

Full Length Research Paper

Genetic characterization of Cape gooseberry (*Physalis peruviana* L.) accessions in selected counties in Kenya using SSR markers

Pauline W. Muraguri^{1*}, Robert M. Gesimba¹, Joseph N. Wolukau¹ and Manfred Miheso²

¹Department of Crops, Horticulture and Soils, Egerton University, P. O. Box 536-20115 Egerton, Kenya.

²Kenya Agricultural and Livestock Research Organization - Food Crops Research Institute, Njoro, Kenya.

Received 12 October, 2020; Accepted 4 December, 2020

Cape gooseberry (*Physalis peruviana* L.) is a neglected high potential crop, knowledge of the genetic diversity of the genotypes domesticated in Kenya is limited. To understand the genetic diversity and structure within and between Cape gooseberry germplasm, 70 accessions from six selected counties were analyzed using 15 pairs of highly polymorphic SSR primers. In this study, a total of 61 polymorphic SSR alleles were identified with mean polymorphic information content (PIC) of 0.43. Analysis of Molecular Variance (AMOVA) revealed that 92.8% of the total genetic variation was within accessions whereas variation among accessions accounted for 7.2% of the total genetic variation. Genetic diversity parameters among the 70 accessions revealed that Cape gooseberry was more diverse than previously recorded. Based on the SSR data, the 70 accessions were classified into five main phylogenetic groups, which corresponded to the county of origin through factorial analysis, principal component analysis (PCA), and phylogenetic analysis. Seven core SSR primer pairs, namely SSR1, SSR2, SSR10, SSR11, SSR123, SSR138, and SSR146 were found to have a wide applicability in genotype identification of cape gooseberry, and thus they are recommended for use in genetic characterization of germplasm collected from other counties not covered by the present study. This study demonstrated the existence of considerable genetic diversity in Cape gooseberry accessions growing in selected counties in Kenya and can therefore be used as a basis for future breeding programs in the development of hybrids with desirable traits. This wider genetic diversity is vital for posterity as it will help cope with unpredictable climatic changes and human needs.

Key words: Simple sequence repeats (SSRs), genetic diversity, germplasm, *Physalis peruviana* L., polymorphic information content (PIC).

INTRODUCTION

Physalis peruviana L. is a species from the family *Solanaceae* and genus *Physalis*, commonly known as

Cape gooseberry, ground cherry, Cape gooseberry, or winter cherry. It contains high amounts of vitamins (A, B,

*Corresponding author. E-mail: pauline.wanjiru65@gmail.com.

and C), micronutrients (iron, phosphorous, and calcium), has anti-inflammatory, antioxidant, and anti-hepatotoxic activities (Wu et al., 2009). The vitamin C content in this fruit is reported to be the highest among all other fruits and plants, thus its reference as a “super fruit”, Cape gooseberry contain up to twenty times vitamin C as that found in oranges (Villacorta and Shaw, 2013). Owing to such concentrated levels of nutrients in this fruit, Villacorta and Shaw (2013) posit that it is useful for medicinal purposes in restoring vitality and boosting immunity by fortifying the liver, supporting cardiovascular activity, strengthening lungs, and enhancing fertility and food absorption.

Investigation of genetic diversity in both wild and domesticated species is essential. Assessment of available genetic diversity is a preliminary stage in genetic improvement in crop plants (Bhandari et al., 2017). Wild populations of different crop species are known to be a potential source of useful genes and traits which could be introduced into the domesticated gene pool (Campisano et al., 2015). There are various ways in which diversity analysis can be carried out including cytological, morphological, biochemical, and molecular approaches. With the advent of genomic tools, assessment of genetic diversity at the molecular level has proven to be more useful as compared to that at the phenotypic level because the latter entails analysis of data on morphological traits, which are generally influenced by environments (Myers et al., 2000). Different molecular marker systems have been proved to be valuable tools in assessing the genetic diversity of plants between and within species regardless of environmental interferences on the phenotype (Demir et al., 2010). Research has shown that different markers reveal different classes of variation (Virk et al., 2000). Simple Sequence Repeat (SSR) markers have become the marker of choice for many researchers as they offer many advantages including technical simplicity, feasibility of automation, even distribution throughout the genome, higher frequency of polymorphism, rapidity, requirement of little and not necessarily high-quality DNA, and no requirement of prior information of any DNA sequence (Mason, 2015). Molecular marker analysis work would be of great help for analyzing genetic diversity, and exploiting genetic resources for identification, isolation, conservation, and utilization (Ravi et al., 2010).

The availability of enough food to meet 95% of the world's requirements is dependent on only a few crop species which are widely and intensively cultivated crops. These have been developed by extensive selection from available large agro-biodiversity pool (Ochatt and Jain, 2007). There is a great need to expand the exploitation of the plant genetic diversity that would broaden the crop diversity for food supply to feed the ever-growing human population and avoid dependence on few food crops. Wild relatives and neglected crops could become an excellent source of useful gene pool.

The average yields of Cape gooseberry are still below the maximum potential mainly due to fruit cracking, small fruits, and premature fruit drop (Ali and Singh, 2016). Also, the poor quality of fruits in some cultivars in terms of the levels of total soluble solids (TSS) and total titratable acidity (TTA) make them unattractive for large scale agriculture (Herrera et al., 2011). No improved cultivars have been developed yet although Leiva-Brondo et al., (2001) reported to have used a simple breeding strategy employed in other Solanaceae crops to develop hybrids with superior yield characteristics by exploiting heterosis. Genetic diversity is important in this context as it serves as the reservoir of many novel traits related to yield, quality as well as tolerance to abiotic and biotic stresses. A thorough understanding of the diversity in the Cape gooseberry genome is necessary before implementation of any breeding program in breaking these yield and quality barriers. The objective of this study was to determine the genetic diversity of Cape gooseberry genotypes in Kenya using SSR markers for use in present and future breeding schemes and conservation programs. This information will contribute to understanding the genetic relationship between and within different genotypes and provide basic information on parental selection for Cape gooseberry breeding material.

MATERIALS AND METHODS

Experimental materials

Seventy dry leaf samples were collected from accessions of Cape gooseberry collected from six selected counties of Kenya (Figure 1). The counties selected were Kiambu (1.0314°S, 36.8681°E), Muranga (0.7839°S, 37.0400°E), Kericho (0.1828°S, 35.4782°E), Nakuru (1.3665°S, 35.3905°E), Nyandarua (0.1804°S, 36.5230°E), and Laikipia (0.3606°N, 37.7820°E). The six counties were selected because they have Cape gooseberry germplasm. The experiments were carried out during the rainy season in May 2019. The young leaves were sampled because they yield DNA better and have low concentrations of phytates which contaminate DNA. The samples were stored on Silica gel (Loba Chemie). The samples were coded based on the county. For instance, a collection was done at Nyandarua County, Olkalou district, Kipipiri division, Miharati location, Mara village and was the seventh Cape gooseberry crop to be sampled; the code assigned was NYD/OLK/KIP/MHT/MAR/07.

