Plant Physiology and Biochemistry 135 (2019) 563–569

Contents lists available at ScienceDirect

Plant Physiology and Biochemistry

journal homepage: www.elsevier.com/locate/plaphy

Research article

UGT76G1 polymorphism in *Stevia rebaudiana*: New variants for steviol glycosides conjugation

Eva Petit[a], Alban Jacques[a], Jean Daydé[a], Veronica Vallejo[b], Monique Berger[a,*]

[a] Equipe Physiologie, Pathologie et Génétique Végétales (PPGV), University of Toulouse, National Polytechnic Institute of Toulouse, Ecole d'ingénieurs de Purpan, 75 voie du TOEC, BP 57611, F-31076, Toulouse Cedex 03, France

[b] PepsiCo Global R&D Agro Discovery, 3 Skyline Dr. Hawthorne, NY, 10532, USA

ARTICLE INFO

Keywords:
Polymorphism
Rebaudioside A enablement locus
Stevia rebaudiana
Steviol glycoside composition
Uridine-diphosphate glycosyltransferase
UGT76G1

ABSTRACT

Steviol glycosides (SVglys) are secondary metabolites derived from terpenoids exhibiting high-sweetening properties produced in Stevia rebaudiana leaves. Their great diversity is due to the number, the position and the nature of glycosylations on the steviol aglycone. Steviol conjugation is mediated by uridine-diphosphate glycosyltransferases (UGTs). Four UGTs have been clearly identified as involved in SVglys metabolism: UGT74G1, UGT85C2, UGT76G1 and UGT73E1. Natural non-functional mutants with nonsense codon have yet been observed for UGT76G1. To investigate the variability of UGT76G1 functionality, natural mutants with low or no content of rebaudioside A and C were identified in a germplasm collection of *Stevia rebaudiana*. These compounds are known to be the direct products of UGT76G1 and their biosynthesis is governed by a single gene at the locus *Rae* (Rebaudioside A enablement). Crosses were done with remarkable accessions including phenotypes with low (0–3%) and high proportions (70%) of rebaudioside A and C, to investigate the functionality of the *Rae* locus in the parents. Seven variants of *UGT76G1* were found, among them 4 lead to a functional protein and 3 lead to non-functional isoforms. Five of these variants are new. We found that non-functionality of UGT76G1 towards SVglys is not due to a premature nonsense codon, which appears to be an extreme case to explain the loss of functionality of an UGT. Variations in steviol glycoside profile in stevia leaves is partly due to *UGT76G1* polymorphism: amino acid substitutions in parts of the protein involved in the substrate specificity can be found by sequence comparison.

1. Introduction

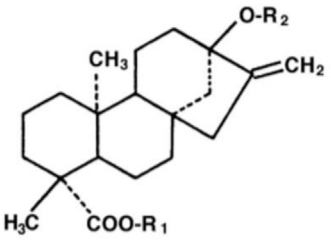
Stevia rebaudiana is a perennial herb used for centuries by Paraguayan Indians for its sweetening properties, due to steviol glycosides (SVglys), a unique secondary metabolite class that accumulates in stevia leaves (Soejarto et al., 1982). SVglys are ent-kaurenoic diterpenoids sharing a part of their biosynthetic pathway with gibberellins (Richman et al., 1999). They are constituted of a aglycone steviol, on which two hydroxyl groups in the C-13 and C-19 position can be glycosylated (Brandle and Telmer, 2007; Ceunen and Geuns, 2013). The biosynthesis of diterpenes begins in the chloroplast by the formation of geranyl-geranyl diphosphate (GGDP) synthesized from the MEP (methyl-erythritol-4-phosphate) pathway (Cheng et al., 2007). GGDP then undergoes two cyclization steps, catalyzed by terpene cyclases to give the ent-kaurene (Brandle and Telmer, 2007). The ent-kaurene is oxidized into kaurenoic acid (KA) by the kaurene oxidase (KO), in endoplasmic reticulum (ER). Kaurenoic acid is the last step of a biosynthetic pathway shared by gibberellins and SVglys. Oxidation of KA leads to gibberellins, while its hydroxylation produces steviol. Steviol glycosylation is catalyzed by UDP (uridine diphosphate)-glycosyltransferases (UGTs). These enzymes achieve their substrate glycosylation by transferring activated sugars on hydroxyl groups of their acceptors, here steviol or SVglys. The nature and number of sugars attached by UGTs is the source of SVglys diversity (Brandle and Telmer, 2007). In *S. rebaudiana*, only four UGTs involved in SVglys biosynthesis have been identified: UGT74G1 and UGT73E1 catalyze β1-(C-19) glucosylations, UGT85C2 and UGT76G1 catalyze respectively β1-2 and β1-3 glucosylations (C-13) (Li et al., 2018; Richman et al., 2005).

In *S. rebaudiana* wild populations, major SVglys are ST, RA, RC and DA (Brandle et al., 1998). A wide phenotypic variability for SVglys

Abbreviations: DA, dulcoside A; RA, rebaudioside A; Rae, rebaudioside A enablement; RC, rebaudioside C; Rce, rebaudioside C enablement; ST, stevioside; SVgly, steviol glycoside; UGT, uridine-diphosphate glycosyltransferase

* Corresponding author.
E-mail address: monique.berger@purpan.fr (M. Berger).

https://doi.org/10.1016/j.plaphy.2018.11.002
Received 3 October 2018; Received in revised form 31 October 2018; Accepted 2 November 2018
Available online 07 November 2018



	R1	R2
Stevioside	Glcβ1-	Glcβ(1-2)Glcβ1-
Rebaudioside A	Glcβ1-	Glcβ(1-2)[Glcβ(1-3)]Glcβ1-
Rebaudioside D	Glcβ(1-2)Glcβ1-	Glc(β1-2)[Glcβ(1-3)]Glcβ1-
Rebaudioside I	Glcβ(1-3)Glcβ1-	Glc(β1-2)[Glcβ(1-3)]Glcβ1-
Rebaudioside M	Glcβ(1-2)[Glcβ(1-3)] Glcβ1-	Glc(β1-2)[Glcβ(1-3)]Glcβ1-
Dulcoside A	Glcβ1-	Rhaα(1-2)Glcβ1-
Rebaudioside C	Glcβ1-	Rhaα(1-2)[Glcβ(1-3)]Glcβ1-

