

Genetic diversity and relationship of husk tomato (*Physalis* spp.) from East Java Province revealed by SSR markers

HALIMATUS SADIYAH^{1,*}, SUMERU ASHARI², BUDI WALUYO², ANDY SOEGIANTO²

¹Department of Agronomy, Faculty of Agriculture, Universitas Jember. Jl. Kalimantan 37, Jember 68121, East Java, Indonesia.

Tel. +62-331-778238, *email: sadiyah@unej.ac.id

²Department of Agronomy, Faculty of Agriculture, Universitas Brawijaya. Jl. Veteran, Malang 65145, East Java, Indonesia

Manuscript received: 23 November 2020. Revision accepted: 18 December 2020.

Abstract. Sadiyah H, Ashari S, Waluyo B, Soegianto A. 2021. Genetic diversity and relationship of husk tomato (*Physalis* spp.) from East Java Province revealed by SSR markers. *Biodiversitas* 22: 184-192. This study aimed to investigate the genetic diversity and relationship of *Physalis* spp. from East Java province, Indonesia. A total of the 33 *Physalis* accessions was analyzed employing 16 SSR markers. AMOVA, UPGMA clustering, and non-parametric ANOVA analyses were applied. The results showed Genetic diversity in this sample showed lower levels ($H_e = 0.171$), as compared to other studies of *Physalis* that used different molecular markers. The dendrogram revealed the presence of five groups, the different species belong to different small groups. The two major groups consist of accessions originated from eastern Java and Madura Island, indicating that there is no significant difference between accessions from both areas although there is geographic isolation in the form of the strait. It is consistent with the low population differentiation and high genetic drift. The AMOVA revealed that 96% of the total variation came from the within-population (among accession), reflect that the accessions used in this study have high variation and valuable for plant improvement through breeding programs. It is recommended that future evaluation studies include more accession from minor accessions detected in the sample of this study to better represent the genetic diversity available in this crop.

Keywords: Diversity, relationship, husk tomato, *Physalis*, SSR

Abbreviations: AMOVA: Analysis of molecular variance; NPMANOVA: Nonparametric multivariate analysis of variance; SSR: Simple sequences repeat; UPGMA: Unweighted pair group method with arithmetic mean.

INTRODUCTION

The husk tomato (*Physalis* spp.) is one of the most important and largest genera in the family of Solanaceae, containing 75-120 species, 65 of them concentrated in Mexico (Garzón-Martínez et al. 2015; Feng et al. 2016). Husk tomato has ornamental and edible value, and a high nutritional content such as protein, carbohydrates, vitamin C, phytic acid, oxalic acid, lysine, valine, and so on (Aliero and Usman, 2016). In recent years, *Physalis* is known to have several medicinal properties including anti-diabetic, anti-bacterial, anti-tumor, and contains antioxidants (Huang et al. 2020). Those phytochemical and nutraceutical characteristics, causing an increase of attention to develop this species.

Although the economic importance of some *Physalis* species has been widely recognized, there is little information known about their genetic diversity at the molecular level (Garzón-Martínez et al. 2015). Genetic diversity provides information on the population structure, allelic richness, reveals patterns of genomic differentiation, which are important for genetic resources and plant breeding applications such as formation of core collections, quantitative trait loci identification, parental lines selection, etc. (Carvalho et al. 2019). Estimation of genetic diversity is important to evaluate, conserve, utilize genetic resources,

and determine the uniqueness or scarcity of a genotype (Hurtado et al. 2012).

The genetic diversity of plants had been investigated using several markers including morphological, biochemical, cytological, and molecular. Now, molecular markers have gained importance for germplasm characterization due to their rapid and quality of data generated (Govindaraj et al. 2015). Genetic diversity estimated using molecular marker allow us to quantify and observe the rate of variation that exists in a species more precisely. Molecular markers can also be used as a tool to separate individuals who have close relationships. Amongst the useful molecular markers for the study of plant germplasm is SSR (Simple Sequences Repeats). SSR is considered to have high efficiency. The main benefit of SSR as a marker for genetic studies is that it is highly polymorphic, inherited codominant, and randomly distributed in the genome, multi-allele genetic markers that are experimentally reproducible and transferable among related species (Mason 2015). Some molecular markers have been applied to the genetic diversity and relationship study on *Physalis* (Usaizan et al. 2014; Zamora-Tavares et al. 2014; Garzón-Martínez et al. 2015; Chacón et al. 2016; Feng et al. 2016; Pagare et al. 2016; Kumar et al. 2018; Yujia et al. 2018; Enciso-Rodríguez et al. 2020); but only two of those mentioned before using SSR markers.

In East Java Province, including eastern Java and Madura Island, husk tomato germplasm can be found at lowland to highlands. To date, there has been no exhaustive molecular study of *Physalis* in Indonesia. The study was carried out only in Bandung, one district of West Java province (Hidayat et al. 2017). No report on the molecular study is available regarding the genetic diversity of husk tomato in East Java, even though this information is important for the establishment of adequate conservation and breeding strategies. Therefore, this study was aimed to investigate the genetic diversity, and relationship among different biotypes of husk tomato collected in the field of East Java Province using SSR markers.

MATERIALS AND METHODS

Plant materials

A total of 33 husk tomato accessions samples were used in this study (Table 1). These accessions were selected largely based on geographic distribution, and representing the six populations. The boundary of each population was determined using the graticule method (Hijmans et al.

2012), as applied in the study of pinus taiwanensis (Gao et al. 2015). Graticule method divided the study area on the map using geographic grid lines, i.e. straight lines perpendicular to each other. The area in a grid is defined as one population.

The seeds of those 33 accessions were planted in Greenhouse from August 2018 to January 2019 (see Sadiyah et al. (2020) for details). Briefly, the seeds were germinated in trays for 13 days, moved to cocopeat medium for 20 days, then transplanted into a polybag containing the 13 kg media, the mixture of compost, soil, and sand, to grow. These husk tomato accessions have been partly evaluated for the morpho-agronomic variable.