DNA extraction and quantification

Total nucleic acid was extracted from the dry young leaf tissue that had been stored in silica gel for one month using a modified CTAB protocol (Porebski et al., 1997). Modifications involved the introduction of the initial wash stage using 2-Mercaptoethanol to remove the aromatic compounds and phytates. Centrifugation time for initial stages was also increased to 10 min to ensure that cell debris and the proteins were well decanted to minimize contamination. Final centrifugation time was reduced to 3 min while centrifugation speed was increased to 14000rpm to avoid pelleting of carry over impurities. Precipitation time was also increased from the recommended time of 2 to 18 h to increase DNA precipitation

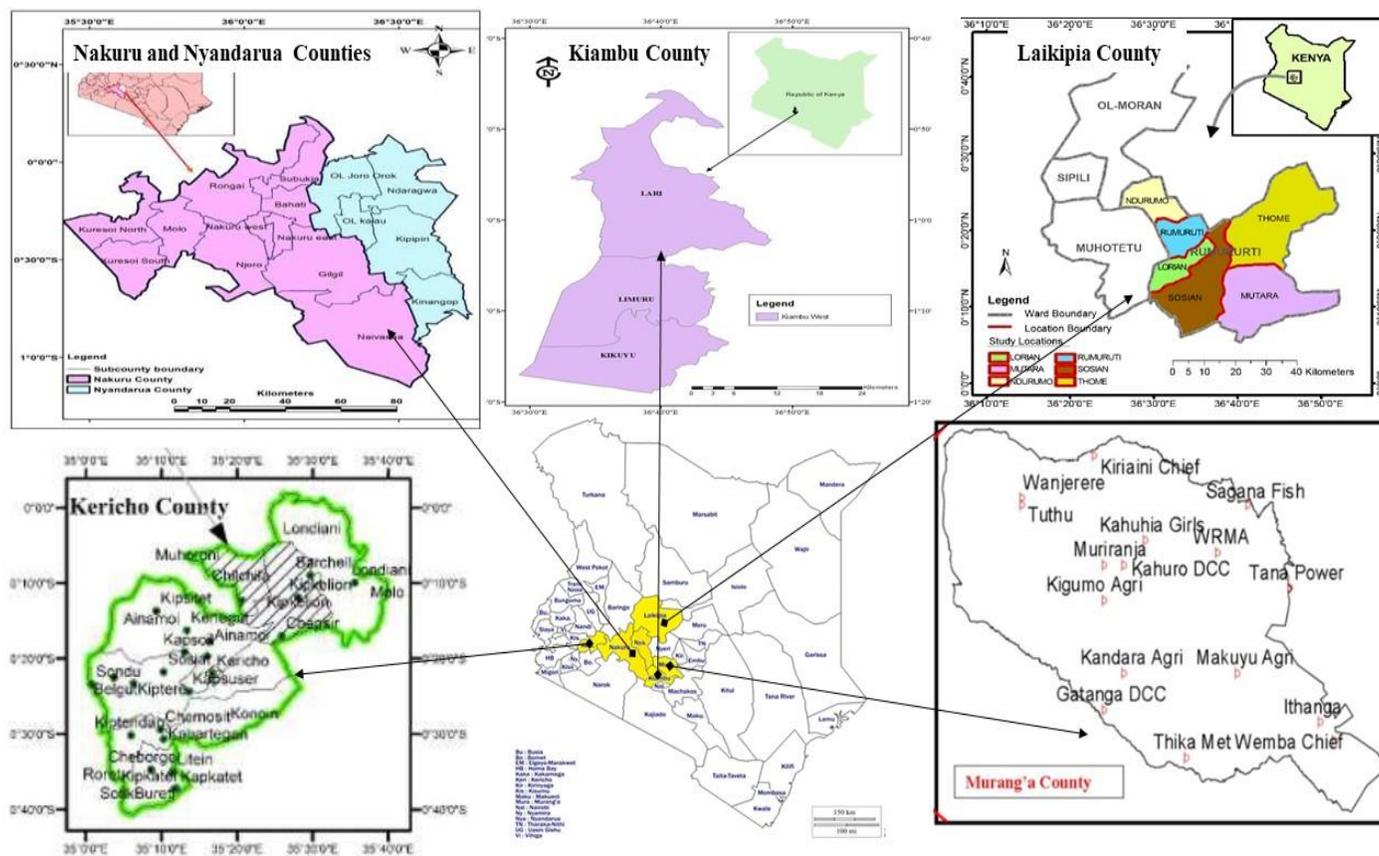


Figure 1. Map showing Cape gooseberry germplasm occurrence in six selected counties in Kenya.

and recovery.

Nanodrop spectrophotometer (Applied Biosystems) and agarose gel electrophoresis assays were used to determine the purity and concentration of DNA in the samples. Nanodrop spectrophotometry involved reading the concentration of DNA from the absorbance of the sample at 260 nm ($1\text{OD (A260)} = 50 \mu\text{g}$ for doubled stranded DNA/ μl). The purity of the DNA sample was determined by A260/A280 ratio (1.6 ± 1.8 for pure DNA). Agarose gel electrophoresis quantification involved resolving the samples in agarose gel 0.8% (0.8g agarose and 100ml Sodium Borate) containing 3ul ethidium bromide staining dye at voltage of 100 volts and 400 mA current for 30 min. The DNA was visualized on a UV transilluminator (Applied Biosystems).

Selection and genotyping of SSR markers

A set of 15 SSR markers selected from earlier published reports (Table 1) were used to determine the diversity of Cape gooseberry collections. The SSR markers were selected based on high polymorphic information content (PIC) values (>0.4) in earlier published studies; the maximum number of alleles detected, genome coverage, and distribution on linkage groups. The primers for the SSR markers were synthesized on contract from Inqaba Biotech in South Africa and genotyping of the SSR markers was carried out at Marker Assisted Breeding Laboratory at Kenya Agricultural and Livestock Research Organization (KALRO) Njoro Center.

