Genet Resour Crop Evol (2020) 67:2057–2069
https://doi.org/10.1007/s10722-020-00959-8

RESEARCH ARTICLE

Evidence of genetic diversity within *Solanum Lycopersicum* L. 'Platense' landrace and identification of various subpopulations

Raúl Amado Cattáneo · Andrés Norman McCarthy · Sergio Enrique Feingold

Received: 16 December 2019 / Accepted: 2 June 2020 / Published online: 14 June 2020

Abstract Local varieties and landraces are traditional crops that have evolved over time through adaptation to their natural and cultural environment. They are presently regarded as a valuable genetic reservoir, given that most of the genetic diversity of domesticated species lies in these and other traditionally used varieties. The 'Platense' tomato landrace is adapted to the local soil and climate of the La Plata horticultural green belt, as a result of the gradual selection of a number of tomato cultivars that were introduced in Argentina towards the end of the nineteenth century following the massive immigration flow, mainly from Europe. In the present study we have evaluated the genetic identity and diversity of this landrace and all its subpopulations registered at the Germplasm Bank of the "La Consulta" Agricultural Experimental Station of INTA. The molecular analysis, based on 14 polymorphic microsatellite markers, presented a mean number of alleles per locus of 2.56 with an average polymorphic information content of 0.28. The genetic information obtained allowed the assessment of the genetic variation between and within entries, as well as the identification of all revealed genotypes with an exclusion power of over 99.99%. Likewise, the diversity was evaluated for the whole landrace, as well as for every subpopulation, establishing notably high diversity values both for the complete landrace and a number of its subpopulations. This information is highly useful in the description of the 'Platense' landrace variability and constitutes a solid approach for the genetic characterization of the available material at Germplasm Bank accessions, as well as for other eventual members of this landrace presently in use by local horticulturists.

Keywords Molecular markers · SSR · Genetic diversity · *S. lycopersicon* · Platense tomato landrace · Identity estimation

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10722-020-00959-8) contains supplementary material, which is available to authorized users.

R. Amado Cattáneo · A. N. McCarthy (✉)
Departamento de Ciencias Biológicas-Facultad de Ciencias Exactas-UNLP, Centro Regional de Estudios Genómicos (CREG), Universidad Nacional de La Plata, CP 1900, Blvd. 120 No. 1459, La Plata, Argentina
e-mail: amccarthy@exactas.unlp.edu.ar

A. N. McCarthy
Comisión de Investigaciones Científicas de La Provincia de Buenos Aires (CICPBA), La Plata, Argentina

S. E. Feingold
Laboratorio de Agrobiotecnología, Instituto Nacional de Tecnología Agropecuaria (INTA), EEA, Balcarce CC 276 (7620), Balcarce, Argentina

Introduction

Landraces are populations of cultivated plants, genetically heterogeneous, with a strong cultural and agronomic identity. They constitute local varieties that are in a constant state of evolution, lacking nevertheless of a formal breeding scheme. This complexity favours agricultural biodiversity and is of prime importance to local production and food sovereignty (Villa et al. 2005). The 'Platense' tomato landrace (*Solanum lycopersicum* L.) is a local variety that was adapted during the 1930′s to climate and soil properties of the horticultural belt of La Plata (Buenos Aires, Argentina). Nevertheless, there is not sufficient information to establish its origin unequivocally. One of the most reliable hypothesis argues that it descends from a French cultivar introduced in Argentina with the name of 'Perdrigeon' (Folquer 1976), while other versions suggest a possible Italian origin (Garat 2002). This landrace was rapidly adopted by farmers based on its hardiness and its notable organoleptic fruit properties, until it became the main variety for the production of round fresh tomato throughout Argentina in the mid-1980s. During the period of maximum adoption, it constituted a typically seasonal product, mainly due to its limited post-harvest conservation. This feature, along with the relatively high frequency of commercially relevant morphological defects such as fasciation and cracking, determined its gradual and all but complete substitution by imported -mainly hibrid- commercial cultivars, before the turn of the century. Nevertheless, thanks to the collaboration between a few researchers from the Universidad Nacional de La Plata (UNLP) and a handful of local horticulturists ("Grupo de Productores de Tomate Platense") who still persevered and cultivate this landrace, it was possible to rescue and revalute the 'Platense' tomato landrace as a real local heirloom (Garat 2002). Since 2005, the UNLP hosts the "Fiesta del tomate Platense" an annual fair where the community and local producers meet yearly every February (Garat and Otero 2012).

Several seed banks within and outside Argentina preserve accessions that correspond to improved lines of this cultivar obtained through formal breeding programs (i.e. Uco Plata INTA). However, to date, a systematic survey of the different local populations existing in the region has not been undertaken. According to the testimony gathered among horticulturists, during the period of greatest commercial dissemination of the 'Platense' tomato landrace, there were up to fifteen different 'Platense' subpopulations in the horticultural belt of La Plata, often named after the surname of the horticulturist responsible for their selection and conservation ('Carcione', 'Bustos', 'Del Manso', 'Gentile', 'Prieto', 'Grasso', 'Molinaro', 'Tomaíno', 'Alborghetti', 'Luna', 'Paoletich', 'Sinópoli', 'Volpi', 'Cataldo' and 'Breccia'). Unfortunately, due to the process of substitution by commercial varieties and the death or retirement of the horticulturists who kept them, many subpopulations were lost. Of all these historical subpopulations, only 9 were preserved in the INTA-La Consulta seed bank, and are the ones studied in the present prospection, namely 'Carcione', 'Bustos', 'Del Manso', 'Gentile', 'Grasso', 'Prieto', 'Paoletich', 'Molinaro' and 'Luna'. Although these subpopulations differ morphologically in characters such as fruit shape, number of flowers per cluster, frequency of commercial defects such as fasciation, cracking, etc. (Nico et al. 2006), the levels of genetic diversity of the landrace and its different subpopulations, as well as their genetic relationships, have not yet been studied. We aim to examine the genetic variability present within the 'Platense' tomato landrace and its different subpopulations to distinguish unequivocally the 'Platense' tomato accessions from those belonging to other landraces or commercial cultivars with similar phenotypic attributes in breeding programs. Additionally, understanding genetic diversity in local traditional tomato varieties is important not only for germplasm management but also to promote future legally protected designation of origin.

