


RESEARCH ARTICLE

Open Access



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

Patrick Cosson¹, Cécile Hastoy^{1,2}, Luis Ernesto Errazu³, Carlos Jorge Budeguer⁴, Philippe Boutié², Dominique Rolin¹ and Valérie Schurdi-Levraud^{1*} 

Abstract

Background: *Stevia rebaudiana* (Asteraceae), native from Paraguay, accumulates steviol glycosides (SGs) into its leaves. These compounds exhibit acaloric intense sweet taste which answers to consumer demands for reducing daily sugar intake. Despite the development of *S. rebaudiana* cultivation all over the world, the development of new cultivars is very recent, in particular due to a colossal lack of (1) germplasm collection and breeding, (2) studies on genetic diversity and its structuring, (3) genomic tools.

Results: In this study, we developed 18 EST-SSR from 150,258 EST from The Compositae Genome Project of UC Davis (<http://compgenomics.ucdavis.edu/data/>). We genotyped 145 *S. rebaudiana* individuals, issued from thirty-one cultivars and thirty-one landraces of various origins worldwide. Markers polymorphic information content (PIC) ranged between 0.60 and 0.84. An average of 12 alleles per locus and a high observed heterozygosity of 0.69 could be observed. The landraces revealed twice as many private alleles as cultivars. The genotypes could be clustered into 3 genetic populations. The landraces were grouped in the same cluster in which the oldest cultivars “Eirete” and “Moritalll” type are also found. The other two clusters only include cultivated genotypes. One of them revealed an original genetic variability. SG phenotypes could not discriminate the three genetic clusters but phenotyping showed a wide range of composition in terms of bitter to sweet SGs.

Conclusion: This is the first study of genetic diversity in *Stevia rebaudiana* involving 145 genotypes, including known cultivars as well as landrace populations of different origin. This study pointed out the structuration of *S. rebaudiana* germplasm and the resource of the landrace populations for genetic improvement, even on the trait of SG's composition.

Keywords: *Stevia rebaudiana*, Genetic diversity, Cultivars, Landraces, Steviol glycosides

* Correspondence: valerie.schurdi-levraud@inra.fr

¹UMR Biologie du Fruit et Pathologie, 1332, INRA Université de Bordeaux, 71 avenue Edouard Bourlaux, 33883 Villenave d'Ornon cedex, France
Full list of author information is available at the end of the article



Background

Stevia rebaudiana, native from Paraguay, is a perennial species that accumulates steviol glycosides (SGs) into its leaves. These natural compounds exhibit acaloric intense sweet taste. Leaves of *S. rebaudiana* were firstly used as a general sweetening agent by native people of Paraguay and Brazil [1]. Consumer demands for reducing daily sugar intake and for healthy products place *S. rebaudiana* production at the crossroad of these needs. Correlated to growing demand, *S. rebaudiana*'s cultivation is increasing all over the world. The species can be grown in a wide range of climatic areas. Nevertheless, the cultivation of *S. rebaudiana* as a crop is very recent and is done on a small scale in the countries of origin. In 1964, its first commercial cultivation was reported in Paraguay (Katayama et al. 1976; Lewis 1992). Afterwards, great efforts were made by Sumida in 1971 to establish *S. rebaudiana* cultivation in Japan (Crammer & Ikan, 1986). Later on, it was introduced as a crop in many countries. In 2016, 80% of world stevia leaves production was coming from China with 50,000–60,000 tons of dry leaves per year [2]. Other significant producing countries are located in Asia (Indonesia, India, Japan, Korea) and in America (Mexico, USA, and Canada) [3]. Recent regulatory approval explains the newly beginning area of production in Europe [4].

This recently developed crop suffers from a lack of high-value adapted and traceable cultivars. Ninety cultivars have been recently listed by Angelini et al. in 2016 [5]. But, farmers' practices show that most of the cultivars that are produced are related to cultivars known as "Eirete", "Criolla" and "Morita" type. They are sold as seeds, often through open-pollinated production as the species is self-incompatible [6]. These genotypes have been mainly bred through mass selection. Besides these currently grown population cultivars, some others are patented as those from the US S&W company [7] or the Malaysian PureCircle. These genotypes are mainly improved for SG's yield and SG's composition producing the sweetest taste.

However numerous other traits such as seed germination rate, flowering date, aerial biomass yield, response to biotic and abiotic stresses are poorly improved. Then, identification of productive genetic resources and breeding is clearly needed.

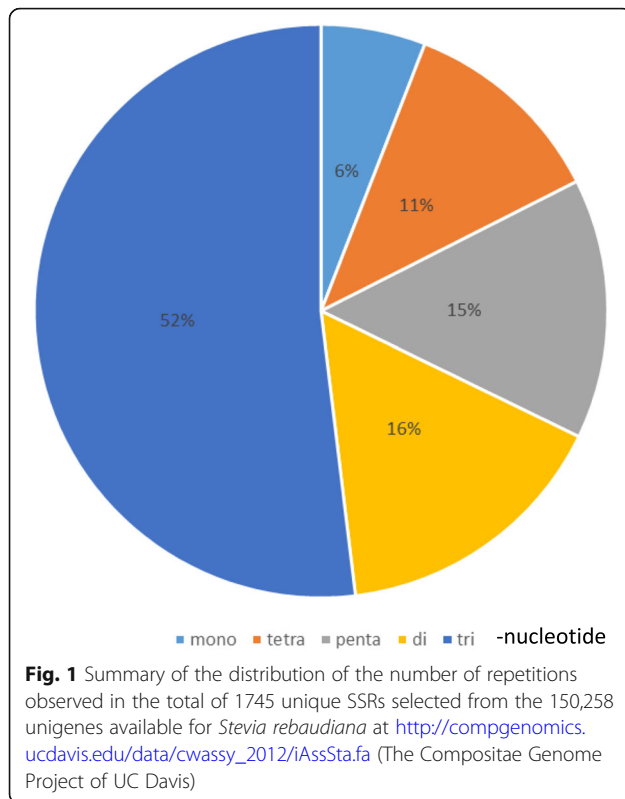
Collecting and studying germplasm is a key pillar to start a breeding program. Diversity in plant genetic resources provides opportunity for plant breeders to develop new and improved cultivars with desirable characteristics and is also fundamental to limit the genetic erosion. No available public collection exists and only private companies own collections that are not available. The phenotypic study of a limited number of genotypes have been conducted in different countries [8–10]. Since many years, the