Polymerase chain reaction (PCR) was done in a 2720, 96 universal gradient thermocycler (Applied Biosystems) in 20 μl final volume containing 10 ± 20 ng DNA template, 5.0 pmol forward and reverse primers, 1x PCR buffer, 2.5mM of each dNTPs, 1.5 mM MgCl and 0.5 U of Taq DNA polymerase (Biolabs). The amplification conditions for PCR profile were: 95°C for 5 min, 35 cycles of 95°C for 30 s, specific annealing temperature for each SSR for 1 min, extension at 72°C for 2 min, and final extension at 72°C for 10 min. The PCR amplicons were run in a 1.8% agarose gel containing ethidium bromide staining dye at a voltage of 80 volts and a current of 400 mA for 1 h and visualized in a UV transilluminator (Applied Biosystems).

Data analysis

SSR marker alleles were scored manually from the gel images using a simple numerical scoring method. When the expected band was present it was scored as 1 while 0 was used to code for absence. Number of alleles, observed heterozygosity (HO) and expected heterozygosity (HE), Shannon's diversity index (I), gene flow (Nm), and gene differentiation coefficient (FST) executed in POPGENE Version 1.32 (Yeh et al., 2000) were used to determine genetic diversity. Chi-squared was used to determine Hardy - Weinberg equilibrium (HWE) while genetic diversity and polymorphic information content (PIC) were computed in Power Marker 3.25 (Liu and Muse, 2005).

Unweighted Pair Group with Arithmetic Mean (UPGMA) algorithm was used to construct an unrooted phylogram from a distance

Table 1. Selected SSR markers used in the study.

Marker	Forward primer	TM	Range of observed bands		Amplicon Size (Bp)	Repeat type	Location
SSR1F SSR1R	AGAGGACTCCATTTGTTTGCT TGAGGGTGTTGGATGTTTTCT	50	170	210	206	AT	39 UTR
SSR2F SSR2 R	CATTGGGTTTTCGCATCCAT AGACAAGCCTAGGGGAAAGG	50	230	250	237	AG	39 UTR
SSR10F SSR10R	GCTTCCTATTGTGTTGCCTGA ACTTTGGGTTTTCGGGAATTG	50	220	240	185	AG	59 UTR
SSR 11F SSR11R	CAGCTGAAATAAGAGAGTGATTGG CCCTCTTTTTCTCCTCCGAGT	50	170	190	180	AT	39 UTR
SSR15F SSR15R	GCTTGTGATCAGCTTTCTTTG TGGATCATAACCTTGCTAATGC	50	180	210	172	AG	39 UTR
SSR54F SSR54R	CGGCTGGTATGCTTACAAAGAT GCACTTCCACTGTTTTAACTTCC	50	160	210	197	AC	59 UTR
SSR72F SSR72R	GTGCTCGCAGTTTCTTCAA CCGCCGTTACTTCTAATCA	50	200	220	158	AT	39 UTR
SSR77F SSR77R	CATACCATAACTCCCATCTCTC TGCCGATTCTGATTTCTTCC	50	160	180	216	AT	39 UTR
SSR112F SSR112R	CTACGCCTACCACTTGCACA CAGTGGAAGCCTCAAGATCC	50	180	230	203	AC	39 UTR
SSR118F SSR118R	AATCAAGGGTCAGAAGAAATGG GCAAGAATGGATGTGGGTGT	50	170	220	180	AT	39 UTR
SSR121F SSR121R	AGCAACCTCCCAATCAGCTA TGGTGAGTAAATGGGGGAAA	52	170	240	170	AG	39 UTR
SSR123F SSR123R	TCAGTGAGCGCGTATATCT GCGATCTCACCAACCTCTC	50	260	330	216	AG	39 UTR
SSR126F SSR126R	TCCAAAAAGAAAACAAAACACT TTGAATGCATGTTTGATGGA	50	190	210	202	AC	39 UTR
SSR138F SSR138R	TCCGATCACTACTCAGCACG CAATTCGGGTTGTGAATCGGGT	50	200	210	138	AG	59 UTR
SSR146F SSR146R	AGGCTAATGAGGACGAAGCA GGTTGCATTACAAAGCACTGA	50	200	210	187	AT	39 UTR

matrix based on Nei's (1973 and 1978) genetic distances, using Darwin 6.0 software (Perrier and Jacquemoud-Collet, 2006) and Power Marker 3.25 (Liu and Muse, 2005).

RESULTS AND DISCUSSION

DNA extraction and PCR

All the samples yielded good quality and quantity of DNA to enable genotyping (Plate 1). Genotyping of markers resulted in the amplification of expected regions resulting in single or multiple bands (Plate 2a). Minimal contamination of samples by proteins and phenolic compounds was observed in this study.

This study evaluated the genetic diversity and population structure of 70 accessions of Cape gooseberry that originated from six selected counties of Kenya

(Nakuru, Laikipia, Nyandarua, Kericho, Murang'a, and Kiambu) by 15 SSR primer pairs. The 15 primer pairs amplified a total of 71 polymorphic SSR alleles (Plate 2b). Every primer pair was able to amplify varying the number of SSR alleles ranging from 100 to 300 bp from all accessions tested, regardless of the county of origin.

Marker polymorphism and genetic diversity of Cape gooseberry accessions

The number of major alleles per primer ranged between 2 (SSR54, SSR15, SSR72, SSR77, SSR112, and SSR 126) and 11(SSR 2). A total of 61 alleles were detected by the 15 primers with a mean of 4.07(Table 2). Allele frequency ranged between 0.26 and 0.96 with a mean of 0.64. The number of observed alleles was 2 for all markers while the number of effective alleles ranged

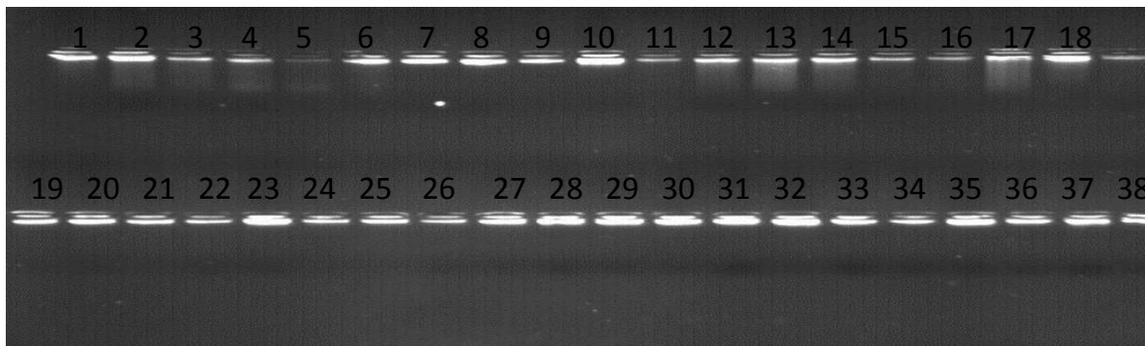


Plate 1. Representative quantification gel image of cape gooseberry accessions.

