



# *Stevia rebaudiana* germplasm characterization using microsatellite markers and steviol glycosides quantification by HPLC

Maria Margarida Ribeiro<sup>1,2,3,4</sup> · Tatiana Diamantino<sup>1</sup> · Joana Domingues<sup>1</sup> · Ílio Montanari Jr<sup>5</sup> · Marcos Nopper Alves<sup>5</sup> · José Carlos Gonçalves<sup>1,2,3</sup>

Received: 4 January 2021 / Accepted: 19 March 2021  
© The Author(s) 2021

## Abstract

*Stevia rebaudiana* Bertoni, Asteraceae, is an herbaceous perennial plant native to Paraguay. This species is considered since ancient times a medicinal plant with important bioactive compounds and pharmacologic and food properties, namely diterpenes glycosides. The high natural sweetener potential stevioside and rebaudioside A produced by *S. rebaudiana* plants are suitable sucrose substitutes, and their obtention is influenced by environmental, phytosociological, and genetic factors. The plants' genetic profile and sweet potential depiction are needed for suitable plant selection for improvement and deployment. Thirty-one *S. rebaudiana* accessions grown in the same plot where leaves samples were collected in early 2019, were genotyped using six microsatellite markers, including two steviol glycosides biosynthesis functionally involved markers. Additionally, an aqueous extract of each sample was obtained in a water bath and purified by SPE for stevioside and rebaudioside A quantification by normal phase HPLC. Stevioside and rebaudioside A contents varied between 0.53–7.36% (w w<sup>-1</sup>) and 0.37–3.60% (w w<sup>-1</sup>), respectively. Two genotypes displayed interesting ratios of rebaudioside A/stevioside (number 3 and 33). The level of genetic similarity between genotypes was tested through a pairwise similarity coefficient, and two groups of individuals had the same fingerprinting. Strong relatedness was found within genotypes, possibly due to cloning, thus, influx of new germplasm ought to be made to prevent mating between relatives, and for further selection and genetic improvement.

**Keywords** Fingerprint · Genetic improvement · Germplasm · Stevioside · Molecular markers · Rebaudioside A

## Introduction

The genus *Stevia* consists of approximately 230 species [1, 2], and *Stevia rebaudiana* features the sweetest essence [3]. *Stevia* genus is distributed worldwide, ranging from the southern parts of the USA to Argentina and Brazil. Nowadays, *Stevia* cultivation has spread worldwide, including

Europe, propagated through seed and cuttings [2]. *S. rebaudiana* Bertoni (2n = 22) is a perennial and branched shrub of the Asteraceae family, native to north-eastern Paraguay [4, 5]. Since ancient times, this plant has been known as a sweetener due to the high content of diterpene glycosides [6], found in the leaves (present in a minor amount in shoots, roots, and flowers) in concentrations from 4 to 20% of dry

✉ Maria Margarida Ribeiro  
mataide@ipcb.pt

Tatiana Diamantino  
tatiana20@ipcb.pt

Joana Domingues  
joana.domingues@ubi.pt

Ílio Montanari Jr  
iliomontanarijr@gmail.com

Marcos Nopper Alves  
nopper@unicamp.br

José Carlos Gonçalves  
jgoncalves@ipcb.pt

<sup>1</sup> Centro de Biotecnologia de Plantas da Beira Interior, Escola Superior Agrária de Castelo Branco, 6001-909 Castelo Branco, Portugal

<sup>2</sup> Instituto Politécnico de Castelo Branco Escola Superior Agrária, 6001-909 Castelo Branco, Portugal

<sup>3</sup> Centro de Recursos Naturais Ambiente e Sociedade (CERNAS) - Instituto Politécnico de Castelo Branco, 6000-084 Castelo Branco, Portugal

<sup>4</sup> Forest Research Centre, School of Agriculture, University of Lisbon, Tapada da Ajuda, 1349-017 Lisbon, Portugal

<sup>5</sup> CPQBA/UNICAMP - Centro Pluridisciplinar de Pesquisas Químicas Biológicas e Agrícolas, Universidade Estadual de Campinas, São Paulo, Brazil

matter [5, 6]. Stevioside and rebaudioside A are the interesting steviol glycosides sugar-substitutes substances, which tend to accumulate in the aging tissues, thus, older leaves contain more sweeteners than younger ones. Conversely, at the beginning of flowering, the concentrations of glycosides in the leaves begin to decrease [1]. In general, stevioside is found in higher amounts than rebaudioside A, thus both substances' production is negatively correlated. Rebaudioside A is a better commercial product than stevioside, due to higher solubility in water and sweeter capacity, and the latter is responsible for aftertaste bitterness [1, 7]. According to Ceunen and Geuns [8], this ratio can be influenced by ontogeny and day length, with larger ratios obtained during the vegetative stage under short days. Rebaudioside A/stevioside ratio in leaves, together with the steviol-glycosides concentration and the leaf dry yield are significant economic parameters for stevia genotypes evaluation [9]. Moreover, in the former study, the ratio of rebaudioside A to stevioside revealed a significant effect of the genotype, in a trial conducted with two genotypes. An interesting approach to overcome the bitter aftertaste and improve the *Stevia* sweetness potential is the biotransformation of stevioside into rebaudioside A. Adari et al. [10] explained the stevioside enzymatic transglycosylation performed via the *Stevia* leaves pre-treatment with cellulose. The steviol-glycosides are currently used as sweet-tasting non-caloric food additives, due to obesity and diabetes rates rise worldwide [4, 7, 11]. Furthermore, displays therapeutic properties, such as antimicrobial, anti-inflammatory, antioxidant, antiviral, anti-jaundice, cardiogenic, and diuretic [7, 12]. The leaves and extracts can be taken raw or cooked, and added to various food products as sweeteners, due to the steviol-glycosides thermostability [13]. Steviol-glycosides were approved for use as a food additive in Europe represented by the code E960, in 2011 [14]. The existence of genetic variability in plant genetic resources provides plant breeders with the possibility to develop new and improved cultivars with desirable characteristics to include stakeholders-preferred traits (e.g., [15]).

