Euphytica (2017) 213:264
https://doi.org/10.1007/s10681-017-2057-3

Comparison of transcriptome-derived simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers for genetic fingerprinting, diversity evaluation, and establishment of relationships in eggplants

Pietro Gramazio · Jaime Prohens · Dionís Borràs · Mariola Plazas · Francisco Javier Herraiz · Santiago Vilanova

Received: 17 July 2017 / Accepted: 30 October 2017 / Published online: 7 November 2017

Abstract Simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers are amongst the most common markers of choice for studies of diversity and relationships in horticultural species. We have used 11 SSR and 35 SNP markers derived from transcriptome sequencing projects to fingerprint 48 accessions of a collection of brinjal (*Solanum melongena*), gboma (*S. macrocarpon*) and scarlet (*S. aethiopicum*) eggplant complexes, which also include their respective wild relatives *S. incanum*, *S. dasyphyllum* and *S. anguivi*. All SSR and SNP markers were polymorphic and 34 and 36 different genetic fingerprints were obtained with SSRs and SNPs, respectively. When combining both markers all accessions but two had different genetic profiles. Although on average SSRs were more informative than SNPs, with a higher number of alleles, genotypes and polymorphic information content (PIC), and expected heterozygosity (H_e) values, SNPs have proved highly informative in our materials. Low observed heterozygosity (H_o) and high fixation index (f) values confirm the high degree of homozygosity of eggplants. Genetic identities within groups of each complex were higher than with groups of other complexes, although differences in the ranks of genetic identity values among groups were observed between SSR and SNP markers. For low and intermediate values of pair-wise SNP genetic distances, a moderate correlation between SSR and SNP genetic distances was observed ($r^2 = 0.592$), but for high SNP genetic distances the correlation was low ($r^2 = 0.080$). The differences among markers resulted in different phenogram topologies, with a different eggplant complex being basal (gboma eggplant for SSRs and brinjal eggplant for SNPs) to the two others. Overall the results reveal that both types of markers are complementary for eggplant fingerprinting and that interpretation of relationships among groups may be greatly affected by the type of marker used.

P. Gramazio · J. Prohens (✉) · D. Borràs · F. J. Herraiz · S. Vilanova
Instituto de Conservación y Mejora de la Agrodiversidad Valenciana, Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain
e-mail: jprohens@btc.upv.es

P. Gramazio
e-mail: piegra@upv.es

D. Borràs
e-mail: dioborpa@etsiamn.upv.es

F. J. Herraiz
e-mail: fraherga@upvnet.upv.es

S. Vilanova
e-mail: sanvina@upvnet.upv.es

M. Plazas
Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universitat Politècnica de València, Camino de Vera 14, 46022 Valencia, Spain
e-mail: maplaav@btc.upv.es

Keywords Eggplants · Simple sequence repeats · Single nucleotide polymorphisms · Fingerprinting · Genetic diversity · Cluster analysis

Introduction

Molecular markers based on DNA polymorphisms are of great utility for different applications in biological and agricultural sciences (Avise 2012; Grover and Sharma 2014). There is a wide array of available molecular markers (e.g., isozymes, RFLPs, RAPDs, ISSRs, AFLPs, SSRs, SNPs, etc.), which have different characteristics depending on their nature and the techniques employed to identify them. Among them, microsatellites or simple sequence repeats (SSRs) have been amongst the most used ones for germplasm management, selection and breeding (Kalia et al. 2011). SSRs are codominant, abundant, robust and highly polymorphic (Varshney et al. 2005). Although a few years ago their identification was relatively expensive and time-consuming through the development of genomic libraries, with the debut of new generation sequencing (NGS) platforms their isolation has become straightforward and cost-effective (Fernandez-Silva et al. 2013; De Barba et al. 2016; Zhan et al. 2016). In fact, thanks to the tremendous advances in sequencing of transcriptomes and genomes, hundreds or thousands of SSRs can be identified (Xiao et al. 2013; Goodwin et al. 2016). However, these large-scale sequencing projects also allow the identification of thousands to millions of molecular markers of single nucleotide polymorphism (SNPs) markers with a reasonable amount of resources (Van Tassell et al. 2008; Davey et al. 2011; Scheben et al. 2017). As occurs with the SSRs, SNPs are codominant, very reliable, ubiquitous and universal molecular markers, and although are generally less informative than SSRs (Yang et al. 2011; Filippi et al. 2015; Gonzaga, 2015), they are much more abundant and easy to automate (Thomson et al. 2014; Kim et al. 2016).

SNPs are more uniformly spread across the genomes, while the SSRs tend to distribute more frequently in heterochromatic regions, although this strongly depends on the species (Li et al. 2002; Hong et al. 2007). In addition, the validation of SSRs is generally more time-consuming and expensive due to the need of detection through an agarose or polyacrylamide gels or capillary sequencing (Jones et al. 2007). In this respect, Yan et al. (2010) estimated that the resources and time needed to build a genetic map using a high-throughput SNP genotyping may be 75% cheaper and 100-fold faster than SSR gel-based methods.

Because of their different characteristics, SSRs and SNPs sample different levels of genetic diversity. In this respect, SSRs have a much higher mutation rate than SNPs (Hamblin et al. 2007; Coates et al. 2009; Fischer et al. 2017), because the mutations can be derived from a variation of the number of repeats, motif length or motif sequence (Ellegren 2004; Kashi and King 2006). In fact, the estimation of mutational rates of SSRs (1×10^{-5}) is several orders of magnitude higher than that of SNPs (1×10^{-9}) (Li et al. 1981; Kruglyak et al. 1998; Martínez-Arias et al. 2001). Although at the transcriptomic level the mutation rate of both SSRs and SNPs are lower than at the genomic level, the mutation rate of SSRs is also much higher than that of SNPs in the expressed sequences (Li et al. 2002).

Cultivated eggplants and their wild relatives are very variable (Vorontsova et al. 2013; Kaushik et al. 2016; Acquadro et al. 2017), with three different eggplant complexes (Daunay and Hazra 2012; Plazas et al. 2014; Syfert et al. 2016), and may be an appropriate material for comparing SSR and SNP markers for fingerprinting, evaluation of genetic identities and distances and for studying their relationships. Although the brinjal (or common) eggplant (*Solanum melongena* L.) is the most economically important, two other cultivated eggplant species exist, namely the gboma (*S. macrocarpon* L.) and the scarlet eggplants (*S. aethiopicum* L.), which are mostly grown in sub-Saharan Africa (Lester et al. 1990; Sunseri et al. 2010). *Solanum melongena* together with its ancestor (*S. insanum* L.) and other close relatives, like *S. incanum* L., are part of the so-called brinjal eggplant complex (Knapp et al. 2013). Similarly, the gboma eggplant together with its ancestor *S. dasyphyllum* Schumach. & Thonn. on one side, and the scarlet eggplant together with its ancestor S. *anguivi* Lam. on the other constitute, respectively, the gboma and scarlet eggplant complexes (Lester and Niakan 1986; Bukenya and Carasco 1994). Among the latter, the scarlet eggplant complex is more variable than the gboma eggplant and four cultivar groups (Aculeatum, Gilo, Kumba, and Shum) are considered within the

cultivated *S. aethiopicum* (Lester and Daunay 2003). In addition, intermediate forms between the wild *S. anguivi* and cultivated *S. aethiopicum* are also common (Plazas et al. 2014).

Although independently of the markers used the three eggplant complexes are often genetically differentiated clearly (Sakata and Lester 1997; Furini and Wunder 2004; Isshiki et al. 2008; Tumbilen et al. 2011; Acquadro et al. 2017), the genetic proximity and phylogenetic relationships among them frequently depend on the markers and plant materials used. In this respect, there is wide discrepancy in the studies over which of the three eggplant complexes is phylogenetically basal or genetically more distant to the two others (Sakata et al. 1991; Sakata and Lester 1997; Furini and Wunder 2004; Levin et al. 2006; Isshiki et al. 2008; Weese and Bohs 2010; Tumbilen et al. 2011; Meyer et al. 2012; Vorontsova et al. 2013; Särkinen et al. 2013; Acquadro et al. 2017). Regarding the relationships within the hypervariable scarlet eggplant complex there have been few molecular studies evaluating them (Sunseri et al. 2010; Adeniji et al. 2013; Acquadro et al. 2017), but in general, they reveal that they present a low genetic differentiation.

Given the lack of studies comparing SSR and SNP markers in *Solanum* crops for genetic resources, breeding and genetic relationships, in this work we have genotyped a collection of accessions from the brinjal, gboma and scarlet eggplants with SSR and SNP markers obtained from the transcriptomes of *S. aethiopicum* and *S. incanum* (Gramazio et al. 2016). Although genomic markers tend to be more informative and accurate because are less prone to selection (Gadaleta et al. 2011), several studies with markers derived from transcriptome have shown that they are useful and reliable for phylogenetics and establishment of relationship among and within populations (Vogel et al. 2006; Castillo et al. 2008; Choudhary et al. 2009). Our objective is to confirm the potential utility of these markers as well as to compare them for fingerprinting, evaluation of genetic identities and distances, and for the establishment of relationships among these three groups.