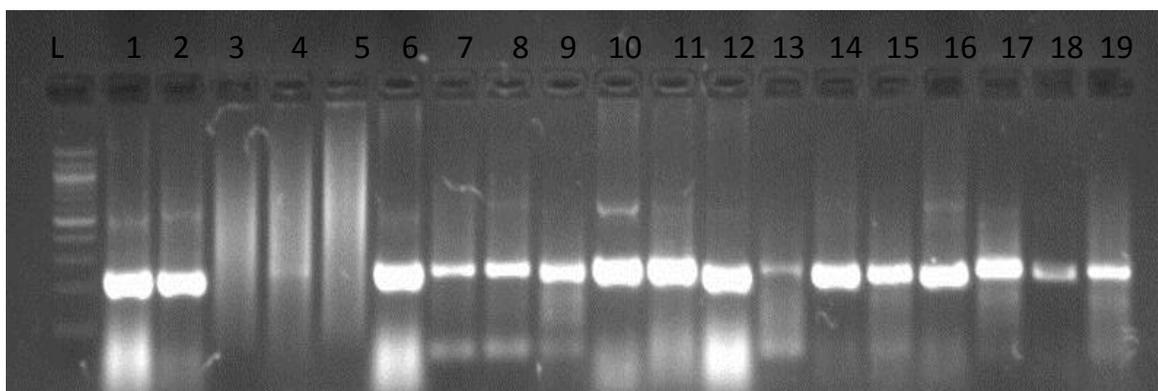


Plate 2a. Representative PCR image of single bands of the cape gooseberry accessions.

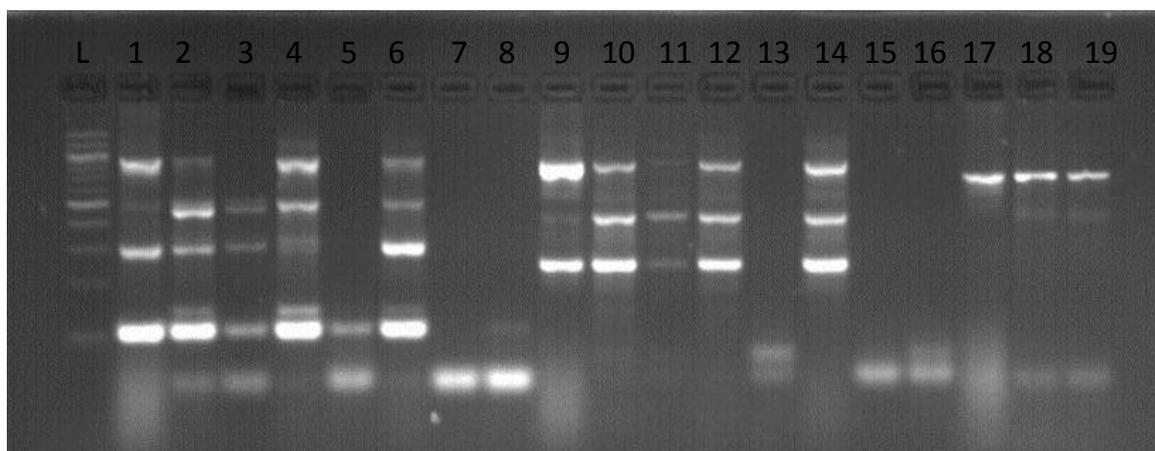


Plate 2b. Representative PCR images of multiple bands of varied sizes of Cape gooseberry accessions.

between 1.10 (SSR112) and 2.00 (SSR1) with a mean of 1.28 (Table 2). The lowest gene diversity was observed for SSR112 (0.08) while the highest was 0.82 for SSR1. All the markers used in the study were polymorphic, SSR1 was the most polymorphic marker (PIC = 0.77) while SSR112 was the least polymorphic (PIC = 0.079)

while the mean polymorphic content was 0.432 (Table 2). The average PIC for the entire population (0.432) classifies the markers used in this study within the range of loci with intermediate polymorphism (0.25 to 0.5) according to Ge et al., (2013). The calculated Shannon's Information indices averaged at 0.25 and ranged between

Table 2. Summary statistics of Marker polymorphism and genetic diversity indices of cape gooseberry accessions.

Locus	Sample size	Allele No.	MAF	*Na	*Ne	H*	I*	PIC	Ht	Hs	Fst	Nm
SSR1	70	7	0.257	2.00	1.9984	0.4996	0.6927	0.793	0.4994	0.0300	0.9399	0.0320
SSR2	70	11	0.386	2.00	1.9935	0.4984	0.6915	0.758	0.4996	0.1214	0.7571	0.1604
SSR10	70	7	0.271	2.00	1.9600	0.4898	0.6829	0.78	0.4983	0.0537	0.8923	0.0604
SSR11	70	6	0.471	2.00	1.9935	0.4984	0.6915	0.662	0.3533	0.0781	0.7789	0.1420
SSR15	70	2	0.814	2.00	1.4336	0.3024	0.4799	0.257	0.2778	0.0000	1.0000	0.0000
SSR54	70	2	0.557	2.00	1.9742	0.4935	0.6866	0.371	0.4941	0.1769	0.6419	0.2789
SSR72	70	2	0.771	2.00	1.5448	0.3527	0.5375	0.291	0.3394	0.0700	0.7938	0.1299
SSR77	70	2	0.771	2.00	1.5448	0.3527	0.5375	0.291	0.3394	0.0700	0.7938	0.1299
SSR112	70	2	0.957	2.00	1.0894	0.0820	0.1769	0.079	0.0740	0.0592	0.2000	2.0000
SSR118	70	3	0.857	2.00	1.3243	0.2449	0.4101	0.239	0.3193	0.0586	0.8166	0.1123
SSR121	70	3	0.657	2.00	1.6897	0.4082	0.5983	0.409	0.4362	0.0221	0.9493	0.0267
SSR123	70	4	0.771	2.00	1.4706	0.3200	0.5004	0.356	0.3333	0.0773	0.7682	0.1509
SSR126	70	2	0.871	2.00	1.2888	0.2241	0.3837	0.199	0.3084	0.0408	0.8676	0.0763
SSR138	70	4	0.714	2.00	1.5817	0.3678	0.5544	0.426	0.4220	0.1322	0.6867	0.2281
SSR146	70	4	0.529	2.00	1.5817	0.3678	0.5544	0.583	0.3288	0.1449	0.5592	0.3941
Mean	70	61	0.257	1.51	1.2673	0.3668	0.2464	0.432	0.4083	0.1089	0.7333	0.1818
St. Dev				0.50	0.3479	0.1877	0.2709		0.1721	0.0321		

Na = Observed number of alleles, Ne = Effective number of alleles (Kimura and Crow, 1964), H* = Nei's (1973) gene diversity, I* = Shannon's Information index (Lewontin, 1972) MAF = Major allele frequency, PIC = polymorphic information content, Ht = genetic distance for the among populations, Hs = genetic distance within the population and Gst = degree of differentiation among populations and Nm = estimate of gene flow from Gstor Gcs. E.g., $Nm = 0.5(1 - Gst)/Gst$; (McDermott and McDonald, 1993).