Fig. 1. Chemical structure of stevioside, rebaudioside A, D, I, and M, dulcoside A and rebaudioside C.

profile depending on the genotype was already highlighted. Indeed, some plants present very high proportion of RA and/or RC, whereas others accumulate preferentially their corresponding substrates ST and DA, respectively (Barbet-Massin et al., 2016; Francisco et al., 2018; Nakamura and Tamura, 1985; Yang et al., 2014). It has been shown that this part of the biosynthetic pathway is mainly governed by two loci: *Rae* (Rebaudioside A enablement) is responsible for the absence/presence of RA with functional (*RAE*) or non-functional alleles (*rae*). Rebaudioside C enablement (*Rce*) is responsible for ST and DA proportions with *Rce* and *rce* alleles (Brandle, 1999). The β1-3 glycosylation (C-13) which yields RA and RC from their respective precursor, ST and DA, is governed by a single gene as their presence or absence are closely correlated (Brandle, 1999), indicating that RA and RC are formed by the same enzyme, UGT76G1 (Richman et al., 2005). The stability of SVglys composition in a genotype is mostly due to genotypic determinism whereas environmental conditions have only a slight effect on determination of these proportions (Barbet-Massin et al., 2016; Gaurav et al., 2008). When UGT76G1 is functional, RA and RC can be used for higher order glycosylations as rebaudioside D (RD), I (RI) or M (RM) (Fig. 1).

Knowledge of the origin of the SVglys profile diversity should help to conduct *S. rebaudiana* breeding according to its SVglys profile. Several studies describe RA content and RA/ST ratio as a selection criteria (Barbet-Massin et al., 2016; Ceunen and Geuns, 2013; Francisco et al., 2018; Kumari et al., 2018). RA presents a major interest for food industry, as its sweetening power is between 150 and 320 times that of sucrose (Phillips, 1989). RA seems to have the most interesting organoleptic properties among the four major SVglys, as it is the least astringent, the least bitter, and had the least persistent aftertaste (DuBois et al., 1985; Tanaka, 2009). Some other SVglys derived from RA exhibit higher sweetness and less bitterness, such as rebaudioside D (RD) and M (RM) (Allen et al., 2013; Prakash et al., 2014), supporting the concept of a selective breeding based on RA pathway functionality.

This study aims to evaluate the allelic diversity of UGT76G1. The wide range of SVglys can be partially explained by the successive action of many UGTs, adding various types of sugars. Moreover, a part of the SVglys variability could also be explained by the existence of multiple UGTs alleles, which may have different substrate specificities, affecting endogenous concentrations and complicating the metabolic pathway of SVglys.

For this purpose, biparental crosses with genotypes chosen for their relative proportion of RA and/or RC were done. Their progenies were then clustered depending on their relative content in UGT76G1 products. This mendelian determination of the functional homo- or heterozygosity at the *Rae* locus was completed by the analysis of allelic variability of the *UGT76G1* cDNA sequences.

2. Material and methods

2.1. Biological material

2.1.1. Biparental crosses

Our germplasm is derived from selection cycles of Criolla and Morita populations. Before crosses, the flowering of the parental genotypes was induced in a growth chamber for a two-weeks period under short days (8 h of daylight). For each cross, 4 plants (two cuttings of each parental genotype) were put under a tent which was sealed to prevent contamination by external pollen. Pollination was made by flies during flowering (4 weeks). Seeds were collected at the stage when they were completely mature, 4 weeks later. Biparental crosses, and their reciprocals, are noted with the mother in first position and the male in second (i.e. A(♀):B(♂) designates a cross with genotype A for the mother plant and genotype B for the father). Seeds from a cross were collected separately from its reciprocal as some crosses produced no viable seed. This absence of seeds is due to a self-incompatibility system in *S. rebaudiana* (Carneiro and Walter, 1990; Oddone, 1997; Sumida, 1980; Yadav et al., 2014).

2.1.2. Growth conditions and sampling

Seeds were sterilized then sown on germination paper moistened with 3 mL water in Petri dishes. Humidity was controlled by adding water as necessary. When cotyledons appeared, seedlings were deposited on 104-well culture plates filled with seed compost. After 8–10 days, seedlings were re-potted in 7 × 7 cm pots.

2.2. Chemical analysis

2.2.1. Sample preparation and extraction

For parental genotypes, a complete stem (from the third to the oldest leaf) was sampled. Leaves were dried in paper bags at 50 °C for 48 h and then ground to a fine powder with a mortar and pestle. Steviol glycosides were extracted from 100 mg of powder in 9 mL of ultra-high-purity water under rotative stirring. For high throughput phenotyping of progenies, three mature half leaves per individual were sampled, dried and ground directly into 2 mL 96-wells deepwell plates, which correspond to 20–30 mg of dry matter. SVglys were extracted in 1.8 mL per well of ultra-high-purity water at 50 °C during 30 min under rotative stirring. Extracts were filtered through a 0.22 µm Minisart® syringe filter (Sartorius AG, Germany) for parental genotypes and 0.22 µm EMD Millipore™ MultiScreen$_{HTS}$ Durapore™ 96-Well Filter Plate (Merck Millipore, USA) for progenies.

2.2.2. HPLC conditions

Liquid chromatography under isocratic conditions was performed on a Dionex UltiMate™ 3000 system from ThermoScientific consisting of a DGP-3600RS pump, a WPS-3000 TRS autosampler, a TCC-3000 SD column compartment maintained at 40 °C and a DAD-3000 diode array detector set to a wavelength of 195 nm. The flow rate and sample injection volume were 0.3 mL min^{-1} and 1 µL, respectively. Steviol glycosides were separated with a reverse phase column Kinetex C18 (150 × 2.1 mm, 1.7 µm) (Phenomenex, USA), maintained at 40 °C. Solvents used as mobile phase were 32% acetonitrile (Carlo Erba Reagents, Italy) and 68% ultra-high-purity water (Milli-Q System from Millipore, USA) acidified with 0,1% formic acid (pH2.6). A mix of 9 SVglys standards was purchased from Chromadex, Inc. (Irvine, CA, USA) and were used to identify stevioside, dulcoside A, rebaudioside A and C.