Study area

Samples of husk tomato collected from 32 sites in East Java Province (see Figure 1, Table 1), which area is 46,428 km². Based on BPS (2020), the average temperature in this province during 2019 was around 28.2°C. The rainfall was as high as 1,862 mm and the number of rainy days was 101 days. The rainy indicator is reduced when compared to 2018.

Table 1. Names and origin of accession evaluated in this study

Pop. code	Accession name	Origin				Species
		District	Island	Latitude-longitude	Altitude	
P1	MDN-TM	Madiun	Eastern Java	7°39'33.8"S 111°31'44.9"E	82	<i>P. angulata</i>
P1	MGT-BR	Magetan	Eastern Java	7°38'38.3"S 111°20'54.3"E	294	<i>P. angulata</i>
P1	NGJ-KS	Nganjuk	Eastern Java	7°36'32.4"S 112°05'10.2"E	48	<i>P. angulata</i>
P1	TBN-KR	Tuban	Eastern Java	6°53'27.4"S 111°54'38.2"E	97	<i>P. angulata</i>
P2	GRS-SM	Gresik	Eastern Java	7°09'10.7"S 112°35'40.1"E	7	<i>P. angulata</i>
P2	JBG-MA	Jombang	Eastern Java	7°34'01.2"S 112°20'11.1"E	35	<i>P. angulata</i>
P2	JBG-SM	Jombang	Eastern Java	7°29'51.6"S 112°10'42.9"E	36	<i>P. pubescens</i>
P2	JBG-BS	Jombang	Eastern Java	7°31'13.4"S 112°10'24.4"E	43	<i>P. minima</i>
P2	LMG-KB	Lamongan	Eastern Java	7°09'33.6"S 112°21'26.2"E	12	<i>P. angulata</i>
P2	MJK-GS	Mojokerto	Eastern Java	7°28'22.8"S 112°26'30.7"E	24	<i>P. angulata</i>
P2	SPG-AS	Sampang	Madura	7°12'22.5"S 113°12'32.2"E	29	<i>P. angulata</i>
P2	SPG-DC	Sampang	Madura	7°12'04.9"S 113°19'53.4"E	35	<i>P. angulata</i>
P2	SPG-GD	Sampang	Madura	7°11'29.6"S 113°15'08.3"E	5	<i>P. angulata</i>
P2	SBY-GY	Surabaya	Eastern Java	7°20'35.0"S 112°43'36.4"E	6	<i>P. angulata</i>
P3	PMK 1	Pamekasan	Madura	7°09'03.2"S 113°28'57.1"E	18	<i>P. angulata</i>
P3	PBL-PT	Probolinggo	Eastern Java	7°42'40.0"S 113°29'54.1"E	6	<i>P. angulata</i>
P3	STB-KP	Situbondo	Eastern Java	7°41'26.1"S 114°03'28.5"E	12	<i>P. angulata</i>
P3	SMN-PG(01)	Sumenep	Madura	7°00'11.6"S 113°52'13.5"E	13	<i>P. angulata</i>
P3	SMN-PG(02)	Sumenep	Madura	7°00'11.6"S 113°52'13.5"E	13	<i>P. angulata</i>
P4	KDR-NL	Kediri	Eastern Java	7°47'30.2"S 111°59'48.9"E	65	<i>P. angulata</i>
P4	KDR-RM(02)	Kediri	Eastern Java	7°51'24.4"S 112°01'31.5"E	88	<i>P. angulata</i>
P4	PCT-BM	Pacitan	Eastern Java	8°10'06.1"S 110°58'46.9"E	332	<i>P. pubescens</i>
P4	TGL-WL	Trenggalek	Eastern Java	8°16'51.8"S 111°43'50.4"E	8	<i>P. angulata</i>
P4	TLG-PL(02)	Tulungagung	Eastern Java	8°14'19.5"S 112°00'01.6"E	234	<i>P. angulata</i>
P5	BLT-SK	Blitar	Eastern Java	8°05'52.3"S 112°08'31.7"E	164	<i>P. angulata</i>
P5	LMJ-PD	Lumajang	Eastern Java	8°07'43.1"S 113°10'29.1"E	96	<i>P. angulata</i>
P5	MLG-TP(01)	Malang	Eastern Java	8°00'51.6"S 112°45'05.8"E	582	<i>P. angulata</i>
P5	PSN-BR	Pasuruan	Eastern Java	7°53'49.7"S 112°54'21.5"E	1886	<i>P. peruviana</i>
P6	BWI-TD	Banyuwangi	Eastern Java	8°33'24.9"S 114°20'28.4"E	14	<i>P. angulata</i>
P6	BWS-JP	Bondowoso	Eastern Java	8°01'54.3"S 114°07'38.5"E	1343	<i>P. peruviana</i>
P6	BWS-WS	Bondowoso	Eastern Java	7°53'52.6"S 113°52'21.1"E	232	<i>P. angulata</i>
P6	JBR-TG	Jember	Eastern Java	8°09'02.5"S 113°43'37.7"E	113	<i>P. angulata</i>
P6	PBL-PK	Probolinggo	Eastern Java	7°49'02.1"S 113°29'56.3"E	109	<i>P. angulata</i>

Note: Part of this table had been presented in Sadiyah et al. (2020)