Among the different molecular markers available, SSRs have gained importance in genetic studies in plants (i.e. population structure, gene flow, evolutionary processes, etc.) and in plant breeding programs (linkage maps, QTLs, marker assisted selection, etc.). We have chosen SSR markers for this study based on their wide genome distribution, reproducibility and co-dominant inheritance (Ditta el al. 2018). Among others, the work of Smulders et al (1997), Bredemeijer et al. (2002), He et al. (2003), Frary et al. (2005), Garcia-Martinez et al. (2006), Song et al. (2006), Ispizúa et al. (2007) and Monte et al. (2018) confirmed the utility of SSRs for studying genetic diversity and variability in the genus *Solanum.*

Materials and methods

Plant material and DNA extraction

In the present study, 13 accessions of cultivated tomato (*Solanum lycopersicum* L.) and one related wild species were assessed. The collection included 9 accessions from the Germplasm Bank of the Agricultural Experimental Station La Consulta (Mendoza, Argentina) of the National Institute of Agricultural Technology (INTA) 1 accession of a commercial hybrid, 3 commercial INTA cultivars and 1 accession of the wild relative *S. pimpinellifolium* L. (LA1589) as an outgroup. Details on accession category and common names are given as supplementary material S1.

Seeds were sown in September (spring), in pots with germination soil mix. Plants were grown in a greenhouse of the "Facultad de Cs. Agrarias y Forestales—UNLP" (photographs of plants in Supplementary file S2). For every accession, a total of 25 individual seeds were planted. From this total, individual samples (100–200 mg of young leaves) were taken from 4 plants (named as a, b, c and d samples), and the remaining 21 plants were bulked and treated as a single sample. Leaf discs from leaves were cut in *eppendorf* tubes and were stored at − 80 °C for lyophilization (Heto FD4, Denmark). The extraction of DNA was performed by the modified method of Doyle and Doyle (1987). Quality determination of DNA extracted from all the samples was performed by electrophoresis in a 1% agarose gel in TBE buffer (89 mM Tris–Borate, 20 mM ethylenediaminetetraacetic acid) and DNA concentration was quantified using a digital spectrophotometer (SmartSpect^{TM} 3000 BIO-RAD, USA) based on 260 nm absorbance and its quality was inferred by 260 nm/280 nm ratio. For all samples, an optimum purity was observed, yielding values between 1.6 and 2.0. The DNA concentration was adjusted to 10 ng/ml with the addition of sterile deionised water.

DNA amplification and detection of microsatellite polymorphisms

The amplification reactions (PCR) were performed in a total volume of 20 µl, containing 25 ng of genomic DNA as template, 50 pmol of each of the primers, 200 µl dNTPs, 2 mM $MgCl_2$, 1 X of Taq polymerase buffer and 1 U Taq DNA polymerase (Promega Madison, WI, USA). Amplification was carried out in 96 well plates by means of the MJ thermocycler (PTC-100 MJ Research, Watertown, Mass., USA) with an initial cycle denaturation of 2 min at 94 °C, 35 cycles with a second step of 15 s at 92 °C, a third step of 15 s at the binding temperature of the primers (Supplementary material S3), and 30 s at 72 °C to extend the fragment. It was finished with a 5 min extension at 72 °C, keeping the extension product at 16 °C.

All amplification products and negative tests were checked by means of a 1% agarose gel electrophoresis in TBE buffer. Electrophoresis in denaturing polyacrylamide gels (6% p/v) was performed and silver-stained as detailed by Benbouza et al. (2006). Band size was determined by comparison with 25 bp DNA ladder (Promega, USA) employing Gel Pro Analyser V3.1 (Media Cybernetics, USA).

Microsatellite loci

The present study was performed using a total of 16 SSR markers, which were selected from literature in order to obtain an even distribution across the whole tomato genome, with at least one marker per chromosome. From a total of 34 markers tested, 14 were polymorphic for all ´Platense´ subpopulations, 15 were monomorphic and 5 resulted non specific. At least one marker candidate showed polymorphism for every chromosome, except for chromosome 6 whose marker candidates resulted monomorphic for all ´Platense subpopulations and were hence discarded. Thus, the selected loci were: LECHI3, Tom152-153, LELEUZIP (Smulders et al. 1997); SSR96, SSR327, 0SSR383 (Frary et al. 2005); Tom47-48, Tom236-237 (Suliman-Pollatschek et al. 2002); TMS63, TMS58, TMS60 TMS42, TMS52 (Areshchenkova and Ganal 2002); TMS42 (Areshchenkova and Ganal 2002); LECHI3 (He et al. 2003) and SSR5, SSR111, SSR43 (Hu et al. 2012). Details on the marker dataset are given in Supplementary material S3.

Data analysis

All the accessions studied were scored with regard to the presence and absence of the corresponding bands among the genotypes. Each microsatellite allele was considered as a dominant marker, which was coded as

'1′ or '0′ to indicate band existence for each allele of the SSR marker respectively.

Genetic distance measures

A marker may be considered as highly informative if any individual chosen at random is likely to be heterozygous for that marker. Markers with many alleles, or highly polymorphic markers, tend to be highly informative. Informativeness can be quantitatively measured by a statistic called the polymorphism information content (*PIC*).

The PIC was calculated according to the expression:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} p_i^2 p_j^2 = 1 - \sum_{i=1}^{n} p_i^2 - \left(\sum_{i=1}^{n} p_i^2\right)^2 + \sum_{i=1}^{n} p_i^4$$

where p_i is the frequency of the marker allele (a_i) and n is the number of different alleles (Hildebrand et al. 1994).

Expected and observed heterozygosity

Values of mean heterozygosity calculated across a number of loci (i.e. H_o: observed frequency of heterozygotes and H_e: expected frequency under Hardy–Weinberg random mating), are valuable parameters used to estimate the degree of genetic variation within a population (or subpopulation). Population structuring occurs when genotype frequencies deviate from the Hardy–Weinberg (HW) expected proportions. If either inbreeding or selection occurs, then populations can be considered "structured" in some way (Sbordoni et al. 2012).

Inbreeding coefficient

Although non random mating has no effect on allele frequency, inbreeding causes genotype frequencies to change, given that the frequency of homozygotes increases and the frequency of heterozygotes decreases. Thus, given that non-random mating only reshuffles genotype frequencies with respect to their HW expectations, we can use the deviation of genotype frequencies from their expected values as a measure of inbreeding. If H_o is the observed frequency of heterozygotes and H_e is the expected frequency under random mating, then the inbreeding coefficient is for any subpopulation is given by:

$$F_{local} = 1 - \frac{H_o}{H_e}$$

If there are no heterozygotes in the population then the inbreeding coefficient is 1.0.