Materials and methods

Plant material

The collection of accessions used in this study encompasses accessions from different origins (African, Asian and European), as indicated in Table 1. All materials are maintained at COMAV germplasm bank (Universitat Politècnica de València, Spain). Thirty-nine accessions belonged to the scarlet eggplant complex and according to Lester and Niakan (1986) classification key and Plazas et al. (2014), morphological characterization, belong to *S. anguivi* (2), to *S. aethiopicum* groups Gilo (16), Kumba (7), Aculeatum (5), Shum (3), or to an intermediate group between *S. anguivi* and *S. aethiopicum* (8). Seven other accessions used in this study corresponded to the gboma complex, of which six are of *S. macrocarpon* and one of *S. dasyphyllum*. Finally, one accession of *S. incanum* and one *S. melongena* were used to represent the brinjal complex. The seeds were germinated following a protocol which is especially recommended for *Solanum* species that may present dormancy (Ranil et al. 2015).

DNA extraction

Total genomic DNA was isolated from leaves of 3–4 true leaves stage plantlets, according to the CTAB protocol (Doyle and Doyle 1987) with slight modifications. The extracted DNA was dissolved in Milli-Q water and general quality was confirmed in agarose gel at 0.8%. After a concentration measurement using a Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, USA), the DNA was diluted at 30 ng/uL for PCR and High Resolution Melting (HRM) amplification.

SSR genotyping

The identification of SSRs from the transcriptomes of one *S. aethiopicum* and one *S. incanum* accessions (Table 2) was reported in Gramazio et al. (2016). The selection of 11 highly reliable polymorphic SSRs was performed through filtering them for quality parameters and checking their coverage and length in the IGV visor (Thorvaldsdóttir et al. 2013). Primers pairs were designed using Primer3 tool (v. 0.4.0, http://bioinfo.ut.ee/primer3-0.4.0/primer3/).

Table 1 Plant materials used for molecular characterization with SNP and SSR markers of a collection of accessions of brinjal, gboma and scarlet eggplant complex, including the species and cultivar group and the country of origin

Accession	Species/group	Origin
Brinjal eggplant complex		
MM577	*S. incanum*	Israel
AN-S-26	*S. melongena*	Spain
Gboma eggplant complex		
BBS117	*S. macrocarpon*	Ivory Coast
BBS168	*S. macrocarpon*	Ivory Coast
BBS171	*S. macrocarpon*	Ivory Coast
BBS178	*S. macrocarpon*	Ivory Coast
MM1153	*S. dasyphyllum*	Uganda
MM1558	*S. macrocarpon*	Malaysia
RNL0367	*S. macrocarpon*	Ghana
Scarlet eggplant complex		
AN05	*S. aethiopicum* group Gilo	Angola
AN39	*S. aethiopicum* group Gilo	Angola
AN67	*S. aethiopicum* group Gilo	Angola
BBS107	*S. aethiopicum* group Kumba	Ivory Coast
BBS110	*S. aethiopicum* group Kumba	Ivory Coast
BBS111	*S. aethiopicum* group Kumba	Ivory Coast
BBS114	Intermediate *S. anguivi-S. aethiopicum*	Ivory Coast
BBS116	Intermediate *S. anguivi-S. aethiopicum*	Ivory Coast
BBS119	*S. anguivi*	Ivory Coast
BBS125	*S. anguivi*	Ivory Coast
BBS131	Intermediate *S. anguivi-S. aethiopicum*	Ivory Coast
BBS135	*S. aethiopicum* group Gilo	Ivory Coast
BBS140	*S. aethiopicum* group Gilo	Ivory Coast
BBS142	*S. aethiopicum* group Gilo	Ivory Coast
BBS147	*S. aethiopicum* group Gilo	Ivory Coast
BBS148	Intermediate *S. anguivi-S. aethiopicum*	Ivory Coast
BBS151	*S. aethiopicum* group Gilo	Ivory Coast
BBS159	*S. aethiopicum* group Gilo	Ivory Coast
BBS170	Intermediate *S. anguivi-S. aethiopicum*	Ivory Coast
BBS180	Intermediate *S. anguivi-S. aethiopicum*	Ivory Coast
BBS181	*S. aethiopicum* group Gilo	Ivory Coast
BBS184	Intermediate *S. anguivi-S. aethiopicum*	Ivory Coast
BBS192	Intermediate *S. anguivi-S. aethiopicum*	Ivory Coast
INRA4	*S. aethiopicum* group Kumba	Senegal
IVIA026	*S. aethiopicum* group Gilo	Spain
MM457	*S. aethiopicum* group Aculeatum	Japan
MM585	*S. aethiopicum* group Kumba	Senegal
MM1207	*S. aethiopicum* group Kumba	Mali
MM1483	*S. aethiopicum* group Aculeatum	Unknown
PI413783	*S. aethiopicum* group Kumba	Burkina Faso
RAREGILO	*S. aethiopicum* group Gilo	Unknown
RNL0022	*S. aethiopicum* group Shum	Benin

Table 1 continued

Accession	Species/group	Origin
RNL0187	*S. aethiopicum* group Aculeatum	Unknown
RNL0252	*S. aethiopicum* group Gilo	Ghana
RNL0288	*S. aethiopicum* group Gilo	Ghana
RNL0340	*S. aethiopicum* group Shum	Zambia
RNL0395	*S. aethiopicum* group Gilo	Liberia
UPV29014	*S. aethiopicum* group Gilo	Algeria
UPV29803	*S. aethiopicum* group Aculeatum	Argelia

The SSRs amplification was performed by PCR in a volume of 12 µL including 7.21 µL water, 1.2 µL 1 × PCR buffer, 0.6 µL MgCl2 50 mM, 0.24 µL dNTPs 10 mM, 0.3 µL 10 µM, 0.06 µL forward primer with M13 tail 10 µM, 0.24 µL fluorochrome (FAM, VIC, NED and PET) 10 µM, 0.15 µL Taq DNA Polymerase (5 U/µL), 2 µL DNA template 20 ng/µL. The PCR program used was the following: 95 °C for 3 min for a denaturation, 30 cycles of 30 s at 95 °C followed by 30 s at 65 °C and of 30 s at 72 °C and finally 72 °C for 5 min for the last step of extension. The PCR products were subsequently diluted in formamide and sequenced by capillary electrophoresis through an ABI PRISM 3100-Avant sequencer (Thermo Fisher Scientific, Waltham, USA) using a 600 LIZ GeneScan size standard (Thermo Fisher Scientific, Waltham, USA). The fragments were analyzed using the GeneScan software (Thermo Fisher Scientific, Waltham, USA) to obtain the electropherograms and polymorphisms were analyzed with Genotyper DNA Fragment Analysis software (Thermo Fisher Scientific, Waltham, USA).

SNP genotyping

The identification of 35 SNPs and primer pairs design (Table 2) was similar to the one for SSRs as indicated above by using *S. aethiopicum* and *S. incanum* transcriptomes (Gramazio et al. 2016). Validation of SNPs was performed through Real-Time PCR in a LightCycler 480 (Roche, Basel, Switzerland). The reactions were performed in a 10 µL volume comprising 5 µL Master Mix 2X, 0.8 µL MgCl2 25 mM, 0.25 µL each primer, 1.7 µL water and 2 µL DNA 30 ng/µL with the following touchdown PCR program: denaturation at 95 °C for 10 min, followed by 55 cycles of 10 s at 95 °C, 15 s at 65 °C (decreasing 1 °C each cycle until 55 °C) and of 15 s at 72 °C, finally the melting step at 1 min at 95 °C, 1 min at 40 °C, 1 s at 60 °C and rising the temperature at 0.02 °C/s until 95 °C.

Data analyses

The molecular marker analysis for SNPs and SSRs was performed using the software packages PowerMarker (Liu and Muse 2005) and GenAlEx 6.5 (Peakall and Smouse 2012). The following parameters were calculated using the PowerMarker package: number of alleles per locus, major allele frequency, number of genotypes, polymorphic information content (PIC) values calculated as

$$\text{PIC} = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2,$$

where n is the total number of alleles detected, p_i the frequency of the i th allele, and p_j the frequency of the jth allele) (Botstein et al. 1980), expected heterozygosity (H_e), calculated as $H_e = 1 - \sum_{i=1}^{n} p_i^2$ (where p_i is the frequency of the ith allele) (Nei 1972), observed heterozygosity (H_o), calculated as the number of heterozygous alleles/number of alleles and fixation index (f), calculated as $f = 1 - (Ho/He)$ (Wright 1965).

