Fruits 74(5), 236–248 | ISSN 0248-1294 print, 1625-967X online | https://doi.org/10.17660/th2019/74.5.4 | © ISHS 2019

Original article

cirad

# Genetic diversity of common guava in Kenya: an underutilized naturalized fruit species

J.C. Chiveu[1,2,a], M. Mueller[3], K.V. Krutovsky[3,4,5,6], K. Kehlenbeck[7], E. Pawelzik[1] and M. Naumann[1]

[1] Department of Crop Sciences, Division of Quality of Plant Products, University of Göttingen, Carl-Sprengel-Weg 1, D-37075 Göttingen, Germany

[2] Department of Seed, Crop and Horticultural Sciences, School of Agriculture and Biotechnology, University of Eldoret, P.O. Box 1125-30100, Eldoret, Kenya

[3] Department of Forest Genetics and Forest Tree Breeding, University of Göttingen, Faculty of Forest Sciences and Forest Ecology, Büsgenweg 2, 37077 Göttingen, Germany

[4] Department of Ecosystem Science and Management, Texas A&M University, 305 Horticulture and Forest Science Building, MS 2138 TAMU, College Station, TX 77843-2138, USA

[5] Laboratory of Population Genetics, N.I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 3 Gubkina Str., Moscow 119333, Russia

[6] Laboratory of Forest Genomics, Genome Research and Education Center, Siberian Federal University, 50a/2 Akademgorodok, Krasnoyarsk 660036, Russia

[7] World Agroforestry Center (ICRAF), P.O. Box 30677, 00100 Nairobi, Kenya (Currently Rhine-Waal University of Applied Sciences, Faculty of Life Sciences, Marie Curie-Straße 1, 47533 Kleve, Germany)

## Summary

*Introduction* – **Common guava (*Psidium guajava* L.) fruit has a significant nutritional and medicinal potential besides its economic importance. Currently, the world guava fruit production is based only on a few cultivars. It is not clear when guava was introduced in Kenya, but the species is currently naturalized. There is no detailed study on guava diversity in Kenya to enable a comparison with other guava-producing countries for purposes of characterization and improvement.** *Objectives* – **The main objective of the study was to analyse the genetic diversity and differentiation of guava accessions from four geographically diverse regions of Kenya.** *Materials and methods* – **The genetic diversity of 177 guava accessions collected from four regions of Kenya (Coast, Eastern, Rift Valley, and Western) was assessed using 13 simple sequence repeat (SSR) markers.** *Results and discussion* – **The neighbour-joining (NJ) phylogenetic tree revealed most accessions generally clustering into multiple weakly supported groups. Only 46 out of 177 accessions were supported by bootstrap values above 50% and clustered in twenty-two groups, each comprising two or three individual accessions only. The principle coordinates analysis (PCoA) did not reveal clear-cut clusters along geographic origins or fruit flesh colour of the samples. The fixation index ($F_{IS}$) was very high ($F_{IS}$ = 0.511), which could be due to a high level of either inbreeding and/or differentiation. The white-fleshed accessions were clustered together with the red-fleshed types, indicative of some degree of genetic similarity, but also pointing to a possibility of shared ancestry between them.** *Conclusion* – **For guava conservation and selection for breeding and utilization purposes in Kenya, sampling of many individual accessions covering the geographical range of the species is recommended.**

## Significance of this study

*What is already known on this subject?*

- Guava fruit is highly nutritious but globally underutilized and only a few cultivars are under production. Initial results have reported genetic clustering mostly based on agroecological regions and flesh colour.

*What are the new findings?*

- Great genetic diversity exists in individual accessions rather than on agroecological basis. Moreover, white and red-fleshed accessions may also cluster together and therefore selection for improvement simply based on morphological tree and fruit traits could lead to genetic bottlenecks. Extensive selection comprising many trees from agroecological regions is recommended.

*What is the expected impact on horticulture?*

- Proper selection for improvement of guava should consider many individuals within agroecological regions. Germplasm exchange among countries in a short term could also help to come up with improved varieties in terms of production and fruit quality.

**Keywords**
domestication, genetic differentiation, microsatellite, *Psidium guajava* L., selection, simple sequence repeats

## Introduction

Common (also known as yellow or lemon) guava (*Psidium guajava* L.) is one of the most important fruit crops domesticated in Mesoamerica and widely cultivated in the tropics and some sub-tropical regions (Gautam *et al.*, 2010; Rodríguez *et al.*, 2010). The fruit is consumed fresh and processed. It is rich in several important nutrients, as the fresh

[a] Corresponding author: josiah.chiveu@agr.uni-goettingen.de.

fruit pulp is high in vitamins, particularly vitamin C, phosphorus, and potassium, as well as many antioxidants and dietary fibres (Jiménez-Escrig *et al.*, 2001; Lukmanji *et al.*, 2008; Flores *et al.*, 2015). Furthermore, not only fruits, but also leaves, flowers, roots and bark are traditionally used in medicine (Gutiérrez *et al.*, 2008). In recent years, the guava fruit has gained popularity in the international trade due to its nutritional value and the diversity of processed products including jam, jelly, and juice (Valera-Montero *et al.*, 2016).