When the frequency of heterozygotes equals the HW expectation, then the inbreeding coefficient is 0. In cases where there is an excess of observed heterozygotes, then the inbreeding coefficient can be negative.

We also calculate the different F indexes (F_{it},F_{is} andF_{st}), in order to measure the deficiency of heterozygotes due to inbreeding within subpopulations:$Fit = 1 - \frac{H_i}{H_t}$; $Fis = 1 - \frac{H_i}{H_s}$; $Fst = 1 - \frac{H_s}{H_t}$

The F_{is} index quantifies the mean deficiency of observed heterozygotes among individuals with respect to that expected across subpopulations, F_{it} index quantifies this deficiency with respect to that expected for the total population, whilst the F_{st} index represents the mean deficiency of expected heterozygotes among subpopulations with respect to that expected for the total population. In order to calculate these three F indexes we first evaluate the heterozygosity indices at different levels of the population structure (H_i, H_s y H_t). The H_i and and H_s are the observed and expected heterozygosity probability for any individual drawn at random from the global population. H_i and H_s differ when sub populations have different genetic structures. The H_t is simply the global expectation of heterozygosity based on the global allele frequencies.

Diversity index

Genetic diversity was calculated as proposed by Weir, according to the expression:

$$D = 1 - \left(\frac{1}{m}\right)\sum_{l=1}^{n}\sum_{u=1}^{n} p_{lu}^2$$

where p_{lu} is the frequency of the uth allele at the lth locus and m is the number of loci (Weir and Cockerham 1996). The frequencies were calculated based on the six loci that resulted informative for the whole of the 'Platense' tomato landrace individuals

studied. The same loci were used for calculating the diversity index of every 'Platense' subpopulation, as well as the total diversity of all accessions studied.

Identity estimation ability

The probability of random identity (*PRI*) for all SSR markers was calculated as follows:

$$PRI = \prod_{i=1}^{n} p_i$$

where p_i is the frequency of allele i and n is the number of evaluated markers.

Based on the *PRI* values, the percentage of excluding power (*EP*) was calculated as follows:

$$EP = (1 - PRI) \times 100$$

which indicates the power to unambiguously identify a given individual or subpopulation among all others in the set, using the 16 SSR markers described in this study (Ribeiro et al. 2017).

Dendrograms

Genetic similarity was estimated from binary matrices using Jaccard's coefficient (Sneath and Sokal 1973). A cluster analysis was performed on all individual and pooled accession samples. Similarity dendrogram based on SSR polymorphisms was conducted on the basis of the binary data using the software NTSys 2.02. (Rohlf 2000) with the unweighted pair group method using arithmetic averages (UPGMA) method.

Principal coordinate analysis

A genetic distance matrix was calculated as Jaccard's similarity coefficient using the software NTSys pc 2.02. (Rohlf 2000). A classical multidimensional scaling (MDS) analysis was performed based on the dissimilarity matrix, and a plot for the first two coordinates was generated. Additionally, confidence ellipses were calculated for each 'Platense' landrace subpopulation in the PCoA graph, assuming normally distributed data along both principal coordinate axes. These error ellipses define the region that contains 95% of all samples belonging to the expected Gaussian distribution.

Results and discussion

Total allele heterozygosity

All the information generated by the molecular markers was transferred to a binary matrix for subsequent analysis. This data is available in worksheet S4 of the supplementary material.

A total of 16 SSR markers were used for studying a total of 13 *S. lycopersicum* accessions, composed by 3 INTA cultivars, one commercial hybrid and 9 subpopulations of the 'Platense' tomato landrace ('Bustos', 'Prieto', 'Grasso', 'Molinaro', 'Paolettich', 'Luna', 'Gentile', 'Carcione' and 'Del Manso'), and a *S. pimpinellifolium* accession (LA1589) as outgroup.

Of the total SSR primer pairs, 87.5% (14 SSR) revealed polymorphisms and only 2 were monomorphic. The average number of bands per locus was 2.71, with a total of 41 alleles for all the accessions studied. Markers SSR111 and Tom236-237 presented the maximum number of alleles, with 5 per locus (Table 1).

Table 1 Expected heterozygosity (H_e), Inbreeding coefficient (F_{local}), Observed heterozygosity (H_o) for 16 SSR loci studied in tomato genotypes (commercial cultivars, landrace subpopulations, hybrid and wild relative)

SSR name	No. of alleles	H_o	H_e	F_{local}	PIC
TMS63	3	0.08	0.31	0.74	0.27
SSR5	2	0	0.13	1	0.12
SSR96	3	0.14	0.34	0.59	0.30
TMS58	2	0	0.13	1	0.12
SSR111	5	0.76	0.59	− 0.29	0.52
Tom47-48	2	0	0.13	1	0.12
SSR43	1	0	0	–	0
Tom152-153	3	0	0.26	1	0.24
TMS60	3	0.93	0.56	− 0.66	0.47
LELEUZIP	2	0.76	0.47	− 0.62	0.36
SSR327	2	0	0.13	1	0.13
Tom236-237	5	0.20	0.56	0.64	0.56
SSR383	2	0.08	0.42	0.81	0.42
LECHI3	2	1	0.50	− 1	0.38
TMS42	2	0	0.24	1	0.41
TMS52	2	0.86	0.50	− 0.72	0.38
Mean	2.56	0.32	0.35		0.28

The PIC values ranged from 0.12 (locus TMS58, Tom47-48 and SSR5) to 0.56 (locus Tom236-237) with an average of 0.28.

Highly informative SSR markers (high PIC) such as SSR111 and Tom236-237 proved very useful both for variety identification and genetic assessment of these tomato accessions.

The degree of polymorphism was found adequate for all subsequent diversity analysis. The highest (0.93) and lowest (0) values of the observed heterozygosity (H_o) were shown by TMS60 and SSR5, TMS58, Tom47-48, SSR43, Tom152-153, SSR327 respectively, with a total average of 0.32.

Likewise, the highest and lowest values of inbreeding coefficient (F_{local}) were shown by SSR383 and TMS52, ranging from 0.81 to -0.72.

