758	0.000
GH01(74)	127241 ± 3600	0.649	124767 ± 3052	0.333
GH01(85)	129269 ± 7817	0.933	123293 ± 2957	0.614
GH02(74)	126419 ± 736	0.387	122332 ± 2048	0.853
GH02(85)	130216 ± 1097	0.605	130011 ± 336	0.014

	Eugenol-hexose SodiumCitrate pH 4.5		**MeSa-hexose** SodiumCitrate pH 4.5		**Guaiacol-hexose** SodiumCitrate pH 4.5	
	Mean ± SD	p-value	Mean ± SD	p-value	Mean ± SD	p-value
Wild Type	13672 ± 1111		7451 ± 98		7496 ± 123	
Empty vector	12486 ± 256		7283 ± 347		7397 ± 474	
Tea-pvd	2594 ± 1597	0.000	1569 ± 442	0.000	2236 ± 1505	0.005
GH01(74)	12442 ± 621	0.915	7097 ± 332	0.540	7076 ± 392	0.417
GH01(85)	13163 ± 1262	0.415	7442 ± 794	0.767	7248 ± 207	0.643
GH02(74)	17308 ± 366	0.000	7571 ± 540	0.481	12125 ± 449	0.000
GH02(85)	17369 ± 1283	0.003	7279 ± 189	0.986	13681 ± 492	0.000

	Eugenol SodiumCitrate pH 4.5		**MeSa** SodiumCitrate pH 4.5		**Guaiacol** SodiumCitrate pH 4.5	
	Mean ± SD	p-value	Mean ± SD	p-value	Mean ± SD	p-value
Wild Type-DG	817 ± 16		9521 ± 415		1815 ± 874	
Wild Type	7120 ± 4447		2446407 ± 81774		170900 ± 39077	
Empty vector	15341 ± 13557		606680 ± 365828		175198 ± 73260	
Tea-pvd	31835 ± 5069	0.120	686436 ± 293363	0.783	220231 ± 8037	0.350
GH01(74)	10978 ± 5198	0.630	657291 ± 121342	0.831	183257 ± 25456	0.866
GH01(85)	17592 ± 6321	0.807	568759 ± 116362	0.872	220179 ± 27985	0.377
GH02(74)	19427 ± 12036	0.716	582708 ± 397307	0.942	189604 ± 37678	0.777
GH02(85)	10292 ± 6925	0.596	550523 ± 151629	0.818	166381 ± 21138	0.851

	Phenylethanol SodiumCitrate pH 4.5		**Methylbutanol** SodiumCitrate pH 4.5		**Benzyl alcohol** SodiumCitrate pH 4.5	
	Mean ± SD	p-value	Mean ± SD	p-value	Mean ± SD	p-value
Wild Type-DG	867 ± 45		8885 ± 3844		2453 ± 1908	
Wild Type	7934 ± 509		7861 ± 1756		3068 ± 1067	
Empty vector	58598 ± 32190		42834 ± 23462		24730 ± 40165	
Tea-pvd	167289 ± 14090	0.012	65507 ± 11667	0.208	134658 ± 54907	0.049
GH01(74)	59525 ± 18941	0.969	36220 ± 12150	0.687	6342 ± 655	0.472
GH01(85)	61003 ± 16808	0.916	32547 ± 8112	0.513	5289 ± 926	0.449
GH02(74)	46375 ± 25464	0.664	24805 ± 5502	0.265	7953 ± 10700	0.523
GH02(85)	56888 ± 24800	0.950	32944 ± 9299	0.535	3867 ± 2046	0.420