use of molecular markers has improved significantly the management and utilization of crop genetic diversity [11]. In *Stevia*, until recently, the lack of genomic information has made genotyping a bottleneck. Some molecular markers as RAPD [12] and ISSR [13] have been used in previous studies to analyze the diversity in very small collection of individuals. RAPD have also been used to construct the first *S. rebaudiana* genetic map in 1999 [14]. Due to their lack of repeatability these markers were abandoned. Despite the huge development of crop sequencing in the recent years, *S. rebaudiana* has just been sequenced in 2017 by the private consortium PureCircle/Coca-Cola/Keygene [15]. The consortium declared to have sequenced one genotype but sequences have not been released yet. The lack of information on SNP markers confirmed the need to develop SSR markers. For some orphan species as *S. rebaudiana* and for genetic populations analysis SSR remain first choice markers [16, 17]. They still have great applicability due to their high polymorphism, relatively easy scoring, testable neutrality, and Mendelian inheritance. Kaur et al. and Bhandawat et al. [18, 19] developed a set of 52 and 17 SSR markers respectively through the screening of 5548 stevia ESTs sequences from leaf tissues retrieved from the NCBI. These markers were used to classify forty randomly chosen genotypes from selection at CSIR Institute (India) or 12 local genotypes from Northern India. These studies were pioneer in the development and use of SSR markers in *S. rebaudiana*. In 2013, the Compositae Genome Project released EST data from 15 reference transcriptome assemblies for Compositae crops or their wild relatives including *S. rebaudiana* [20]. This large amount of sequences allowed to screen for SSR patterns and develop EST-SSR markers. To our knowledge, in *S. rebaudiana* no studies addressed the question of developing molecular markers to classify germplasm and to question population structure. The purpose of this work was 1) to develop and assess the applicability of EST-SSRs developed for *S. rebaudiana* as markers in thirty-one landraces genotypes and one hundred and fourteen cultivated genotypes issued from thirty-one cultivars; 2) to identify genetic diversity and population structure in *S. rebaudiana*'s landraces and cultivars and 3) to check the link between genetic variability and its structure and phenotypic SG variability in cultivated and landraces populations.

Results

EST-SSR genotyping

Of the 150,258 unigenes, 3401 SSR pattern could be detected, 1745 being unique. These 1745 SSR are divided into 6% of mono-nucleotide, 16% of di-nucleotide, 52% of tri-nucleotide, 11% of tetra-nucleotide and 15% of penta-nucleotide (Fig. 1 and Additional file 2: Table S1). Among these 1745, 1060 were considered suitable for



primer design (Table 1). More than 60% are tri-nucleotide repeat type. Ninety-four with 10 to 26 repeats were selected for primer design (Additional file 1: Figure S1). Screened with 5 genotypes, 19 did not produce any amplification, 5 generated PCR products over 300 bp size and 17 monomorphic or multiloci profiles. Fifty-three primer pairs amplified a polymorphic fragment. Eighteen were selected for population studies, based on the multiplexing possibility (Additional file 2: Table S2).

Table 1 Information summary on the number and percentage of each SSR selected through the pipeline Additional file 1: Figure S1

Searching items	Numbers	
Total number of unigene	150,258	
Total number of SSR detected	3401	
Total number of unique SSR (Blast <80%)	1745	
Total number of SSR suitable for primer design ^a	1060	
Repeat type	Numbers	Percentage
mono-nucleotide	56	5.28
di-nucleotide	107	10.09
tri-nucleotide	655	61.79
tetra-nucleotide	108	10.19
penta-nucleotide	134	12.64
Total	1060	100.00

^a minimum 100 pb each side

Genetic variability

A panel consisting of 145 individual plants (Table 2 and Additional file 2: Table S3) was genotyped with the 18 selected markers. A total of 213 alleles were detected in the 145 individual plants analyzed (Table 3). The number of alleles detected per locus ranged between 5 and 19 (stvia021; stvia048 respectively) with an average of 12 alleles per locus. Markers' PICs ranged between 0.60 (stvia021) and 0.84 (stvia025 and stvia051). The average observed and expected heterozygosity across markers were $H_o = 0.69$ (min-max: 0.46 (stv018)-0.87 (Stv036)) and $H_e = 0.78$ (min-max: 0.66-0.86). The allelic richness calculated as the mean number of private alleles ranged from 1.5 for cultivated genotypes to 3 for landraces (Additional file 1: Figure S2).

Population structure

The change rate in the log-likelihood between successive K values (DK) inferred with STRUCTURE revealed three clusters with a relatively high ΔK value at $K = 3$ (Fig. 2a). We used STRUCTURE membership coefficients inferred at $K = 3$ to define the populations used in subsequent analyses. For analyses hereafter, genotypes were assigned to a given population if their membership coefficient for that population was ≥ 0.80 . Thus, genotypes with an identity value under 80% probability of belonging to a given subpopulation were considered as admixed. Based on this, cluster 1 had 24 accessions, cluster 2 had 48 accessions and cluster 3, 54. Nineteen accessions were admixed (Additional file 2: Table S4). At $K = 3$ (Fig. 2a), the cultivated *Stevia* were shown to belong to the three clusters (named cluster 1, 2 and 3) whereas the landraces from Argentina and Cuba belonged to cluster 2. In this cluster 2, cultivated *stevia* belong to the oldest selections known as "EireteI", "EireteII" and "MoritaIII" types but also with the "C" and "D" cultivars more recently selected in Germany and belonging to the EUSTAS collection.

Genetic variation and differentiation among the three *stevia* clusters

We built a Neighbor-Joining tree of the 126 non-admixed individuals (Additional file 2: Table S4) based on dissimilarity scores. The dendrogram showed three major clades (Fig. 2b). Structure of the dendrogram was in agreement with the clustering inferred with STRUCTURE, distinguishing three clades corresponding to clusters 1 (in red), 2 (in blue) and 3 (in orange) except for some genotypes, CULT75_GER, CULT02_CAN and CULT40_FRA. Dendrograms of clusters provided an interesting pattern with a clear differentiation of cultivated *stevia* cluster 1. The PCA (Fig. 3) revealed a similar pattern as inferred with STRUCTURE, with a clear differentiation of cultivated *stevia* cluster 1, clearly separated from the other *stevia* clusters.