The consensus tree was calculated from the genetic similarity to illustrate the level of relatedness between the accessions using the UPGMA method (unweighted pair-group method using arithmetic averages) (Sneath and Sokal 1973) and it was reconstructed using the software TreeView (Page 2001). Branch support on the phenogram was tested by bootstrap analysis with 1000 replications using the PHYLIP version 3.67 software (Felsenstein 2007). Bootstrap values of 50%

Table 2 Characteristics of the SRR and SNP markers used, including the chromosome in which they are situated, the *S. incanum* and *S. aethiopicum* unigenes to which they correspond (Gramazio et al. 2016), the corresponding scaffold in the eggplant genome (Hirakawa et al. 2014), and the forward and reverse primers

Marker	Chr.	*S. incanum* unigene	*S. aethiopicum* unigene	Scaffold eggplant genome	Forward primer	Reverse primer
SSR markers						
SSR_38227	1	SIUC38227_TC01	SAUC18592_TC02	Sme2.5_00192.1	TGCATAACCAACATTCAAACC	CCCTCCAAGCTCAAAGAAAG
SSR_37353	2	SIUC37353_TC01	SAUC65284_TC01	Sme2.5_01898.1	AGGGGTTCGCTGAAACAAG	GTGTCACTCATCCCCTGTTG
SSR_37602	4	SIUC37602_TC01	SAUC23061_TC01	Sme2.5_01233.1	AAAAATGGAGGAATTAGAGAAGAAG	CAGCACCAGAAACAACATGAG
SSR_31851	5	SIUC31851_TC01	SAUC57317_TC01	Sme2.5_00211.1	CATGGGAACTTAGCCATGC	TTTTTGGTCTTTGGTTTTAGCTG
SSR_29668	6	SIUC29668_TC01	SAUC22524_TC01	Sme2.5_00163.1	GTGTTGAGCCAGTTGCAGAG	CCCAAAACCAAGATCCAAAC
SSR_37966	7	SIUC37966_TC01	SAUC48208_TC01	Sme2.5_00502.1	GCAACTTCCTTCTCCATCTCC	TGAAACGGTGGCTTTAGCTC
SSR_21086	8	SIUC21086_TC01	SAUC64285_TC01	Sme2.5_00001.1	CGAGCCTGAAGAAGTTGTTG	AGACCCATCATCCAAATTCC
SSR_18542	9	SIUC18542_TC01	SAUC66830_TC02	Sme2.5_02687.1	CAAAATGGATGGGAGAGGAG	TTTCATGTGTGGGAGGGAAC
SSR_03112	10	SIUC03112_TC04	SAUC25971_TC01	Sme2.5_00173.1	GGTGGCATTGATCCAAGAAC	TCCCATTTTAGCAGCCTCAC
SSR_18317	11	SIUC18317_TC02	SAUC52997_TC01	Sme2.5_03669.1	CCTTTGGCAGTCACCATTTAG	TCATATGAGAAGCAACACTTTGG
SSR_37681	12	SIUC37681_TC01	SIUC37681_TC01	Sme2.5_00226.1	ACCGATGAAATCGACTCTGG	TCCATTCCTTCTGGACCATC
SNP markers						
SNP_38959	1	SIUC38959_TC01	SAUC82582_TC01	Sme2.5_03132.1	TGTGTGACTAGGACTTCATCCTC	GCCCTAGAAGGAGCTTTCATC
SNP_38905	1	SIUC38905_TC01	SAUC83481_TC01	Sme2.5_00529.1	TGAAGGAGAAGGACCAGCAG	TCAGCCCATATCAGATCTTGC
SNP_11564	2	SIUC11564_TC01	SAUC38669_TC01	Sme2.5_02714.1	AGGAGAATTGCAGAGTGATGC	TCGCAGCTCATAGCCATATTC
SNP_19191	2	SIUC19191_TC03	SAUC71621_TC01	Sme2.5_05438.1	AACCTCCCTAAAACCCCAAC	TGGCTCTGACAACTGGAAATC
SNP_38971	2	SIUC38971_TC01	SAUC54450_TC02	Sme2.5_01610.1	GATGGTGGTTCTGCGGTATC	TAGGTTCACCAGGCTCCATC
SNP_34715	3	SIUC34715_TC01	SAUC15119_TC01	Sme2.5_02159.1	CTAAGGGGCAGAGCTTCTTG	GACGCCAATAGTTAATAGAACTGC
SNP_00907	3	SIUC00907_TC04	SAUC70117_TC02	Sme2.5_01604.1	GGGAAAGAAAGGAGGAATGG	AAATTTTGGATTTCCATCATCTTC
SNP_27060	3	SIUC27060_TC05	SAUC62129_TC02	Sme2.5_04268.1	TGTTCCTCACTCAATGTGTCG	AGGTGCACCGATTCTTTCC
SNP_13379	3	SIUC13379_TC02	SAUC75690_TC01	Sme2.5_04555.1	CTTCTCTCCCACCAGGCTAC	TGGCAGCATACCAAATAGGC
SNP_23081	4	SIUC23081_TC01	SAUC07358_TC01	Sme2.5_06002.1	AAGTACCTCTGCAGCAACAGC	TCATCACCAAATCTCCATCG
SNP_39035	4	SIUC39035_TC01	SAUC20071_TC01	Sme2.5_03086.1	CCCGTTACTTCAAGGGGATG	CACTGCCTTTCCAAATGAGG
SNP_23613	4	SIUC23613_TC02	SAUC38638_TC01	Sme2.5_00505.1	AAATCCAATTCACAGACATTGC	TGTTGATATCACCGACAACG
SNP_15567	4	SIUC15567_TC01	SAUC21364_TC01	Sme2.5_09958.1	TCAAATGAATGTGAGGAACAGG	TGGAAGAGGAAGAGGCTGAG
SNP_00676	5	SIUC00676_TC02	SAUC66850_TC01	Sme2.5_00697.1	CTCGGGGTCCAGAACTAGAA	CCTACTCCAGGGCTTCCTTC
SNP_19562	5	SIUC19562_TC02	SAUC02478_TC02	Sme2.5_09181.1	GCATCTCAATGTAAAAGCTTCC	GCCTTTGAGTCCGAGTTCAG
SNP_10686	6	SIUC10686_TC01	SAUC40320_TC01	Sme2.5_04479.1	GCACAATTAGCTGGTGTTGG	AAGAGATTGTTGAAGAAAGACGTG
SNP_32044	6	SIUC32044_TC02	SAUC45004_TC02	Sme2.5_02660.1	GACCTAGGCAAGAACGAAGG	TGTAGGACGCTATCCCATTG
SNP_37940	6	SIUC37940_TC01	SAUC79062_TC01	Sme2.5_00001.1	CGGCTATGTACTTCATAACAGC	CCCAGAAATGATTTGCGAAG

Table 2 continued

Marker	Chr.	*S. incanum* unigene	*S. aethiopicum* unigene	Scaffold eggplant genome	Forward primer	Reverse primer
SNP_30643	6	SIUC30643_TC05	SAUC05686_TC01	Sme2.5_00016.1	CACGGTCACTGCTTTCTCTG	AGATGGTGAGCCTTCCTACG
SNP_31222	7	SIUC31222_TC01	SAUC23026_TC01	Sme2.5_00673.1	CCTCCACCTACCCTCAACTC	AAGAAGCGCGAGTTGTTCAG
SNP_01600	7	SIUC01600_TC01	SAUC78921_TC01	Sme2.5_02807.1	GGGAGGGTGGTAAAGGAGTG	GGTTTTCACTCAGCCGCTAC
SNP_23399	7	SIUC23399_TC02	SAUC11413_TC01	Sme2.5_00423.1	TAGAGATGGCCTCGGGAAG	GGAAGATAGATCAAAACGAGCTG
SNP_17586	8	SIUC17586_TC01	SAUC46883_TC01	Sme2.5_00013.1	CTCCGGAATAAATGCAAACC	CCTGTCAATGGAGATGTTCG
SNP_02438	8	SIUC02438_TC01	SAUC15038_TC01	Sme2.5_00858.1	GAGACAGGGGATGATGAAGG	GATGGCACATTGCACCTAAC
SNP_38436	8	SIUC38436_TC01	SAUC26635_TC02	Sme2.5_00391.1	TTGATGCAATAAAGGAAGTGG	AGGCAGATGGGACACTCTTC
SNP_32294	9	SIUC32294_TC04	SAUC36933_TC01	Sme2.5_00104.1	TAGCAAGCTTACGGCTGGTC	CAACTGAAGTGGCATGATGG
SNP_14499	9	SIUC14499_TC01	SAUC71552_TC02	Sme2.5_00504.1	CGGAACAAAAAGCTTTCAACC	ATGCTTCTTTGGGGCTAGAG
SNP_13910	9	SIUC13910_TC01	SAUC15875_TC01	Sme2.5_00488.1	CCCATAAGGTCGCGTAATTC	TCACCCGCAAACCTACTCTC
SNP_14015	10	SIUC14015_TC01	SAUC34299_TC01	Sme2.5_00190.1	TGCCCCATTTCTTCAACTTC	CAGCCATCTTCTCCTGGTAG
SNP_39817	10	SIUC39817_TC01	SAUC37197_TC01	Sme2.5_07454.1	TGTTGTGGACACGGCTACTC	CAAATGTTCTAGGCCCATTCC
SNP_14306	11	SIUC14306_TC01	SAUC48392_TC01	Sme2.5_00442.1	TGGCATCAGCAGTCGTTG	CATGGGGAATTGAATTTTGG
SNP_39475	11	SIUC39475_TC01	SAUC08994_TC01	Sme2.5_09299.1	CACATTGGTGAAAGCCATTG	CTGGCTGCCTCTTGTTGAG
SNP_29844	11	SIUC29844_TC04	SAUC64313_TC07	Sme2.5_00432.1	GCTCGCTTAGGATGAATTTCC	GCATATGGTGGAGGTGGTTC
SNP_14718	12	SIUC14718_TC01	SAUC85575_TC01	Sme2.5_00381.1	TCATGGGTTGCATTGTGAAC	TGCCGACGTAAAGGTCAATC
SNP_30456	12	SIUC30456_TC03	SAUC63329_TC01	Sme2.5_07259.1	ACTGGCCAAGCTTTTGCTAC	GTGTGGGCTCTAAGGGAATG

or higher were used to indicate support for the phenogram topology at a node (Highton 1993). The genetic distance matrix (Nei 1972) among the different accessions was calculated with the GenAlEx 6.5 software package for both for SSRs and SNPs. The correlation between pair-wise genetic distances calculated for both markers was investigated by the Mantel test (Mantel 1967) of matrix correspondence.