2.2.3. Data analysis

Steviol Glycosides peak areas obtained from Chromeleon™ software for ST, RA, DA, RC, were summed to get total area, representing total SVglys as only the steviol moiety is detected at 195 nm. Proportion of each SVgly was calculated relatively to the total area. Relative

proportions of RA and RC were summed to get the index %RA + RC, used to determine UGT76G1 functionality. Ranking between minimum and maximum %RA + RC observed in progeny from a cross was calculated and divided into 12 classes. These classes have an equal amplitude of 1/12 between the minimum and the maximum of %RA + RC. Individuals from the progeny were clustered in the 12 created classes depending on their %RA + RC. Profile segregations were confirmed with Chi square test.

2.3. UGT76G1 sequencing

2.3.1. Sample preparation and total RNA extraction

For *UGT76G1* sequencing, one young leaf (second or third node from the top) was sampled for each plant. Leaf samples were frozen in liquid nitrogen immediately after sampling. Samples were ground into fine powder in a mixer mill MM400 (Retsch, Haan, Germany) with stainless steel balls (2 mm Ø) and then stabilized in extraction buffer (2% CTAB, 2% PVPP, 300 mM Tris, 25 mM EDTA, 2 M NaCl (pH 8.0), 2% β-mercaptoethanol) for 10 min at 65 °C. The lysate was washed twice with one volume of chloroform/isoamylic alcohol (24/1, v/v). Nucleic acids were precipitated by 2-h freezing with 0.6 volume of isopropanol and 0.1 volume of 3 M sodium acetate (pH 5.0). Nucleic acid pellets were washed with 70% ethanol and solubilized in sterile RNAse-free water. The RNA concentration and purity were measured with Nanodrop (ThermoFisher Scientific). Residual genomic DNA was eliminated by a DNAse I treatment (Thermo Fisher Scientific Inc., USA).

2.3.2. RT-PCR

Reverse transcription was performed with Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Fisher Scientific Inc., USA). Primers for *UGT76G1* coding sequence were designed from accession AY345974 (Richman et al., 2005) with OligoPerfect™ Designer (Forward: CGTGTAAACGTCAGTCAAACCC; Reverse: CATGCAATCCAAGT GCTTGA; melting temperature: 53,5 °C; product size: 1436 bp). PCR reactions were conducted in a MyCycler™ (Bio-Rad Laboratories Pty Ltd., Gladesville, Australia) with Platinum™ Taq DNA Polymerase High Fidelity (Invitrogen, Carlsbad, CA, USA) according to manufacturer specifications.

2.3.3. Cloning and sequencing

Purified PCR products were cloned into pGEM®-T vector (Promega, Madison, WI, USA). JM109 competent cells were transformed with produced vectors. Recombinants were purified from 5 mL liquid cultures with Wizard® Plus SV Minipreps DNA Purification System (Promega, Madison, WI, USA) and tested by PCR with UGT76G1 primers. The sequencing of *UGT76G1* was done using T7 and SP6 primers.

2.3.4. Sequencing data analysis

The *UGT76G1* nucleotide sequences were aligned with MUSCLE alignment (3.8) tool. Putative protein sequences were obtained with Web Expasy translate tool. Correct reading frames were chosen according to accession AY345974.1 (Richman et al., 2005). Phylogenetic analysis was conducted with Phylogeny.fr (Dereeper et al., 2008). Pairwise and multiple alignments of proteins were done with EMBOSS Needle and Clustal Omega (1.2.4), respectively. Variants identified as functional for β1-3 glycosylation by UGT76G1 were annotated *RAE*, whereas non-functional ones were annotated *rae*.

3. Results

3.1. Genetic determinism of UGT76G1 functionality by biparental crosses

Steviol glycoside profiles were established for 7 genotypes (Fig. 2). Profile stability was confirmed by analysis of 2–10 plants grown 2 or 3 years in field or in greenhouse. Several plants per genotype were sampled to confirm profiles stability. Some genotypes (C, D, F) were analyzed only 2 times due to cutting mortality. Genotypes A, B, C and D can produce RA and/or RC, meaning that they have one or two functional alleles (e.g. *raeRAE* or *RAERAE*). Proportions of RA and RC are not equivalent depending on these genotypes, confirmed by discrepancies between RA/ST and RC/DA ratios (i.e $RA/ST_{genotype\ B} = 0.5 \pm 0.1$ and $RC/DA_{genotype\ B} = 1.7 \pm 0.3$). However, successive glycosylations of RA and RC could happen after β1-3 glycosylation by UGT76G1.

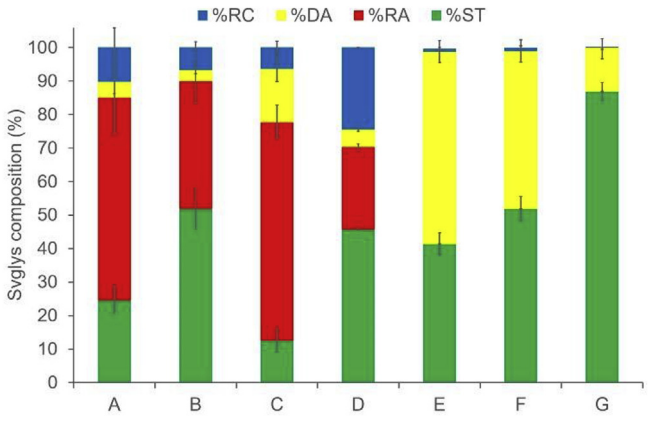


Fig. 2. Relative compositions of steviol glycosides (SVglys) for genotypes A (n = 9), B (n = 9), C (n = 5), D (n = 2), E (n = 7), F (n = 6), G (n = 9).

Genotypes E, F and G do not produce RA nor RC thus are consequently homozygous for UGT76G1 non-functionality (*raerae*). For genotype E and F, a slight activity of UGT76G1 towards DA can be observed, however composition of RC (0.3 ± 0.4 and 1.0 ± 0.2 for E and F, respectively) is too low to consider UGT76G1 as an active enzyme, compared to *raeRAE*- or *RAERAE*-genotypes.