There are probably more than 400 guava cultivars around the world, but only a few are under common cultivation (Pommer and Murakami, 2009). Cultivars grown are widely diverse regarding tree size, bearing habit, and yield, as well as fruit size, shape, ripening season and quality in terms of nutrient composition (Pommer and Murakami, 2009; Sharma *et al.*, 2010). Irrespective of the morphological and chemical diversities observed in these cultivars, several reports indicated that selection of the accessions was based on a few important morphological traits such as fruit size and shape, flesh thickness, skin and flesh colour (Mehmood *et al.*, 2013, 2015; Galli *et al.*, 2015; Valera-Montero *et al.*, 2016). The important chemical traits included total soluble solids (TSS), titratable acidity (TA), and ascorbic acid content (Mehmood *et al.*, 2013). Recent studies focused also on other traits such as fruit aroma (*e.g.*, Moon *et al.*, 2018). The focus on a few morphological and chemical traits means that much of the variation is left untapped. This is likely to lead to genetic vulnerability of the crop (Nogueira *et al.*, 2014), especially with respect to climate change.

With regard to Kenya, guava is found in all the agroecological regions apart from very arid areas and highlands. It is, however, not clear when it was introduced in Kenya, but has been naturalized and occurs in the wild and on farmer's fields. Most guava fruits are collected for home consumption, although lately fresh fruits could be found being marketed in major Kenyan towns. According to the Horticultural Crops Directorate, HCD (2014), the trees are mainly unattended, growing from seeds dispersed unintentionally. The HCD report cited the lack of suitable superior varieties, limited knowledge of agronomic and postharvest practices, and limited value addition as constraints in guava production in Kenya. The starting point for guava improvement in Kenya is, therefore, to collect germplasm for characterization and conservation, and synthesize knowledge of the existing genetic diversity and production situation.

The objective of this study was to analyse the genetic diversity and differentiation of guava accessions collected in four different regions of Kenya. It was hypothesized that due to diverse range of agroecological conditions of guava, the accessions are highly differentiated. In addition, white and red-fleshed accessions would cluster separately. Accurate knowledge of the genetic diversity and the origin of the accessions would assist in the selection of parental materials in breeding programmes. Such information will consequently eliminate the possibility of redundant collection of identical individuals for conservation and improvement, thereby enhancing cost effective use of land, space and time regarding field gene bank establishment and breeding activities.

## Materials and methods

### Study site selection and sampling procedure

Four known major guava-producing regions in Kenya were selected for guava sampling based on the horticulture-validated report data (HCD, 2014). The sites for sampling within these regions included Meru (Eastern region), Uasin-Gishu and Elgeyo-Marakwet (Rift Valley region), Homabay, Siaya, Kakamega, and Vihiga (Western region), and Kwale, Kilifi, and Mombasa (Coastal region) (Figure 1). Most sampled trees were found on individual farmer's fields, but also one prison fruit farm and one commercial fruit farm were included in the sampling. A majority of the sampled trees (27%) growing on farmers' fields were found either growing in fruit orchards together with other fruit trees or in crop fields. About 25% of the sampled trees were found in the farmers' homestead fence and 11% were growing in uncultivated farm parts together with other wild trees and shrubs. The remainder of the trees were found growing as shade trees in farmers' compounds, along rivers, and in fallow fields. The trees were sampled randomly in cases where more than 10 trees occurred on the same farm, though in most cases all guava trees within the farm were sampled.

### Leaf material sampling

Leaves from a total of 177 guava trees (here also referred to as accessions or sample) were sampled (Supplementary Table S1). At least five young fully developed healthy leaves were picked at random from each of the 177 accessions (72 in Western, 48 in Rift Valley, 38 in Coastal, and 19 in Eastern regions). The leaves were then briefly dried under a shade in the field and placed in the sealable polythene bags containing silica gel for complete drying and preservation. Afterwards, the leaf samples were taken to the laboratory for DNA isolation and subsequent fragment analysis.