0.1769 (SSR1) and 0.6927 (SSR112) (Table 2). Seven SSR primer pairs (SSR1, SSR2, SSR10, SSR11, SSR123, SSR138, and SSR146) produced more than five alleles among the 70 Cape gooseberry accessions. The seven SSR primer pairs would have a priority of choice in evaluating cape gooseberry because they were more informative in their ability to segregate between the accessions and had PIC values above 0.4. The highest number of SSR loci detected in this study contained dinucleotide (two nucleotide units) and hendecanucleotide (eleven nucleotide units) repeats characteristic of markers located along untranslated regions (UTR). This is in line with the hypothesis by Morgante et al.

(2002) which posit that in most plants the SSR loci are found along the UTR's and may have accounted for the markers with low PIC because SSRs found in untranslated regions have been reported to be less polymorphic than genomic markers (Ellis and Burke, 2007). The 61 polymorphic SSR alleles detected in this study were higher than the 6 alleles reported by Chacón et al. (2016) and the 30 polymorphic alleles reported by Simbaqueba et al. (2011). These differences may be due to different accessions used in previous studies or the stringency of scoring. The differences could also be due to the relatively narrow genetic base of commercial Cape gooseberry varieties used in the previous

studies.

Population structure of Cape gooseberry accessions

Genetic distances between gooseberry accessions within counties

Plant breeding applications such as germplasm collections, selection of parental materials, identification of quantitative trait loci, linkage, and association mapping are dependent on previous genetic diversity information (Rao and Hodgkin, 2002; Zhu et al., 2008; Rauf et al., 2010). In this study, the application of SSR markers on the

Table 3. Average genetic distances of cape gooseberry accessions within counties.

County	Ho	He	Mean Fst	HWE
Nakuru	0.500	0.1021	0.0031	
Nyandarua	0.456	0.0832	0.1778	
Kericho	0.488	0.1018	0.0076	
Kiambu	0.394	0.0973	0.0668	
Murang'a	0.384	0.0968	0.0662	
Laikipia	0.277	0.0375	0.5874	
HWE	-	-	-	70.00

Ho = Observed heterozygosity, He = Expected heterozygosity, Fst = gene differentiation coefficient, HWE = Hardy Weinberg equilibrium as calculated from Chi square.

Table 4. Genetic diversity of cape gooseberry accessions between counties.

County	Nakuru	Nyandarua	Kericho	Kiambu	Murang'a	Laikipia
Nakuru	0.000					
Nyandarua	0.0007					
Kericho	0.0000	0.0007				
Kiambu	0.0001	0.0008	0.0001			
Murang'a	0.0001	0.0003	0.0001	0.0003		
Laikipia	0.0067	0.0048	0.0066	0.0063	0.0061	

whole collection of *P. peruviana* from six selected counties of Kenya revealed a moderate to high genetic diversity. Average distances between individuals in the same cluster (expected heterozygosity) ranged from 0.03 in Laikipia county to 0.102 in Nakuru county indicating that Cape gooseberry accessions in Nakuru county have higher genetic diversity than those in Laikipia county. Genetic distance value recorded in this study ($H_e=0.3668$) is higher than those reported by Bonilla et al. (2008) ($H_e = 0.2559$) and lower than those reported by Garzón-Martínez et al. (2015) ($H_e = 0.40$). However, it is noteworthy that Bonilla et al. (2008) used RAM markers which are dominant in nature and therefore have the inherent limitation of underestimating allelic diversity as compared to the co-dominant SSR markers used in this study; while Garzón-Martínez et al. (2015) used genomic SNP markers which have a higher resolution than SSR markers.

Mean fixation index (F_{ST}) ranged between 0.003 and 0.58 for Nakuru and Laikipia respectively. Cape gooseberry accessions in Nakuru and Kericho counties exhibited insignificant genetic differentiation ($F_{st} < 0.05$) indicating that the accessions in the two counties are interbreeding freely. Cape gooseberry accessions in Nyandarua, Kiambu, Murang'a, and Laikipia showed significant genetic differentiation ($F_{st} < 0.05$). This great genetic differentiation is a sign of geographic isolation and a high inbreeding rate (Table 3). Overall, genetic differentiation reported in the present study is very high ($F_{st} = 0.7333$). This may be attributed to the fact that

Kenyan Cape gooseberry cultivars have no history of domestication. Lagos et al. (2008) reported that *P. peruviana* is more than 53% outcrossing and its domestication from the wild did not involve a long process as compared to fruit-bearing relatives such as tomato (Labate et al., 2009; Sim et al., 2012). It is therefore probable that natural selection is still an important factor in retaining heterogeneous populations with broad genetic adaptability and variability (Rauf et al., 2010).

Genetic distances of Cape gooseberry accessions among the six selected counties

For this study, the allele frequency of Cape gooseberry between counties was significantly lower than diversity within the counties (Table 4). Allele-frequency divergence among counties was the highest between Laikipia and Nakuru (0.0067) and the lowest between Kericho and Nakuru (0.00). Overall there was a very small allele divergence observed in Cape gooseberry across counties. This shows that the gooseberry accessions from the selected counties share most of the alleles evaluated, and this is an indicator of low geographic differentiation among these accessions. The findings of this study show that the population showed a slight deviation from the Hardy Weinberg equilibrium ($HWE = 0.7$). This may be due to the significant effects of natural selection resulting from the limited domestication of Cape gooseberry

Table 5. Analysis of molecular variance (AMOVA) for the 70 cape gooseberry accessions from 6 select counties of Kenya.