Mean subpopulation heterozygosity

Every 'Platense' subpopulation studied shows a negative mean Inbreeding coefficient, ranging from -0.05 for 'Paolettich' to -1 for 'Molinaro' and 'Bustos'. This means that an average excess heterozygosity is observed for each of the subpopulations studied (Table 2).

These cultivated tomato varieties would be expected to fixate their alleles in homozygote state, mainly due to preferential autogamy of this species. This unexpectedly high degree of heterozygosity could be partially explained by various alternative or concurrent factors. As regards genetic variability between generations, since the SSR mutation rates (10^{-2}–10^{-6} events per locus per generation) are very high, as compared with the rates of point mutation at coding gene loci (Li et al. 2002), the appearance of new alleles cannot be ruled out as a possible factor. Additionally, there are various factors that interfere with the expected autogamy of tomato reproduction, which in principle is expected to show a minor, although not negligible (5% of cross pollination), degree of insect mediated pollination (Argerich and Gaviola 1995). In the first place, it has recently been proven that tomato stigma exsertion may be induced by high ambient temperatures (Pan et al. 2019), which would expectedly enhance cross pollination. Furthermore, an additional factor could be the fact that local horticulturists frequently select the so-called "Florones" as the fruits from which to extract seeds (Ahumada et al. 2011). The "Florón" is the tomato fruit that is formed from fasciated (fused) flowers, in which the reproductive parts frequently show higher exposition to the ambient.

Table 2 Mean allele values of observed heterozygosity (H_o), expected heterozygosity (H_e) and Inbreeding coefficient (F_{local}) for every 'Platense' tomato subpopulation

Subpopulation	H_o	H_e	F_{local}
'Prieto'	0.19	0.12	− 0.60
'Paolettich'	0.20	0.19	− 0.05
'Molinaro'	0.25	0.13	− 1
'Luna'	0.28	0.16	− 0.8
'Grasso'	0.31	0.21	− 0.5
'Gentile'	0.19	0.11	− 0.64
'Del Manso'	0.30	0.22	− 0.37
'Carcione'	0.27	0.17	− 0.6
'Bustos'	0.31	0.16	− 1

Subpopulation genetic differentiation

In order to measure the extent of genetic differentiation among subpopulations, we evaluated the subpopulation Fixation index (F_{st}), which quantifies the mean reduction in H of a subpopulation (relative to the total population) due to genetic drift among subpopulations, as well as the subpopulation Inbreeding coefficient (F_{is}) and the Overall Fixation Index (F_{it}). The results are shown in Table 3, and allow us to evaluate sub-population differentiation within the total platense landrace.

Although F_{st} has a theoretical range of 0 to 1.0, the observed maximum is usually much less than 1.0. As shown in Table 3, in the case of the 'Platense' and its subpopulations, F_{st} has a value of 0.33. As suggested by Wright (1978), values of F_{st} above 0.25 indicate very great genetic differentiation. That is, 'Platense'

Table 3 Mean individual expected and observed heterozygosity (H_i y H_s), mean expected global heterozygosity (H_t) and F Indexes (F_{it},F_{st},F_{is}) for the 'Platense' tomato population

H_i	H_s	H_t
0.26	0.16	0.24
F_{is}	F_{it}	F_{st}
− 0.63	− 0.08	0.33

landrace shows a substantial genetic differentiation among all the subpopulations presently studied. A clear way to resume this result is the following: 33% of the total variation is distributed among subpopulations, whilst 67% is distributed among individuals within subpopulations.

Diversity analysis

All 14 accessions were included in the Diversity Index analysis (9 "La Consulta" platense accessions, 3 commercial INTA accessions, 1 commercial Hybrid accession and 1 wild tomato relative accession). The number of individual samples used was four for each platense accession and one for every other accession (i.e. all individual samples), giving a total of 41 samples, which allowed for a valid estimation of the corresponding allele frequencies used in the calculation. The values of the diversity index (*D*) were calculated for every 'Platense' subpopulation accession, and the index ranged from 0.08 to 0.41, with a mean value of 0.23 (Table 4). The minimum *D* value was found for the 'Molinaro' subpopulation (0.08), which could indicate that this particular accession has undergone a higher degree of selection than its other landrace counterparts, arguably indicating that it may be regarded more as a cultivar than a diverse subpopulation from a genetic point of view. On the other hand, the highest intra subpopulation diversity (0.41) was shown by the 'Del Manso' accession, almost as high as the diversity value shown by the whole Platense landrace (0.42).

The diversity index for the 'Platense' landrace (calculated on the basis of all the platense accessions studied) was only slightly lower than the diversity index calculated for all accessions in the present study (0.46). These results indicate that the 'Platense' tomato landrace is comprised by a general population in which substantial diversity is maintained. Nevertheless, when regarding the subpopulation level, very different degrees of diversity are observed.

Table 4 Diversity Index (*D*) for all accessions, the whole 'Platense' landrace and each 'Platense' subpopulation

Cluster/landrace/subpopulation	*D*
All accessions	0.46
Whole 'Platense' landrace	0.42
'Del Manso'	0.41
'Paolettich'	0.35
'Grasso'	0.31
'Luna'	0.25
'Carcione'	0.23
'Bustos'	0.17
'Prieto'	0.15
'Gentile'	0.14
'Molinaro'	0.08

Identity estimation ability

The set of 14 SSR informative markers generated a unique allelic combination for each accession evaluated, with the probability of random identity values ranging from 10 e−11 and 10 e−15, for commercial hybrid Elpida and *S. pimpinellifolium*, to 10 e−4 for UCO Plata INTA and 'Molinaro'. This corresponds to values of exclusion power of over 99.99% (Table 5) for every accession studied.

In order to evaluate all SSR markers used, the value of PRI per marker was calculated, consequently obtaining the average PRI per marker for all

Table 5 Probability of random identity (*PRI*) and the percentage of exclusion power (*EP*) of the informative 14 SSR markers for each of the accessions evaluated. ('Platense' tomato subpopulations, INTA cultivars, hybrid and wild relative)

Accession	*PRI*	*EP* (%)
Uco Plata INTA	1.3784e−04	99.986216
Uco 14 INTA	4.5577e−08	99.999995
Uco 18 M INTA	1.4137e−05	99.998586
'Prieto'	2.3172e−05	99.997683
'Paolettich'	3.7225e−07	99.999963
'Molinaro'	1.0452e−04	99.989548
'Luna'	1.1840e−05	99.998816
'Grasso'	9.0614e−07	99.999909
'Gentile'	5.2492e−05	99.994751
'Del Manso'	7.6519e−07	99.999923
'Carcione'	1.1717e−06	99.999883
'Bustos'	1.4768e−05	99.998523
Elpida F_1	5.6952e−11	> 99.999999
S. pimpinellifolium	5.3686e−15	> 99.999999

accessions, thus establishing a marker rank from the highest to the lowest average PRI. The result is available in supplementary material (S5).