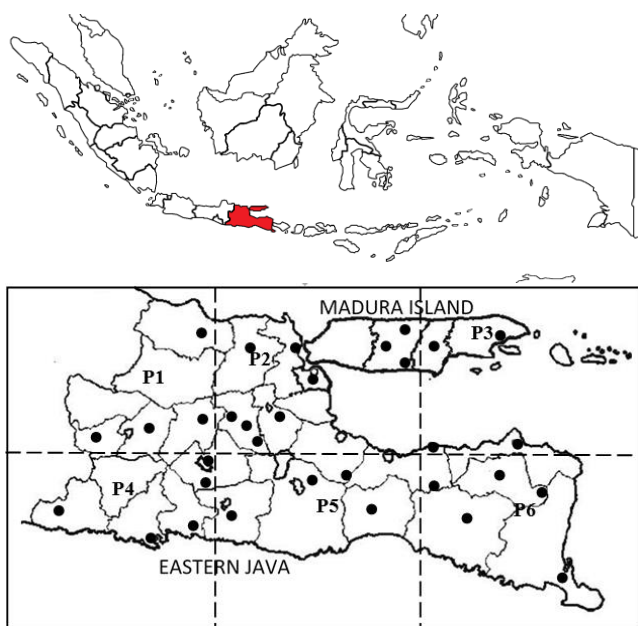


Figure 1. Origin of accession in any districts in East Java Province, Indonesia

Molecular marker analysis

The DNA was extracted from the leaves sample, 55 days after sowing, using Invitrogen purelink genomic plant DNA purification kit. A total of 16 SSRs markers that proved to be polymorphic in a previous study of Simbaqueba et al. (2011) on cape gooseberries (*P. peruviana*) were used. The SSR loci were: SSR1, SSR2, SSR9, SSR13, SSR18, SSR20, SSR54, SSR55, SSR68, SSR77, SSR107, SSR121, SSR123, SSR126, SSR138, and SSR146. Those SSR were chosen because it was proven to have amplicons in *P. peruviana* (Chacón et al. 2016), on tomatoes (Wei et al. 2012), or both. The three SSR, i.e. SSR13, SSR55, and SSR107 were chosen randomly.

PCR amplification was carried out with the following conditions: initial denaturation at 95°C for 3 min, followed by 35 cycles of denaturation at 95°C for the 30s, annealing at $T_m - 5^\circ\text{C}$ (depending on the particular primer) for 30s and extension at 72°C for 1 min; and a final extension at 72°C for 10 min. The detailed information about primers and PCR conditions of each primer can be found in Simbaqueba et al. (2011). After amplification, PCR products were separated using 1% agarose at 50 volts for 60 min in buffer TAE 1x and ethidium bromide (EtBr) at 10-15 min along with 100 bp DNA ladder. The gel was viewed using Bio-Rad gel documentation. The band size was determined using GelAnalyzer 2010a. The data were saved in excel for further analysis.

Statistical analysis of molecular data

The analysis of diversity was conducted at the locus and population level using the fact that SSR is codominant marker. The average number of alleles per locus (N_a), effective number of alleles (N_e), observed heterozygosity (H_o), expected heterozygosity (H_e), inbreeding coefficient (F), percentage of polymorphic loci (PPL), Shannon's information index (I), Gene flow (N_m) were computed by

GenAIEx v.6.5 (Peakall and Smouse 2012). The polymorphic information content (PIC) for each locus were computed by PowerMarker (Liu and Muse, 2005). $PIC_i = 1 - \sum_j P_{ij}^2$; where P_{ij} is the frequency of the j th allele for the i th locus. The power of polymorphism of SSR was measured based on the PIC, evaluated as follows: PIC values less than 0.25 are considered uninformative; PIC values between 0.25 and 0.5 are considered moderately informative, and PIC values above 0.5 indicate highly polymorphic loci (Botstein et al. 1980).

For further estimation of the population variability and structure, the analysis of molecular variance (AMOVA) was performed using GenAIEx v. 6.5 (Peakall and Smouse 2012), with 999 permutations. The AMOVA, including among populations and among individuals within populations. The F -statistic (Wright 1950) is also computed. Wright used an estimation approach to three types of statistics, namely: the deficiency or excess of average heterozygotes in each population (F_{IS}), the degree of genetic differentiation among populations in terms of allele frequencies (F_{ST}), and the deficiency or excess of average heterozygotes in a group of populations (F_{IT}). When F_{ST} is equal to zero, this signifies that the populations are identical; when the value is 1, they are completely different (de Vicenta 2003).

Cluster analysis was performed to obtain a dendrogram based on the similarity coefficient. Clustering carried out by the Unweighted Pair Group Method using Arithmetic Averages (UPGMA). A dendrogram, including all of the 33 accessions, was created using the Nei-Li and Jaccard coefficient of similarity. The Nei-Li and Jaccard coefficient were suitable for SSR as a codominant marker (de Vicente et al. 2003). The Nei-Li coefficient has complete meaning in terms of DNA similarity. It counts the proportion of shared bands among two individuals. Bands that are present in both is given more weight and consider that absence has less biological significance. Jaccard coefficient treated double absence as missing data, and only counts bands present for either accession (i or i'). The correlation between the dissimilarity matrices described above was also evaluated. The correlations and their significance were tested by using the Mantel test with 1000 permutations, calculated using Genstat 18.2 software. This procedure allowed us to assess the resemblance between the two coefficients of similarity matrices.

After clustering, the optimal number of the groups was defined by nonparametric multivariate analysis of variance (NPMANOVA), a non-parametric test of significant difference between two or more groups. There is any similarity/dissimilarity measure that can be used for this purpose. NPMANOVA is useful for ecological data, where groups of samples are to be compared. NPMANOVA calculates an F -value in analogy with ANOVA. Its significance is computed by the permutation of a group membership. For the NPMANOVA test, this study using a one-way procedure with 999 permutations. Both NPMANOVA and cluster analysis were calculated using PAST 3.25 software.

RESULTS AND DISCUSSION

Molecular characterization and assessment of genetic diversity

Genetic diversity is usually thought of as the amount of genetic variability among the population of a species, or individuals of a variety. It results from the many genetic differences between the individual or the population of a species and maybe evident in differences in a DNA sequence, in physiological properties, in biochemical or morphological characters (Rao and Hodgkin 2002). The study of genetic diversity is essential as it constitutes the backbone of any crop improvement and breeding program, for developing desired crop varieties for present and future needs. The SSR would be a valuable resource for those purposes.

A total of 33 husk tomato accessions, the 28 of which were identified as *Physalis angulata*, whether the other three species were represented with either one or two accessions. The unbalance number of among species sampled could be used as an indication of the richness of one species, i.e. *P. angulata* in East Java province comparing to others, or it was probably caused by the use of convenience sampling method, a kind of nonprobability sampling. Although nonprobability sampling has limitations due to the subjectivity in choosing the sample and therefore representativeness of the population is not good, but it is useful especially when randomization is impossible like when there is no sampling frame (Etikan, 2016), that is in case.