Five biparental crosses were done with these 7 genotypes to determine functionality of their *UGT76G1* alleles. Mendelian segregations were expected to confirm monogenic determinism of locus *Rae*. Crosses D:G and G:D gave 50 plants, divided in two equal groups, one producing RA and/or RC (dark green), and the other having null or very low content in RA and RC (light green) (Fig. 3A). This segregation supports the hypothesis of a cross between *raerae* and *raeRAE* genotypes. Considering genotype G status of UGT76G1 functionality (*raerae*), genotype D is therefore *raeRAE*.

Genotype D was crossed with genotype A. No viable seed was obtained for the cross D:A. Segregation of A:D progeny gave 66 plants, all having residual content of ST and DA, indicating that UGT76G1 substrates are not limiting for enzyme functionality (Fig. 3B). All plants produce RA and/or RC, meaning that genotype A is mandatory *RAERAE*. Moreover, plants are divided into two groups that may correspond to *raeRAE* and *RAERAE* plants (but separation is not clear, preventing statistical tests).

Genotype B was crossed with genotype A. Only 4 seeds from cross B:A were able to give viable plants. There were 59 plants in progeny from the A:B cross, which all have a residual content of ST and DA, indicating that UGT76G1 substrates are not limiting for enzyme functionality (Fig. 3C). If B had been *raeRAE*, two groups would have been observed, with 50% of *raeRAE* plants (medium content in RA and RC) and 50% *RAERAE* plants (high content in RA and RC). Here, all plants produce RA and/or RC but no segregation can be observed, meaning that genotype B is most likely *RAERAE*.

Genotype A was crossed with genotype C and 265 plants were obtained (Fig. 3D). Only 7 plants convert the whole DA, indicating a possible limiting activity ahead of UGT76G1 in these plants. Analysis of UGT76G1 functionality was done by removing them, considering n = 265 as a sufficient population sample. All plants produce % RA + RC but no segregation was observed, indicating that genotype C may be *RAERAE*.

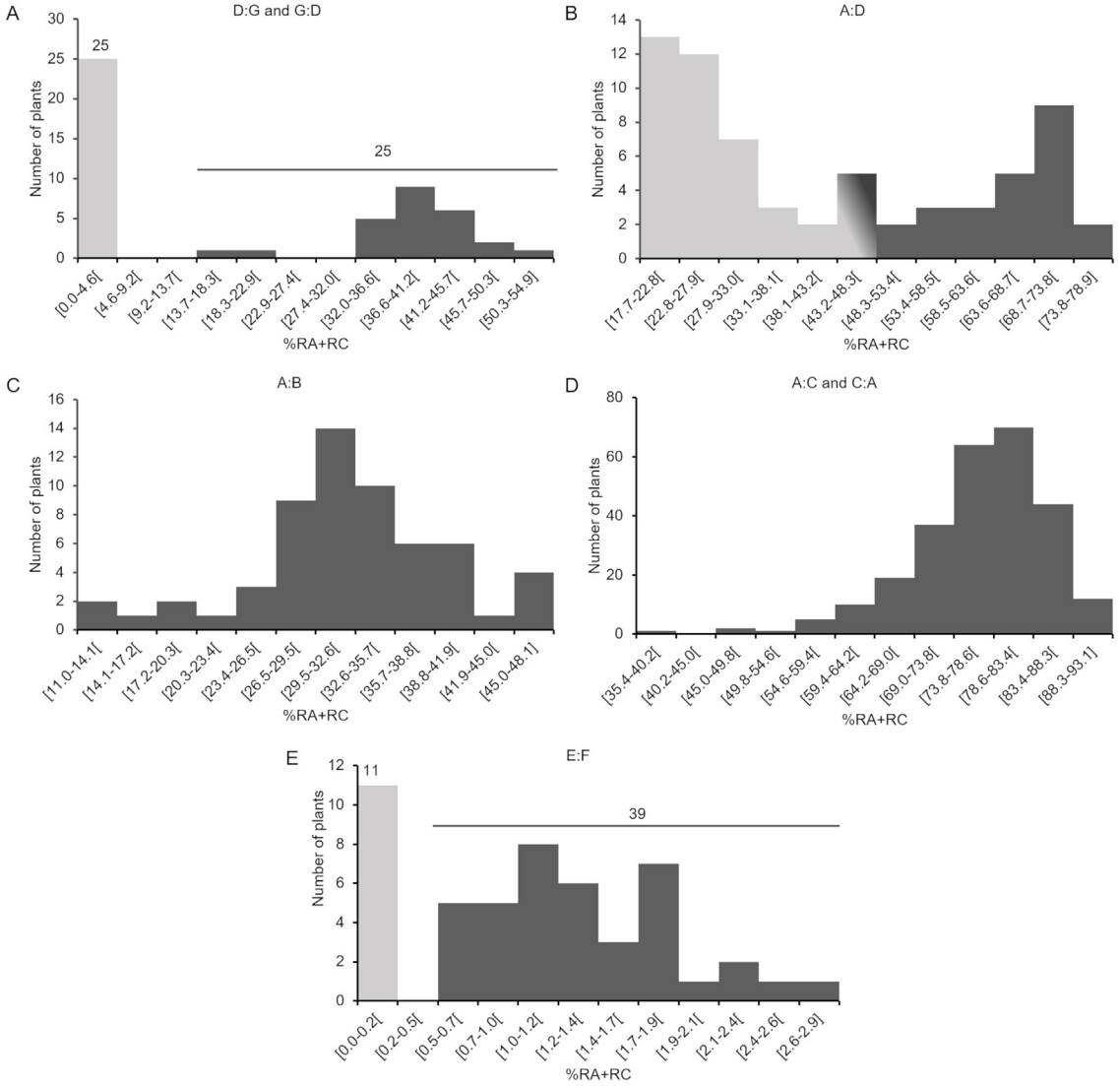


Fig. 3. Segregation of progenies according to %RA + RC for the five biparental crosses (A) D:G and G:D, n = 50 (B) A:D, n = 66 (C) A:B, n = 59 (D) A:C and C:A, n = 265 (E) E:F, n = 50. Grey tones represent groups.