Results and discussion

Validation of the SSR and SNP markers for fingerprinting

All SSR and SNP markers tested amplified and were polymorphic in the collection evaluated, although in a few cases there were missing data. In other studies in which eggplant SSR markers were developed de novo, the levels of polymorphism were not as high. For example, Vilanova et al. (2012) found an 85.5% of polymorphism in SSRs derived from a genomic library of *S. melongena*. A lower level of polymorphism (56.7%) was detected in the genetic SSRs derived from a library developed by Nunome et al. (2009), where 598 out of 1054 markers were polymorphic. When comparing the SSR and SNP profile of the two accessions used for obtaining the transcriptomes of *S. aethiopicum* (accession BBS135) and *S. incanum* (accession MM577) (Gramazio et al. 2016), only one SNP marker (SNP_14499) was found to be monomorphic among the two accessions. This confirms the high quality of the transcriptome sequences and in silico analysis performed by Gramazio et al. (2016) and reveals that both SSR and SNP markers obtained from in silico analyses of transcriptomes in materials of cultivated and wild eggplants are reliable and transferable to related species for being used in genotyping and fingerprinting. Other studies confirmed the reliability of the molecular markers discovered in silico from transcriptomes. For example, a set of SNPs identified in silico from a pepper (*Capsicum annuum* L.) transcriptome were validated in 43 pepper lines and accessions resulting in a rate of 89.9% polymorphic markers (Ashrafi et al. 2012). Also 86.7% of a subset of SSRs identified in silico in a zucchini (*Cucurbita pepo* L.) transcriptome resulted polymorphic in a set of ten accessions of genus *Cucurbita*, nine of which were representative of the diversity within *C. pepo* and one accession *C. moschata* accession (Blanca et al. 2011). Up to now, few transcriptomes have been sequenced in genus *Solanum* and in just a few of them molecular markers have been identified and primers pairs designed in silico for a subsequent validation in a wide range of related materials. An example is the *S. dulcamara* L. transcriptome where the SSR identified in silico were validated in seven plants and all of them resulted polymorphic (D'Agostino et al. 2013).

When considering the 11 SSRs, a total of 34 different genetic profiles were found among the 48 accessions evaluated, while for the 35 SNPs, the number of different genetic profiles was 36 (Table 3). This confirms the highest discrimination potential of SSRs compared to SNPs for genetic fingerprinting (Hamblin et al. 2007; Varshney et al. 2007; Yang et al. 2011), as the number of different profiles obtained is almost the same with 11 SSRs or 35 SNPs. In this respect, several authors considered that in order to obtain a similar genetic power to discriminate individuals from different populations the number of SNPs required might be 8–15 times the number of SSRs (Hess and Matala 2011; Yu et al. 2009).

When considering a combination of both SNP and SSR markers 47 unique genetic fingerprints were obtained for the 48 accessions, and only two of them (both from *S. aethiopicum* Gilo group) shared the same genetic profile for the markers that amplified in both of them. Although in general (Hu et al. 2011; Nandha and Singh 2014; Thiel et al. 2003) and in the particular case of eggplant (Muñoz-Falcón et al. 2011) genomic SSRs and SNPs are frequently more polymorphic than transcriptome-derived SSRs and SNP, in our case we have found a considerable level of polymorphism in the markers we tested. Particularly, the combination of both SSR and SNP markers has been highly efficient for genetic fingerprinting. Probably the fact that both types of markers sample different levels of genomic diversity (van Inghelandt et al. 2010) increases the efficiency of fingerprinting when combining both types of markers.

In all cases, accessions sharing a single SSR or SNP profile corresponded to the same eggplant complex (brinjal, gboma, or scarlet eggplant complexes) (Table 3), indicating that both types of markers provide consistent results. The number of accessions having a single genetic profile ranged between two and six for SSRs, between two and seven for SNPs, and only two when combining both SSR and SNP markers.

Table 3 Accessions having a shared genetic profile with 11 SSR or 35 SNP markers, or using all of them (11 SNP plus 35 SNP markers)

Shared genetic profiles	Accessions	Species/group
Using 11 SSRs		
GP_SSR1	BBS 168, RNL0367	*S. macrocarpon*
GP_SSR2	BBS119	*S. anguivi*
	BBS116	Intermediate *S.anguivi-S-aethiopicum*
	BBS111	*S. aethiopicum* gr. Kumba
GP_SSR3	MM1483, RNL0187	*S. aethiopicum* gr. Aculeatum
GP_SSR4	UPV29803	*S. aethiopicum* gr. Aculeatum
	BBS148, BBS192	Intermediate *S.anguivi-S-aethiopicum*
	BBS159, RNL0252	*S. aethiopicum* gr. Gilo
	BBS110	*S. aethiopicum* gr. Kumba
GP_SSR5	BBS151, BBS181, RNL0288, RNL0395	*S. aethiopicum* gr. Gilo
	MM585	*S. aethiopicum* gr. Kumba
GP_SSR6	BBS107, MM1207	*S. aethiopicum* gr. Kumba
Using 35 SNPs		
GP_SNP1	BBS117, MM1558	*S. macrocarpon*
GP_SNP2	BBS180	Intermediate *S.anguivi-S-aethiopicum*
	BBS140, BBS159, RNL0252	*S. aethiopicum* gr. Gilo
GP_SNP3	MM1483	*S. aethiopicum* gr. Aculeatum
	RAREGILO	*S. aethiopicum* gr. Gilo
	MM1207	*S. aethiopicum* gr. Kumba
GP_SNP4	AN39, IVIA026, UPV29014	*S. aethiopicum* gr. Gilo
	BBS111, MM585, PI413783	*S. aethiopicum* gr. Kumba
	RNL0022	*S. aethiopicum* gr. Shum
Using 11 SSRs plus 35 SNPs		
GP_SSRSNP1	BBS159, RNL0252	*S. aethiopicum* gr. Gilo

When considering only SSR or SNP markers, accessions from different groups of the scarlet eggplant complex (i.e., *S. anguivi*, the Intermediate *S. anguivi-S. aethiopicum* group, and the four groups of *S. aethiopicum*) often shared a same profile. Our data are in agreement with those of other authors who used AFLPs and SSRs (Sunseri et al. 2010), RAPDs (Aguoru et al. 2015), and SNPs (Acquadro et al. 2017) and also found that accessions of *S. aethiopicum* did not cluster according to the cultivar group. This is an additional indication that these groups, which are distinguished on the basis of morphology (Lester 1986; Lester and Daunay 2003; Plazas et al. 2014), are not genetically differentiated.

SSR and SNP diversity statistics

In general, the diversity statistics for the SSRs had higher average values than those of the SNPs (Tables 4, 5). This is common because due to its nature, the potential variability of SSRs is larger than that of SNPs (Ellegren 2004; Kashi and King 2006; Fischer et al. 2017). In our study, the number of alleles obtained with SSRs ranged between three and seven, with an average value of 5.00 (Table 4), while for SNPs it ranged between two and four with an average value of 2.34 (Table 5). Also, the major allele frequency was generally lower with SSRs, ranging from 0.531 to 0.857, with an average value of 0.689, while for SNPs the values were considerably higher, with a range from 0.510 to 0.989, and an average value of 0.851. These values reveal that, despite a having a great diversity, there is always a major allele with a

Table 4 Genetic diversity statistics, including major allele frequency, number of genotypes, number of alleles, expected heterozygosity (H_e), observed heterozygosity (H_o), polymorphic information content (PIC), and coefficient of inbreeding (f) for the 11 SSR markers evaluated in a collection of brinjal, gboma and scarlet eggplants accessions