This marker rank was subsequently used to calculate the EP vs number of markers used. The results are shown in Fig. 1. From this representation it may be readily concluded that, in addition to the fact that the use of the 14 informative markers allows us to identify all the accessions with an EP of at least 99.99%, the use of the 9 highest ranked markers renders identity at an EP of over 99.5%, and the use of only the first 6 ranked markers allows for identification with an EP of over 95%. Thus, this set of 14 SSR markers confers a high distinctiveness capacity and could be recommended as a molecular database for the 'Platense' tomato landrace subpopulations evaluated in this study, either for seed bank management or participative breeding program purposes. Additionally, a subset of these markers could be used in eventual non exhaustive 'Platense' landrace identification assays.

Dendrogram

The results shown in Fig. 2 are in general accordance with the diversity values obtained for all accessions studied, as well as for the whole 'Platense' landrace and its subpopulations. Namely, three of the subpopulations with lowest diversity index ('Molinaro', 'Bustos' and 'Prieto') appear closely grouped in the dendrogram, whilst the subpopulation with highest diversity index ('Del Manso') appears completely ungrouped, showing likewise almost complete unclustering for the subpopulation with the second highest diversity index ('Paolettich'). The other three subpopulations with mean diversity index values ('Grasso', 'Luna' and 'Carcione'Luna') reasonably show an intermediate clustering behavior.

Regarding the non Platense accessions, although the Uco 18 M INTA round tomato cultivar appears in the middle of the 'Platense' cluster, it is nevertheless connected at a great distance (0.79 Jaccard coefficient) to the closest accession of this landrace, almost at the root of the Platense cluster (0.74 Jaccard coefficient). Moreover, this distance is 1,615 times greater than the greatest distance separating any two Platense accessions within the cluster (i.e. 0.87 Jackart coefficient for 'Paolettich' and 'Del Manso' subpopulations). On the other hand, the commercial cultivar Uco Plata INTA shows the exact same SSR allele pattern as the b-'Luna' individual (1.00 Jaccard coefficient) and sister to the 'Luna' subpopulation, indicating a maximum degree of genetic similarity with this accession. This comes as no surprise given that INTA describes this cultivar as a tall platense type tomato variety (Lepez 2020). Moreover, this result confirms the general 'Platense' pedigree of UCO Plata INTA,