Source	SS	Df	MS	F	Prob > F
Between accessions	10.36	5	2.07	92.80	0.0000
Within acce	1.429	64	0.02	7.20	
Total	11.79	69	0.17		

SS = sum of squares, MS= expected mean squares, F = F-statistics, >F = Significance level.

Table 6. Factorial analysis of the Gooseberry accessions from the five select counties.

County	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5
Nakuru	31.906	40.492	1.271	2.584	0.641
Nyandarua	0.116	0.010	0.062	0.113	-0.048
Kericho	0.038	0.098	0.101	-0.100	-0.014
Kiambu	-0.056	0.099	0.034	0.034	0.041
Murang'a	-0.063	0.047	-0.128	0.008	-0.010
Laikipia	-0.227	-0.212	0.083	-0.007	-0.016
Proportion of Variance	31.714	40.533	1.422	2.632	0.592
Cumulative variance	31.714	72.247	73.67	76.303	76.89
Eigen values	0.013	0.012	0.008	0.005	0.002
% Inertia	27.74	26.27	17.62	10.35	4.37

collections in Kenya. This value is slightly higher than those reported by Tian et al. (2008) (HWE = 0.5481) though Tian's study used cultivated varieties which may have higher domestication influence. Partitioning of the genetic variation of the Cape gooseberry accessions was done using Analysis of Molecular Variance (AMOVA) to determine whether variation observed between the accessions was due to genetic makeup or microclimatic factors. A total of 92.80% of the variation was found among the accessions while a total of 7.20% of the variation was revealed within the accessions (Table 5). This is a further proof that the Cape gooseberry accessions in Kenya have a broad genetic diversity.

Factorial analysis

Factorial analysis was performed to analyze the genetic relationship and population structure of the accessions within and between the counties. A dissimilarity matrix calculated using raw data from the SSR "1" and "0" matrix was used for factorial analysis using Darwin 6.0.21 software. The first five axes accounted for 76.81% of all the variance observed in the test samples, with 41.2, 52.71, and 1.85% explained by PC axes 1, 2, and 3, respectively. The highest variance was observed in Nakuru county (31.90) indicating that cape gooseberry accessions in this county have high genetic diversity and the lowest variance was recorded in Laikipia (-0.21) showing comparatively genetic diversity in the accessions

in this county (Table 6).

The high percentage of variation explained by the first three components in the factorial analysis shows that the differentiation of most of the individuals was well captured. However, it is noteworthy to consider the use of a larger number of high-resolution markers and platforms such as SNPs and genotyping by sequencing (GBS) due to the outcrossing nature of the species (Zhu et al., 2008).

Factorial analysis grouped the accessions into five distinct clusters depending on the county of origin. Accessions from Kiambu and Murang'a counties, however, clustered together showing that they were genetically more identical (Figure 2). Individuals were distinct within clusters with little overlap between individuals indicating that the collections are genetically diverse. The collections show a higher level of diversity within and across the clusters (Figure 2).

Phylogenetic analysis

A distance tree was constructed in Darwin 6 using the UPGMA method. The robustness of the node of the phylogenetic tree was assessed from 1000 bootstrap replicates. In this study, the minimum dissimilarity value for the phylogenetic tree was 0.027 while the maximum value was 1. This high dissimilarity value is further proof of high genetic diversity found in the Kenyan gooseberry accessions. Tree length varied between 0 for duplicates

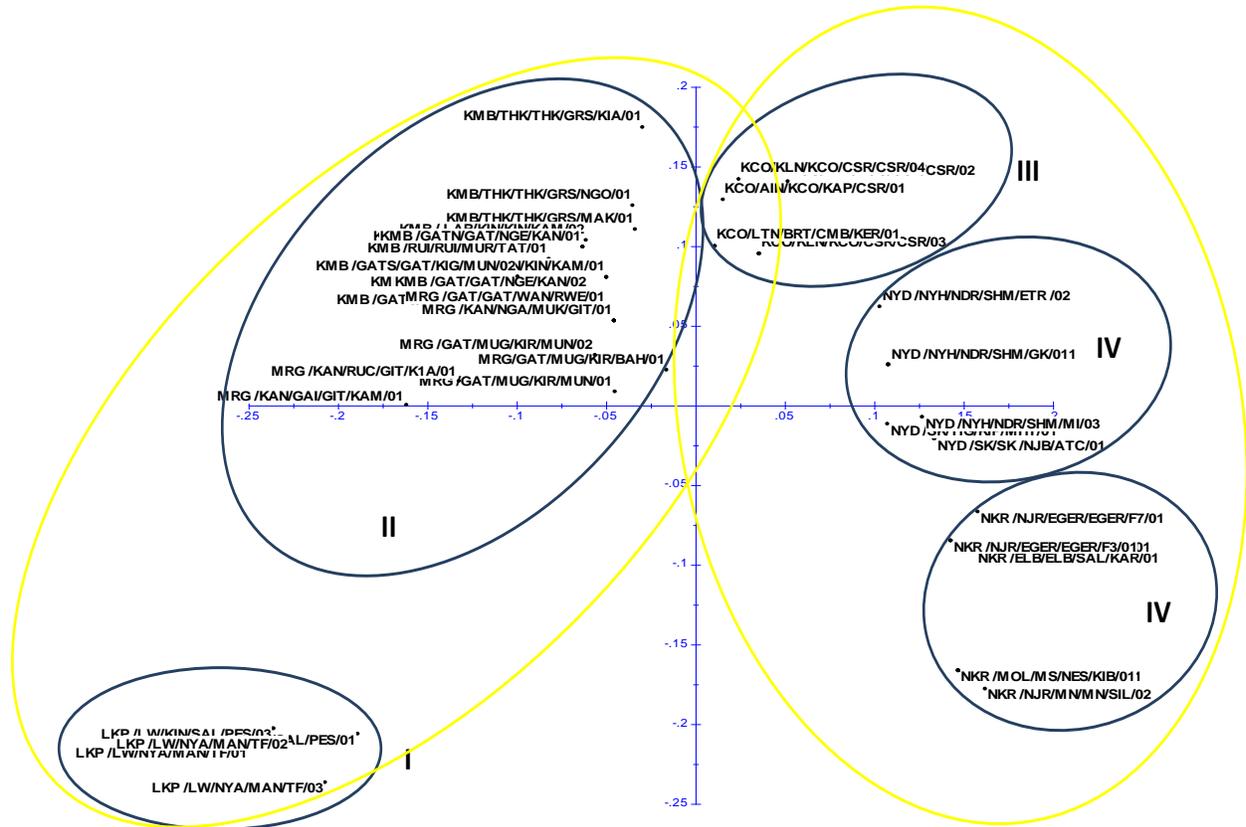


Figure 2. Factorial analysis of cape gooseberry accessions from the six selected counties in Kenya; Names in the circles correspond to individual accession names.

and 0.161 for distinct individuals, the edge length sum was 2.04 indicating that most of the cultivars are distinct. The collections were grouped into two major clusters I and II. Cluster I contained accessions sampled from Nyandarua (1.A), Nakuru (I.B), Kericho (I.C), and Laikipia (I.D) counties while cluster II contained accessions from Kiambu (II.A) and Murang'a (II.B). The phylogenetic tree further clustered the accessions into six sub-clusters (I.A, I.B, I.C, I.D, II.A, and II.B) based on the county of origin (Figure 3).