Some genetic studies were reported in this species using markers with several limitations, such as RAPDs [16–18], AFLPs, and ISSRs [19–21], due to their dominance and to RAPDs reproducibility problems [22]. Dominance leads to recessive allele frequency underestimation biasing genetic diversity estimates [23, 24]. Thus, for fingerprinting purposes, microsatellite markers (SSRs) are the markers of choice, owing to their high polymorphism, multiallelism, codominance, high reproducibility, and uniform distribution in the genome [25–27]. These markers can competently be used in the genotyping process with a good genome sampling, which makes them suitable to estimate genotypes relatedness that requires multilocus genotype identification [28]. Kaur et al. [29] and Bhandawat et al. [30] were the pioneers in microsatellite markers development and use in

*S. rebaudiana* and, recently, from freely available expressed sequence data (ESTs), and Cosson et al. [31] also developed molecular markers (EST-SSRs) for population genetic and germplasm characterization studies. Recently, also a group of *S. rebaudiana* genotypes from Paraguay were studied with SSR and ISSR [32], and analysed as dominant markers. Additionally, molecular data can be pooled with biochemical data to detect stevia genotypes suitable to start a breeding program [18].

The use of microsatellite loci and the development of statistical tools made it possible inferring kin relationships from molecular data, in particular 'relatedness', meaning the estimation and assignment of pairs or groups of individuals to categories of relationship or a measure of the fraction of alleles shared identical by descent among individuals [33]. Indeed, to ensure that levels of co-ancestry and inbreeding among selected plants are kept to a minimum, it is advantageous to know their relatedness [34]. Additionally, besides avoiding related individuals (in particular for outcrossed species), it is important to accurately identify the genotypes and to distinguish the new cultivars for registration purposes. The germplasm fingerprint and cultivar identification with molecular markers-based grew in importance, due to generated data quality and speed, and also because morphological characteristics are influenced by environmental factors [22, 35–37].

In addition to the genetic and environmental factors that can influence the steviol-glycosides content, the analytical methodology selection is important to obtain an efficient protocol for plant bioactive compounds present extraction, purification, and quantification. The *Stevia* leaves steviol-glycosides were generally obtained subsequently to hot water application, followed sometimes by solid-phase extraction. However, extraction methods using other solvents, such as ethanol and methanol or even supercritical fluid extraction, were also described [38]. These solvents, namely methanol, appear to be used in the extraction process, presumably to improve extraction efficiency and facilitate the individual steviol-glycosides separation. However, regarding industry food safety, the use of water for extraction should be favoured over methanol or ethanol use [11]. Additionally, the water extraction seems the extraction method to be used in this case, given that rebaudioside A water solubility is higher compared to stevioside [11]. Several HPLC methodologies used to quantify the steviol-glycosides in *S. rebaudiana* leaves samples were reported in the literature [38 and references therein, 39]. Matrix load dramatically shortens chromatographic columns' lifetime, thus the analytical purification method, including solid phase extraction (SPE), must be applied [7, 38]. Finally, the HPLC methods was considered as the most reliable and simple method for the quantification of glycosides [40].

The genotypes collection of *S. rebaudiana* established in the Pluridisciplinary Center for Chemical, Biological and Agricultural Research, State University of Campinas, São Paulo, Brazil (CPQBA/UNICAMP) were studied for breeding purposes. The aims of the present study were: (i) to genotype 31 individuals with microsatellite markers to compute the genetic diversity, the fingerprint, and the relatedness, and (ii) to identify the genotypes that produce the highest rebaudioside A/stevioside ratio.

## Material and methods

### Plant material

We have studied 31 individuals from the *S. rebaudiana* germplasm plot established in the campus of the CPQBA/UNICAMP (Pluridisciplinary Center for Chemical, Biological and Agricultural Research, State University of Campinas, São Paulo, Brazil). The *S. rebaudiana* seeds were previously collected in the species' region of origin, but the exact location is unfortunately unknown. Four to five leaves sampled in each genotype from different parts of the plant were collected from January to February 2019. The genotypes were grown in the same plot with 1×0.5 m spacing, and 5 to 10 replications per genotype and the sampling was made just before flowering. Afterward, the leaves were lyophilized and frozen at -80 °C.

### DNA extraction and amplification

DNA was extracted from 100 mg of fresh leaves of each *S. rebaudiana* genotype, using the CTAB method as described by Doyle and Doyle [41], and a high concentration of pure genomic DNA was obtained. The genomic DNA from all individuals was amplified using six microsatellites (SSR)

markers selected from literature (Table 1). The two selected SUGMS (*Stevia* UniGene derived MicroSatellites) primers (Table 1) are putatively related to the steviol biosynthesis [30]. All amplifications were conducted separately for each primer pair and forward primers were 6-FAM fluorescently labelled. PCR reactions with the SUGMS primers (Table 1) were performed in 10 µL a total reaction volume, containing 50–60 ng of template DNA, 0.2 U Supreme NZYTaQ 2× Colourless Master Mix® separate MgCl<sub>2</sub> (Nzytech, Lisbon, Portugal), 2.5 mM MgCl<sub>2</sub>, and 1.0 µM of each primer. The amplifications were performed on a UNO96 Gradient thermocycler (VWR®, Leuven, Belgium). The PCR protocol consisted of an initial denaturation step of 4 min at 94 °C, followed by 35 amplification cycles composed of denaturation (1 min at 94 °C), annealing (1 min at optimal annealing temperature for each pair, see Table 1) and polymerizing (1 min at 72 °C). After the amplification cycles, a final extension step was performed for 7 min at 72 °C.

The PCR reactions with the 'gi' primers (Table 1) were performed in a 10 µL total reaction volume, containing 50–60 ng of template DNA, 0.8 U Supreme NZYTaQ 2× Colourless Master Mix® separate MgCl<sub>2</sub> (Nzytech, Lisbon, Portugal), 2.5 mM MgCl<sub>2</sub>, and 1.5 µM of each primer. The amplifications were performed in the same thermocycler, programmed with an initial denaturation step of 5 min at 95 °C, followed by 40 amplification cycles composed of denaturation (1 min at 95 °C), annealing (1 min at optimal annealing temperature for each pair, see Table 1) and polymerizing (2 min at 72 °C). The PCR reaction with the stvia036 primer was performed in a 10 µL total reaction volume, containing 50–60 ng of template DNA, 0.5 U Supreme NZYTaQ 2× Colourless Master Mix® separate MgCl<sub>2</sub> (Nzytech, Lisbon, Portugal), 2.5 mM MgCl<sub>2</sub>, and 2.0 µM of each primer (Table 1). The amplifications were performed in the same thermocycler, programmed with an initial denaturation step of 15 min at 95 °C, followed by 35 amplification cycles