Marker	Number of alleles	Major allele frequency	Number of genotypes	PIC	H_e	H_o	f
SSR_1	4	0.786	4	0.319	0.354	0.020	0.944
SSR_2	6	0.490	6	0.636	0.679	0.000	1.000
SSR_4	5	0.765	7	0.338	0.379	0.122	0.683
SSR_5	3	0.724	4	0.366	0.422	0.020	0.953
SSR_6	4	0.796	4	0.304	0.339	0.041	0.882
SSR_7	5	0.531	8	0.570	0.625	0.102	0.840
SSR_8	5	0.439	6	0.615	0.672	0.020	0.970
SSR_9	7	0.602	8	0.549	0.587	0.020	0.966
SSR_10	5	0.776	6	0.342	0.373	0.041	0.893
SSR_11	5	0.857	6	0.249	0.259	0.041	0.845
SSR_12	6	0.816	6	0.315	0.326	0.000	1.000
Mean ± SE	5.00 ± 0.33	0.689 ± 0.044	5.91 ± 0.44	0.419 ± 0.043	0.456 ± 0.046	0.039 ± 0.012	0.907 ± 0.028

frequency over 50%, both for SSRs and SNPs. The number of genotypes, as occurs for the alleles, was also higher for SSRs, with a range from four to eight and an average of 5.91 (Table 4), than for SNPs, which presented a range from two to five, with an average of 3.14 (Table 5). As a consequence of the larger number of alleles and lower frequency of the major allele, the PIC values were generally higher for SSRs, with a range from 0.249 to 0.636 and an average of 0.419, than for SNPs, which ranged from 0.021 to 0.386 and an average value of 0.191. Our results are consistent with previous works in eggplants, which find similar values to ours. For example, Vilanova et al. (2012), using genomic SSRs in a collection of 20 accessions of *S. melongena* plus one accession of *S. aethiopicum* and *S. macrocarpon* found similar levels of number of alleles (4.72) and PIC (0.47) than us. Otherwise, other authors found lower values of average PIC rate and number of alleles per locus. For example, Stàgel et al. (2008) found a mean PIC rate of 0.38 and number of alleles per locus of 3.1 when assessed 11 EST-SSRs in 38 *S. melongena* accessions. Also, Nunome et al. (2009) found lower values for the diversity statistics when they genotyped eight lines of *S. melongena* using 1054 genomic SSRs (mean PIC value = 0.27, mean number of alleles = 2.2) and 66 EST-SSRs (mean PIC value = 0.13, mean number of alleles = 1.4).

The expected heterozygosity (H_e) was, on average, higher for SSRs (0.456) than for SNPs (0.224), and much higher than the observed heterozygosity (H_o), which had average values of 0.039 for SSRs and 0.025 for SNPs. The much higher values for H_e compared to H_o result in high levels for the fixation index (f), with average values of 0.907 for SSRs and 0.837 for SNPs. In fact, for some SSR and SNP markers, the *f* value was 1, and so all materials were homozygous for these loci. Our values are similar to those obtained by others with SSRs and SNPs in common eggplant collections (Muñoz-Falcón et al. 2009; Vilanova et al. 2012, 2014, Ge et al. 2013; Augustinos et al. 2016). Although it is known that cultivated brinjal eggplant is fundamentally autogamous (Arumuganathan and Earle 1991; Pessarakli and Dris 2004; Daunay and Hazra 2012) our results also provide evidence that gboma and scarlet eggplants present a reproductive system similar to that of brinjal eggplant.

SSR and SNP-based genetic relationships

Genetic identity values among groups within each of the eggplant complexes had high values, both with SSR and SNP markers (Table 6). Not surprisingly, low values of genetic identities have been obtained between *S. incanum* and *S. aethiopicum* groups, as markers were selected for polymorphism between the

Table 5 Genetic diversity statistics, including number of alleles, major allele frequency, number of genotypes, polymorphic information content (PIC), expected heterozygosity (H_e), observed heterozygosity (H_o), and coefficient of inbreeding (f) for the 35 SNP markers evaluated in a collection of brinjal, gboma and scarlet eggplants accessions

Marker	Number of alleles	Major allele frequency	Number of genotypes	PIC	H_e	H_o	f
SNP_38959	2	0.926	3	0.128	0.138	0.021	0.849
SNP_38905	2	0.714	3	0.325	0.408	0.041	0.902
SNP_11564	3	0.723	4	0.386	0.433	0.043	0.904
SNP_19191	3	0.929	4	0.130	0.135	0.020	0.851
SNP_38971	2	0.949	3	0.092	0.097	0.020	0.793
SNP_34715	3	0.959	3	0.078	0.079	0.000	1.000
SNP_00907	3	0.911	3	0.157	0.165	0.000	1.000
SNP_27060	2	0.739	3	0.311	0.386	0.043	0.890
SNP_13379	2	0.949	3	0.092	0.097	0.020	0.793
SNP_23081	2	0.531	3	0.374	0.498	0.041	0.920
SNP_39035	3	0.724	4	0.373	0.425	0.020	0.953
SNP_23613	2	0.854	3	0.218	0.249	0.042	0.836
SNP_15567	2	0.698	3	0.333	0.422	0.021	0.952
SNP_00676	2	0.696	3	0.334	0.423	0.043	0.899
SNP_19562	2	0.724	3	0.320	0.399	0.020	0.950
SNP_10686	4	0.888	5	0.198	0.206	0.061	0.709
SNP_32044	2	0.745	3	0.308	0.380	0.020	0.947
SNP_37940	2	0.745	3	0.308	0.380	0.020	0.947
SNP_30643	2	0.967	3	0.061	0.063	0.022	0.662
SNP_31222	3	0.918	3	0.148	0.153	0.000	1.000
SNP_01600	3	0.750	4	0.325	0.385	0.042	0.894
SNP_23399	2	0.980	2	0.039	0.040	0.000	1.000
SNP_17586	2	0.949	3	0.092	0.097	0.020	0.793
SNP_02438	2	0.969	3	0.058	0.059	0.020	0.662
SNP_38436	3	0.929	4	0.131	0.135	0.020	0.852
SNP_32294	2	0.946	3	0.098	0.103	0.022	0.793
SNP_14499	2	0.989	2	0.021	0.021	0.021	0.000
SNP_14015	2	0.792	3	0.275	0.330	0.042	0.876
SNP_39817	2	0.949	3	0.092	0.097	0.020	0.793
SNP_14306	3	0.878	4	0.202	0.219	0.082	0.634
SNP_39475	2	0.980	2	0.039	0.040	0.000	1.000
SNP_29844	2	0.948	3	0.094	0.099	0.021	0.793
SNP_14718	3	0.949	4	0.096	0.098	0.020	0.796
SNP_30456	2	0.510	2	0.375	0.500	0.000	1.000
SNP_13910	2	0.966	3	0.064	0.066	0.023	0.661
Mean ± SE	2.34 ± 0.09	0.851 ± 0.022	3.14 ± 0.11	0.191 ± 0.021	0.224 ± 0.027	0.025 ± 0.003	0.837 ± 0.031

transcriptomes of two accessions of these taxa (Gramazio et al. 2016). When considering each of the eggplant complexes, the genetic identities within eggplant complexes have been larger than the between eggplant complexes identities. For example for the brinjal eggplant complex, the within complex genetic

Table 6 Nei (1972) genetic identities among the different groups of the brinjal, gboma and scarlet eggplant complexes based on SSR (above the diagonal) and SNP (below the diagonal) markers

	Brinjal eggplant complex		Gboma eggplant complex		Scarlet eggplant complex					
	S. incanum	*S. melongena*	*S. dasyphyllum*	*S. macrocarpon*	*S. anguivi* (*S. an.*)	Intermediate *S. an.-S. ae.*	*S. aethiopicum* (*S. ae.*)			
Complex/species/groups							Aculeatum	Gilo	Kumba	Shum
Brinjal eggplant complex										
S. incanum		0.810	0.150	0.152	0.293	0.192	0.145	0.169	0.179	0.098
S. melongena	0.588		0.050	0.045	0.439	0.273	0.217	0.310	0.275	0.195
Gboma eggplant complex										
S. dasyphyllum	0.343	0.500		0.935	0.103	0.132	0.101	0.117	0.103	0.103
S. macrocarpon	0.402	0.565	0.911		0.174	0.229	0.199	0.207	0.196	0.192
Scarlet eggplant complex										
S. anguivi	0.162	0.441	0.694	0.722		0.834	0.704	0.887	0.853	0.750
Intermediate *S. an.-S.ae.*	0.165	0.434	0.680	0.743	0.924		0.884	0.955	0.957	0.900
S. aethiopicum Aculeatum	0.172	0.470	0.644	0.734	0.878	0.929		0.901	0.925	0.889
S. aethiopicum Gilo	0.107	0.409	0.644	0.712	0.922	0.981	0.954		0.970	0.922
S. aethiopicum Kumba	0.159	0.436	0.636	0.748	0.859	0.948	0.965	0.976		0.940
S. aethiopicum Shum	0.148	0.443	0.652	0.753	0.904	0.953	0.963	0.973	0.970	