Genotypes E and F were crossed; the F:E cross did not produce viable seeds but 50 plants were obtained for E:F (Fig. 3E). These plants do not produce RA, and the residual activity towards DA is conserved from parental genotypes, as RC is produced in very low proportions (0.0–2.9%). These results confirm UGT76G1 non-functionality in both E and F. Segregation of %RC suggests two different groups with 39 plants at 0.5–2.9 %RC and 11 plants without RC. Mendelian 3:1 segregation is typical of crosses between heterozygous individuals, yielding two groups, the first with 25% non-functional UGT76G1 and the other with 75% functional UGT76G1. Chi-square test ($\chi^2 = 0.24$, DF = 1; p = 0.62) fits this hypothesis, meaning that E and F are both *raerae* and heterozygous at molecular level with one allele providing a residual activity towards DA, for each genotype.

3.2. *UGT76G1 sequences polymorphism*

3.2.1. *UGT76G1 sequencing*

After RT-PCR, a band at 500 bp was amplified, corresponding to tubulin, used as control. A band was obtained from each cDNA at around 1.4 kb, corresponding to the coding sequence of UGT76G1 (Fig. 4). UGT76G1 was amplified even in genotypes homozygous for non-functionality, meaning that source mRNAs were probably not degraded. UGT76G1 seems to be less expressed in genotype G.

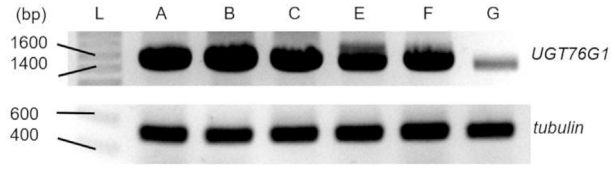


Fig. 4. UGT76G1 products (complete CDS) amplified in cDNAs of the 6 genotypes. L: ladder (bp).

Seven sequences were obtained from these 6 genotypes (Table 2). Sequencing of *UGT76G1* in genotype A provides only one sequence, meaning that this genotype may be homozygous. Genotypes B, C, E, F and G are heterozygous, confirming our hypothesis previously done. The seven alleles present two different length (1377 and 1380 bp), but insertions observed in allele 2 and 5 are differently located. Alleles 2, 5 and 6 can be found in several genotypes.

3.2.2. *UGT76G1 protein sequences*

The protein sequences obtained from these alleles were aligned and compared with each other and with available UGT76G1 sequences. The results support the allelic status for *Rae* locus in each genotype, that

Table 1
Summary of functionality of locus *Rae* for the 7 parental genotypes.

Genotypes	Locus *Rae*
A	*RAERAE*
B	*RAERAE*
C	*RAERAE*
D	*raeRAE*
E	*raerae*
F	*raerae*
G	*raerae*

Table 2
General features of sequenced alleles.

Alleles	Length (bp)	Genotypes
1	1377	A
2	1380	B, C
3	1377	B
4	1377	C
5	1380	E, F, G
6	1377	E, F
7	1377	G

were deduced from chemical profiles and segregations. Genotype A is *RAERAE* (Table 1), meaning that allele 1 provides a functional protein. This result is confirmed by the fact that allele 1 has 100% of identity with accession AAR06912.1, which was shown to produce a functional protein (Richman et al., 2005). Genotype A may be *RAERAE* with two identical alleles giving a protein of 458 residues as no other sequence was found. Alleles 3 and 4 have 100.0% and 99.8% of similarity with allele 1, respectively and they both share 99.8% amino acid identity with the allele 1 (Table S.1). Considering that alleles 3 and 4 are nearly identical to allele 1, and the functionality of the locus *Rae* in the genotypes B and C, these alleles are most likely functional. Genotypes B and C are *RAERAE* and heterozygous on a molecular point of view. Allele 2 is identical to accession ACT33422.1 and to the functional isoform UGTSr studied by Madhav et al. (2013), confirming functionality of allele 2 (the sequence of *UGTSr* was reconstituted thanks to the SNPs table provided by the authors).

The protein sequences obtained from the seven alleles were then compared with supplemental protein sequences, Z04 and Z05, which were identified as functional and non-functional, respectively (Yang et al., 2014). Genotypes E and F both have a copy of alleles 5 and 6; these genotypes being *raerae*, alleles 5 and 6 producing non-functional proteins. Moreover, protein of allele 6 has 99.8% of identity with the Z05 sequence. The only difference between these sequences is the deletion of the R235/236, but this modification is observed in the functional Z04 isoform so does not nullify activity. Genotype G is *raerae* and has a copy of allele 5, confirming that this allele provides a non-functional protein. Due to genotype G's phenotype, allele 7 is also non-functional.

A phylogenetic tree was built with protein sequences of alleles and public accessions, showing genetic divergence between the clade containing UGTSr, ACT33422.1, Z04, Z05, allele 2, 5, 6 and the other one with AAR06912.1, alleles 1, 3, 4 and 7 (Fig. 5). This tree and the pairwise comparisons (Table S.1) show that non-functionality of UGT76G1 is not due to a divergence event as *rae* sequences are not grouped in a single clade. Indeed, allele 7 is *rae* but its sequence is more distant from the *rae* alleles 5 and 6 (around 91% and 94% of identity and similarity, respectively) than from the *RAE* alleles 1, 3 and 4 (around 95% and 97% of identity and similarity, respectively) so point mutations that provoke loss of functionality for allele 7 are probably different from those of alleles 5 and 6.

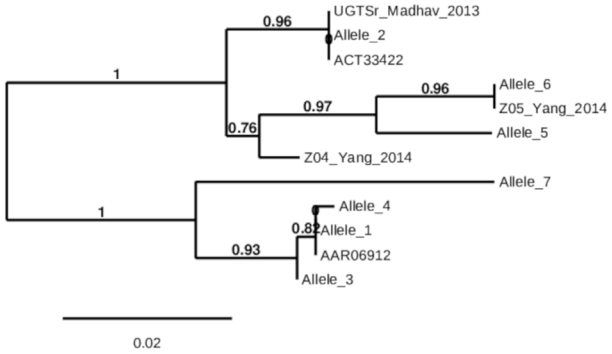


Fig. 5. Phylogenetic tree representing link between UGT76G1 versions.