**Table 1** Primers used for the SSRs analysis

Primer	Sequence 5'–3'	Size range (bp)	Annealing temp. (°C)	References
SUGMS28	F: CAAATTGGGAATTGCAGCTT R: GACAAACAAGCCGAGAGAGG	210–310	55	Bhandawat et al. [30]
SUGMS43	F: CCAATCTACAATTGCCACAAGA R: TTTTCCGAGGTTTTTGGTTG	225–255	55	Idem
gi18465444	F: ATGAAAGCGAGCCTGATGAT R: TCAAGCAACGATTCTTTCCA	100–610	56	Kaur et al. [29]
gi16949765	F: CAAGGCTTGCTCCGAAATAC R: TCATCTGCAAGTGCTTCCTC	680–900	56	Idem
gi18465673	F: CGGGTTAGAAGGAAACGTGA R: AAGTTCCACCAACCCATCA	500–800	56	Idem
stvia036	F: TGTCTCTGACAAAATTATACGG R: TTGTCTGTCCACCCTGTGG	144–180	55	Cosson et al. [31]

composed of denaturation (30 s at 94 °C), annealing (45 s at 55 °C, see Table 1) and polymerizing (1 min at 72 °C). After the amplification cycles, a final extension step was performed for 5 min at 72 °C. All PCR products were diluted in 50 µL of Milli-Q water and an aliquot of 3.0 µL of each dilution was mixed with 10 µL of formamide and 0.2 µL of LIZ-600 size standard. Genotyping was performed with an ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA), the fragment analysis was performed using GeneMapper 4.0 software (Applied Biosystems, Foster City, CA, USA), and the data manually scored.

### Aqueous extracts of *S. rebaudiana*

Lyophilized leaf samples were grounded with a mortar and pestle. Sample extracts were prepared by weighting 0.4 g of dried leaf powder and adding 8 mL of distilled water into 15 mL centrifuge tubes, afterward mixed in the vortex and placed in a 90 °C water bath for 30 min. Subsequently, the extracts were cooled at ambient temperature and centrifuged at 4400 rpm for 30 min at 4 °C (CLINIconic 75003623, Thermo Scientific, Massachusetts, EUA). From each centrifuged extract, the supernatant was transferred to a 25 mL volumetric flask. The pellet was reused for two more extractions and the supernatants transferred to the same volumetric flask. Finally, the volumetric flask was filled to 25 mL with distilled water.

### Stevioside and rebaudioside A extracts purification and HPLC quantification

Both steviol-glycosides were quantified according to the improved HPLC method described by Woelwer-Rieck et al. [38], with some modifications. The extracts were purified two times by solid-phase extraction (SPE) with the C18-E column (100 mg, 1 mL, 55 µm, 70 Å, Strada, Phenomenex, California, USA), with a constant vacuum pressure of 15 inHg. Columns were first conditioned with 3 mL of methanol and 3 mL of Milli-Q water and subsequently charged with 400 µL of extract, followed by washing with 3 mL of Milli-Q water and 5 mL of acetonitrile: Milli-Q water (20:80 v/v). Columns were left to dry for 3 min and then, steviol-glycosides were eluted with 2 mL of acetonitrile:Milli-Q water (75:25 v/v) and recovered in a 5 mL volumetric flask.

Seven stevioside (Cayman Chemical, Michigan, USA) and rebaudioside A (Extrasynthese, Genay, France) standard solutions were prepared in acetonitrile:Milli-Q water (75:25 v/v) in the concentration range of 5–800 µg mL<sup>-1</sup>. Both standard solutions and purified extracts were filtered with a 0.45 µm nylon filter (0.4 mm diameter, Technocroma, Barcelona, Spain) before loading in HPLC.

The linearity was assessed by injecting standard solutions of stevioside and rebaudioside A from 10 to 800 µg mL<sup>-1</sup>, using six different concentrations, and all standard solutions were injected in two different days. Standards were loaded in triplicate and each purified extract in duplicate into HPLC (Agilent 1100 series, with a quaternary pump, autosampler, diode array detector, degasser, thermostat, and data system), equipped with an NH<sub>2</sub> column (Purospher® STAR NH<sub>2</sub>, 250×4 mm, 5 µm) under the following conditions: 20 µL of injection volume, 210 nm wave-length, mobile phase: HPLC grade acetonitrile:Milli-Q water (75:25 v/v), 1 mL min<sup>-1</sup> constant flow rate, 20 min at 36 °C. The qualitative determination was achieved by comparing the stevioside and rebaudioside A standard solutions retention times with those of the samples. The quantification was possible by applying the computed calibration curves equations. The linear regression curves of stevioside and rebaudioside A are presented below were obtained (both adjusted the interception in 0):

$$\text{Stevioside: } Y = 4.7091 X \quad (r^2 = 0.9994)$$

$$\text{Rebaudioside A: } Y = 4.1668 X \quad (r^2 = 0.9998)$$

where Y is the peak area (mAU s<sup>-1</sup>) and X is concentration (µg mL<sup>-1</sup>) (unpublished results). The stevioside and the rebaudioside A concentration were represented in % w w<sup>-1</sup> (steviol glycoside/dry leave).

### Genetic data analysis

The genetic diversity parameters Na (number of alleles), Ne (effective number of alleles), Ho (observed heterozygosity), He (unbiased expected heterozygosity), and F (fixation index) were obtained with the GenAlEx software version 6.502 [42]. To test the genotypes pairwise relatedness the Lynch and Ritland [43] kinship coefficient was used (hereafter LR), and this coefficient was selected for it was considered the most accurate in a simulation with a set of relatedness coefficients validated with an elite group genetic similarity of *Eucalyptus globulus* with known pedigree [44]. The dendrogram (UPGMA method) was performed using the kinship matrix and the software NTSyS-PC version 2.11 [45]. The genetic relationship among genotypes was further analyzed by principal coordinate analysis (PCoA) using GenAlEx, based on the same pairwise relatedness matrix. The maximum deviation method was used for the factorial rotation and the main components were extracted, with the first main component and the second main component being the leading coordinates. The LR pairwise matrix, the genotypes multilocus comparison analysis, were computed in the GenAlEx.

## Results

### SSRs analysis

The analysis of the genetic diversity parameters (Table 2) revealed that the number of alleles per locus ( $N_a$ ) ranged from 3 to 6, averaging 4 per locus. The effective number of alleles per locus ( $N_e$ ) was 3.02, lower than the average  $N_a$ , but reflecting more accurately genetic diversity. Indeed, three loci displayed the same  $N_a$  (3) and two had higher  $N_e$  than the other one (gi16949765), also reflected in higher genetic diversity values: the SUGMS43 and Stevia36 loci (Table 2).