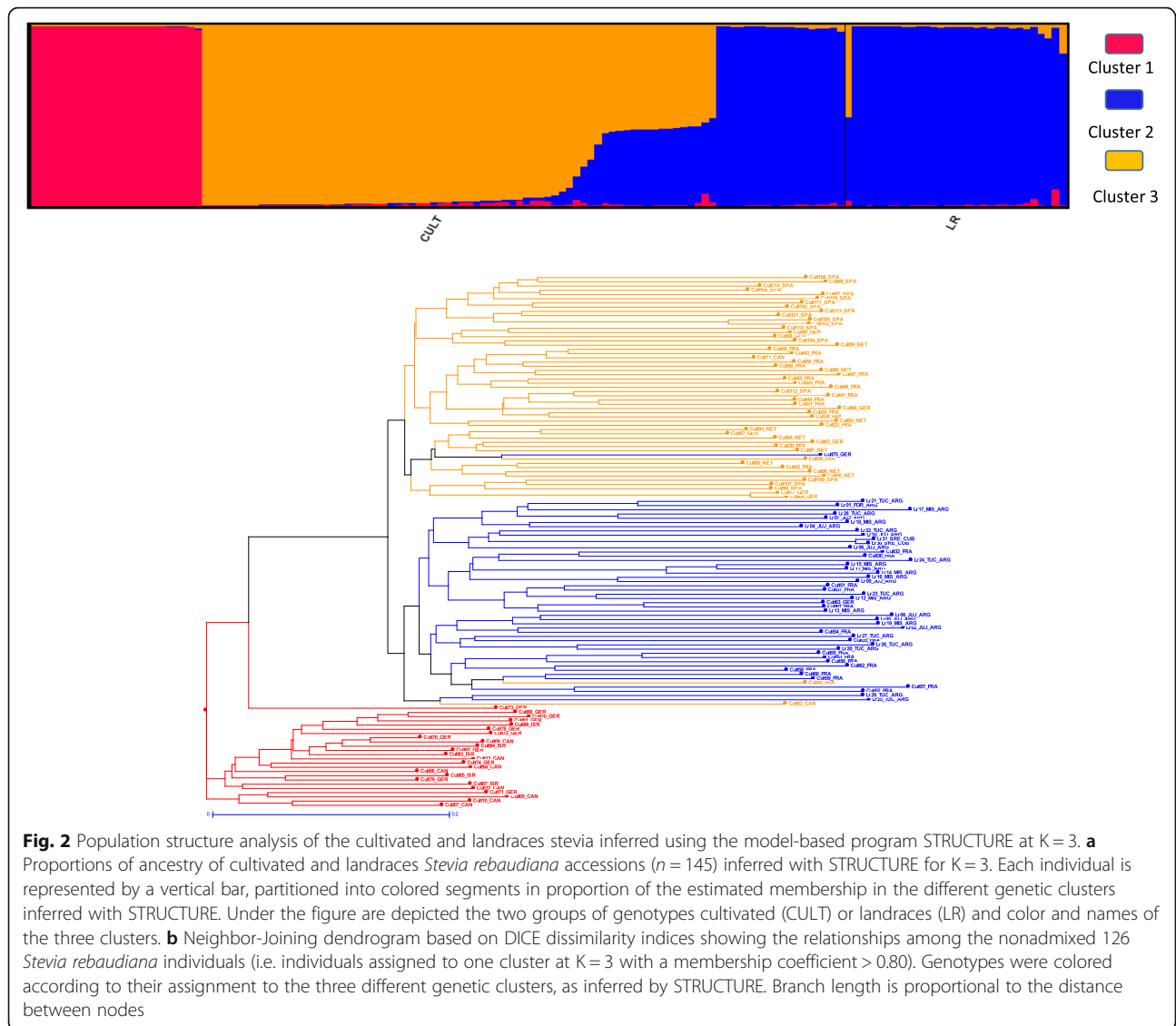
Table 2 List of *Stevia rebaudiana* cultivated and landraces genotypes used in the current study

ID ^a	Type of material ^b	Origin ^c	Number of cultivars ^d	Total number of genotypes
Cult_CAN	Cultivar	Canada	3	13
Cult_CHI	Cultivar	China	2	15
Cult_FRA	Cultivar	France	12	34
Cult_GER	Cultivar	Germany	8	20
Cult_ISR	Cultivar	Israël	1	5
Cult_NET	Cultivar	Netherlands	2	9
Cult_SPA	Cultivar	Spain	3	18
Lr_FOR	Landrace	Argentina		1
Lr_JJJ	Landrace	Argentina		9
Lr_MIS	Landrace	Argentina		9
Lr_TUC	Landrace	Argentina		10
Lr_CUB	Landrace	Cuba		2

^aIdentification^bCultivars refer to sold genotypes through commercial providers. They were provided as seed lots^cCountry of the provider or origin of the landrace; Landraces from Cuba were provided by the New York Botanical Garden Herbarium, catalog number 1687090 and 1,687,091, collection number 5353, collected in Cuba in 1927 and 1931^dDetails of ID of each genotype per cultivar is given in Additional file 2: Table S1**Table 3** Polymorphism analysis at 18 loci of *Stevia rebaudiana* genome for 145 genotypes

Markers Name	Allele Number	Product size expected	Allele size range	Major Allele Frequency	PIC ^a	Gene Diversity (He)	Heterozygosity (Ho) ^b	Fis W&C ^c
stvia004	17	185	166–226	0.32	0.81	0.83	0.79 ns	0.047
stvia018	13	121	89–134	0.22	0.82	0.84	0.46***	0.452
stvia021	5	119	97–121	0.38	0.60	0.67	0.47***	0.299
stvia024	16	205	173–220	0.25	0.83	0.84	0.69***	0.18
stvia025	15	224	203–243	0.21	0.84	0.86	0.70***	0.193
stvia036	15	163	144–180	0.28	0.83	0.85	0.87***	−0.021
stvia044	9	208	203–221	0.36	0.70	0.74	0.49***	0.341
stvia048	19	153	132–163	0.31	0.82	0.84	0.84 ns	0.002
stvia051	13	194	182–224	0.23	0.84	0.86	0.81***	0.062
stvia057	16	236	190–265	0.28	0.77	0.80	0.68***	0.156
stvia071	8	187	168–191	0.36	0.74	0.77	0.74 ns	0.038
stvia072	10	239	231–251	0.56	0.63	0.65	0.61***	0.057
stvia079	13	173	153–198	0.29	0.80	0.82	0.79***	0.035
stvia084	11	159	142–178	0.49	0.70	0.72	0.70 ns	0.041
stvia093	12	221	205–247	0.38	0.78	0.80	0.79***	0.017
stvia096	8	204	174–213	0.29	0.75	0.78	0.76***	0.033
stvia099	6	111	88–121	0.34	0.66	0.71	0.66 ns	0.076
stvia107	7	125	125–143	0.52	0.61	0.65	0.61 ns	0.062

^aPIC Polymorphic information content^bChi-square test for Hardy-Weinberg equilibrium; ns = not significant; *** $P < 0.001$ ^cFIS W&C FIS Weir and Cockerham (1984) [47]



We computed population genetic statistics for the three genetic populations (Table 4). The inbreeding coefficients (F_{is}) were all low but positive. This result can be linked to the significant Chi-square test for Hardy-Weinberg equilibrium on most of the markers (Table 3), related to the putative kinship within the seeds lots. High genetic diversity was found in the cultivated populations clusters 1 and 3 with $H_e = 0.725$ and 0.753 respectively and in the mixed landraces and cultivated cluster 2 ($H_e = 0.801$).