3.2.3. SNPs and amino acid modifications between rae and RAE variants

Variants 5, 6 and 7 were aligned with *RAE* variants to identify modifications that can discriminate functional alleles from non-functional ones. As expected, modifications of allele 7 are different from those of alleles 5 and 6 and no nucleotide modification gave rise to a premature termination codon in any of the three alleles. However, several mutations are observed, shared between N− and C-terminal domains (Table 3). Insertion of a single residue (86insS) was observed for allele 5 and at least 11, 9, 15 substitutions were counted for allele 5, 6 and 7 respectively (Table 3). Each of these modifications can be involved in lack of functionality towards ST and/or DA, but some of them are suspected to have a more important role in this loss.

Several residues have been identified to be highly conserved between UGTs of *S. rebaudiana* involved in steviol or SVglys glycosylation, among them twelve amino acids are suspected to be involved in steviol or SVglys recognition as substrates (Richman et al., 2005). Some of these residues are mutated, as K39E and N386T in allele 7. Moreover, a serine is inserted next to the fully conserved L85 in allele 5. Substitutions of such amino acids or modification of their chemical environment can directly affect UGT functionality toward its regular substrates. Other substitutions are located along the sequences, where important residues could be involved in UGTs activity.

4. Discussion

Segregation study via biparental crosses allowed us to confirm the hypothesis of a monogenic determination for the conversion of ST to RA and DA to RC (Brandle, 1999). Chemical profiles of SVglys and segregation ratios enabled determination of UGT76G1 functionality in each genotype. Genotypes A, B and C are *RAERAE* whereas genotypes E, F and G have both two non-functional alleles. Moreover, non-functionality of UGT76G1 in *raerae*-genotype was confirmed by HPLC-MS analysis of several plants of the progeny from the cross E:F (Figures S.2 and S.3). Indeed, we never observed SVglys that could derive from RA by their common structure, as RD, RI and RM (Figures S.2 and S.3). This result suggests that a large part of the variability is likely due to UGT76G1 polymorphism and not to consumption of RA and RC for higher order glycosylations. Amplification of UGT76G1 in cDNAs was possible for the *raerae*-genotypes and respective coding sequences do not harbor premature termination codon, meaning that mRNAs are probably not degraded by nonsense-mediated decay (NMD) as suggested Yang et al. (2014). Presumed functionality of the 7 alleles was confirmed by sequencing, whose result highlighted similarities with sequences already known to be functional or not.

Compared to *RAE* alleles, several substitutions have been identified in *rae*, located both in C− and N-terminal domains. N-terminal domain provides some residues interacting with the sugar donor (UDP-sugar) which are involved in determining sugar donor specificity. It also provides almost entirely residues forming the sugar acceptor pocket (Osmani et al., 2009; Wang, 2009). N-terminal domain modifications

Table 3
SNPs and amino acid substitutions differentiating non-functional UGT76G1 sequences compared to functional versions.

Allele 5		Allele 6		Allele 7	
Protein	*SNPs*	*Protein*	*SNPs*	*Protein*	*SNPs*
86insS	256insAGT	P80S	238C > T	A/T5R	14C > G
A87T	259G > A	K192Q	574A > C	E6Q	16G > C
R103Q	308G > A	C/S193F	578G/C > T	R13Q	38G > A
K/I200Q	598A > C	K/I200Q	A598C	Y37S	110A > C
Y/H/L201G	601T/C > G	Y/H/L201G	601T/C > G	K39E	115A > G
	602A/T > G		602A/T > G	F46L	136T > C
E291A	**872A > C**	**P327A**	**979C > G**	N78H	232A > C
W322H	**963T > C**	***D378A***	***1133A > C***	P/L92A	274C > G
	964G > A	***G380A***	***1139G > C***	I94F	280A > T
	965G > C	***L381F***	***1141C > T***	R142Q	426G > A
D378A	***1133A > C***			Y/H/L201V	601T/C > G
G380A	***1139G > C***			**R299C**	**895C > T**
L381F	***1141C > T***			**L309I**	**925T > A**
R430S	**1290A > T**			**L385V**	**1153T > G**
S455A	**1363T > G**			**N386T**	**1157A > C**
N-terminal domain		**C-terminal domain**		***PSPG-box***	

can be observed in the three *rae* variants. Consequently, modifications in N-terminal may be critical for secondary and tertiary structure conservation, substrate specificity (*i.e.* nature of sugar) and correct acceptor positioning, which are prerequisites for UGTs activity (Osmani et al., 2009).

The C-terminal domain of UGTs provides the majority of amino acids involved in sugar donor pocket formation (Osmani et al., 2009; Wang, 2009). Several modifications are common in alleles 5 and 6, among them, three are in the C-terminal domain and more precisely in the PSPG-box, which is known to be highly conserved in UGTs and flanks one side of the sugar donor pocket. Several residues of the PSPG-box directly interact with the donor sugar while the others are suspected to stabilize intramolecular interactions, all of them contributing to sugar donor specificity (Osmani et al., 2009). In *S. rebaudiana*, residues essential for activity of UGTs involved in steviol or SVglys glycosylation, are not yet identified. However, the substitution of one of them may modify the nature and/or the size of lateral chains that constitute the acceptor and/or donor pocket, thus, impairing the recognition mechanism of substrates and regioselectivity, leading to loss of functionality towards original substrates.

Most of the time, few substitutions are enough to lose UGT function. In *Dorotheantus bellidiformis*, substitution of two highly conserved residues of UDP-glucose-dependent betanidin 5-O-glucosyltransferase caused the total loss of functionality of the enzyme (Hans et al., 2004). In other cases, substitution can lead to a gain of function. For instance, Kubo et al. (2004) demonstrated that the galactosyltransferase ACGaT from *Aralia cordata* was also active with a glucose donor, following substitution of a single residue. In *Medicago truncatula*, some residues were identified by site-directed mutagenesis as essential for regioselectivity of UGT71G1 for quercetin glycosylation. Other mutations in UGT71G1 were responsible for the modification of relative specificity for tested sugar donors (He et al., 2006). *In vitro* activity assays conducted on UGT76G1 showed the wide range of potential substrates for this enzyme (Dewitte et al., 2016). As mRNAs are not degraded, UGT76G1 isoforms non-functional towards ST and DA could be produced and may have a different glycosylation activity, with other sugar donors or acceptors. Mutational analysis would be interesting to identify essential residues for β1-3 glycosylation of SVglys by UGT76G1.