However, the loci with higher  $N_a$ , in general, had higher  $N_e$  and  $H_o$  values (gi18465444 and gi18465673), which

**Table 2** Genetic diversity parameters by locus for the 31 genotyped *S. rebaudiana* individuals

Locus	$N_a$	$N_e$	$H_o$	$H_e$	$F$
<b>gi18465444</b>	<b>5</b>	<b>4.06</b>	<b>0.84</b>	<b>0.77</b>	-0.11
gi16949765	3	1.60	0.45	0.38	-0.21
<b>gi18465673</b>	<b>6</b>	<b>3.59</b>	<b>0.87</b>	<b>0.73</b>	-0.21
SUGMS28	4	3.42	0.77	0.72	-0.09
SUGMS43	3	2.66	0.84	0.63	-0.34
Stvia36	3	2.81	0.81	0.65	-0.25
Mean	4	3.02	0.76	0.65	-0.20
SE		0.35	0.06	0.06	0.04

Values in bold correspond to primers with the highest genetic diversity parameters:  $N_a$ ,  $N_e$ ,  $H_o$  and  $H_e$

$N_a$  number of alleles,  $N_e$  effective number of alleles,  $H_o$  observed heterozygosity,  $H_e$  expected unbiased heterozygosity,  $F$  fixation index

were found in putative non-coding regions, values in bold in Table 2. However, the SUGMS43 locus, found in the coding region, had appreciable genetic diversity ( $H_o = 0.84$ ), though with only three alleles, yet with even frequencies (data not shown).

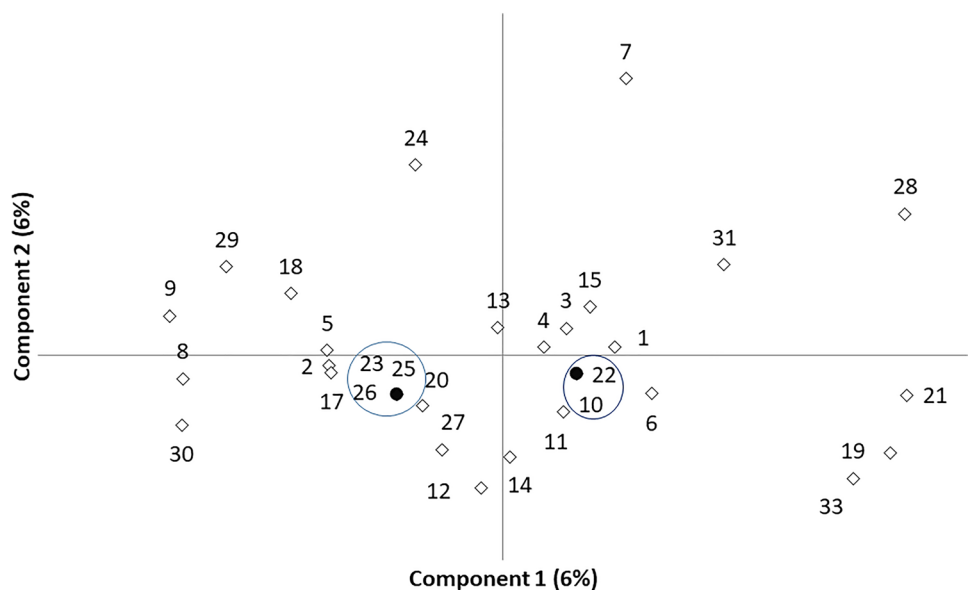
The total observed heterozygosity ( $H_o$ ) ranged from 0.45 to 0.87, with an average of 0.76. There was an excess of heterozygous in all loci since the observed is higher than that expected heterozygosity ( $H_o > H_e$ ), while a lower difference is found in the SUGMS28 locus (Table 2). However, the analyzed germplasm is a group of individuals unlikely in Hardy–Weinberg equilibrium.