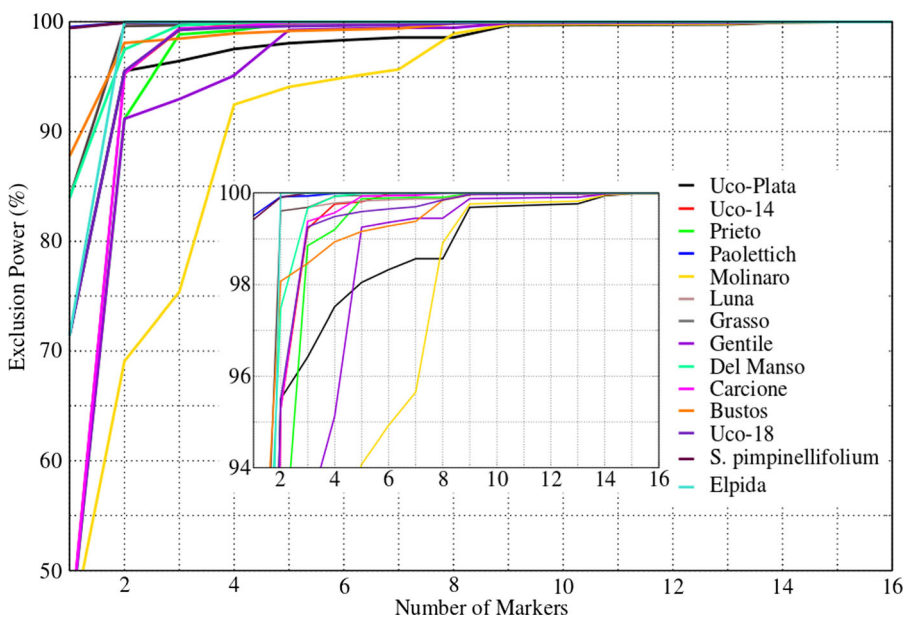

Fig. 1 Exclusion power vs. the number of highest ranked SSR markers used

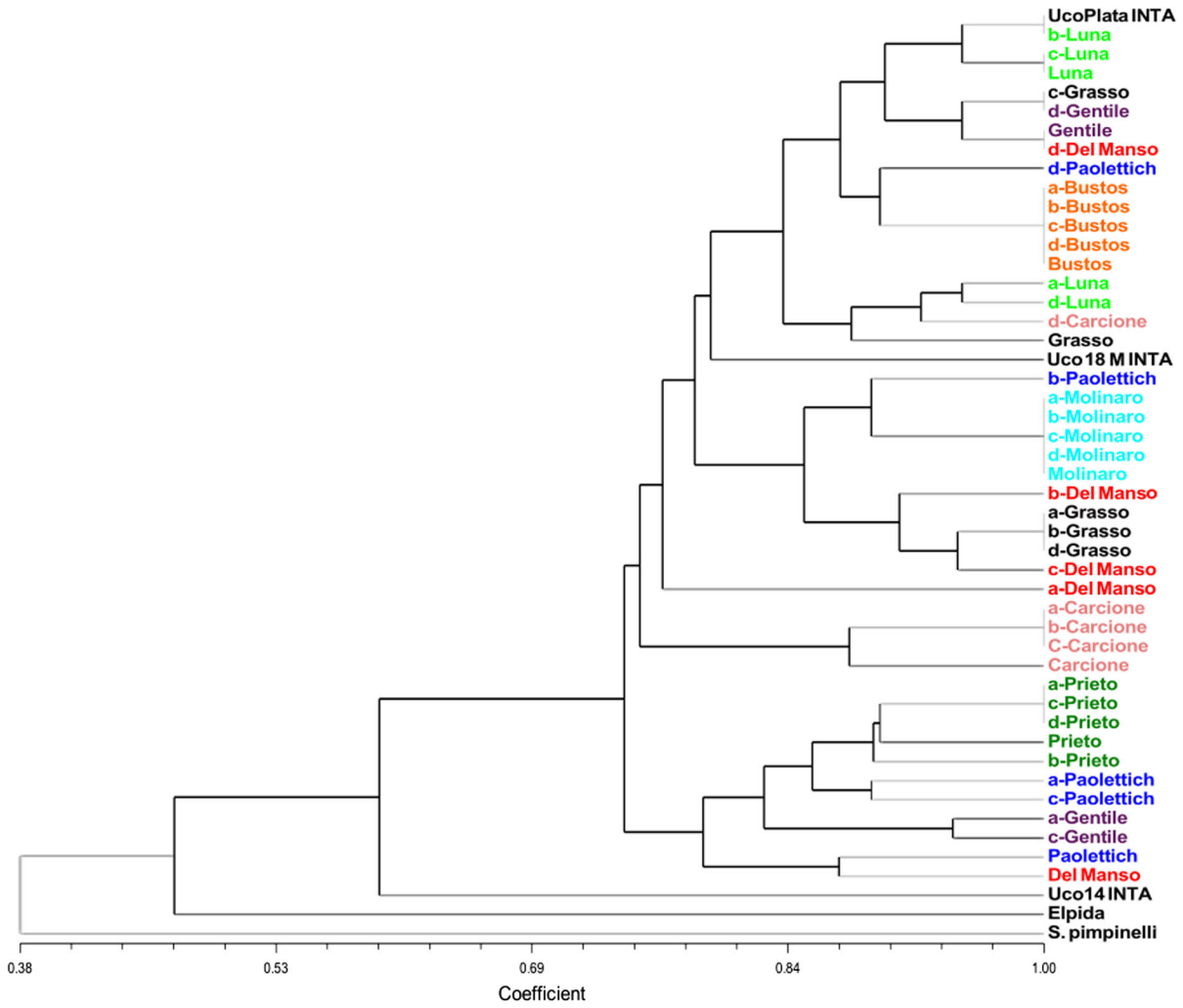

Fig. 2 Dendrogram for the 'Platense' tomato subpopulations and individual samples, INTA cultivars, hybrid and *S. pimpinellifolium* accessions, constructed using the UPGMA method. The scale is based on the Jaccard's similarity coefficient

and situates the 'Luna' subpopulation as its closest relative among the 'Platense' accessions available in the INTA 'La Consulta' seed bank. Finally, Uco 14 INTA (an ellipsoid tomato) reasonably appears as sister to the whole platense cluster, whilst the Elpida commercial hybrid is sister to all the latter, and *S. pimpinellifolium* expectedly appears as outgroup.

Principal coordinates analysis

A Principal Coordinates Analysis (PCoA) was performed for all individuals and subpopulations of the 14 accessions studied. Results are shown in Fig. 3. The first two coordinates of the PCoA analysis account for 36.8% of the total variation revealed by the 14 polymorphic SSR markers among all the accessions studied, with the first and second coordinates capturing 23.1% and 13.7% of the total variation, respectively (Fig. 3). Additionally, a confidence ellipse is shown for each 'Platense' landrace subpopulation in the PCoA graph, assuming 2D normally distributed data. These error ellipses represent an iso-contour of the Gaussian distribution for each subpopulation, and allow for the visualization of a 2D 95% confidence interval, defining the region that contains 0.95 of all samples that could be drawn from an underlying Gaussian distribution.

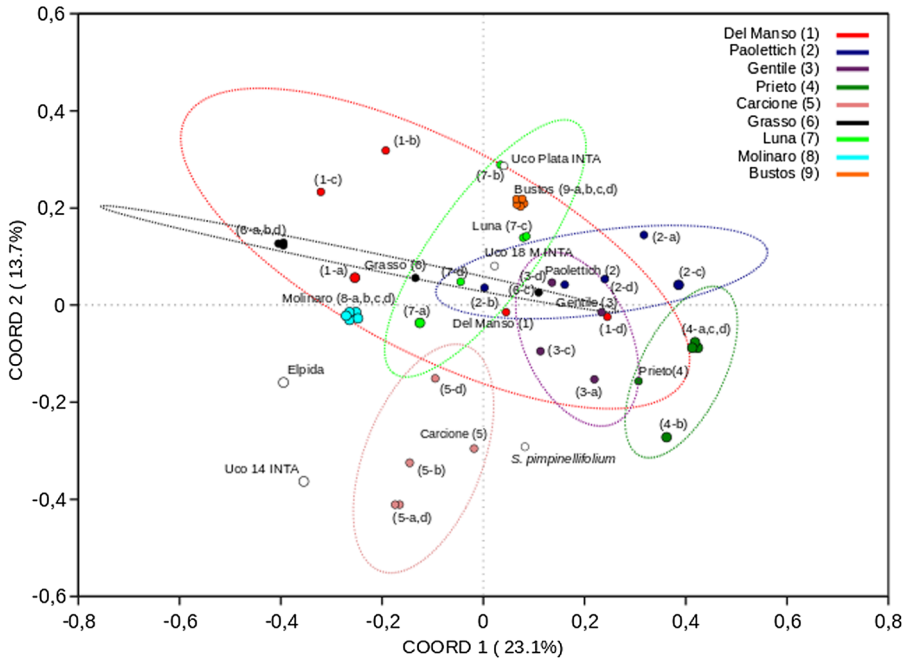

Fig. 3 Principal coordinates analyses based on the SSR data for every individual and subpopulation accession studied. The first and second coordinates account for 23.1% and 13.7% variation, respectively

The area encircled by each oval gives an idea of the variability captured by the first two Principal Coordinates for each subpopulation. The largest confidence ellipse is shown by 'Del Manso', followed by 'Paoletich', 'Grasso' and 'Luna', whist 'Molinaro and 'Bustos' show zero variability in this plane, and 'Carcione', 'Prieto and 'Gentile' show intermediate values. This confirms the general observations previously established by the diversity analysis of Table 1, as well as the clustering analysis of Fig. 2; i.e. that the 'Platense' subpopulations show very different degrees of variability, spanning from very genetically diverse cases (such as 'Del Manso' or 'Paoletich') to other very homogenous ones (such as 'Molinaro and 'Bustos').