Our polymorphism study emphasizes sequence divergences that may explain loss of functionality of rae isoforms towards ST and DA, the original substrates of UGT76G1. These results can be used to derive molecular markers to be used in markers-assisted selection of *S. rebaudiana* to accelerate the production of plants with high proportions of RA or even RA-derived SVglys as RD. This work shows the importance of the allelic diversity in a plant UGT. UGTs constitute a complex and large multi-gene family (Coutinho et al., 2003; Mackenzie et al., 1997; Ross et al., 2001; Vogt and Jones, 2000). In *Arabidopsis thaliana*, more than a hundred UGT genes involved in the modification of secondary metabolites have been identified (Li et al., 2001). It has been demonstrated that UGTs are not specific to a particular substrate, but rather regioselective or regiospecific (Jones et al., 1999; Vogt and Jones, 2000), which allows a single UGT to glycosylate several substrates, giving rise to numerous modified metabolites. The variability of UGT sequences observed in a population confers an additional level of complexity of metabolic profiles by the impact of this variability on substrate specificity, on endogenous metabolic concentrations, or even on the biosynthesis of new metabolites. The natural allelic multiplicity of the UGTs superfamily combined with regioselectivity highlights the complexity and richness of plant metabolomes.

CRediT authorship contribution statement

Eva Petit: Formal analysis, Writing – original draft.

Acknowledgments

This work was funded by PepsiCo Global R&D. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of PepsiCo Inc. The authors are grateful to Lauriane Heissat for production and culture of *Stevia rebaudiana* and to Apolline Champenois for her help in cloning experiments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.plaphy.2018.11.002.

References

Allen, A.L., McGeary, J.E., Hayes, J.E., 2013. Rebaudioside A and Rebaudioside D bitterness do not covary with Acesulfame K bitterness or polymorphisms in TAS2R9 and TAS2R31. Chemosens. Percept 6. https://doi.org/10.1007/s12078-013-9149-9.

Barbet-Massin, C., Giuliano, S., Alletto, L., Daydé, J., Berger, M., 2016. Towards a semi-perennial culture of *Stevia rebaudiana* (Bertoni) Bertoni under temperate climate: effects of genotype, environment and plant age on steviol glycoside content and composition. Genet. Resour. Crop Evol. 63, 685–694. https://doi.org/10.1007/s10722-015-0276-9.

Brandle, J., 1999. Genetic control of rebaudioside A and C concentration in leaves of the sweet herb, *Stevia rebaudiana*. Can. J. Plant Sci. 79, 85–91. https://doi.org/10.4141/P98-048.

Brandle, J.E., Telmer, P.G., 2007. Steviol glycoside biosynthesis. Phytochemistry 68,