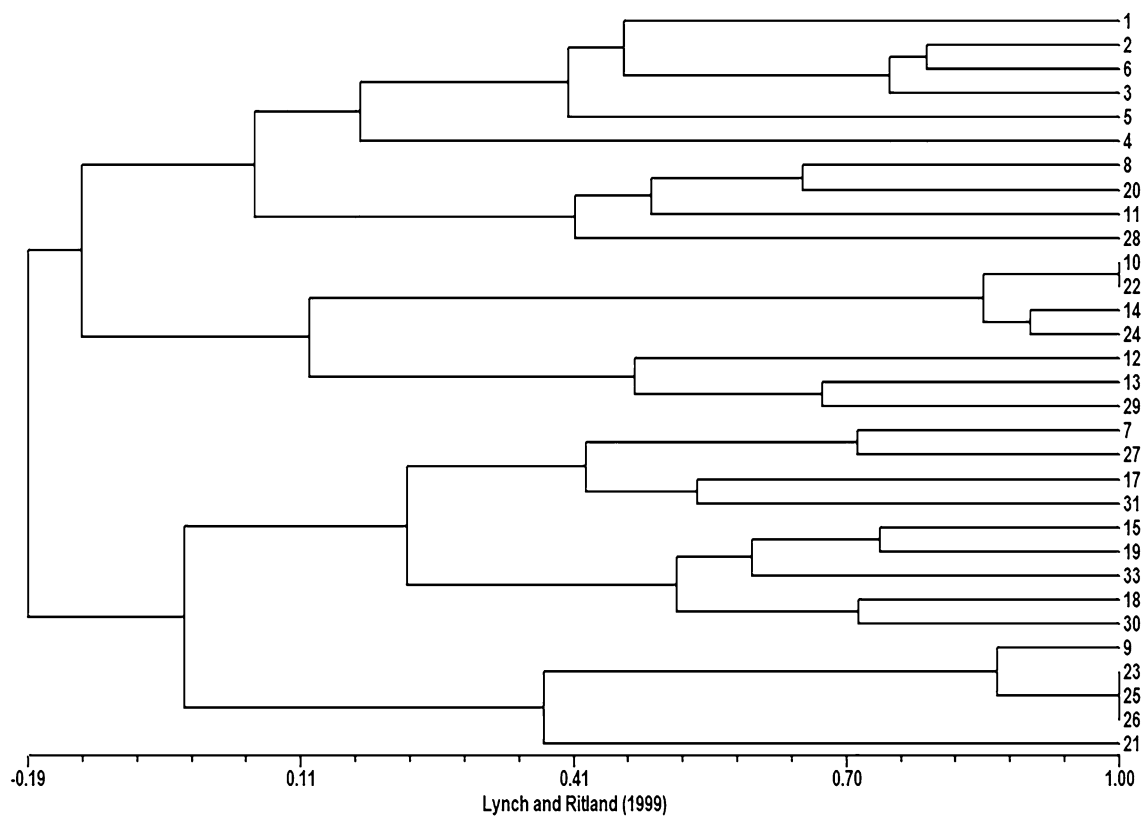
### Genetic distances and clustering

Genotypes were grouped according to the level of genetic similarity using the LR index. In the PCoA analysis, based on the kinship distance (LR), individuals with the same genotype plotted together (black circles, Fig. 1), but the remaining ones were, in general, well individualized, except in the case of genotypes 2, 5 and 17. Both the first and the second component explained  $\approx 6\%$  of the total variation and the total variation explained was 12%.

The dendrogram (UPGMA method) obtained with the kinship matrix (LR) (Fig. 2) reveals the relatedness among *S. rebaudiana* genotypes. This analysis separated the genotypes into two groups, with a very low degree of similarity (group 1: 4, 7, 9, 15, 17, 18, 19, 21, 23, 25, 26, 27, 30, 33, and 31; group 2: 1, 2, 3, 4, 5, 6, 8, 10, 11, 12, 13, 14, 20, 22, 24, 28 and 29). According to the literature for a value of LR coefficient  $> 0.25$  the genotypes may be considered half-sibs, resulting from cross-fertilization; for  $LR > 0.5$  the genotypes may be full-sibs, sharing both mother and father,

**Fig. 1** The principal component analysis was based on the LR pairwise kinship matrix. Black circles embodied individuals with the same genotype, all the others were represented by white diamonds. Both the first and the second components account for 6% of the total variation.





**Fig. 2** Accessions' dendrogram using the UPGMA clustering methods and the Lynch and Ritland [43] pairwise kinship matrix

for  $LR > 0.67$  they may be the result of self-pollination, and for  $LR = 1$  for the same genotype or clones [33, 44]. Indeed, two sets of individuals shared the same multilocus genotype (1st set: 10 and 22, and 2nd set: 23, 25, and 26), meaning that they share the same genotype, possibly a sampling/labeling error or clones. The fingerprinting of the two sets that displayed the same genotype are listed in Table S1.

### HPLC quantification of steviol-glycosides

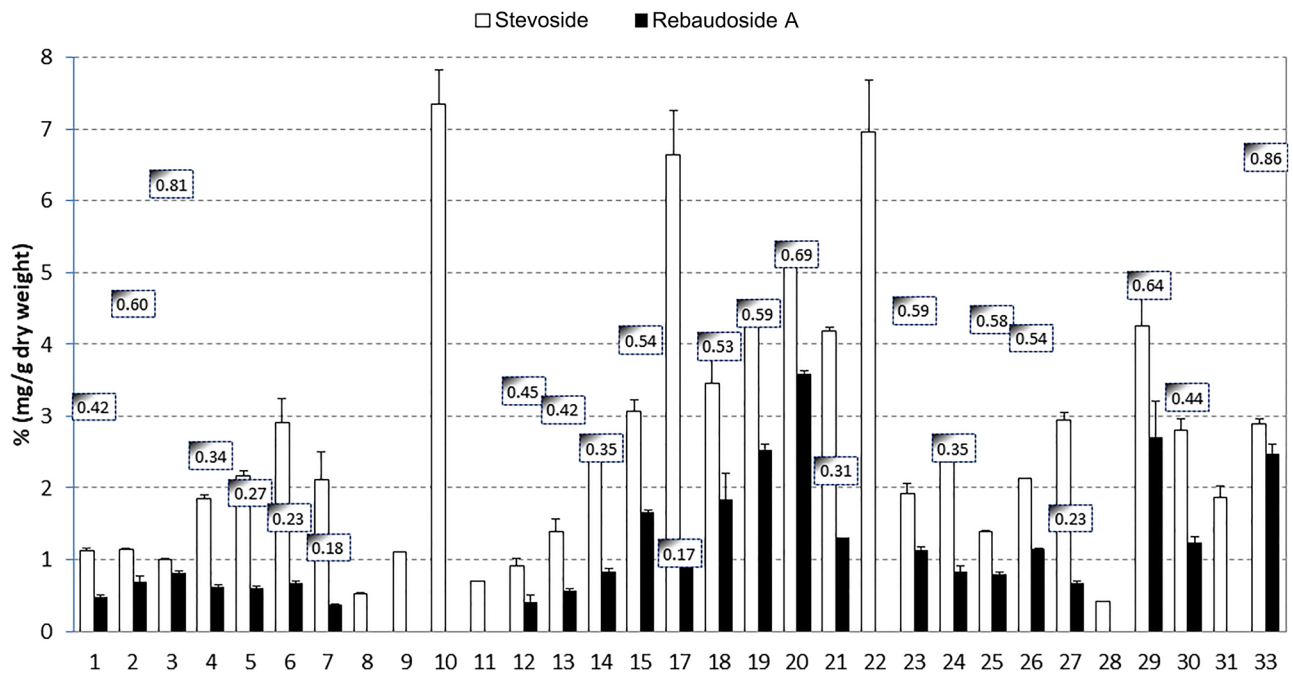
The contents of stevioside and rebaudioside A in the different genotypes were determined using calibration curves obtained by authentic standards. The linearity corresponds to the method's ability to provide results directly proportional to the analyzed compound concentration, within a given interval. The calibration curves were linear in the range of  $10\text{--}800\ \mu\text{g mL}^{-1}$  for both steviol glycosides standards and the linearity of the HPLC method was excellent ( $r^2 > 0.99$ ).