It can be seen that all of the examined SSR markers were successfully amplified in all four species of husk tomato used. Among 16 SSR markers, 15 of which were polymorphic (93.75%). SSR55 is identified as a monomorphic locus and will be excluded for further diversity analysis. The other 15 SSR were polymorphic and could be applied within husk tomato and provided data on polymorphisms among related species. The previous study of Wei et al. (2012) using 25 SSR markers found that all of them can be amplified in 36 accessions of six husk tomato species.

The DNA amplification produced the alleles of size ranges from 161 bp to 280 bp. Bands whose size was above 500 bp were not included in the subsequent analysis because the target of the markers used is not those bands. Since, according to Simbaqueba et al. (2011), the expected and observed size of that SSR ranged from 130 to 330 bp.

The presence of a band located above the target size in this study was thought to be due to the somatic type of chromosome ($2n = 24$) so that the resulting banding pattern becomes more numerous. Representation of the results of DNA amplification using SSR primers on agarose gels was presented in Figure 2. This study using Agarose to separate amplified PCR due to its lower cost and easier routine application. Sánchez-Pérez et al. (2006) found that Agarose was able to produce only a slightly different clustering of genotypes despite the big differences observed in the DNA fingerprinting.

Table 2 showed that almost all of the markers used were produced no band, or only one band (which shows that the genotype is homozygous). The average number of alleles per locus (N_a) ranged from 1.167 to 2.00 and the grand mean was 1.522. The average number of alleles per locus provides complementary information to that of polymorphism (de Vicente et al. 2003). The effective number of alleles (N_e) varied from 1.037 (SSR121) to 1.765 (SSR54) with an average of 1.290. The PIC values for the 15 SSR loci were observed to differ, from 0.057 (SSR121, SSR123, SSR146) to 0.393 (SSR107), and the average was 0.195. Almost all of the locus showed low PIC. Only three loci showed moderately high PIC values, ranging from 0.347 to 0.393. Based on the value of N_a , and PIC suggested that SSR54 is the best allele to assess the molecular diversity of husk tomato.

The N_a and average PIC in this study were slightly lower than the findings of Chacón et al. (2016), in the study on *P. peruviana*, where the average number reported within two to four alleles per locus (average=3.059), and the average PIC value equivalent to 0.215. Garzón- Martínez et al. (2015) reported PIC values ranged from 0.094 and 0.663 in the analysis of the *P. peruviana* and related taxa with 24 InDel markers. The PIC values showed the level of polymorphism or informativeness of the molecular marker. The expected heterozygosity (H_e) differed among the markers and it ranged from 0.03 (SSR121) to 0.416 (SSR54) with an average of 0.171. These values were lower than the finding of Chacón et al. (2016) which was the expected heterozygosity ranged from 0.025 to 0.568, with an average of 0.215. The discrepancy in PIC and H_e may be explained by the differences in accessions used i.e. the origin and the sample size, the kind of marker used, and the number of loci analyzed.

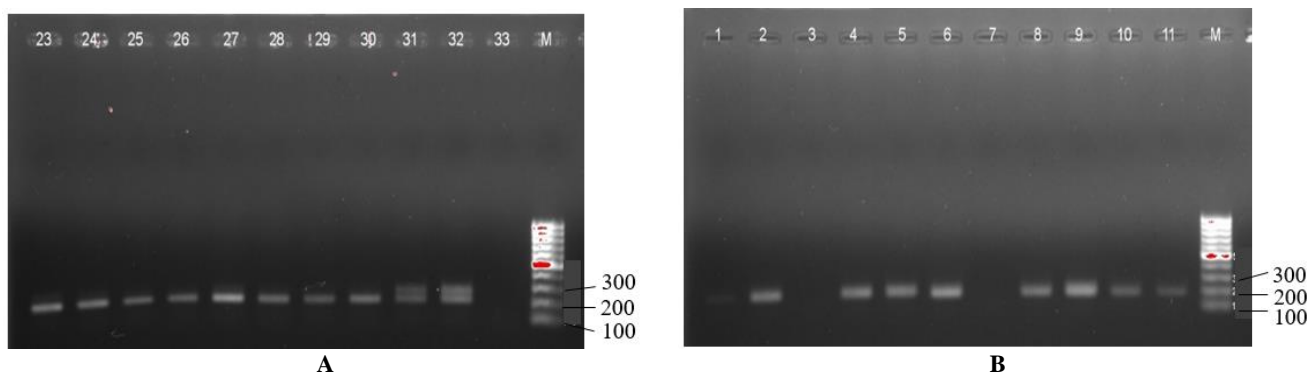


Figure 2. A. Polymorphism among accessions 23-33 using SSR20; B. Polymorphism among accessions 1-11 using SSR68; M ladder 100bp

Table 2. Allelic diversity revealed by 15 polymorphic SSR markers

Locus	Na	Ne	I	He	F	PIC
SSR1	1.667	1.293	0.315	0.199	1.000	0.190
SSR2	1.500	1.299	0.279	0.186	1.000	0.224
SSR9	1.167	1.078	0.083	0.053	1.000	0.107
SSR13	1.333	1.254	0.206	0.143	1.000	0.278
SSR18	1.833	1.478	0.456	0.302	1.000	0.278
SSR20	1.500	1.097	0.141	0.074	0.256	0.112
SSR54	2.000	1.765	0.600	0.416	1.000	0.369
SSR68	2.000	1.617	0.487	0.299	1.000	0.347
SSR77	1.500	1.215	0.231	0.146	1.000	0.152
SSR107	2.000	1.623	0.531	0.343	1.000	0.393
SSR121	1.167	1.037	0.054	0.030	1.000	0.057
SSR123	1.167	1.078	0.083	0.053	1.000	0.057
SSR126	1.333	1.157	0.167	0.107	1.000	0.107
SSR138	1.500	1.282	0.253	0.167	1.000	0.190
SSR146	1.167	1.078	0.083	0.053	1.000	0.057
mean	1.522	1.290	0.265	0.171	0.965	0.195