Pairwise F_{ST} among the three stevia populations were all significant (Table 5). They indicated a differentiation between the three populations. AMOVA was used to estimate the variance among and within populations (Table 6). The results indicated that the majority of genetic variation was due to a remarkable degree of within population variation (98%; Table 6). Only 2% of the

genetic variation was attributed to differences among populations.

Variation of steviol glycosides composition among *Stevia* populations

The analyses above revealed high genetic diversity in cultivated as in stevia landraces. In order to evaluate the classification of the different populations according to their SG composition, we estimated the ratio of steviol (ST), RebA, sweet SG as the sum of RebM, RebD and RebBF and bitter SG as the sum of RebC, DulA, Rub and SB. Expectedly, the graph of the variables contrasted the sweet-flavored SGs such as RebA and sweet SG with stevioside and bitter SG (Fig. 4). The graph of the individuals showed a distribution along the first dimension separating the genotypes presenting the most sweetie taste SG until those presenting a dominant bitterness.

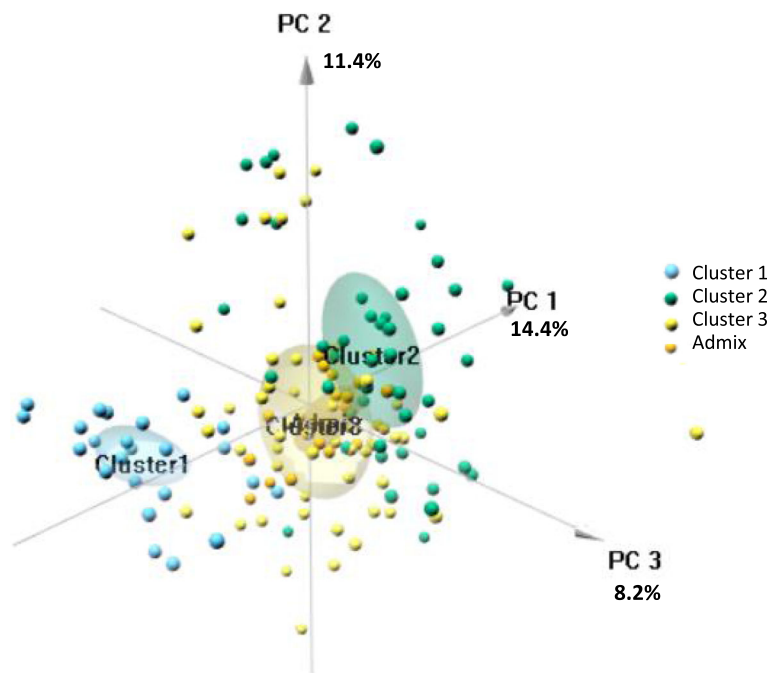


Fig. 3 Principal component analysis of the 145 *Stevia rebaudiana* genotypes for steviol glycosides proportions. Analysis was based on 9 SGs. Graph of variable (left) shows steviol proportion (PST), RebA proportion (PRebA), sweet SG proportion (Psweet; sum of RebM, RebD and RebBF) and bitter SG proportion (Pbitter; sum of RebC, DulA, Rub and SB). Graph of individuals (right) shows the distribution along dimension 1 which explains 95.28% of the variance. The names of the genotypes present at the ends (surrounded in blue) are indicated. The 3 clusters cannot be distinguished (in blue).

Interestingly, the 3 genetic populations appear little differentiated for the content trait in the different SGs, the 3 clusters being grouped in the center of the graph of the individuals. Clusters 1 and 3 appear superimposed. Nevertheless, it should be noted that the cluster 2, stands out slightly towards the highest compositions in SG of sweet taste. Thus, the genotypes that explain the most the composition of sweet taste SG are 3 genotypes

of populations of Argentina (Lr15_MIS_ARG, Lr12_MIS_ARG and Lr26_TUC_ARG) as well as the cultivated genotype Cult75_GER which is the improved genotype “C” of the EUSTAS collection (Hastoy et al., 2019). The genotypes that explain the most the composition in bitter SG are essentially cultivated genotypes such as Cult69_GER, Cult12_CAN and Cult34_FRA.

Table 4 Summary statistics of genetic variation among the three *Stevia rebaudiana* populations detected with STRUCTURE

Pop		N	Na	Ne	Ho	He	Fis
Cluster1	Mean	24	6.444	3.833	0.623	0.725	0.140
	SE		0.372	0.203	0.040	0.016	0.054
Cluster2	Mean	48	10.778	5.626	0.723	0.801	0.100
	SE		0.823	0.433	0.031	0.017	0.031
Cluster3	Mean	54	8.611	4.441	0.690	0.753	0.083
	SE		0.652	0.307	0.031	0.020	0.032

N = sample size
 Na = No. of Different Alleles
 Ne = No. of Effective Alleles = $1/(\sum p_i^2)$
 Ho = Observed Heterozygosity = No. of Heterozygotes/N
 He = Expected Heterozygosity = $1 - \sum p_i^2$
 He = Unbiased Expected Heterozygosity = $(2N / (2N - 1)) * He$
 Fis = Fixation Index = $(He - Ho) / He = 1 - (Ho / He)$
 Where p_i is the frequency of the i^{th} allele for the population & $\sum p_i^2$ is the sum of the squared population allele frequencies

Discussion

The gathering and in-depth study of genetic resources is an essential step in setting up a breeding program. *S. rebaudiana* has a recognized interest in health and the associated industry is now highly developed and world-wide. Nevertheless, the optimized improvement of Stevia is in its very early stages. Breeding programs are relatively new. In particular, they suffer from the critical lack of in-depth information on genetic resources associated with a lack of

Table 5 Pairwise Population Matrix of Fst Values for Total

	Cluster1	Cluster2	Cluster3
Cluster1	0.000	0.001	0.046
Cluster2	0.020	0.000	0.004
Cluster3	0.013	0.012	0.000

Fst values below the diagonal. Probability, $P(\text{rand} >= \text{data})$ based on 999 permutations is shown above diagonal

Table 6 Analysis of molecular variance (AMOVA) using SSR data for the three subpopulations of *S. rebaudiana*

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among populations	2	31.395	0.10901	1.54
Within populations	249	1737.256	6.97693	98.46
Total	251	1768.651	7.08594	

genomic tools. On the basis of the gathering of 145 genotypes of origin as diverse as possible, our study has for objective (1) to develop robust SSR markers (2) to study the pattern of genetic diversity and genetic structure of 114 cultivated genotypes from 31 cultivars but also 31 landraces coming from different regions of Argentina and Cuba.