The stevioside and rebaudioside A retention time (RT) were  $5.28 \pm 0.05$  min and  $6.91 \pm 0.10$  min (mean  $\pm$  standard deviation) respectively, with the HPLC method. Figure 3 shows the stevioside and rebaudioside A mean concentrations (% w w<sup>-1</sup> dry leaf) for each sample and the respective rebaudioside A/stevioside ratio. The rebaudioside A-concentrations were lower in all the genotypes compared to the

stevioside ones. From the results, significant among sample differences existed in both steviol glycoside concentrations. Samples 10, 17, and 22 had the highest stevioside content (6.64–7.36%), while samples 19, 20, and 29 had the highest rebaudioside A value (2.53–3.60%). Samples 8, 11, and 28 showed a lower concentration of both steviol glycosides compared to the other samples. Several samples revealed undetectable rebaudioside A content (8, 9, 10, 11, 22, 28, and 31), considering the HPLC conditions linearity method. The rebaudioside A/stevioside ratio should be as high as possible, given the rebaudioside A better sweetening properties than stevioside, as previously referred. The ratio ranged from 0.17 to 0.86, and the highest rebaudioside A/stevioside ratio was observed in samples 3 and 33, rating 0.81 and 0.86, respectively (Fig. 3). Decreasing ratio values were found in samples 2, 20, and 29, with 0.60, 0.69, and 0.64, respectively.

### Discussion

The *S. rebaudiana* is undergoing domestication [31], notwithstanding that the species was used as a sweetener as earlier as the 1940s, and cultivated worldwide [1 and references therein]. Nevertheless, few studies reported both the



**Fig. 3** Stevioside (white bars) and rebaudioside A (black bars) content (% w w<sup>-1</sup> dry leaf). White labels show the *S. rebaudiana* genotypes rebaudioside A/stevioside ratio. Two genotypes had the highest ratio level (3 and 33).

species genetic information and the steviol glycosides content [31 and references therein], and no study stated both the genotypes' genetic relatedness and their fingerprint. In this study, microsatellite markers were used to reveal the genetic diversity and the genetic relationship among the 31 individuals analyzed and to further detect highly similar genotypes. In the study made by Bhandawat et al. [30] the authors used 40 selected genotypes with a set of SUGMS markers and they found that the observed diversity was significantly higher than the expected one (0.80 and 0.63, on average, respectively), similarly to the results from the current study (0.76 and 0.65). In both studies, the expected diversity was slightly lower than the observed one, resulting in a negative fixation index. But the plants used in the two studies had different origins making a genotype collection improbably in Hardy–Weinberg equilibrium. Possibly this could result in a “melting pot”, meaning a mixture of individuals from different origins, and putatively with very different allelic compositions [46]. Simultaneously, the average  $H_e$  value (0.78) obtained by Cosson et al. [31], was higher than the value from the current study, but the authors surveyed a much larger germplasm set (145 genotypes) with a wider origin range and a high number of markers. In a meta-study surveying species with different mating systems and using nuclear microsatellite markers to evaluate genetic diversity, the observed and expected estimates were, on average, 0.65 and 0.63, respectively [23], for the group of outcrossed species. The high level of genetic variability obtained in

*S. rebaudiana* is most likely due to allogamy in this species mating system [31]. Moreover, the use of an artificial population constituted by individuals from different origins could explain a high value of observed genetic diversity in our study.

Estimators are useful to identify relatives and to minimize consanguinity in breeding populations in the absence of pedigree information [44]. Different relatedness estimators exist in the literature and they respond differently to the available sample [47]. The Lynch and Ritland [43] coefficient (LR) was selected as the appropriate kinship estimator in this study, as it proved to be impartial, accurate, with a low percentage of overlapping values between kinship groups and low percentiles of exact confidence [44]. The genotypes were initially divided into two groups and distributed according to the level of similarity in the dendrogram. Considering the values of LR, two sets of individuals had the same genotype, possibly for they were the same individual due to cloning (LR = 1.0). Vegetative propagation was used to maintain the collection, and mislabeling is possible in these cases (Ílio Montanari Jr., personal communication). In the other cases, the LR values grouping the genotypes was very high, detecting appreciable relatedness in the group of individuals studied. Indeed, cultivar mislabelling is possible [36], and the rapid development of DNA-based molecular markers are helpful to overcome this problem.

The steviol-glycosides content and variability determination are important for the plant producers and industries to

obtain performant cultivars through selection and breeding. Therefore, all individuals from the current study were grown in the same conditions and collected just before flowering, when steviol-glycosides content is expected to be highest. The steviol-glycosides concentrations in the current study varied between 0.53–7.36% (w w<sup>-1</sup>) for stevioside, and 0.37–3.60% (w w<sup>-1</sup>) for rebaudioside A. In some samples, the rebaudioside A remained undetected, probably due to either steviol-glycoside production being lower than needed for quantification by the linearity HPLC conditions or to the genotype lack of production. In general, the steviol-glycosides concentrations observed in the thirty-one samples are following the values found in the literature, with stevioside content always higher than the rebaudioside A content [7, 11, 40]. Woelwer-Rieck et al. [38] reported that plants growing in a different cultivated field also showed a higher stevioside content than rebaudioside A values, with mean values of 7.90% (w w<sup>-1</sup>) and 4.93% (w w<sup>-1</sup>), respectively. Similar steviol-glycosides values were also obtained in another study, by water extraction, with stevioside values ranging between 3.84 and 9.36% (w w<sup>-1</sup>), and the rebaudioside A values ranging between 2.59 and 7.77% (w w<sup>-1</sup>) [48].

Despite the glycosides great solubility in water, some studies reported extraction with other solvents (e.g., [11]). Methanolic extracts had a range 4.10–8.20% (w w<sup>-1</sup>) and 0.60–4.44% (w w<sup>-1</sup>) of stevioside and rebaudioside A, respectively [39]. Analogous values have been reported in hydroethanolic extracts with a range of stevioside and rebaudioside A concentration between 3.78–9.84% (w w<sup>-1</sup>) and 1.62–7.27% (w w<sup>-1</sup>), respectively [9]. Regardless of the steviol-glycosides absolute quantity, the quality is defined by a high rebaudioside A/stevioside ratio. Indeed, the rebaudioside A/stevioside ratios obtained in the current study are under the numbers reported in the literature, with values ranging from 0.34 to 0.77 [9, 38]. Actually, in this study two genotypes showed interesting levels of rebaudioside A/stevioside ratio, the genotypes 3 and 33, with 0.81 and 0.86, respectively. In wild plants, this ratio is less than 0.5 [1]. Besides, those two genotypes were found in the two different clusters of the dendrogram, with very low LR, and they are unrelated individuals.