### DNA isolation and quantification

Nuclear DNA from silica gel dried-leaf samples (about 300–500 mg) was extracted using the DNeasy 96 Plant Kit (QIAGEN, Hilden, Germany). DNA quality and quantity were checked on a 3% (w/v) agarose gel by comparing it with a known λ DNA concentration. The stock DNA preps were diluted accordingly with molecular-grade water and then stored at -20 °C for eventual analyses.

### Primer selection for polymerase chain reaction (PCR) amplification

The PCR primers, previously designed for guava by Risterucci *et al.* (2005) and successfully used in assessing guava diversity (Valdés-Infante *et al.*, 2010; Sitther *et al.*, 2014), were used. Twenty primer pairs used by Risterucci *et al.* (2005) were tested, and the best 13, which were also good for multiplexing, were selected. The primers were labelled with fluorescent dyes; and those primers that amplified alleles with non-overlapping fragment lengths were pooled to save on the PCR cost and time (Supplementary Table S2). Table S2 also shows the allele size ranges in base pairs (bp) observed for each primer pair in our guava accessions.

The PCR amplification was conducted in a 14 µL volume containing 1 µL of genomic DNA (20 ng $\mu L^{-1}$), 1.5 µL PCR buffer (0.8 M Tris-HCl pH 9.0, 0.2 M $[NH_4]_2SO_4$, 0.2% w/v Tween-20; Solis BioDyne, Tartu, Estonia), 1.5 µL $MgCl_2$, 1 µL dNTPs, 2 µL fluorescent dye-labelled forward and reverse primers, 0.2 µL *Taq* DNA polymerase (HOT FIREPol DNA Polymerase, Solis BioDyne, Tartu, Estonia), and 6.8 µL distilled water. The amplification procedure included an initial denaturation step of 95 °C for 15 min followed by 35 cycles of 94 °C for 1 min (denaturation), 50 °C for 1 min (annealing), 72 °C for 1 min (extension), and a final extension step of 72 °C for 20 min. The PCR reactions were conducted in a T-Professional thermocycler (Biometra, Analytik Jena, Germany).

In preparation for fragment analysis, the PCR products were diluted with water in a ratio of 1:100. Next, 2 µL of the diluted PCR product comprising of 12 µL of Hi-Di Formamide and 0.6 µL of internal size standard Genescan 500 Rox (Applied Biosystems Inc.) was denatured at 95 °C for 3 min in a thermocycler. The fragments were then analysed in an ABI 3130xl Genetic Analyzer (Applied Biosystems, Waltham, MA, USA). The sizes of the microsatellite fragments were determined, and the microsatellite loci were genotyped using the GeneMapper software v. 4.0 (Applied Biosystems Inc.).

### Data analysis

The number of alleles per locus ($N_a$), observed ($H_o$) and expected ($H_e$) heterozygosities were computed using the GenAlEx 6.5 software (Peakall and Smouse, 2012). The duplicates were checked by multi-locus matching. The fixation index ($F_{IS}$) in the entire sample was computed following the definition of Wright (1965) using Genepop software v. 4.0 (Rousset, 2008). Nei's chord distance (1983) matrix between accessions was generated using microsatellite analyser (Dieringer and Schlötterer, 2003) with 10,000 bootstrappings. The distances were then used to generate a phylogenetic tree using the neighbour-joining (NJ) method of clustering (Saitou and Nei, 1987) available in PHYLIP (Felsenstein, 1993), which was visualized using the Geneious software v. 10.1.3. (www.geneious.com; Kearse *et al.*, 2012).

A PCoA with covariance standardization available in GenAlEx 6.5 was used to determine the spatial distribution of the samples based on their genetic distances. Moreover, analysis of molecular variance (AMOVA) was performed from a triangular distance matrix with 1,000 permutations to quantify genetic variation within and among regions. The SSR data was also subjected to a Bayesian cluster analysis using the STRUCTURE software v. 2.3.2 (Pritchard *et al.*, 2000)