Note: Na, the average number of alleles per locus; Ne, effective number of alleles; I, Shannon's information index; He, expected heterozygosity; PIC, polymorphic information content; F, inbreeding coefficient

Shannon's information index, reflecting the heterozygous nature of the population, varied from 0.054 in SSR121 to 0.6 in SSR54. While the inbreeding coefficients (F) for these loci were all highly positive values, with the average 0.965, indicating the deficiency of heterozygotes for these loci. This is probably because these plants have the capacity for sexual reproduction through self-pollination. Based on an earlier study, it was known that *Physalis* species have an autogamous reproduction (*P. angulata*: Morales-Saavedra et al. 2019; Figueiredo et al. 2020; *P. minima*: Djakbé et al. 2017). Lagos et al. (2008) found that *P. peruviana* has mixed pollination, but it was the greenhouse study, not the natural population. There is a very limited study on other *Physalis* species.

When comparing the accession from different populations (Table 3), it can be seen that among the six populations, P2 has the highest value in terms of diversity index (Na, Ne, I, He, and PPL), whether P6 was the lowest. The PPL for all of the six populations of husk tomato has an average of 47.78%. The highest PPL was shown by P2 (80%) and the lowest was population 1 and 6 (26.67%). In

P. philadepica using ISSR markers, the PPL for the total population was 91% (Zamora-Tavares et al. 2014), higher than what resulted in this study.

Genetic variation is the basis for survival and adaptation, makes it possible for the survival and adaptive processes that can be viewed in terms of space, time, and fitness. The process of extinction can be due to biotic or abiotic stresses, caused by factors such as competition, parasitism, disease, and climatic change. The description of the distribution of genetic diversity, and of how it is structured, are essential to determine what, where, and how to conserve (Rao and Hodgkin 2002).

The AMOVA for six populations showed a population structure with a genetic variation within accessions contributed with 2% to total genetic variation, while the variance among populations and within-population accounted for 2% and 96% of the total variation, respectively (Table 4). It means that there is high variation among accessions, which can be attributed to the accessions belonging to different species, therefore, the greatest variation is among accessions.

Table 3. Genetic diversity parameters of six husk tomato populations

Population	Na	Ne	I	Ho	He	F	PPL
P1	1.267	1.187	0.159	0	0.108	1	26.67%
P2	2	1.428	0.419	0.007	0.255	0.969	80.00%
P3	1.467	1.28	0.257	0	0.171	1	46.67%
P4	1.667	1.383	0.337	0.013	0.215	0.877	60.00%
P5	1.467	1.307	0.271	0	0.183	1	46.67%
P6	1.267	1.156	0.145	0	0.096	1	26.67%
Mean	1.522	1.29	0.265	0.003	0.171	0.965	47.78%

Note: N, number of accessions; Na, the average number of alleles; Ne, effective number of alleles; I, Shannon's information index; Ho, observed heterozygosity; He, expected heterozygosity; F, inbreeding coefficient; PPL, percentage of polymorphic loci

Table 4. Analysis of molecular variance (AMOVA) for 6 populations of husk tomato

Source	df	Sum squares	Mean squares	Estimated variation	% Variation
Among populations	5	18.635	3.727	0.034	2.00
Among accessions	27	90.850	3.365	1.667	96.00
Within accession	33	1.000	0.030	0.030	2.00
Total	65	110.485		1.731	100.00

Meanwhile, the low among populations variation means that the accession variation in each population is almost similar. Variation of a different pattern has been reported in a previous study by Zamora-Tavares et al. (2014) using ISSR markers, resulted in high variations between individuals (44.3%) and among populations (46.81), and by Garzón-Martinez et al. (2015) which variation among populations and among accessions were 23% and 4%, respectively, revealed by using the InDels markers. The high value of variation within the population means that the rate of genetic change within the population is greater than among populations. This aspect must be considered for conservation and breeding strategies, where further variations would be desirable.

The genetic diversity of a species is hierarchically structured. Genetic differentiation may exist between individuals within a population, between populations, or between regions. Several methods can be used to study genetic differentiation. The F statistic is one that can be used to assess differentiation between populations.

The AMOVA revealed a strong and significant genetic structure at the accession level ($F_{IS} = 0.982$, p-value = 0.001), and produced a low and not significant F_{ST} value ($F_{ST} = 0.020$, p-value = 0.198). The low F_{ST} among natural wild husk tomato populations suggesting that many wild husk tomato populations are genetically not distinct from each other. This result was consistent with the low value of the among-population variation of AMOVA. Chacón et al. (2016) found the F_{ST} of small but significant ($F_{ST} = 0.058$, p-value = 0.001) when comparing different biological statuses using the same marker on one of the husk tomato species. A small F_{ST} value is in line with a study from Garzón-Martinez et al. (2015) where the F_{ST} values are low to moderate (SNPs = 0.038 and InDels = 0.143).

In a complementary manner to the estimation of genetic structure was the indirect measure of gene flow (Nm). Nm reflects the degree of connectedness among populations, preventing population differentiation caused by genetic drift. In this study, the gene flow was high (Nm = 12.528) with low F_{ST} , contrary to what was reported by Zamora-Tavares et al. (2014) on *P. philadelphica* which produces low gene flow (Nm = 0.642) and high F_{ST} (0.449). The low among population variation revealed by AMOVA, low levels of differentiation among population, and the high value of gene flow were probably caused by the germplasm exchange resulted from human activities.

Genetic relationship analysis

Genetic similarities were calculated from the scoring matrix using the Nei-Li and Jaccard similarity coefficient.

From the similarity matrices, genetic distance was calculated for each pair of varieties (distance = 1 - similarity) and used for clustering. The dendrogram from the cluster analysis based on 15 polymorphic SSR markers was shown in Figure 3. Similarity among accessions that were calculated by these two methods, resulted in two similarity matrices.