Vertical and horizontal lines separate the different eggplant complexes

identity (i.e., between *S. incanum* and *S. melongena*) has been 0.810 for SSRs and 0.588 for SNPs, while the range for genetic identities with other eggplant complex groups has been between 0.045 (*S. melongena* vs. *S. macrocarpon*) and 0.439 (*S. melongena* vs. *S. anguivi*) for SSRs and between 0.107 (*S. incanum* vs. *S. aethiopicum* group Gilo) and 0.565 (*S. melongena* vs. *S. macrocarpon*) for SNPs (Table 6). Amazingly, in this case, the genetic identity between *S. melongena* and *S. macrocarpon* has been largest with SNPs and lowest with SSRs, reflecting that different levels of genetic diversity are sampled by both markers (Ellegren 2004; Kashi and King 2006; Fischer et al. 2017). In the case of the gboma eggplant complex, the within complex identity has been very high, with a value of 0.935 for SSRs and 0.911 for SNPs, and it has ranged between 0.045 (*S. melongena* vs. *S. macrocarpon*) and 0.229 (*S. macrocarpon* vs. Intermediate *S. anguivi*-*S. aethiopicum*) for SSRs and between 0.343 (*S. incanum* vs. *S. dasyphyllum*) and 0.753 (*S. macrocarpon* vs. *S. aethiopicum* group Shum). For this complex group the genetic identities with other complex groups have been much larger with SNPs than with SSRs (Table 6). Finally, for the scarlet eggplant complex the within complex identity has ranged between 0.704 (*S. anguivi* vs. *S. aethiopicum* group Aculeatum) and 0.970 (*S. aethiopicum* group Gilo vs. *S. aethiopicum* group Kumba) for SSRs, and between 0.859 (*S. anguivi* vs. *S. aethiopicum* group Kumba) and 0.976 (*S. aethiopicum* group Gilo vs. *S. aethiopicum* group Kumba) for SNPs; when considering genetic identities with other groups it has ranged between 0.098 (*S. incanum* vs. *S. aethiopicum* group Shum) and 0.439 (*S. melongena* vs. *S. anguivi*) for SSRs and between 0.107 (*S. incanum* vs. *S. aethiopicum group* Gilo) and 0.753 (*S. macrocarpon* vs. *S. aethiopicum* group Shum) (Table 6). In this case, the genetic identities of scarlet eggplant with the brinjal eggplant have been larger than with the gboma eggplant with SSRs, while the contrary occurred with SNPs. Overall, these results provide evidence that important differences exist among SSR and SNP markers for genetic identities among groups. This may have important consequences for the establishment of relationships among eggplant complexes based on genetic profiles, as depending on the markers used the results may be very different.

SSR and SNP pair-wise genetic distances among all individual accessions displayed a moderate correlation ($r^2 = 0.529$; $b = 0.341$; $P < 0.0001$) (Fig. 1). However, it is evident from Fig. 1 that there is a difference in the relationship between both markers depending on the value of the SNP genetic distance. In this way, for values of SNP genetic distance below 70, the correlation is much higher ($r^2 = 0.592$; $b = 0.534$; $P < 0.0001$) than for values for SNP genetic distance above 70 ($r^2 = 0.080$; $b = 0.045$; $P = 0.0047$). This shows that, for the materials used, when genetic distances based on SNPs are low to intermediate (i.e., within complexes), SNPs and SSRs provide similar levels of information, while when the genetic distances are higher (i.e., among complexes), SNPs provide better resolution, as SSR-based genetic distances seem to reach a saturation. This different performance of both types of markers is very likely due to the much higher rate of mutation of SSRs compared to SNPs (Hamblin et al. 2007; Coates et al. 2009). In this way, when certain levels of phylogenetic distance are reached, the SSR alleles seem to have diverged so much that they are not good to establish relationships based on phylogenetic distance. This has important implications for phylogenetic and germplasm conservation studies.

Cluster analyses

The topology of the SSR-based and SNP-based phenograms is different (Fig. 2). In other studies, the results obtained on the relationships among the three eggplant complexes depended on the markers used (Sakata et al. 1991; Furini and Wunder 2004; Levin et al. 2006; Isshiki et al. 2008; Weese and Bohs 2010; Meyer et al. 2012; Särkinen et al. 2013; Vorontsova et al. 2013; Acquadro et al. 2017). Although in both cases three major clusters are identified, corresponding to each of the three eggplant complexes, in the SSR-based phenogram the gboma eggplant complex cluster is basal to the brinjal and scarlet eggplant complexes, while in the SNP-based phenogram the brinjal eggplant cluster is basal to the gboma and scarlet eggplant complexes. Amazingly, Acquadro et al. (2017), using genomic SNPs found that the scarlet eggplant complex cluster is basal to the brinjal and gboma eggplant complexes. In our case, because markers used here were selected for polymorphism between expressed sequences of one accession of each of *S. incanum* and *S. aethiopicum* (Gramazio et al. 2016), it was expected that the largest distance in the

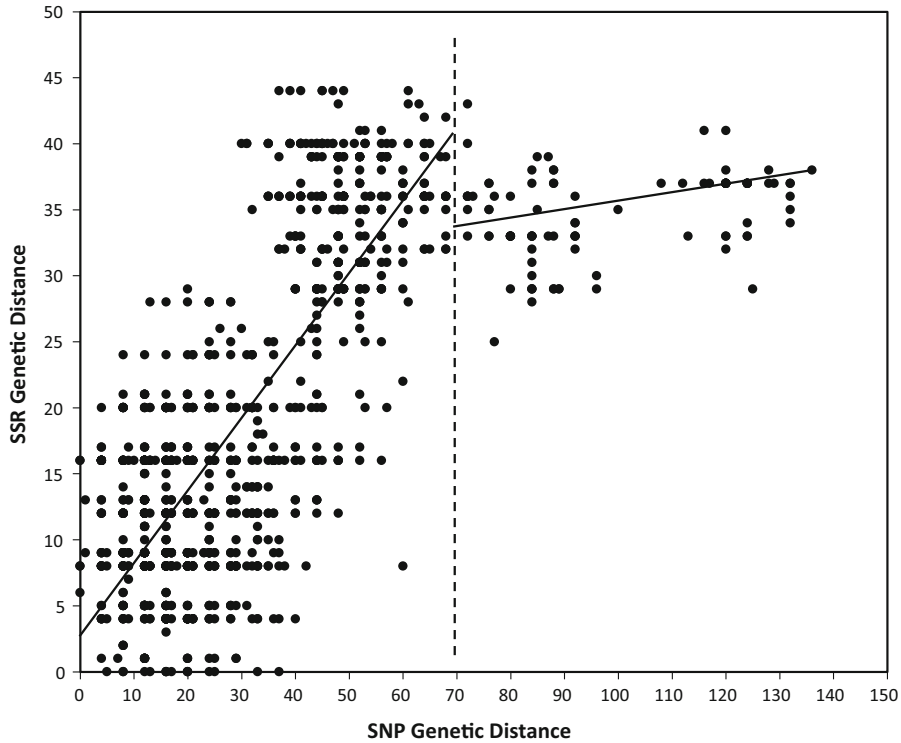

Fig. 1 Relationship between SNP (*X*-axis) and SSR (*Y*-axis) pair-wise genetic distances among 48 individual accessions of the brinjal, gboma and scarlet eggplant complexes

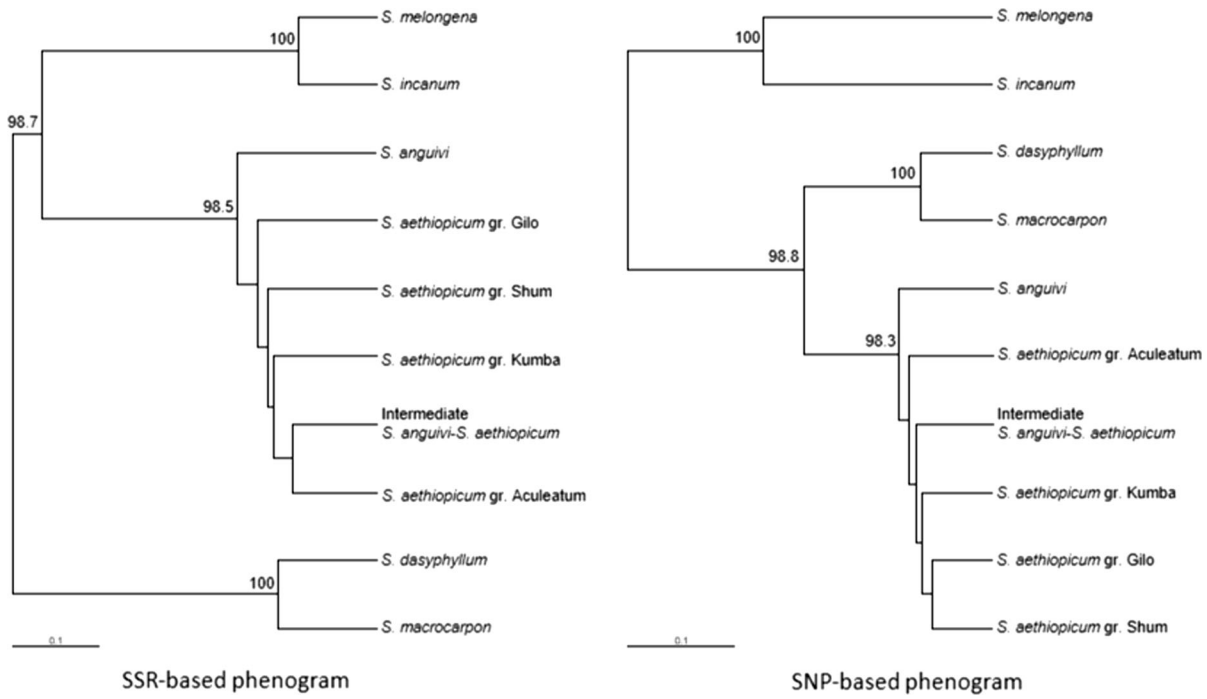

Fig. 2 UPGMA hierarchical clustering consensus phenograms based on Nei (1972) genetic distances for the brinjal (*S. melongena* and *S. incanum*), gboma (*S. macrocarpon* and *S. dasyphyllum*) and scarlet eggplant (*S. aethiopicum* and *S. anguivi*) complex groups according to SSR (left) and SNP markers (right). Bootstrap values (based on 1000 replications; expressed in percentage) greater than 50% are indicated at the corresponding nodes

tree should have been between the clusters of brinjal and scarlet eggplants, but this was only true for SNP markers (Fig. 2). This provides evidence that SNP markers may provide better resolution than SSRs when evaluating phylogenetic relationships among *Solanum* taxa not belonging to the close primary genepool. When considering the scarlet eggplant complex, in both phenograms the wild *S. anguivi* is basal to the other groups, although the other groups present a different clustering pattern depending on the markers used (Fig. 2).