Comparing the genetics with the chemical results, it is worth noting that individuals 8, 11, and 28 revealed close genetic relatedness and similar chemical profile, with low steviol-glycosides concentration. Individuals 10 and 22 had displayed the same genetic fingerprint and the same chemical profile, both in stevioside concentration and absence or very low rebaudioside A synthesis. Individuals 23, 25, and 26 had the same fingerprint, and they also revealed a very similar chemical profile, being probably the same genotype/clone. Indeed, since the late 1980s, DNA fingerprinting has

become an immensely important instrument for genotype identification in wild plant and cultivated species. Additionally, when the morphological characters are mostly quantitative, correspondence with DNA marker estimates is generally quite high as compared to qualitative characters, which are more likely to reflect only a small number of mutation events [37]. But, due to this group of individuals' high relatedness, a new influx of unrelated genotypes should be included in this collection to further improve breeding and selection and avoiding relatedness. Additionally, other studies should be undertaken to reveal, e. g., genotypes' leaves biomass production and other plant characteristics that would increase steviol-glycosides production (see Ref. [1] for details).

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11033-021-06308-x>.

**Author contributions** JD, MMR and MNA designed experiments; TD and JD performed the experiments; MMR, JD and TD analyzed the data and wrote the original draft paper; MMR and MNA supervised the study; JCG and MNA were involved in funding acquisition; JD, MMR, TD, JCG, MNA and IM Jr. revised and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding** This research was funded by the Forest Research Centre, a research unit funded by Fundação para a Ciência e a Tecnologia I.P. (FCT), Portugal (UIDB/00239/2020), and by the research unit CER-NAS-IPCB also funded by FCT (UIDB/00681/2020).

**Data availability** All the data and material will be available upon request to the authors.

## Declarations

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

**Ethical approval** Sampling of *Stevia* samples was non-destructive. No specific permission was required to sample *Stevia* species in Portugal and Brazil. Field studies did not involve endangered or protected species. No humans or animals were involved in the experiments.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Yadav AK, Singh S, Dhyani D, Ahuja PS (2011) A review on the improvement of *Stevia* [*Stevia rebaudiana* (Bertoni)]. *Can J Plant Sci* 91(1):1–27
- Gantait S, Das A, Banerjee J (2018) Geographical distribution, botanical description and self-incompatibility mechanism of genus *Stevia*. *Sugar Tech* 20(1):1–10. <https://doi.org/10.1007/s12355-017-0563-1>
- Soejarto DD, Compadre CM, Medon PJ, Kamath SK, Kinghorn AD (1983) Potential sweetening agents of plant origin. II. Field search for sweet-tasting *Stevia* species. *Econ Bot* 37(1):71–79. <https://doi.org/10.1007/bf02859308>
- Soejarto DD, Addo EM, Kinghorn AD (2019) Highly sweet compounds of plant origin: from ethnobotanical observations to wide utilization. *J Ethnopharmacol* 243:112056. <https://doi.org/10.1016/j.jep.2019.112056>
- 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(4):685–694. <https://doi.org/10.1007/s10722-015-0276-9>
- Ghanta S, Banerjee A, Poddar A, Chattopadhyay S (2007) Oxidative DNA damage preventive activity and antioxidant potential of *Stevia rebaudiana* (Bertoni) Bertoni, a natural sweetener. *J Agric Food Chem* 55(26):10962–10967. <https://doi.org/10.1021/jf071892q>
- Lemus-Mondaca R, Vega-Gálvez A, Zura-Bravo L, Ah-Hen K (2012) *Stevia rebaudiana* Bertoni, source of a high-potency natural sweetener: a comprehensive review on the biochemical, nutritional and functional aspects. *Food Chem* 132(3):1121–1132
- Ceunen S, Geuns JMC (2013) Influence of photoperiodism on the spatio-temporal accumulation of steviol glycosides in *Stevia rebaudiana* (Bertoni). *Plant Sci* 198:72–82. <https://doi.org/10.1016/j.plantsci.2012.10.003>
- Munz S, Präger A, Merkt N, Claupein W, Graeff-Hönninger S (2018) Leaf area index, light interception, growth and steviol glycoside formation of *Stevia rebaudiana* Bertoni under field conditions in southwestern Germany. *Ind Crops Prod* 111:520–528. <https://doi.org/10.1016/j.indcrop.2017.11.021>
- Adari BR, Alavala S, George SA, Meshram HM, Tiwari AK, Sarma AVS (2016) Synthesis of rebaudioside-A by enzymatic transglycosylation of stevioside present in the leaves of *Stevia rebaudiana* Bertoni. *Food Chem* 200:154–158. <https://doi.org/10.1016/j.foodchem.2016.01.033>
- Abou-Arab AE, Abou-Arab AA, Abu-Salem MF (2010) Physico-chemical assessment of natural sweeteners steviosides produced from *Stevia rebaudiana* Bertoni plant. *Afr J Food Sci* 4(5):269–281
- Chen T-H, Chen S-C, Chan P, Chu Y-L, Yang H-Y, Cheng J-T (2005) Mechanism of the hypoglycemic effect of stevioside, a glycoside of *Stevia rebaudiana*. *Planta Med* 71(02):108–113. <https://doi.org/10.1055/s-2005-837775>
- Serio L (2010) La *Stevia rebaudiana*, une alternative au sucre. *Phytothérapie* 8(1):26–32. <https://doi.org/10.1007/s10298-010-0526-4>
- EU (2011) Regulation No 1131 Commission Regulation 11 November amending Annex II to Regulation (EC) No 1333/2008 of the European Parliament and of the Council with regards to steviol glycosides, pp 205–211
- Guan C, Zhang Y, Zhang P, Chachar S, Wang R, Du X, Yang Y (2020) Germplasm conservation, molecular identity and morphological characterization of persimmon (*Diospyros kaki* Thunb.) in the NFGP of China. *Sci Hortic* 272:109490. <https://doi.org/10.1016/j.scienta.2020.109490>
- Yao Y, Ban M, Brandle J (1999) A genetic linkage map for *Stevia rebaudiana*. *Genome* 42(4):657–661. <https://doi.org/10.1139/g98-161>
- Chester K, Tamboli ET, Parveen R, Ahmad S (2013) Genetic and metabolic diversity in *Stevia rebaudiana* using RAPD and HPTLC analysis. *Pharm Biol* 51(6):771–777. <https://doi.org/10.3109/13880209.2013.765898>
- Dyduch-Siemińska M, Najda A, Gawroński J, Balant S, Świca K, Żaba A (2020) *Stevia rebaudiana* Bertoni, a source of high-potency natural sweetener—biochemical and genetic characterization. *Molecules* 25:767. <https://doi.org/10.3390/molecules25040767>
- Heikal AH, Badawy O, Hafez A (2008) Genetic relationships among some *Stevia* (*Stevia rebaudiana* Bertoni) accessions based on ISSR analysis. *Res J Cell Mol Biol* 2:1–5
- Othman H, Osman M, Zainuddin Z (2018) Genetic variabilities of *Stevia rebaudiana* Bertoni cultivated in Malaysia as revealed by morphological, chemical and molecular characterisations. *AJAS*. <https://doi.org/10.17503/agrivita.v40i2.1365>
- Luz GC, Strioto DK, Mangolin CA, Machado M (2020) ISSR markers to assess genetic diversity of cultivated populations from artificial selection of *Stevia rebaudiana* (Bert.) Bertoni. *Breed Sci* 70(4):508–514. <https://doi.org/10.1270/jsbbs.20014>
- Ahmad R, Anjum MA, Balal RM (2020) From markers to genome based breeding in horticultural crops: an overview. *Phyton Int J Exp Bot* 89(2):183–204. <https://doi.org/10.32604/phyton.2020.08537>
- Nyblom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol Ecol* 13(5):1143–1155
- Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Mol Ecol* 3:91–99
- Hodel RGJ, Segovia-Salcedo MC, Landis JB, Crowl AA, Sun M, Liu X, Gitzendanner MA, Douglas NA, Germain-Aubrey CC, Chen S, Soltis DE, Soltis PS (2016) The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. *APPS* 4(6):1600025. <https://doi.org/10.3732/apps.1600025>
- Goldstein DB, Schlötterer C (Eds) (1998) *Microsatellites. Evolution and applications*, Oxford University Press, New York
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotechnol* 23(1):48–55. <https://doi.org/10.1016/j.tibtech.2004.11.005>
- Selkoe KA, Toonen RJ (2006) *Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers*. *Ecol Lett* 9(5):615–629. <https://doi.org/10.1111/j.1461-0248.2006.00889.x>
- Kaur R, Sharma N, Raina R (2015) Identification and functional annotation of expressed sequence tags based SSR markers of *Stevia rebaudiana*. *Turk J Agric For* 39(3):439–450. <https://doi.org/10.3906/tar-1406-144>
- Bhandawat A, Sharma H, Nag A, Singh S, Singh Ahuja P, Sharma RK (2015) Functionally relevant novel microsatellite markers for efficient genotyping in *Stevia rebaudiana* Bertoni. *J Genet* 94(1):75–81. <https://doi.org/10.1007/s12041-014-0406-8>
- Cosson P, Hastoy C, Errazzu LE, Budeguer CJ, Boutie P, Rolin D, Schurdi-Levraud V (2019) Genetic diversity and population structure of the sweet leaf herb, *Stevia rebaudiana* B., cultivated and landraces germplasm assessed by EST-SSRs genotyping and steviol glycosides phenotyping. *BMC Plant Biol*. <https://doi.org/10.1186/s12870-019-2061-y>
- Bogado-Villalba L, Nakayama Nakashima H, Britos R, Iehisa JCM, Flores Giubi ME (2020) Genotypic characterization and steviol glycoside quantification in a population of *Stevia rebaudiana*