SSR polymorphism

SSR markers are useful for studying genetic diversity because they are highly polymorphic, multi-allelic and their development requires a limited genomic information. Therefore, they are the easiest and more informative molecular markers. The development of molecular markers in *S. rebaudiana* is longstanding. In 1999, one hundred and eighty-three RAPD markers were developed and used to produce the first partial genetic map from a pseudo test-cross F1 [14]. RAPD and ISSR

markers were then used in three studies to study the relationships between 6 Egyptian and Indonesian accessions in 2008 and 2011 respectively [13, 21] and 12 accessions from India in 2016 [22]. As underlined in [19], the dominance and low reproducibility rate of such markers lead to the development of EST-SSR markers. EST-SSR were developed from 2977 unigenes predicted from 5548 publicly available ESTs of *S. rebaudiana* by [18, 19]. Fifty-two and seventeen EST-SSR were developed respectively. They were used to study the genetic variability of forty randomly genotypes from CSIR, India [19] and twelve accessions from North India [18, 22]. These works reported a number of alleles per locus ranging between 2 and 15, with an average of 4.7 allele per SSR locus which is lower than what we found. This could be explained by the areas of the genome targeted by SSRs and/or by a narrower genetic diversity in Bhandawat et al. studies

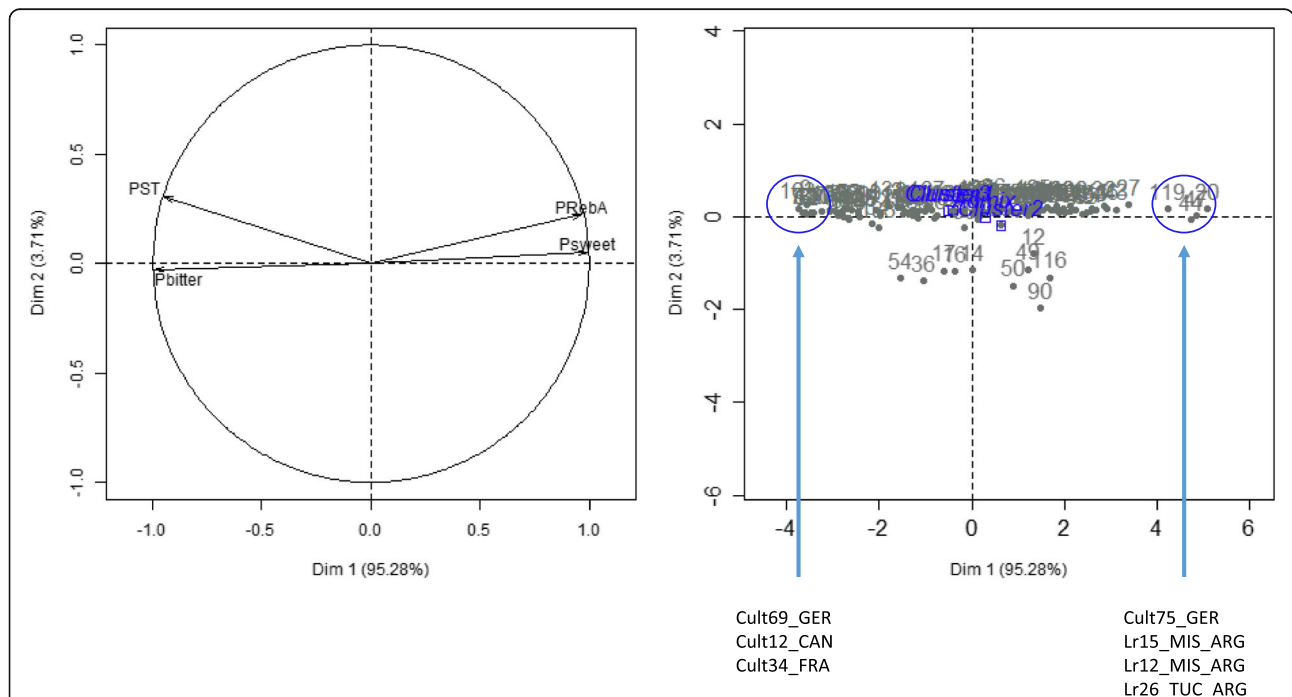


Fig. 4 Principal component analysis of the 145 *Stevia rebaudiana* genotypes for steviol glycosides proportions. Analysis was based on 9 SGs. Graph of variable (left) shows steviol proportion (PST), RebA proportion (PRebA), sweet SG proportion (Psweet; sum of RebM, RebD and RebBF) and bitter SG proportion (Pbitter; sum of RebC, DulA, Rub and SB). Graph of individuals (right) shows the distribution along dimension 1 which explains 95.28% of the variance. The names of the genotypes present at the ends (surrounded in blue) are indicated. The 3 clusters cannot be distinguished (in blue)

[19] and in Kaur et al. studies [18], focused on Indian genotypes. It has also been shown in durum wheat [23], in eggplant [24] or in opium poppy [25] that genomic SSRs are more polymorphic than EST-SSRs.

Nevertheless, in both studies and our, average H_o was high, 0.80 and 0.69 respectively, in accordance with the reported outcross mating system of *S. rebaudiana* [6].

Population structure and genetic diversity

The population structure and dendrogram analyses divided the accessions into three clusters/subpopulations, although the analyses are based on different methods. The low number of admixed genotypes (19/145) can be explained by the composition of the collection, 78% (114/145) of which are cultivated varieties and 22% (31/145) of landraces. No wild genotypes could be analyzed. In addition, the varieties grown are often produced by crossing between two identified parents. The use of a limited number of parents may also explain the low number of admixed.