Conclusions

The transcriptome-derived SSR and SNP markers have been highly polymorphic in the eggplants collection evaluated and proved useful for genetic fingerprinting. As expected, SSRs were on average more informative than SNPs, but a similar number of fingerprints were obtained with 11 SSRs and 35 SNPs. Both SSR and SNP markers confirmed the high fixation index of the eggplant materials and clearly distinguished the three eggplant complexes. However, different results depending on the type of marker were obtained for the relationships among eggplant complexes, indicating that they sample different levels of genetic variation. In this respect, SSRs and SNPs presented a moderate correlation for low to intermediate values of SNP pair-wise genetic distance, but a low correlation for high SNP genetic distances. This suggests that both markers are complementary in the information provided, although SNPs seem more appropriate to evaluate materials genetically distant in the eggplant complexes. This information will be useful for eggplants germplasm management, phylogenetic and relationships studies, as well as for genetic fingerprinting and breeding.

Acknowledgements This work has been funded by European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 677379 (G2P-SOL project: Linking genetic resources, genomes and phenotypes of Solanaceous crops) and by Spanish Ministerio de Economía y Competitividad and Fondo Europeo de Desarrollo Regional (Grant AGL2015-64755-R from MINECO/FEDER). Pietro Gramazio is grateful to Universitat Politècnica de València for a pre-doctoral contract (Programa FPI de la UPV-Subprograma 1/2013 call). Mariola Plazas is grateful to Spanish Ministerio de Economía, Industria y Competitividad for a post-doctoral grant within the Juan de la Cierva-Formación programme (FJCI-2015-24835).

References

Acquadro A, Barchi L, Gramazio P et al (2017) Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. PLoS ONE 12:e0180774. https://doi.org/10.1371/journal.pone.0180774

Adeniji O, Kusolwa P, Reuben S (2013) Morphological descriptors and micro satellite diversity among scarlet eggplant groups. Afr Crop Sci J 21(1):37–49

Aguoru C, Omoigui L, Olasan J (2015) Molecular characterization of *Solanum* species (*Solanum aethiopicum* complex; *Solanum macrocarpon* and *Solanum anguivi*) using multiplex RAPD primers. J Plant Stud 4:27–34. https://doi.org/10.5539/jps.v4n1p27

Arumuganathan K, Earle E (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9(3):208–218

Ashrafi H, Hill T, Stoffel K et al (2012) De novo assembly of the pepper transcriptome (*Capsicum annuum*): a benchmark for in silico discovery of SNPs, SSRs and candidate genes. BMC Genom 13:1–15. https://doi.org/10.1186/1471-2164-13-571

Augustinos AA, Petropoulos C, Karasoulou V et al (2016) Assessing diversity among traditional Greek and foreign eggplant cultivars using molecular markers and morphometrical descriptors. Span J Agric Res 14:e0710. https://doi.org/10.5424/sjar/2016144-9020

Avise JC (2012) Molecular markers, natural history and evolution. Springer Science & Business Media, Berlin. https://doi.org/10.1007/978-1-4615-2381-9

Blanca J, Cañizares J, Roig C et al (2011) Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). BMC Genom 12:104. https://doi.org/10.1186/1471-2164-12-104

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32(3):314–331

Bukenya Z, Carasco J (1994) Biosystematic study of *Solanum macrocarpon—S. dasyphyllum* complex in Uganda and relations with *Solanum linnaeanum*. East Afr Agric For J 59(3):187–204

Castillo A, Budak H, Varshney RK et al (2008) Transferability and polymorphism of barley EST-SSR markers used for phylogenetic analysis in *Hordeum chilense*. BMC Plant Biol 8:97. https://doi.org/10.1186/1471-2229-8-97

Choudhary S, Sethy NK, Shokeen B, Bhatia S (2009) Development of chickpea EST-SSR markers and analysis of allelic variation across related species. Theor Appl Genet 118:591–608. https://doi.org/10.1007/s00122-008-0923-z

Coates BS, Sumerford DV, Miller NJ et al (2009) Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis. J Hered 100:556–564. https://doi.org/10.1093/jhered/esp028

D'Agostino N, Golas T, van de Geest H et al (2013) Genomic analysis of the native European *Solanum species, S. dulcamara*. BMC Genom 14:356. https://doi.org/10.1186/1471-2164-14-356

Daunay MC, Hazra P (2012) Eggplant. In: Peter KV, Hazra P (eds) Handbook of Vegetables. Studium Press, Houston, pp 257–322

Davey J, Hohenlohe P, Etter P et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12:499–510. https://doi.org/10.1038/nrg3012

De Barba M, Miquel C, Lobréaux S et al (2016) High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. Mol Ecol Resour 17(3):492–507. https://doi.org/10.1111/1755-0998.12594

Doyle J, Doyle J (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435–445. https://doi.org/10.1038/nrg1348

Felsenstein, J (2007). PHYLIP (Phylogeny Inference Package) Version 3.67. Department of Genome Sciences, University of Washington, Seattle, WA, USA

Fernandez-Silva I, Whitney J, Wainwright B (2013) Microsatellites for next-generation ecologists: a post-sequencing bioinformatics pipeline. PLoS ONE 8(2):e55990

Filippi CV, Aguirre N, Rivas JG et al (2015) Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. BMC Plant Biol 15:52. https://doi.org/10.1186/s12870-014-0360-x

Fischer MC, Rellstab C, Leuzinger M et al (2017) Estimating genomic diversity and population differentiation—an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. BMC Genom 18:69. https://doi.org/10.1186/s12864-016-3459-7

Furini A, Wunder J (2004) Analysis of eggplant (*Solanum melongena*)-related germplasm: morphological and AFLP data contribute to phylogenetic interpretations and germplasm utilization. Theor Appl Genet 108:197–208. https://doi.org/10.1007/s00122-003-1439-1

Gadaleta A, Giancaspro A, Zacheo S et al (2011) Comparison of genomic and EST-derived SSR markers in phylogenetic analysis of wheat. Plant Genet Resour 9:243–246. https://doi.org/10.1017/S147926211100030X

Ge H, Liu Y, Jiang M et al (2013) Analysis of genetic diversity and structure of eggplant populations (*Solanum melongena* L.) in China using simple sequence repeat markers. Sci Hortic 162:71–75. https://doi.org/10.1016/j.scienta.2013.08.004

Gonzaga ZJ (2015) Evaluation of SSR and SNP Markers for Molecular Breeding in Rice. Plant Breed Biotechnol 3:139–152. https://doi.org/10.9787/PBB.2015.3.2.139

Goodwin S, McPherson J, McCombie W (2016) Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 17(6):333–351

Gramazio P, Blanca J, Ziarsolo P et al (2016) Transcriptome analysis and molecular marker discovery in *Solanum incanum* and *S. aethiopicum*, two close relatives of the common eggplant (*Solanum* melongena) with interest for breeding. BMC Genom 17:300. https://doi.org/10.1186/s12864-016-2631-4

Grover A, Sharma PC (2014) Development and use of molecular markers: past and present. Crit Rev Biotechnol 8551:1–13. https://doi.org/10.3109/07388551.2014.959891

Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. PLoS ONE 2:e1367. https://doi.org/10.1371/journal.pone.0001367

Hess JE, Matala AP (2011) Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. Mol Ecol Resour. https://doi.org/10.1111/j.1755-0998.2010.02958.x

Highton R (1993) The relationship between the number of loci and the statistical support for the topology of UPGMA trees obtained from genetic distance data. Mol Phylogenet Evol 2:337–343

Hirakawa H, Shirasawa K, Miyatake K, Nunome, T et al (2014) Draft genome sequence of eggplant (*Solanum melongena* L.): the representative *solanum* species indigenous to the old world. DNA Res 21:649–660. https://doi.org/10.1093/dnares/dsu027

Hong CP, Piao ZY, Kang TW et al (2007) Genomic distribution of simple sequence repeats in *Brassica rapa*. Mol Cells 23:349–356.