The clustering analysis applied was able to detect a geographical distribution pattern. This observation may be due to lack of frequent gene flow through the exchange of seeds among the counties by humans as is often the case in heavily domesticated species. This finding is in disagreement with the findings of Garzon-Martínez et al. (2015) who failed to deduce any geographical clustering of Columbian *P. peruviana* varieties using SSR and InDels markers. This may be because the study by Garzón-Martínez et al. (2015) used well defined domestic commercial gooseberry accessions. Both factorial and phylogenetic population analyses show that the whole Cape gooseberry population has two different genetic populations. Using the PCA approach, two different populations within the *P. peruviana* were

identified.

CONCLUSION AND RECOMMENDATION

This is the first study in Kenya that used SSR markers to genetically characterize Cape gooseberry. The study established that Cape gooseberry in the six target counties of Kenya have a broad genetic diversity. Based on the SSR data, the 70 accessions were classified into two main phylogenetic groups and six sub-clusters which corresponded to the county of origin through factorial analysis, principal component analysis (PCA), and phylogenetic analyses. The study also established that seven SSR primer pairs with higher polymorphism namely, SSR1, SSR2, SSR10, SSR11, SSR123, SSR138, and SSR146 have a wide applicability in genotype identification and characterization of the population structure of Cape gooseberry. The information generated by this study contributes to understanding diversity and population structure and enhances the management of Cape gooseberry genetic resources in Kenya.

This study enhances understanding of levels of genetic variations among Kenyan Cape gooseberry germplasm

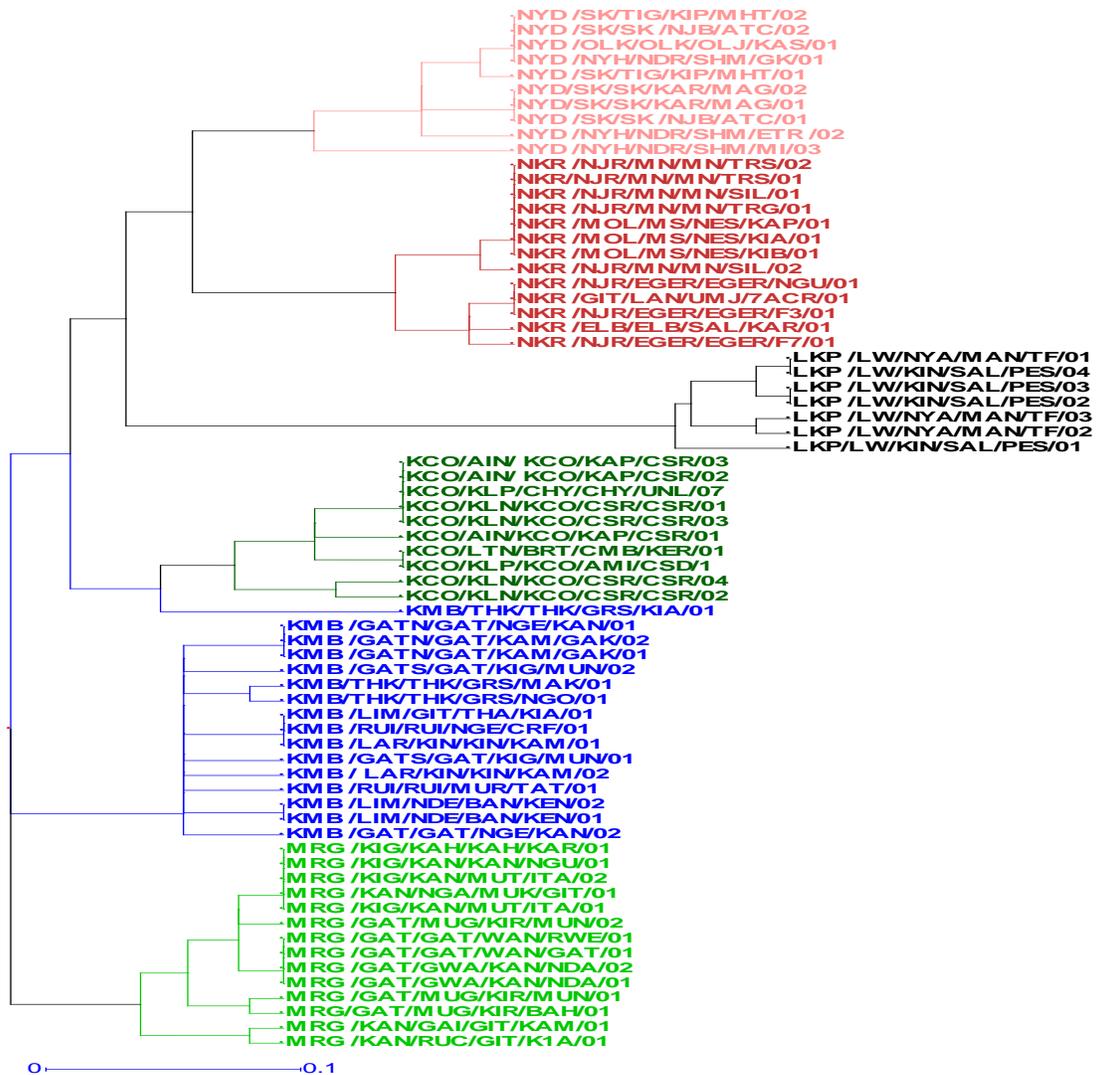


Figure 3. Phylogenetic trees of cape gooseberry accessions based on SSR data.

and informs the need to introduce commercial Cape gooseberry varieties as sources of genetic variation for breeding and hybridization purposes. The findings of the study also inform on the need to use more advanced molecular platforms such as genome-wide sequencing to establish more diversity in wild and cultivated Cape gooseberry in Kenya.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGMENTS

This work was funded by RUFORUM in collaboration with Egerton University TAGDev program under the nurturing

Grant award, Kenya. I would like to thank the staff at KALRO-Njoro (Food Crops Institute) for the assistance and insight accorded to me during my lab work, colleagues, and staff from the Department of Crops, Horticulture and Soils, Egerton University for their support during the research.