The union of all subpopulation ellipses accounts for the total 'Platense' landrace variability captured by the first two Principal Coordinates, and encompasses almost half of the PCoA plane. Furthermore, the intersection between ellipses gives an idea of the degree of genetic similarity between the different subpopulations. Notably, all nine 'Platense' subpopulations constitute a common cluster. Nevertheless, variable degrees of connectivity between each other are observed, spanning from loosely connected subpopulations ('Carcione' connected only to 'Del Manso' and 'Luna', or 'Prieto' connected to 'Paoletich', 'Gentile' and 'Del Manso') to a totally connected one ('Del Manso'), with the rest showing intermediate connectivity. Finally, the least diverse subpopulations ('Molinaro' and 'Bustos') appear within the 'Platense' cluster, although without a confidence ellipse, given that all their individuals appear exactly at the same point of the PCoA plane.

Regarding the non 'Platense' accessions presently studied, although Uco 18 M INTA appears almost in the middle of the main 'Platense' cluster, it is nevertheless nonadjacent to any particular platense accession. On the other hand, Uco Plata INTA appears in the same position as the b-'Luna' individual. Contarily to these, the last three non 'Platense' accessions appear clearly outside of the 'Platense' regions of the PCoA plane, as is expected by the results shown in the diversity and clustering analysis sections.

Conclusions

Despite the fact that a great number of local landraces have historically been cultivated in the La Plata region (Nico et al. 2006), there are few studies focused on the genetic characterization of such collections. In the present study we have genetically identified the 'Platense' tomato landrace and its subpopulation accessions presently available at the 'La Consulta' INTA Germplasm Bank.

Tomato is generally considered to present a narrow genetic base, even among autogamous cultivated species. Nevertheless, tomato landraces are typically considered to contain higher genetic and phenotypic variability than commercial cultivars and hybrids (Park et al. 2004; Mazzucato et al. 2008), thus constituting a valuable resource for improvement programs. In this respect, the present study has determined the 'Platense' tomato landrace as a highly diverse resource, with some of its subpopulations retaining most of the total diversity. On the other hand, few subpopulations presented genetic homogeneity (for the markers used) and thus may be regarded as cultivars). In general, seed replacement is common among farmers. Moreover, seed exchange is the most common way by which highly self-pollinating annual crops become disseminated (Zeven 1999). Such practices generally make it difficult to identify clear "genetic" boundaries, being the 'Platense' tomato landrace no exception to the rule.

Additionally, these results may be used as reference for future studies on 'Platense' germplasm samples of diverse origin, such as other landrace subpopulations currently in use by local horticulturists, commercial seeds, or 'Platense' accessions presently available in other seed banks around the world. This alternative may prove of relevance given that during the last few years the local production and commercialization of this landrace has been socio-culturally revalued for both its organoleptic quality and its adaptability to grow under agroecological conditions (Garat and Otero 2012).

Since the early 20th century local farmers of the La Plata horticultural belt progressively performed concurrent selection processes on diverse tomato varieties, most probably from French or Italian origin (Folquer 1976; Garat 2002), thus yielding the diverse yet characteristic 'Platense' tomato landrace. This complex and irregular procedure may nevertheless be regarded as an informal selection and improvement process. Such rich horticultural heritage, combined with the results here presented, open the possibility of a novel participative improvement scheme, integrating local horticulturists, consumers and researchers under a common goal: the consolidation of the 'Platense' landrace. Concurrently, these results reaffirm the role of in situ conservation by local horticulturists in the preservation of tomato genetic resources, allowing for both natural and human selection to concur synergically.

Acknowledgments We wish to thank Dra. McCarthy, C.B. (CREG-UNLP) for the molecular marker selection, Dra. Peralta, I. (Fac. Cs. Agrarias-UNCuyo) for sharing *S. pimpinellifolium* LA1589 seeds and Dra. Gabriela Massa, Dra. Natalia Norero y Lic. Martín Carboni from the Laboratorio de Agrobiotecnología—EEA Balcarce, INTA. We also want to recognize Ing. Agr. Garat, J.J. (Fac. Cs. Agrarias y Forestales-UNLP) for the rescue and conservation of the 'Platense' tomato subpopulation seeds. In memory and recognition of Guillermo Gallardo, tomato breeder of the UCO 's cultivars from INTA La Consulta.

References

Ahumada J, Otero J, Y Garat JJ (2011) Las hortalizas típicas del Cinturón Verde de la Plata (no son cualquier verdura). La Plata, Argentina: Universidad Nacional de La Plata. ISBN:978-987-33-0047-9

Areshchenkova T, Ganal MW (2002) Comparative analysis of polymorphism and chromosomal location of tomato microsatellite markers isolated from different sources. Theor Appl Genet 104(2–3):229–235

Argerich CA, Gaviola JC (1995) Manual de producción de semillas hortícolas. Tomate. Instituto Nacional de Tecnología Agropecuaria. Argentina. https://inta.gob.ar/documentos/manual-de-produccion-de-semillas-horticolas.-tomate

Bauchet G, Causse M (2012) Genetic diversity in tomato (*Solanum lycopersicum*) and its wild relatives. In: Genetic diversity in plants, vol 133. p 162

Benbouza H, Jacquemin JM, Baudoin JP, Mergeai G (2006) Optimization of a reliable, fast, cheap and sensitive silver staining method to detect SSR markers in polyacrylamide gels. Biotechnol Agron Soc Environ 10(2):77–81

Bredemeijer G, Cooke R, Ganal M, Peeters R, Isaac P, Noordijk Y, Dijcks M (2002) Construction and testing of a microsatellite database containing more than 500 tomato varieties. Theor Appl Gen 105(6–7):1019–1026

Corrado G, Caramante M, Piffanelli P, Rao R (2014) Genetic diversity in Italian tomato landraces: implications for the development of a core collection. Sci Hortic 168:138–144

D'Angelo M, Zanor MI, Sance M, Cortina PR, Boggio SB, Asprelli P, Valle EM (2018) Contrasting metabolic profiles of tasty Andean varieties of tomato fruit in comparison with commercial ones. J Sci Food Agric 98(11):4128–4134