Except for the two Chinese cultivars belonging to admix cluster, no regional aggregation could be observed. Cluster 1, in particular, gathers genotypes from Canada, Germany and Israël, indicating large seed and alleles exchanges in Stevia. Cluster 1 and Cluster 3 consisted essentially of cultivated varieties. Cluster 1 seems to reveal original diversity. It contains the genotype “Gawi” selected in Germany [26, 27] but of unknown origin [28]. Nevertheless, in our previous study on the phenotype of 15 genotypes from various origin in southwestern France field conditions, “Gawi” was shown to belong to a different morphotype from “Eirete” on the basis of foliar biomass production [28]. Cluster 3 also consist of cultivated varieties, some of them being sold as “Criolla” type. In a very interesting way, the Argentine landraces as well as the old samples from Cuba are all grouped in the cluster 2. Landraces revealed a mean number of private allele per locus around twice the mean of cultivated genotypes. This cluster is shared with cultivated varieties, type “Eirete” and “MoritaIII”, known to be varieties formerly improved and very widespread in the world.

AMOVA results showed that the genetic differentiation was overwhelmingly due to within population variation. As our study is the first large study on *S. rebaudiana* genetic resources, there is no reference study. But, similar results have been observed in other outcrossing perennials such as alfalfa (*Medicago sativa*) [29], Dalmatian Pyrethrum (*Tanacetum cinerariifolium*) [30] or agronomic switchgrass (*Panicum virgatum*) [31]. The high level of genetic variability observed within populations of such species is most likely due to their partial or completely allogamous reproductive systems.

SG phenotyping

Phenotyping of SGs does not allow to structure the different populations. The three genetic populations differed very

little based on this trait. Surprisingly, it is the landraces populations that appear to draw towards the quality of sweetness in SG, while cluster 1 and 3, composed only of cultivars cannot be distinguished. They are composed of cultivars producing very varied SGs, from bitter to sweet, which may seem surprising considering that the ratio SG soft/bitter is the first selection criterion in Stevia [5].

Conclusions

Our study is the first analysis of genetic diversity in *S. rebaudiana* involving 145 genotypes, including known cultivars as well as landrace populations of different origin. This study generated 18 new highly polymorphic and robust microsatellite markers. These markers are of great potential for genetic diversity evaluation and germplasm managing which is a crucial step of breeding. These markers could also be used for genotypes traceability. The 145 genotypes of *S. rebaudiana* were successfully genotyped. They revealed three genetic populations with a remarkable variation within population. Landraces revealed their allelic richness and their interest in term of sweet SG phenotype. They harbor valuable genetic variation for further improvement via breeding as shown through their phenotyping in Argentina [32].

Methods

Plant materials

A panel consisting of 145 individual plants belonging to *S. rebaudiana* was collected from different origins (Table 2 and Additional file 2: Table S3). One hundred and fourteen genotypes called “cultivated” issued from thirty-one cultivars were obtained from different providers as seed lots and distributed as follow: 13 from Canada, 15 from China, 34 from France, 20 from Germany, 5 from Israël, 9 from The Netherlands and 18 from Spain. Thirty-one are landraces. Twenty-nine are originated from North Argentina, regions of Formosa, Tucuman, Jujuy and Misiones and belong to INTA collection [32]. Two, catalog number 1687090 and 1, 687,091, collection number 5353, were provided by the New York Botanical Garden Herbarium. They were collected in Cuba in 1927 and 1931, respectively.

SSR genotyping

Identification of EST-SSR markers

Microsatellite or SSRs were developed from 150,258 EST generated on 454 sequencing and downloaded from The Compositae Genome Project of UC Davis (http://compgenomics.ucdavis.edu/data/cwassy_2012/iAssSta.fa) according to the pipeline described in Additional file 1: Figure S1. The program Sputnik (<http://abajian.net/sputnik/>) were used to identify EST containing SSR. For SSR identification, the minimum motif repeats were defined as 20 repeats for a mononucleotide unit, 10 repeats for a dinucleotide unit, 7 repeats for trinucleotide unit, 5 repeats for tetranucleotide

unit and 4 repeats for pentanucleotide unit. Primer pairs flanking the SSRs were designed using Primer3Plus [33] with melting temperatures 52–57 °C, primer lengths 18–24 bp, expected fragment size 100–300 bp. All the designed primers were screened on five samples, Cult32_FRA, Cult33_FRA, Cult34_FRA, Cult36_FRA, Cult37_FRA. The primers producing clear and polymorphic bands were subsequently used for genetic diversity assessments.

DNA extraction and molecular genotyping

Genomic DNA was extracted using a modified previously published protocol [34]. Young leaves were dehydrated in the oven at 55 °C for 48 h, and then ground to fine powder in a Grinobender. One ml of buffer (0.1 M Tris-HCl pH 8, 0.7 M NaCl, 0.04 M EDTA, 1% HATMAB, plus 1% β -mercaptoethanol and 50 μ g/ml proteinase K added just before use) was added to 30 mg of powder and incubated for 60 min at 65 °C, gently mixing by inversion every 15 min. After Chloroform: IsoAmylic Alcohol extraction, isopropanol precipitation and 70% ethanol washing, DNA was resuspended in 100 μ l of pure water. Genomic DNA was quantified on Epoch Microplate Spectrophotometer (BioTek).

PCR amplifications were performed in 15 μ l reaction volume containing 10 ng of template DNA, 1 X PCR buffer (10 mM Tris-HCl pH 8.3; 50 mM KCl), 2.5 mM MgCl₂, 0.2 μ M of each primer, 200 μ M of each dNTP's and 0.5 U SurePRIME™ DNA polymerase (MP Biomedicals). The amplifications were performed in a Mastercycler Pro (Eppendorf) with the following PCR protocol: 15 min initial denaturation at 95 °C and 35 cycles of 30 s at 94 °C, 45 s at 55 °C, 60 s at 72 °C, followed by a final extension for 10 min at 72 °C.

PCR amplicons were separated on denaturing polyacrylamide gels consisting of 4.5% polyacrylamide (acrylamide: bis-acrylamide 19: 1) and 7 M urea in 1 \times TBE buffer. After run, amplified fragments were visualized using silver staining protocol [35].

Fragments were genotyped at each locus, and the allele sizes were scored using 10–330 bp ladder (Invitrogen).