- Bertoni from Paraguay. *J Crop Sci Biotechnol*. <https://doi.org/10.1007/s12892-020-00066-1>
33. Blouin MS (2003) DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *TREE* 18(10):503–511
  34. Ballou JD, Lacy RC (1995) Identifying genetically important individuals for management of genetic variation in pedigreed populations. In: Ballou JD, Gilpin M, Foose TJ (eds) *Population management for survival and recovery. Analytical methods and strategies in small population conservation*. Columbia University Press, New York, pp 76–111
  35. Govindaraj M, Vetriventhan M, Srinivasan M (2015) Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet Res Int* 2015:431487. <https://doi.org/10.1155/2015/431487>
  36. Scarano D, Rao R, Masi P, Corrado G (2015) SSR fingerprint reveals mislabeling in commercial processed tomato products. *Food Control* 51:397–401. <https://doi.org/10.1016/j.foodcont.2014.12.006>
  37. Nybom H, Weising K, Rotter B (2014) DNA fingerprinting in botany: past, present, future. *Investig Genet* 5(1):1
  38. Woelwer-Rieck U, Lankes C, Wawrzun A, Wüst M (2010) Improved HPLC method for the evaluation of the major steviol glycosides in leaves of *Stevia rebaudiana*. *Eur Food Res Technol* 231(4):581–588
  39. Gardana C, Scaglianti M, Simonetti P (2010) Evaluation of steviol and its glycosides in *Stevia rebaudiana* leaves and commercial sweetener by ultra-high-performance liquid chromatography-mass spectrometry. *J Chromatogr* 1217(9):1463–1470. <https://doi.org/10.1016/j.chroma.2009.12.036>
  40. Kovylyayeva G, Bakaleinik G, Strobykina IY, Gubskaya V, Sharipova R, Al'Fonsov V, Kataev V, Tolstikov A (2007) Glycosides from *Stevia rebaudiana*. *Chem Nat Compd* 43(1):81–85
  41. Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
  42. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28(19):2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
  43. Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics* 152(4):1753–1766
  44. Ribeiro MM, Sanchez L, Ribeiro C, Cunha F, Araújo J, Borralho N, Marques C (2011) A case study of *Eucalyptus globulus* fingerprinting for breeding. *Ann For Sci* 68(4):701–714. <https://doi.org/10.1007/s13595-011-0087-x>
  45. Rohlf FJ (2000) *NTSYS-pc: numerical taxonomy and multivariate analysis system, version 2.1*. Exeter Publishing, New York
  46. Bucci G, Gonzalez-Martinez SC, Le Provost G, Plomion C, Ribeiro MM, Sebastiani F, Alia R, Vendramin GG (2007) Range-wide phylogeography and gene zones in *Pinus pinaster* Ait. revealed by chloroplast microsatellite markers. *Mol Ecol* 16(10):2137–2153. <https://doi.org/10.1111/j.1365-294X.2007.03275.x>
  47. Milligan BG (2003) Maximum-likelihood estimation of relatedness. *Genetics* 163(3):1153–1167
  48. Kaplan B, Turgut K (2019) Improvement of rebaudioside A diterpene glycoside content in *Stevia rebaudiana* Bertoni using clone selection. *Turk J Agric For* 43:232–240

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.