1855–1863. https://doi.org/10.1016/j.phytochem.2007.02.010.
Brandle, J.E., Starratt, A.N., Gijzen, M., 1998. *Stevia rebaudiana*: its agricultural, biological, and chemical properties. Can. J. Plant Sci. 78, 527–536. https://doi.org/10.4141/P97-114.
Carneiro, P., Walter, J., 1990. *Stevia rebaudiana* (Bert) Bertoni: Produção de Sementes (Universidade Estadual de Maringá. Valdecir Antoninho Dalpasquale, José Walter Pedroza Carneiro, Raimundo Pinheiro Neto, Agostinho Baldin, Carmen Torresan).
Ceunen, S., Geuns, J.M.C., 2013. Steviol glycosides: chemical diversity, metabolism, and function. J. Nat. Prod. 76, 1201–1228. https://doi.org/10.1021/np400203b.
Cheng, A.-X., Lou, Y.-G., Mao, Y.-B., Lu, S., Wang, L.-J., Chen, X.-Y., 2007. Plant terpenoids: biosynthesis and ecological functions. J. Integr. Plant Biol. 49, 179–186. https://doi.org/10.1111/j.1672-9072.2006.00395.x.
Coutinho, P.M., Deleury, E., Davies, G.J., Henrissat, B., 2003. An evolving hierarchical family classification for glycosyltransferases. J. Mol. Biol. 328, 307–317. https://doi.org/10.1016/S0022-2836(03)00307-3.
Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., et al., 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36, W465–W469. https://doi.org/10.1093/nar/gkn180.
Dewitte, G., Walmagh, M., Diricks, M., Lepak, A., Gutmann, A., Nidetzky, B., Desmet, T., 2016. Screening of recombinant glycosyltransferases reveals the broad acceptor specificity of stevia UGT-76G1. J. Biotechnol. 233, 49–55. https://doi.org/10.1016/j.jbiotec.2016.06.034.
DuBois, G.E., Bunes, L.A., Dietrich, P.S., Stephenson, R.A., 1985. Diterpenoid sweeteners. Synthesis and sensory evaluation of stevioside analogs with improved organoleptic properties. J. Med. Chem. 28, 93–98. https://doi.org/10.1021/jm00379a017.
Francisco, F., Pereira, G.P., Machado, M.P., Kanis, L.A., Deschamps, C., 2018. Characterization of *Stevia rebaudiana* Bertoni accessions cultived in southern Brazil. J. Agric. Sci. 10, 353. https://doi.org/10.5539/jas.v10n3p353.
Gaurav, S.S., Singh, Y.P., Sirohi, S.P.S., 2008. Genetic variability for yield and quality traits in *Stevia rebaudiana* (Bertoni). Progress. Res. 3, 95–96.
Hans, J., Brandt, W., Vogt, T., 2004. Site-directed mutagenesis and protein 3D-homology modelling suggest a catalytic mechanism for UDP-glucose-dependent betanidin 5-O-glucosyltransferase from *Dorotheanthus bellidiformis*. Plant J. 39, 319–333. https://doi.org/10.1111/j.1365-313X.2004.02133.x.
He, X.-Z., Wang, X., Dixon, R.A., 2006. Mutational analysis of the medicago glycosyltransferase UGT71G1 reveals residues that control regioselectivity for (Iso)flavonoid glycosylation. J. Biol. Chem. 281, 34441–34447. https://doi.org/10.1074/jbc.M605767200.
Jones, P.R., Møller, B.L., Høj, P.B., 1999. The UDP-glucose:p-Hydroxymandelonitrile-O-glucosyltransferase that catalyzes the last step in synthesis of the cyanogenic glucoside Dhurrin in *Sorghum bicolor*. isolation, cloning, heterologous expression, and substrate specificity. J. Biol. Chem. 274, 35483–35491. https://doi.org/10.1074/jbc.274.50.35483.
Kubo, A., Arai, Y., Nagashima, S., Yoshikawa, T., 2004. Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. Arch. Biochem. Biophys. 429, 198–203. https://doi.org/10.1016/j.abb.2004.06.021.
Kumari, N., Raina, R., Sharma, Y.P., 2018. Interpopulation variations in morphochemical characteristics of *Stevia rebaudiana* Bertoni. Turk. J. Bot. 42. https://doi.org/10.3906/bot-1712-26.
Li, W., Zhou, Y., You, W., Yang, M., Ma, Y., Wang, M., Wang, Y., Yuan, S., Xiao, Y., 2018. Development of photoaffinity probe for the discovery of steviol glycosides biosynthesis pathway in *Stevia rebaudiana* and rapid substrate screening. ACS Chem. Biol. 13, 1944–1949. https://doi.org/10.1021/acschembio.8b00285.
Li, Y., Baldauf, S., Lim, E.-K., Bowles, D.J., 2001. Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. J. Biol. Chem. 276, 4338–4343. https://doi.org/10.1074/jbc.M007447200.
Mackenzie, P.I., Owens, I.S., Burchell, B., Bock, K.W., Bairoch, A., Belanger, A., Gigleux, S.F., Green, M., Hum, D.W., Iyanagi, T., et al., 1997. The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. Pharmacogenetics Genom. 7, 255. https://doi.org/10.1097/00008571-199708000-00001.
Madhav, H., Bhasker, S., Chinnamma, M., 2013. Functional and structural variation of uridine diphosphate glycosyltransferase (UGT) gene of *Stevia rebaudiana*–UGTSr involved in the synthesis of rebaudioside A. Plant Physiol. Biochem. 63, 245–253. https://doi.org/10.1016/j.plaphy.2012.11.029.
Nakamura, S., Tamura, Y., 1985. Variation in the main glycosides of stevia (*Stevia rebaudiana* Bertoni). Jpn. J. Trop. Agric. 29, 109–115. https://doi.org/10.11248/jsta1957.29.109.
Oddone, B., 1997. How to Grow Stevia. Technical Manual. Pawtucket, CT.
Osmani, S.A., Bak, S., Møller, B.L., 2009. Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochemistry 70, 325–347. https://doi.org/10.1016/j.phytochem.2008.12.009.
Phillips, K.C., 1989. Stevia: steps in developing a new sweetener. In: Developments in Sweeteners, Grenby, T.H., London, UK, pp. 1–43.
Prakash, I., Markosyan, A., Bunders, C., 2014. Development of next generation stevia sweetener: rebaudioside M. Foods 3, 162–175. https://doi.org/10.3390/foods3010162.
Richman, A., Swanson, A., Humphrey, T., Chapman, R., McGarvey, B., Pocs, R., Brandle, J., 2005. Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of *Stevia rebaudiana*. Plant J. 41, 56–67. https://doi.org/10.1111/j.1365-313X.2004.02275.x.
Richman, A.S., Gijzen, M., Starratt, A.N., Yang, Z., Brandle, J.E., 1999. Diterpene synthesis in *Stevia rebaudiana*: recruitment and up-regulation of key enzymes from the gibberellin biosynthetic pathway. Plant J. Cell Mol. Biol. 19, 411–421. https://doi.org/10.1046/j.1365-313X.1999.00531.x.
Ross, J., Li, Y., Lim, E., Bowles, D.J., 2001. Higher plant glycosyltransferases. Genome Biol. 2https://doi.org/10.1186/gb-2001-2-2-reviews3004. REVIEWS3004.
Soejarto, D.D., Kinghorn, A.D., Farnsworth, N.R., 1982. Potential sweetening agents of plant origin. III. Organoleptic evaluation of stevia leaf herbarium samples for sweetness. J. Nat. Prod. 45, 590–599. https://doi.org/10.1021/np50023a013.
Sumida, T., 1980. Studies on *Stevia rebaudiana* Bertoni as a new possible crop for sweetening resource in Japan. J. Cent. Agric. Exp. Stn. 30, 1–17.
Tanaka, O., 2009. Improvement of taste of natural sweeteners. Pure Appl. Chem. 69, 675–684. https://doi.org/10.1351/pac199769040675.
Vogt, T., Jones, P., 2000. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. Trends Plant Sci. 5, 380–386. https://doi.org/10.1016/S1360-1385(00)01720-9.
Wang, X., 2009. Structure, mechanism and engineering of plant natural product glycosyltransferases. FEBS Lett. 583, 3303–3309. https://doi.org/10.1016/j.febslet.2009.09.042.
Yadav, A.K., Singh, S., Rajeev, 2014. Self-incompatibility evidenced through scanning electron microscopy and pollination behaviour studies in *Stevia rebaudiana*. Indian J. Agric. Sci. 84, 93–100.
Yang, Y., Huang, S., Han, Y., Yuan, H., Gu, C., Zhao, Y., 2014. Base substitution mutations in uridinediphosphate-dependent glycosyltransferase 76G1 gene of *Stevia rebaudiana* causes the low levels of rebaudioside A: mutations in UGT76G1, A key gene of steviol glycosides synthesis. Plant Physiol. Biochem. 80, 220–225. https://doi.org/10.1016/j.plaphy.2014.04.005.