Ditta A, Zhou Z, Cai X, Wang X, Okubazghi K, Shehzad M, Wang K (2018) Assessment of genetic diversity, population structure, and evolutionary relationship of uncharacterized genes in a novel germplasm collection of diploid and allotetraploid Gossypium accessions using est and genomic ssr markers. Int J Mol Sci 19(8):2401

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19:11–15

Folquer F (1976) El tomate: estudio de la planta y su producción comercial. Editorial Hemisferio Sur, Buenos Aires, Argentina

Frary A, Xu Y, Liu J, Mitchell S, Tedeschi E, Tanksley S (2005) Development of a set of PCR-based anchor markers encompassing the tomato genome and evaluation of their usefulness for genetics and breeding experiments. Theor Applied Gen 111(2):291–312

Garat JJ (2002) El tomate platense en La Plata, Argentina. Biodiversidad 34:19–21

Garat J, Otero J (2012) ¿Desde dónde intervenimos cuando intervenimos? El caso del "proyecto Tomate Platense". Mundo Agrario, Revista de estudios rurales 12(24). ISSN: 1515-5994

García-Martínez S, Andreani L, Garcia-Gusano M, Geuna F, Ruiz JJ (2006) Evaluation of amplified fragment length polymorphism and simple sequence repeats for tomato germplasm fingerprinting: utility for grouping closely related traditional cultivars. Genome 49(6):648–656

He C, Poysa V, Yu K (2003) Development and characterization of simple sequence repeat (SSR) markers and their use in determining relationships among *Lycopersicon esculentum* cultivars. Theor Appl Gen 106(2):363–373

Hildebrand CE, David C, Torney C, Wagner P (1994) Informativeness of polymorphic DNA markers. Los Alamos Sci 20:100–102

Hu X, Wang H, Chen J, Yang W (2012) Genetic diversity of Argentina tomato varieties revealed by morphological traits, simple sequence repeat, and single nucleotide polymorphism markers. Pak J Bot 44(2):485–492

Ispizúa VN, Guma IR, Feingold S, Clausen AM (2007) Genetic diversity of potato landraces from Northwestern Argentina assessed with simple sequence repeats (SSRs). Genet Resour Crop Evol 54:1833–1848

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177(3):309–334

Lepez R (2020) Semillas hortícolas. Tomate (*Lycopersicon esculentum*). https://inta.gob.ar/sites/default/files/script-tmp-inta_semillas_hortcolas__tomate__lycopersicon_esculen.pdf

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11(12):2453–2465. https://doi.org/10.1046/j.1365-294x.2002.01643.x

Lopez-Lambertini P, Biderbost E, Di Feo LY, Mollinedo V (2003) Relación entre la concentración viral y la tolerancia al TSWV determinada por el gen "Platense" en tomate. Fitopatología 38:23–31

Mazzucato A, Papa R, Bitocchi E, Mosconi P, Nanni L, Negri V, Picarella ME, Siligato F, Soressi GP, Tiranti B (2008) Genetic diversity, structure and marker-trait associations in a collection of Italian tomato (*Solanum lycopersicum* L.) landraces. Theor Appl Genet 116(5):657–669

Monte MN, Rey Burusco MF, Carboni MF, Castellote MA, Sucar S, Norero NS, Colman SL, Massa GA, Colavita ML, Feingold SE (2018) Genetic diversity in Argentine Andean Potatoes by means of functional markers. Am J Potato Res 95(3):286–300

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci 76(10):5269–5273

Nico A, Garat J, Castro A, Gramuglia S, Ahumada A, Gamboa S, Sarandón S (2006) Recursos genéticos de plantas hortícolas alóctonas en el cinturón verde metropolitano (La Plata, Buenos Aires, Argentina). Plant Genetic Resuorces Newsletter, N°148 pp 44–48. ISSN 1020-3362

Pan C, Yang D, Zhao X, Jiao C, Yan Y, Lamin-Samu AT, Wang Q, Xu X, Fei Z, Lu G (2019) Tomato stigma exsertion induced by high temperature is associated with the jasmonate signalling pathway. Plant Cell Environ 42(4):1205–1221. https://doi.org/10.1111/pce.13444

Park YH, West MAL, St. Clair DA, (2004) Evaluation of AFLPs for germplasm fingerprinting and assessment of genetic diversity in cultivars of tomato (*Lycopersicon esculentum L*). Genome 47(3):510–518

Ribeiro CAG, Pinto MDO, Maciel TEF, Pastina MM, Barros EGD, Guimarães CT (2017) Universal tail sequence-SSR applied to molecular characterization of tropical maize hybrids. Sci Agricola 74(2):163–168

Rohlf FJ (2000) NTSYSpc. Numerical taxonomy and multivariate analysis system-version 2.lOb. Applied Biostatistics Inc., New York

Sbordoni V, Allegrucci G, Cesaroni D (2012) Population Structure. In: Encyclopedia of Caves. Academic Press, Chennai, pp 608–618. ISBN: 978-0-12-383832-2

Smulders MJM, Bredemeijer G, Rus-Kortekaas W, Arens P, Vosman B (1997) Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other Lycopersicon species. Theor Appl Gen 94(2):264–272

Sneath PHA, SoIkal RR (1973) Numerical taxonomy: the principles and practice of numerical classification. W. H. Freeman, San Francisco

Song J, Chen J, Chen HY, Liu Y, Zhuang TM (2006) Research of genetic diversity of tomato using SSR markers. J Shanghai Jiaotong Univ 24:524–528

Suliman-Pollatschek S, Kashkush K, Shats HADAS, Hillel JOSSI, Lavi U (2002) Generation and mapping of AFLP, SSRs and SNPs in Lycopersicon esculentum. Cell Mol Biol Lett 7(2A):583–598

Villa TCC, Maxted N, Scholten M, Ford-Lloyd B (2005) Defining and identifying crop landraces. Plant Gen Resour 3(3):373–384

Von der Pahlen A (1970) Herencia de la resistencia del tomate (cv. Platense) a la peste negra. Revista de Investigaciones Agropecuarias 64:129–134

Weir BS, Cockerham C (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer Assoc. Inc., Sunderland, MA, USA. ISBN-13: 978-0878939022

Wright S (1978) Evolution and the genetics of populations vol 4. Variability within and among natural populations. Univ. of Chicago Press, Chicago