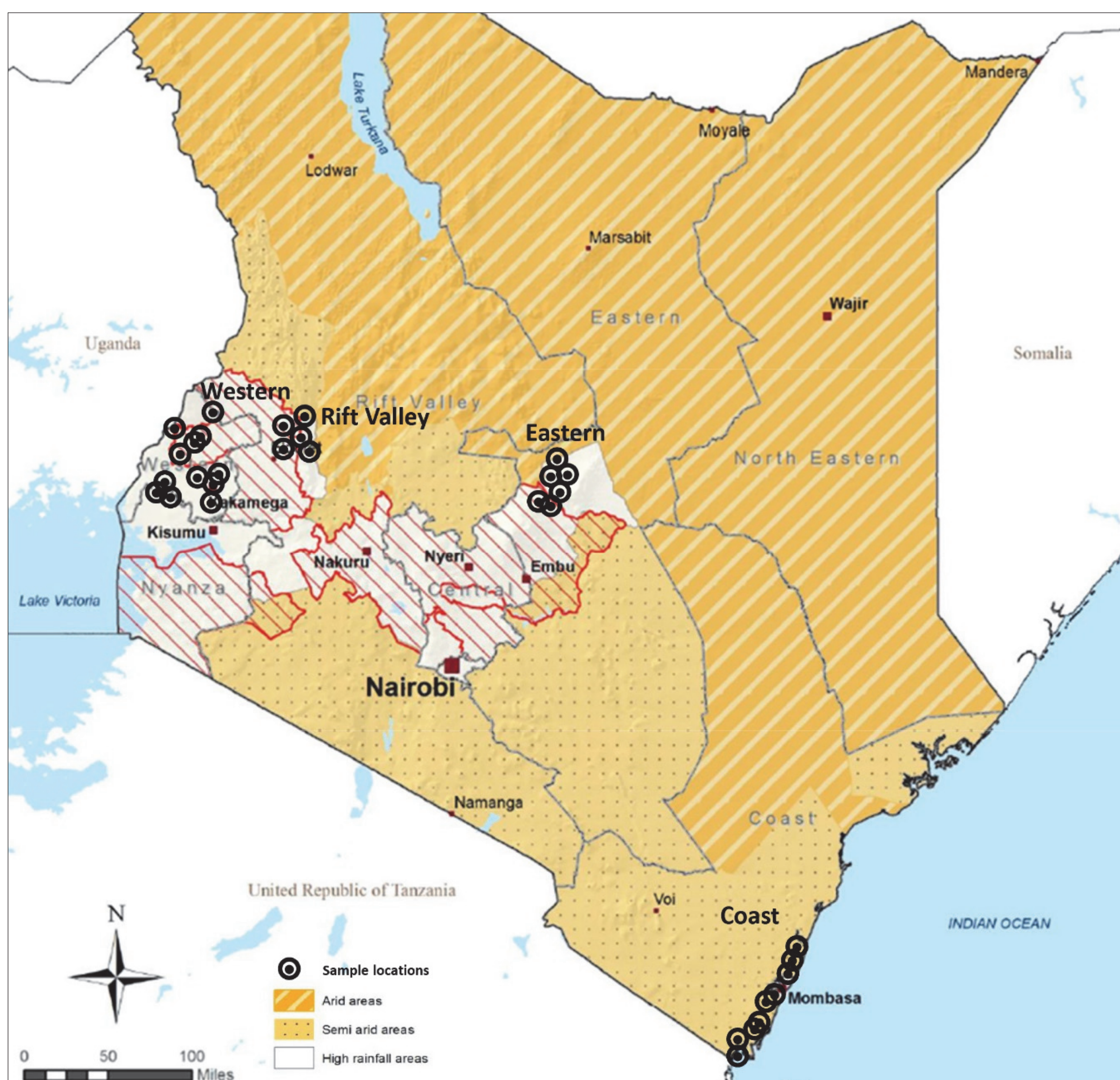



**Figure 1.** Sample collection locations for the guava accessions (circles) in four regions of Kenya (Coastal, $n = 38$; Eastern, $n = 19$; Rift Valley, $n = 48$; and Western, $n = 72$). The map was adapted from the International Fund for Agricultural Development (IFAD) report (https://www.ifad.org/documents/10180/42d2d8ea-644f-4bc3-a977-c3edb103b148, accessed December 3, 2017. See also Supplementary Table S1).

in order to infer the most likely number of subpopulations or groups (clusters) in the sample. The admixture model was applied without assigning individual trees to particular groups or geographic regions *a priori*; and the samples were tested for number of potential clusters ($K$) ranging from 1 to 10. Ten runs per each $K$ were performed, each consisting of a burn-in of 100,000, followed by 1,000,000 Monte Carlo Markov Chain iterations. The $\Delta K$ value approach (Evanno *et al.*, 2005) was used to determine the most probable number of clusters using the STRUCTURE Harvester program (Earl and von Holdt, 2012).

## Results

All the PCR primers generated fragments in all samples, and all the amplified alleles were polymorphic. No samples with genotypes identical for all markers (supposedly duplicates) were found in the collected 177 guava accessions. The 13 primer pairs amplified 84 alleles in the studied guava accessions in total. The highest number of alleles (13) was found in locus mPgCIR10, while the least number of alleles (four) in loci mPgCIR08, mPgCIR11, mPgCIR13, and mPgCIR21. The expected heterozygosity ($H_e$) values ranged from 0.507 to 0.843 with an average of 0.630, while the observed heterozygosity ($H_o$) values ranged from 0.192 to 0.497 with an average of 0.312. The fixation index ($F_{IS}$) among the accessions for the entire sample ranged from 0.410 to 0.621 for different markers with an average of 0.511 (Table 1).