Genetic variation

Allele frequencies and genetic diversity measures were calculated using PowerMarker 3.25 [36] and GenAlix 6.5 [37]. These measures included number of alleles (Na), major allele frequency, polymorphic information content (PIC), expected heterozygosity or Gene Diversity (He), observed heterozygosity (Ho). Inbreeding coefficient (Fis) was calculated using SPAGED1 1.3 [38] and verified with GENETIX v4.05 [39]. Private allelic richness was computed with ADZE software to adjust for sample size differences [40]. We further explored the

genetic differentiation and relationships among samples using an unweighted Neighbor-Joining tree constructed using simple matching dissimilarity indices of Jaccard's coefficient method and bootstrap values over 2000 replicates as implemented in the DARWIN software package v6.0.010 [41]. Among-population FST and Nei's indices were estimated using ARLEQUIN v3.5 [42]. The significance of FST was assessed by random resampling of the genotypic data through 1000 permutations.

Population structure

Representation of the genetic relationships among individuals was explored with a principal component analysis (PCA) performed with GENETIX v4.05 [39]. We also used the individual based Bayesian clustering method implemented in STRUCTURE 2.3.3 [43] to investigate population subdivision. We ran STRUCTURE from K = 2 to K = 10 using admixture and correlated allele frequencies assuming no prior population information. Burn-in and number of Markov chain Monte Carlo iterations were set to 10,000 and 100,000, respectively. Ten independent runs were carried out for each K, and outputs were processed with CLUMPP V1.1.2 [44]. STRUCTURE barplots were displayed using DISTRUCT 1.1 [45]. We examined the distribution of Δ K, plotted with STRUCTURE harvester (<http://taylor0.biology.ucla.edu/structureHarvester/>) according to Evanno et al. (2005) [46].

Steviol glycosides extraction and quantification

Steviol glycosides extraction and quantification was done as described in Hastoy et al. (2019) [28]. Nine SGs were detected at 202 nm (RebD, RebM, ST, RebA, RebC, DulA, Rub, RebB, SB) and previously identified by purified SG standard (Chromadex, USA). For each SG, a standard range between 5 and 1000 ng/ μ L of purified standard was used to quantify each amount. Results were expressed as content per unit of leaf dry weight (% w/w) for each SG and total SGs, and as a proportion (%) of the content of each SG to total SG content.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12870-019-2061-y>.

Additional file 1: Figure S1. Summary of the pipeline for the selection of the 18 SSRs used in the current study. **Figure S2.** Mean number of private alleles per locus and mean number of distinct allele per locus for the cultivated and landraces genotypes computed with ADZE software

Additional file 2: Table S1. Type and number of repeat pattern of the 1745 unique SSR among the 150,258 unigenes available for *Stevia rebaudiana* at http://compgenomics.ucdavis.edu/data/cwassy_2012/iAssSta.fa (The Compositae Genome Project of UC Davis). **Table S2.** List of the 18 SSR and related primers and characteristics used in this study.

Table S3. List of *Stevia rebaudiana* cultivated and landraces groups studied. **Table S4.** Distribution of the studied genotypes in the different clusters and admix following the analysis by Structure

Acknowledgements

Authors thank the Director of the Herbarium and Dr. Tom Zanoni from The New York Botanical Garden, Bronx, NY 10458-5126 (USA) for providing the samples from Cuba, Coralie Chesseron (INRA, Bordeaux, France) for plant production and Pierre Jannot for scientific support (Rouages, Agen, France). Grateful thanks to Dr. S. Mariette (INRA, UMR BIOGECO, France) for critical reading of the manuscript.

Authors' contributions

VSL conceived and designed the experiments with the contribution of PC and DR; PC developed the EST-SSR. PC and CH performed the experiments. PB, LEE and CJB contributed to material supply; VSL and PC analyzed the data. VSL wrote the paper with proofreading from DR, LEE and CJB. All authors read and approved the final version of the manuscript.

Funding

C. Hastoy was supported by ANRT n° 2014/0915 funding and Oviatis SA, France. Nouvelle-Aquitaine region supported the work through Cifre specific fundings.

Availability of data and materials

All data generated or analysed during this study are included in this published article as supplementary information files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹UMR Biologie du Fruit et Pathologie, 1332, INRA Université de Bordeaux, 71 avenue Edouard Bourlaux, 33883 Villenave d'Ornon cedex, France. ²Oviatis SA, Le Bourg, 47150 Lacauzade, France. ³INTA INTA-EEA Famaillá CP 4132 Famaillá, Tucumán, Argentina. ⁴FAZ, National University of Tucumán, Tucumán, Argentina.