A Mantel test of 100 permutations, showed the high and significant association between the Nei-Li and Jaccard similarity matrices (0.996) where percent permutations with the equal or greater association were 0.00. The dendrogram resulted from both similarity coefficient was identical. So, we will discuss only one dendrogram for further analysis.

The dendrogram (Figure 3) revealed that there were mixed accessions from eastern Java and Madura Island in one group, even if 14 groups at a Nei-Li similarity of 0.92 was used. The genetic similarity of the accessions ranged from 0.522 to 1.000. The lowest similarity values were observed between JBG-BS and PCT-BM (0.522), JBG-BS, and TGL-WL (0.545), JBG-BS, and KDR-NL (0.571), and JBG-BS and SPG-AS (0.583). Genetically, JBG-BS has the lowest similarity with some other accessions. This is probably because JBG-BS is the only species of *P. minima*. The highest similarity values were observed between PMK1, TBN-KR, and JBR-TG, KDR-RM(02) and LMJ-PD, MGT-BR, and SPG-DC, LMG-KB and PBL-PK, MLG-TP(01) and STB-KP. Those pairs of accession were genetically exactly similar i.e. similarity = 1 (data not shown). In general, those pairs whose highest similarity is coming from the same species, i.e. *P. angulata*.

The optimal number of groups was assessed by NPMANOVA (Figure 4), using four different similarity/dissimilarities coefficients. Crossa and Franco (2004) suggested the use of graphs for different numbers of groups and observed the point of maximum increase as a criterion for defining the optimal number of groups. In the present study, as in Sadiyah et al. (2020), we used the opposite criterion, i.e. using the maximum decrease. Generally, these four coefficients showed that the steepest decrease of F-value was from five to seven groups. It can be concluded that the optimal number for grouping the 33 accessions were five groups. The F-values calculated by the Nei-Li, Jaccard, correlation, and Gower similarity/dissimilarities for the five groups were 16.26, 11.41, 14.85, 14.29 (Figure 4), successively, and their probabilities were 0.001. The F-value for six and eight groups could not be investigated because it was not possible to divide the accessions into six and eight groups based on the dendrogram.

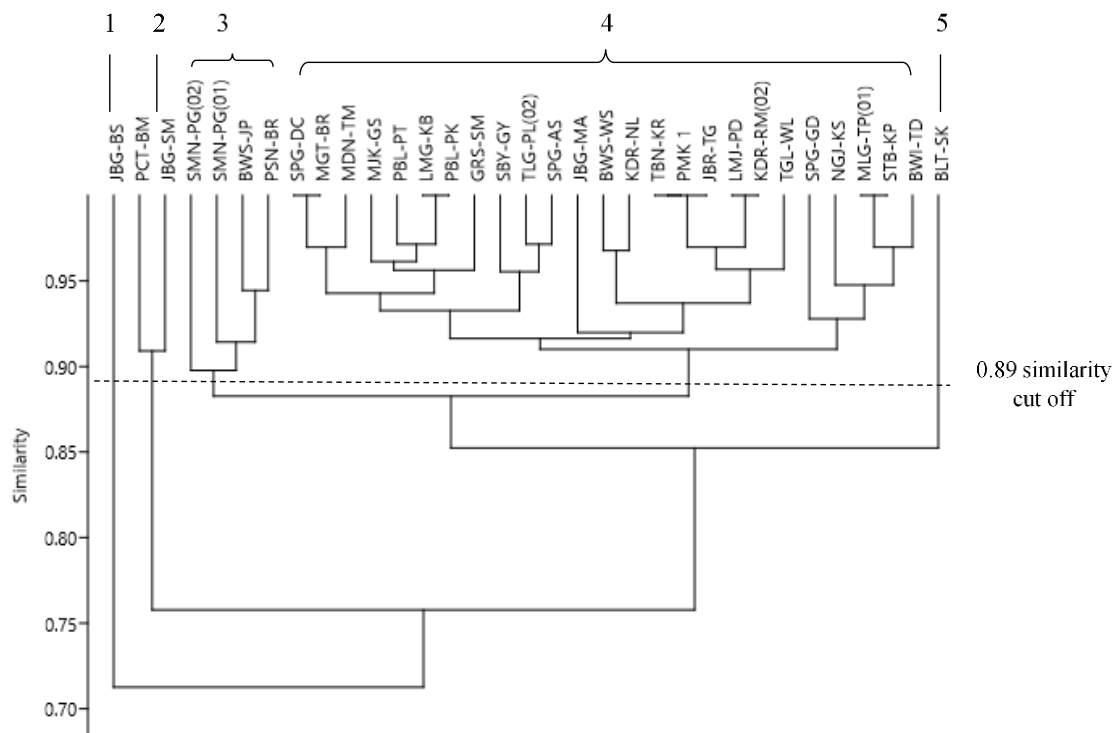


Figure 3. Dendrogram of 33 accessions based on 15 polymorphic SSR markers using UPGMA cluster analysis and Nei-Li similarity

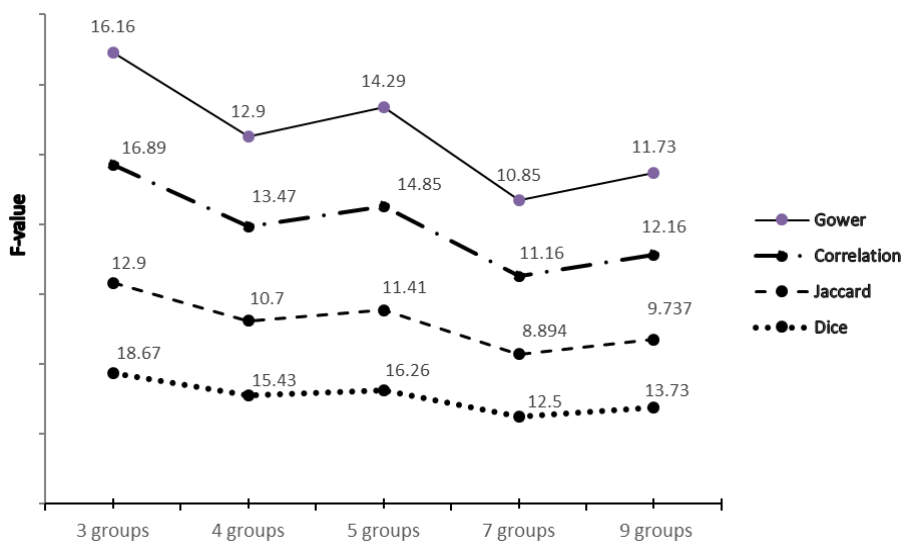


Figure 4. The F-value of the NPMANOVA test on dendrogram of 33 accessions based on 15 polymorphic SSR markers using UPGMA cluster analysis and Nei-Li similarity