REFERENCES

- Ali A, Singh P (2016). Studies on production potential of cape gooseberry (*Physalis peruviana* L.) in sodic soil under varying agronomic manipulations. *Journal of Applied and Natural Science* 8(1):368-374.
- Bhandari R, Nishant Bhanu A, Srivastava K, Singh M, Shreya H (2017). Assessment of genetic diversity in crop plants. An overview. *Advances in Plant and Agriculture Research* 7:00255.
- Bonilla M, Piedrahíta E, Mauricio A, Terranova P, Amariles V, Eduardo J, Flórez M (2008). Molecular characterization of 43 accessions of cape gooseberry from six departments of Colombia. *Acta Agronomica* 57:109-115.

- Campisano A, Pancher M, Puopolo G, Puddu A, Lòpez-Fernàndez S, Biagini B, Pertot I (2015). Diversity in endophyte populations reveals functional and taxonomic diversity between wild and domesticated grapevines. *American Journal of Enology and Viticulture* 66(1):12-21.
- Chacón S, Isabel M, Sánchez Y, Barrero M, Stella L (2016). Genetic structure of a Colombian cape gooseberry (*Physalis peruviana* L.) collection by means of microsatellite markers. *Agronomía Colombiana* 34(1):5-16.
- Demir K, Bakir M, Sarikamis G, Acunalp S (2010). Genetic diversity of eggplant (*Solanum melongena*) germplasm from Turkey assessed by SSR and RAPD markers. *Genetics and Molecular Research* 9(3):1568-1576.
- Ellis J, Burke M (2007). EST-SSRs as a resource for population genetic analyses. *Heredity* 99(2):125-132.
- Garzón-Martínez G, Osorio-Guarín J, Delgadillo-Durán P, Mayorga F, Enciso-Rodríguez F, Landsman D, Barrero L (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.
- Herrera M, Ortiz D, Fischer G, Chacón I (2011). Behavior in yield and quality of 54 cape gooseberry (*Physalis peruviana* L.) accessions from north-eastern Colombia. *Agronomía Colombiana* 29(2):189-196.
- Kimura M, Crow F (1964). The number of alleles that can be maintained in a finite population. *Genetics* 49(4):725-738.
- Labate A, Robertson D, Wu F, Tanksley D, Baldo M (2009). EST, COSII, and arbitrary gene markers give similar estimates of nucleotide diversity in cultivated tomato (*Solanum lycopersicum* L.). *Theoretical and Applied Genetics* 118:1005-1014.
- Lagos T, Vallejo F, Criollo H, Muñoz J (2008). Sexual reproduction of the cape gooseberry. *Acta Agronomica* 57(2):81-87.
- Leiva-Brondo M, Prohens J, Nuez F (2001). Genetic analyses indicate superiority of performance of cape gooseberry (*Physalis peruviana* L.) hybrids. *Journal of New Seeds* 3(3):71-84.
- Lewontin C (1972). The apportionment of human diversity. *Evolutionary Biology* 6:381-398.
- Liu K, Muse V (2005). Power Marker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128-2129.
- Mason A (2015). SSR Genotyping. In *Plant Genotyping*. Humana press, New York, USA. pp. 77-89.
- McDermott M, McDonald A (1993). Gene flow in plant pathosystems. *Annual review of phytopathology* 31(1):353-373.
- Morgante M, Hanafey M, Powell W (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature genetics* 30(2):194-200.
- Myers N, Mittermeier A, Mittermeier G, Da Fonseca A, Kent J (2000). Biodiversity hotspots for conservation priorities. *Nature* 403(6772):853-858.
- Nei M (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* 70(12):3321-3323.
- Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89(3):583-590.
- Ochatt S, Jain S (2007). *Breeding of Neglected and Under-Utilized Crops, Spices, and Herbs*. Science Publishers, Inc.: New Hampshire, NH, USA.
- Perrier X, Jacquemoud-Collet P (2006). DARwin software.
- Porebski R, Bailey S, Baum R (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter* 15:8-15.
- Rao R, Hodgkin T (2002). Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell Tissue Organ Culture* 68:1-19.
- Rauf S, Teixeira da Silva J, Khan A, Naveed A (2010). Consequences of plant breeding on genetic diversity. *International Journal of Plant Breeding* 4:1-21.
- Ravi B, Hrideek K, Kumar K, Prabhakaran R, Mal B, Padulosi S (2010). Mobilizing neglected and underutilized crops to strengthen food security and alleviate poverty in India. *Indian Journal of Plant Genetic Resources* 23(1):110-116.
- Sim C, Van Deynze A, Stoffel K, Douches S, Zarka D, Ganai W, Chetelat T, Hutton F, Scott W, Gardner G, Panthee R, Mutschler M, Myers R, Francis M (2012). High-density SNP genotyping of tomato (*Solanum lycopersicum* L.) reveals patterns of genetic variation due to breeding. *PLoS One* 7(9):e45520.
- Simbaqueba J, Sanchez P, Sanchez E, Zарantes N, Chacon M, Barrero S, Mariño-Ramírez L (2011). Development and characterization of microsatellite markers for the Cape gooseberry *Physalis peruviana*. *PLoS one* 6(10):e26719.
- Thies E (2000). Promising and underutilized species, crops and breeds. Available at: <https://cgspace.cgiar.org/bitstream/handle/10568/3682/thies.pdf?sequence=1>
- Villacorta M, Shaw J (2013). *Peruvian Power Foods: 18 Superfoods, 101 Recipes, and Anti-aging Secrets from the Amazon to the Andes*. Health Communications, Inc.
- Virk S, Zhu J, Newbury J, Bryan J, Jackson T, Ford-Lloyd V (2000). Effectiveness of different classes of molecular marker for classifying and revealing variation in rice (*Oryza sativa*) germplasm. *Euphytica* 112(3):275-284.
- Wu J, Chang P, Lin L, Wang S, Hou F, Ng T (2009). Supercritical carbon dioxide extract of *Physalis peruviana* induced cell cycle arrest and apoptosis in human lung cancer H661 cells. *Food and Chemical Toxicology* 47(6):1132-1138.
- Yeh F, Yang R, Boyle T (2000). POPGENE Version 1.32: Software for Population Genetic Analysis.
- Zhu C, Gore M, Buckler S, Yu J (2008). Status and prospects of association mapping in plants. *Plant Genome* 1:5-20.