Received: 28 April 2019 Accepted: 29 September 2019

Published online: 21 October 2019

References

- Soejarto DD. Botany of *Stevia* and *S. rebaudiana*. *Stevia: The genus Stevia*. 2002;Taylor an:18–39.
- Sun J. Development of *Stevia* and steviol glycosides industry in China. *Proc 9th Stevia Symp 2016, EUSTAS, Sweden, 15-16th Sept 2016, Jan MC Geuns Ed.* 2016;145–50.
- Gantait S, Das A, Banerjee J. Geographical distribution, botanical description and self-incompatibility mechanism of genus *Stevia*. *Sugar Tech.* 2018;20:1–10.
- Commission Regulation (EU). No 1131/2011 of 11 November 2011 amending Annex II to Regulation (EC) No 1333/2008 of the European Parliament and of the Council with regard to steviol glycosides. *Off J Eur Union.* 2011;amending A:L295/205 L295/211.
- Angelini LG, Martini A, Passera B, Tavarini S. Cultivation of *S. rebaudiana* Bertoni and associated challenges. In: *Sweeteners*. Cham: Springer International Publishing; 2016. 1–52.
- Yadav AK, Dhyani D, Ahuja PSSS. A review on the improvement of *Stevia rebaudiana* (Bertoni). *Can J Plant Sci.* 2011;91:1–27.
- Parris CA, Shock CC, Qian M. Dry leaf and steviol glycoside productivity of *Stevia rebaudiana* in the Western United States. *HortScience.* 2016;51:1220–7.
- Barbet-Massin C, Giuliano S, Alletto L, Daydé J, Berger M. 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.* 2016;68:85–94.
- Grevsen K, Sorensen JN. 2nd year cultivation results of the Danish “Green Stevia” project—a natural sweetener for organic food products. *From Field to Fork, Proc 9th Stevia Symp 2016, EUSTAS, Sweden, 15-16th Sept 2016, Jan MC Geuns Ed.* 2016;115–26.
- Lankes C, Grosser P. Evaluation of *Stevia rebaudiana* genotypes at a location in the Alentejo region in Portugal. *Proc 8th Stevia Symp 2015, EUSTAS, Bonn, Ger Geuns JMC, Ceunen S, Eds.* 2015;167–76.
- Govindaraj M, Vetriventhan M, Srinivasan M. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet Res Int.* 2015;215:14.
- Chester K, Tamboli ET, Parveen R, Ahmad S. Genetic and metabolic diversity in *Stevia rebaudiana* using RAPD and HPTLC analysis. *Pharm Biol.* 2013;51:771–7.
- Hadia HA, Badawy O, Hafez AM. Genetic relationships among some *Stevia (Stevia rebaudiana* Bertoni) accessions based on ISSR analysis. *Res J Cell Mol Biol.* 2008;2:1–5.
- Yao Y, Ban M, Brandle J. A genetic linkage map for *Stevia rebaudiana*. *Genome.* 1999;42:657–61.
- Schauer S, Huang T, Samuel P, Khazi F, Markosyan A. Insights from the sequencing and annotation of the *Stevia rebaudiana* genome and their application in agronomy and health. *Ann Nutr Metab.* 2017;71:1316–7.
- Hodel RG, Segovia-Salcedo MC, Landis JB, Crowl AA, Sun M, Liu X, et al. The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. *Appl Plant Sci.* 2016;4.
- Carneiro Vieira ML, Santini L, Diniz AL, de Freitas Munhoz C. microsatellite markers: what they mean and why they are so useful. *Genet Mol Biol.* 2016; 39:312–28.
- Kaur R, Sharma N, Raina R. Identification and functional annotation of expressed sequence tags based SSR markers of *Stevia rebaudiana*. *Turkish J Agric For.* 2015;39:439–50.
- Bhandawat A, Sharma H, Nag A, Singh S, Singh Ahuja P, Kumar SR. Functionally relevant novel microsatellite markers for efficient genotyping in *Stevia rebaudiana* Bertoni. *J Genet.* 2014;93:75–81.
- Hodgins KA, Lai Z, Oliveira LO, Still DW, Scascitelli M, Barker MS, et al. Genomics of Compositae crops: reference transcriptome assemblies and evidence of hybridization with wild relatives. *Mol Ecol Resour.* 2014;14:166–77.
- Dyah S, Rina Sri K, Budi SD. Genetic diversity of *Stevia (Stevia rebaudiana* Bertoni) Bertoni) based on molecular characters. *Int Conf Biosci Biotechnol Proc.* 2011;1:37–43.
- Sharma N, Kaur R, Era V. Potential of RAPD and ISSR markers for assessing genetic diversity among *Stevia rebaudiana* Bertoni accessions. *Indian J Biotechnol.* 2016;15:95–100.
- Eujayl I, Sorrells M, Baum M, Wolters P, Powell W. Assessment of genotypic variation among cultivated durum wheat based on EST-SSRS and genomic SSRS. *Euphytica.* 2001;119:39–43.
- Muñoz-Falcón JE, Vilanova S, Plazas M, Prohens J. Diversity, relationships, and genetic fingerprinting of the Listada de Gandia eggplant landrace using genomic SSRs and EST-SSRs. *Sci Hortic.* 2011;129:238–46.
- Şelale H, Çelik I, Gültekin V, Allmer J, Doğanlar S, Frary A. Development of EST-SSR markers for diversity and breeding studies in opium poppy. *Plant Breed.* 2013;132:344–51.
- Zabala UM. Optimierung von Wachstum und Ertrag (Süßstoffbildung) bei *Stevia rebaudiana* Bertoni unter mitteleuropäischen Standortbedingungen. *Rheinischen Friedrich-Wilhelms-Universität Bonn*; 2011.
- Lankes C, Zabala U. Evaluation of *Stevia rebaudiana* genotypes. *Proc 5th Stevia Symp 2011, EUSTAS, Belgium, Jan MC Geuns Ed.* 2011;75–87.
- Hastoy C, Cosson P, Cavaignac S, Boutié P, Waffo-Teguo P, Rolin D, et al. Deciphering performances of fifteen genotypes of *Stevia rebaudiana* in southwestern France through dry biomass and steviol glycoside evaluation. *Ind Crop Prod.* 2019;128:607–19.
- Yin S, Wang Y, Nan Z. Genetic diversity studies of alfalfa germplasm (*Medicago sativa* L. subsp sativa) of United States origin using microsatellite analysis. *Legum Res.* 2018;41:202–7.
- Grdisa M, Liber Z, Radosavljevic I, Carovic-Stanko K, Kolak I, Satovic Z. Genetic diversity and structure of Dalmatian pyrethrum (*Tanacetum cinerariifolium* Trevir./Sch/Bip., Asteraceae) within the Balkan Refugium. *PLoS One.* 2014;9.
- Nageswara-Rao M, Stewart CN, Kwit C. Genetic diversity and structure of natural and agronomic switchgrass (*Panicum virgatum* L.) populations. *Genet Resour Crop Evol.* 2013;60:1057–68.
- Moreno FE, Salomon V, Erazzú LE, Budeguer CJ, Maldonado LM. *Stevia rebaudiana* (Bertoni) Bertoni: Determinación de Estevisido y Rebaudisido A en genotipos promisorios de una colección en Tucumán. *Novena Reun Prod Veg y Séptima Prod Anim del NOA S M Tucumán.* 2016:393–8.
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 2007;35:1–4.
- Doyle J. Isolation of plant DNA from fresh tissue. *Focus (Madison).* 1990;12:13–5.

35. Bassam BJ, Caetano-Anollés G, Gresshoff PM. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal Biochem.* 1991;196:80–3.
36. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics.* 2005;21:2128–9.
37. Peakall R, Smouse PE. GenALEX 6.5: genetic analysis in excel. Population genetic software for teaching and research-an update. *Bioinformatics.* 2012;28:2537–9.
38. Hardy OJ, Vekemans X. SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes.* 2002;2:618–20.
39. Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France). 2004.
40. Szpiech ZA, Jakobsson M, Rosenberg NA. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics.* 2008;24:2498–504.
41. Perrier X, Jacquemoud-Collet JP. DARwin software. 2006. <http://darwin.cirad.fr/>.
42. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. *Mol Ecol Resour.* 2010;10:564–7.
43. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000;155:945–59.
44. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics.* 2007;23:1801–6.
45. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes.* 2004;4:137–8.
46. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol.* 2005;14:2611–20.
47. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution (N Y).* 1984;38:1358–70.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