According to the dendrogram (Figure 3), when 33 accessions were divided into five groups, the similarity between groups was approximately 0.89. The number of accessions per group ranged from one in group 1 and 5 to 25 in group 3. Groups 1 and 5 were composed of JBG-BS and BLT-SK, respectively. JBG-BS was distinct because based on the morphology, it was identified as the only *P.*

minima in this study. BLT-SK was distinct although it was identified as *P. angulata* species.

Group 2 was composed of two accessions, i.e. JBG-SM and PCT-BM. They were identified as *P. pubescence*, the annual plant which has pubescence in their stem and leaves.

Group 3 was composed of four accessions that can be divided into two subgroups. The first subgroup consists of SMN-PG(01) and SMN-PG(02), identified as *P. angulata*, originating from the same district, i.e. Sumenep (Madura Island), around 13 m.s.a.l. Whether the second subgroup consists of BWS-JP and PSN-BR, identified as *P. peruviana*, the perennial crop which has pubescence in its stem and leaves, grow on the plateau and mountain area, more than 1300 m a.s.l.

Group 4 was the biggest, which consists of 25 accessions. All of which were the *P. angulata* species. This group can be divided into 3 subgroups. There were accessions from eastern Java and Madura Island in each subgroup. This fact showed that there was no significant difference between accessions from eastern Java and those from Madura Island, even though there was geographic isolation in the form of the strait. It is consistent with the low population differentiation among those two islands ($F_{ST} = 0.04$) and low among population variation, i.e. 4% (data not shown).

In general, accessions of different species form separate groups from most other accessions. It indicated that the SSR from *P. peruviana* species used in this research can be applied to distinguish the different species of husk tomato, especially *P. angulata*, *P. minima*, *P. pubescens*, and *P. peruviana*.

The results presented in this paper are the first data on the assessment of husk tomato based on the molecular marker, especially those from East Java province. The diversity of husk tomato species is relatively moderate. Most of the SSR markers used in this study were polymorphic and can be used as tools to help species distinction for official registration, breeding, and conservation program of husk tomato, and SSR54 is the best for this purpose. The information found here evidenced the relatively moderate diversity detected using SSR markers from husk tomato accessions and also pointed out the higher within-population than among-population variation. The NPMANOVA test showed that five groups were the optimum number. Generally, different species belong to different groups or subgroups. Some groups consist of accession from two separated islands, which showed that accession from those islands was not significantly different.

Enciso-Rodrigues et al. (2020) can differentiate *P. peruviana* and related taxa i.e. six *Physalis* species using thousands of SNPs. The use of lower markers in this recent study can recover the same differences in the dendrogram analysis. This result shows that the SSR marker has great value when a sequences analysis is not possible. The use of 16 SSR markers is relatively enough based on the simulation study of Li et al. (2009) which concluded that when loci number exceeded 15, the value of genetic parameter tends to be stable. It is recommended that future evaluation studies include more accession from minor accessions detected in the sample of this study to better represent the genetic diversity available in this crop. Further, a comparison of molecular and morpho-agronomical based study is required for a more comprehensive decision for the breeding programs.

ACKNOWLEDGEMENTS

Thanks to The Indonesian Education Scholarship (LPDP) for its financial support to the research reported in this article.

REFERENCES

- Aliero AA, Usman H. 2016. Leaves of ground cherry (*Physalis angulata* L.) may be suitable in alleviating micronutrient deficiency. *Food Sci Technol* 4 (5): 89-94.
- 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 Gen* 32: 314-331.
- BPS. 2020. Jawa Timur Province in Figures 2020. BPS Statistic of Jawa Timur, Surabaya.
- Carvalho YGS, Vitorino LC, de Souza UJB, Bessa LA. 2019. Recent trends in research on the genetic diversity of plants: Implications for conservation. *Diversity* 11 (4): 1-21.
- Chacon MI, Sanchez Y. del P, Barrero LS. 2016. Genetic structure of a Colombian Cape Gooseberry (*Physalis peruviana* L.) collection by means of microsatellite markers. *Agron Colomb* 34(1): 5-16.
- Crossa J, Franco J. 2004. Statistical methods for classifying genotypes. *Euphytica* 137 (1): 19-37.
- de Vicente MC, Lopez C, Fulton T, Molina L. 2003. Genetic Diversity Analysis with Molecular Marker Data: Learning Module. International Plant Genetic Resources Institute (IPGRI), Rome and Cornell University, New York.
- Djakbe JD, Ngakou A, Christian W, Faibawe E, Tchuenguem NFF. 2017. Pollination and yield components of *Physalis minima* (Solanaceae) as affected by the foraging activity of *Apis mellifera* (Hymenoptera: Apidae) and compost at Dang (Ngaoundéré, Cameroon). *Intl J Agron Agric Res* 11 (3): 43-60.
- Enciso-Rodríguez FE, Osorio-Guarín JA, Garzón-Martínez GA, Delgadillo-Duran P, Barrero LS. 2020. Optimization of the genotyping-by-sequencing SNP calling for diversity analysis in cape gooseberry (*Physalis peruviana* L.) and related taxa. *PLoS ONE* 15 (8): e0238383. DOI: 10.1371/journal.pone.0238383.
- Etikan I. 2016. Comparison of convenience sampling and purposive sampling. *Am J Theor Appl Stat* 5 (1): 1-4.
- Feng S, Jiang M, Shi Y, Jiao K, Shen C. 2016. Application of the ribosomal DNA ITS2 region of *Physalis* (Solanaceae): DNA barcoding and phylogenetic study. *Front Plant Sci* 7: 1047. DOI: 10.3389/fpls.2016.01047.
- Figueiredo MCC, Passos AR, Hughes FM, dos Santos KS, da Silva AL, et al. 2020. Reproductive biology of *Physalis angulata* L. (Solanaceae). *Sci Hortic (Amsterdam)* 267: 1-10.
- Gao X, Shi L, Wu Z. 2015. Cluster analysis of *Pinus taiwanensis* for its ex situ conservation in China. *Genet Mol Res* 14 (2): 5845-5855.
- Garzón-Martínez GA, Osorio-Guarín JA, Delgadillo-Durán P, Mayorga F, Enciso-Rodríguez FE, et al. 2015. Genetic diversity and population structure in *Physalis peruviana* and related taxa based on InDels and SNPs derived from COSII and IRG markers. *Plant Gene* 4: 29-37.
- Govindaraj M, Vetriventhan M, Srinivasan M. 2015. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genet Res Intl* 2015: 431487. DOI: 10.1155/2015/431487.
- Hidayat T, Yosnata LJ, Wulandari LT, Priyandoko D, Aisyah S. 2017. Random amplified polymorphism DNA method to authenticate Indonesian medicinal plant ciplukan (*Physalis angulata*; Solanaceae). *Pertanika J Sci Technol* 25 (S): 13-22.
- Hijmans R, Guarino L, Mathur P. 2012. DIVA-GIS Version 7.5 Manual. : 71. https://www.diva-gis.org/docs/DIVA-GIS_manual_7.pdf.
- Huang M, He JX, Hu HX, Zhang K, Wang XN. 2020. Withanolides from the genus *Physalis*: a review on their phytochemical and pharmacological aspects. *J Pharm Pharmacol* 72: 649-669.
- Hurtado M, Vilanova S, Plazas M, Gramazio P, Fonseka HH, Fonseka R, Prohens J. 2012. Diversity and relationships of eggplants from three geographically distant secondary centers of diversity. *PLoS One* 7 (7): e41748. DOI: 10.1371/journal.pone.0041748.