Hu J, Wang L, Li J (2011) Comparison of genomic SSR and EST-SSR markers for estimating genetic diversity in cucumber. Biol Plant 55:577–580. https://doi.org/10.1007/s10535-011-0129-0

Isshiki S, Iwata N, Khan MMR (2008) ISSR variations in eggplant (*Solanum melongena* L.) and related Solanum species. Sci Hortic 117:186–190. https://doi.org/10.1016/j.scienta.2008.04.003

Jones ES, Sullivan H, Bhattramakki D, Smith JSC (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). Theor Appl Genet 115:361–371. https://doi.org/10.1007/s00122-007-0570-9

Kalia RK, Rai MK, Kalia S et al (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177:309–334

Kashi Y, King DG (2006) Simple sequence repeats as advantageous mutators in evolution. Trends Genet 22:253–259. https://doi.org/10.1016/j.tig.2006.03.005

Kaushik P, Prohens J, Vilanova S et al (2016) Phenotyping of eggplant wild relatives and interspecific hybrids with conventional and phenomics descriptors provides insight for their potential utilization in breeding. Front Plant Sci 7:677

Kim C, Guo H, Kong W et al (2016) Application of genotyping by sequencing technology to a variety of crop breeding programs. Plant Sci 242:14–22

Knapp S, Vorontsova MS, Prohens J (2013) Wild relatives of the eggplant (*Solanum melongena* L.: Solanaceae): new understanding of species names in a complex group. PLoS ONE 8:e57039

Kruglyak S, Durrett RT, Schug MD, Aquadro CF (1998) Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proc Natl Acad Sci USA 95:10774–10778. https://doi.org/10.1073/pnas.95.18.10774

Lester RN, Daunay MC (2003) Diversity of African vegetable *Solanum* species and its implications for a better understanding of plant domestication. Schriften zu Genetischen Ressourcen 22:137–152

Lester RN, Niakan L (1986) Origin and domestication of the scarlet eggplant, *Solanum aethiopicum*, from *S. anguivi* in Africa. In: D'Arcy WG (ed) Solanaceae: biology and systematics. Columbia University Press, New York, pp 433–456

Lester RN, Jaeger PML, Bleijendaal-Spierings BHM et al (1990) African eggplants-a review of collecting in West Africa. Plant Genet Resour Newsl 81:17–26

Levin R, Myers N, Bohs L (2006) Phylogenetic relationships among the 'spiny solanums' (*Solanum* subgenus Leptostemonum, Solanaceae). Am J Bot 93(1):157–169

Li WH, Gojobori T, Nei M (1981) Pseudogenes as a paradigm of neutral evolution. Nature 292:237–239

Li YC, Korol AB, Fahima T et al (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11:2453–2465

Liu K, Muse S (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220. https://doi.org/10.1038/214637b0

Martínez-Arias R, Calafell F, Mateu E et al (2001) Sequence variability of a human pseudogene. Genome Res 11:1071–1085. https://doi.org/10.1101/gr.167701

Meyer RS, Karol KG, Little DP et al (2012) Phylogeographic relationships among Asian eggplants and new perspectives on eggplant domestication. Mol Phylogenet Evol 63:685–701. https://doi.org/10.1016/j.ympev.2012.02.006

Muñoz-Falcón J, Prohens J, Vilanova S, Nuez F (2009) Diversity in commercial varieties and landraces of black eggplants and implications for broadening the breeders' gene pool. Ann Appl Biol 154(3):453–465

Nandha PS, Singh J (2014) Comparative assessment of genetic diversity between wild and cultivated barley using gSSR and EST-SSR markers. Plant Breed 133:28–35. https://doi.org/10.1111/pbr.12118

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292. https://doi.org/10.1086/282771

Nunome T, Negoro S, Kono I et al (2009) Development of SSR markers derived from SSR-enriched genomic library of eggplant (*Solanum melongena* L.). Theor Appl Genet 119:1143–1153. https://doi.org/10.1007/s00122-009-1116-0

Page R (2001) TreeView. Glasgow University, Glasgow

Peakall P, Smouse R (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research an update. Bioinformatics 28:2537–2539

Pessarakli M, Dris R (2004) Pollination and breeding of eggplants. J Food Agric Environ 2:218–219

Plazas M, Andújar I, Vilanova S et al (2014) Conventional and phenomics characterization provides insight into the diversity and relationships of hypervariable scarlet (*Solanum aethiopicum* L.) and gboma (*S. macrocarpon* L.) eggplant complexes. Front. Plant Sci 5:318

Ranil R, Niran H, Plazas M et al (2015) Improving seed germination of the eggplant rootstock *Solanum torvum* by testing multiple factors using an orthogonal array design. Sci Hortic 193:174–181. https://doi.org/10.1016/j.scienta.2015.07.030

Sakata Y, Lester RN (1997) Chloroplast DNA diversity in brinjal eggplant (*Solanum melongena* L.) and related species. Euphytica 97:295–301. https://doi.org/10.1023/A:1003000612441

Sakata Y, Nishio T, Matthews PJ (1991) Chloroplast DNA analysis of eggplant (*Solanum melongena*) and related species for their taxonomic affinity. Euphytica 55:21–26

Särkinen T, Bohs L, Olmstead RG, Knapp S (2013) A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. BMC Evol Biol 13:214. https://doi.org/10.1186/1471-2148-13-214

Scheben A, Batley J, Edwards D (2017) Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. Plant Biotechnol J 15:149–161

Sneath P, Sokal R (1973) Numerical taxonomy. The principles and practice of numerical classification. W H Freeman Limited, San Francisco

Stàgel A, Portis E, Toppino L et al (2008) Gene-based microsatellite development for mapping and phylogeny studies in eggplant. BMC Genom 9:357. https://doi.org/10.1186/1471-2164-9-357

Sunseri F, Polignano GB, Alba V et al (2010) Genetic diversity and characterization of African eggplant germplasm collection. Afr J Plant Sci 4:231–241

Syfert MM, Castañeda-Álvarez NP, Khoury CK et al (2016) Crop wild relatives of the brinjal eggplant (*Solanum melongena*): poorly represented in genebanks and many species at risk of extinction. Am J Bot 103:635–651. https://doi.org/10.3732/ajb.1500539

Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor Appl Genet 106:411–422. https://doi.org/10.1007/s00122-002-1031-0

Thomson MJ, Alfred J, Dangl J et al (2014) High-throughput SNP genotyping to accelerate crop improvement. Plant Breed Biotechnol 2:195–212. https://doi.org/10.9787/PBB.2014.2.3.195

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. https://doi.org/10.1093/bib/bbs017

Tumbilen Y, Frary A, Daunay MC, Doganlar S (2011) Application of EST-SSRs to examine genetic diversity in eggplant and its close relatives. Turk J Biol 35:125–136. https://doi.org/10.3906/biy-0906-57

van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. Theor Appl Genet 120:1289–1299. https://doi.org/10.1007/s00122-009-1256-2

Van Tassell CP, Smith TPL, Matukumalli LK et al (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Methods 5:247–252. https://doi.org/10.1038/nmeth.1185

Varshney R, Graner A, Sorrells M (2005) Genic microsatellite markers in plants: features and applications. Trends Biotechnol 23(1):48–55

Varshney RK, Chabane K, Hendre PS et al (2007) Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. Plant Sci 173:638–649. https://doi.org/10.1016/j.plantsci.2007.08.010

Vilanova S, Manzur JP, Prohens J (2012) Development and characterization of genomic simple sequence repeat markers in eggplant and their application to the study of diversity and relationships in a collection of different cultivar types and origins. Mol Breed 30:647–660. https://doi.org/10.1007/s11032-011-9650-2

Vilanova S, Hurtado M, Cardona A (2014) Genetic diversity and relationships in local varieties of eggplant from different cultivar groups as assessed by genomic SSR markers. Not Bot Horti Agrobo Cluj-Napoca 42:59–65

Vogel JP, Gu YQ, Twigg P et al (2006) EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*. Theor Appl Genet 113:186–195. https://doi.org/10.1007/s00122-006-0285-3

Vorontsova MS, Stern S, Bohs L, Knapp S (2013) African spiny *solanum* (subgenus Leptostemonum, Solanaceae): a thorny phylogenetic tangle. Bot J Linn Soc 173:176–193. https://doi.org/10.1111/boj.12053

Weese TL, Bohs L (2010) Eggplant origins: out of Africa, into the Orient. Taxon 59:49–56. https://doi.org/10.2307/27757050

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395–420. https://doi.org/10.2307/2406450

Xiao M, Zhang Y, Chen X et al (2013) Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. J Biotechnol 166:122–134. https://doi.org/10.1016/j.jbiotec.2013.04.004

Yan J, Yang X, Shah T et al (2010) High-throughput SNP genotyping with the Goldengate assay in maize. Mol Breed 25:441–451. https://doi.org/10.1007/s11032-009-9343-2

Yang X, Xu Y, Shah T et al (2011) Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. Genetica 139:1045–1054. https://doi.org/10.1007/s10709-011-9606-9

Yu J, Zhang Z, Zhu C et al (2009) Simulation appraisal of the adequacy of number of background markers for relationship estimation in association mapping. Plant Genome 2:63. https://doi.org/10.3835/plantgenome2008.09.0009

Zhan L, Paterson I, Fraser B (2016) MEGASAT: automated inference of microsatellite genotypes from sequence data. Ecol Resour, Mol. https://doi.org/10.1111/1755-0998.12561