- Kumar, V, Sahay S, Singh RS, Mir H, Rashmi K. 2018. Molecular marker based genetic diversity analysis in cape gooseberry (*Physalis peruviana* L.). *Curr J Appl Sci Technol* 31 (3): 1-6.
- Lagos TCB, Cabrera FAV, Escobar HC, Flórez JEM. 2008. Sexual reproduction of the cape gooseberry. *Acta Agron* 57 (2): 81-87.
- Li O, Zhao Y, Guo N, Lu C, Sun X. 2009. Effects of sample size and loci number on genetic diversity in wild population of grass carp revealed by SSR. *Zool Res* 30 (2): 121-130.
- Liu K, Muse SV. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21 (9): 2128-2129.
- Mason AS. 2015. SSR genotyping. In: Batley J (eds.). *Plant Genotyping: Methods and Protocols*. Springer, New York.
- Medina-medrano, JR, Almaraz-abarca N, Sifuentes-rincón AM, Molinamoret S. 2016. Potential of ISSR to discriminate among species of *Physalis* (Solanaceae). *J Chem Biol Phys Sci* 6 (4): 1184-1195.
- Morales-Saavedra J del C, Zaragoza FAR, Toledo DC, Hernández CVS, Vargas-Ponce O. 2019. Agromorphological characterization of wild and weedy populations of *Physalis angulata* in Mexico. *Sci Hortic (Amsterdam)* 246: 86-94.
- Pagare S, Mishra RP, Tripathi N, Rathore M, Kumar B. 2016. Assessment of genetic diversity among different biotypes of *Physalis minima*. *Indian J Weed Sci* 48 (4): 417-420.
- Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* 28 (19): 2537-2539.
- Rao VR, Hodgkin T. 2002. Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell Tissue Organ Cult* 68: 1-19.
- Sadiyah H, Soegianto A, Waluyo B, Ashari S. 2020. Short communication: Preliminary characterization of groundcherry (*Physalis angulata*) from east Java Province, Indonesia based on morpho-agronomic traits. *Biodiversitas* 21 (2): 759-769.
- Sánchez-Pérez R, Ballester J, Dicenta F, Arús P, Martínez-Gómez P. 2006. Comparison of SSR polymorphisms using automated capillary sequencers, and polyacrylamide and agarose gel electrophoresis: Implications for the assessment of genetic diversity and relatedness in almond. *Sci Hortic (Amsterdam)* 108: 310-316.
- Simbaqueba J, Sánchez P, Sanchez E, Núñez Zarantes VM, Chacon MI. 2011. Development and characterization of microsatellite markers for the cape gooseberry *Physalis peruviana*. *PLoS One* 6(10): e26719. DOI: 10.1371/journal.pone.0026719.
- Usaizan N, Ashikin N, Abdullah P, Saleh G. 2014. Assessment of genetic diversity of *Physalis minima* L. (Solanaceae) based on ISSR marker. *J Appl Sci Agric* 9 (18): 18-25.
- Vargas-Ponce O, Sánchez Martínez J, Zamora Tavares M del P, Valdivia Mares LE. 2016. Traditional management of a small-scale crop of *Physalis angulata* in Western Mexico. *Genet Resour Crop Evol* 63 (8): 1383-1395.
- Wei J, Hu X, Yang J, Yang W. 2012. Identification of single-copy orthologous genes between *Physalis* and *Solanum lycopersicum* and analysis of genetic diversity in *Physalis* using molecular markers. *PLoS One* 7 (11): e50164. DOI: 10.1371/journal.pone.0050164
- Wright S. 1950. Genetical structure of populations. *Nature* 166: 247-249.
- Zamora-Tavares P, Vargas-Ponce O, Sánchez-Martínez J, Cabrera-Toledo D. 2014. Diversity and genetic structure of the husk tomato (*Physalis philadelphica* Lam.) in Western Mexico. *Genet Resour Crop Evol* 62 (1): 141-153.