Most samples in the cluster analysis were not supported by bootstrap values above 50% (Supplementary Figure S1). Only 46 samples had bootstrap values above 50% and clus-

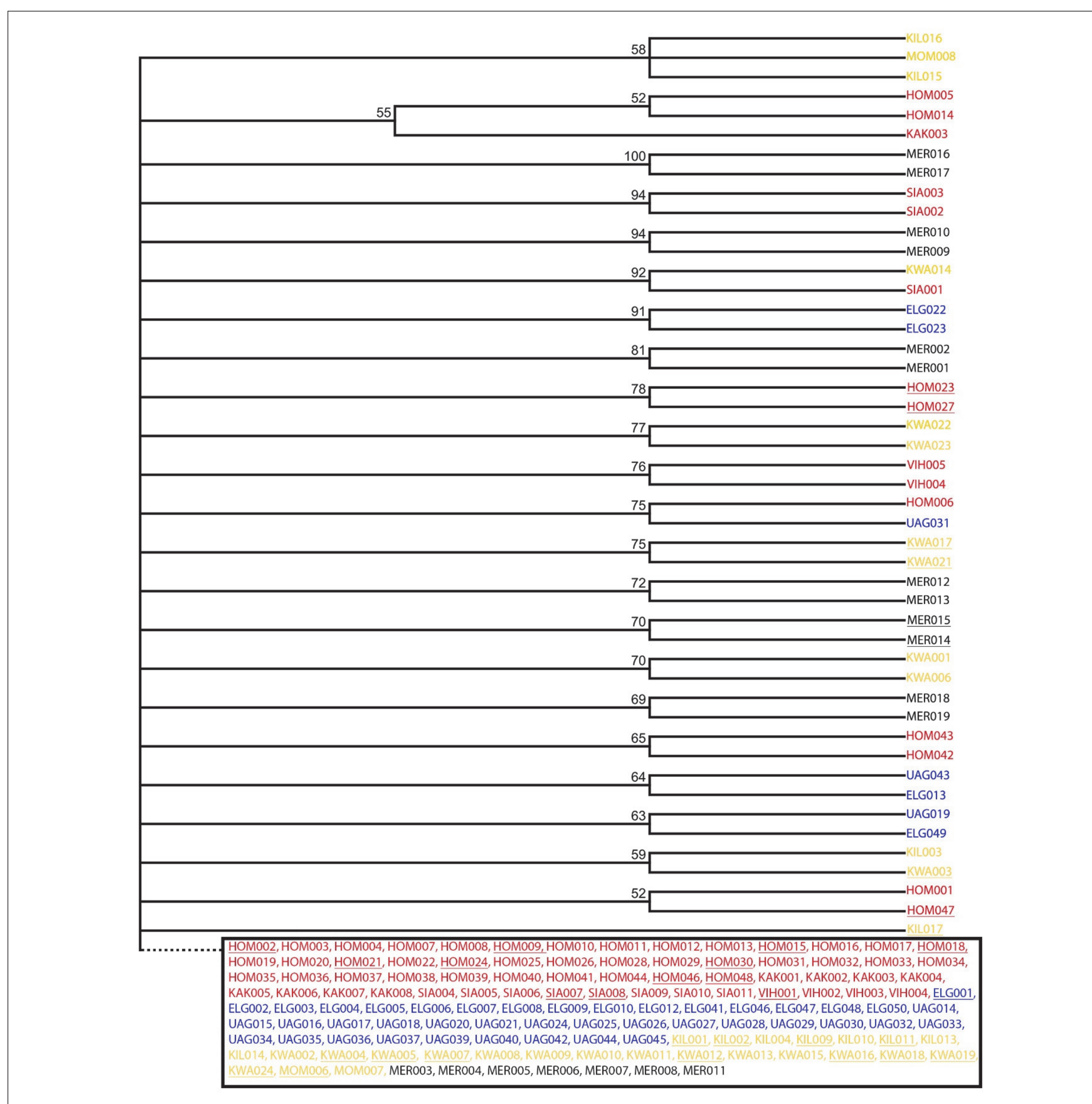



**FIGURE 2.** The neighbour-joining phylogenetic tree of 177 guava accessions collected from four regions of Kenya. The accession colour codes depict the region of collection (Red = Western, Blue = Rift Valley, Gold = Coast, Black = Eastern). The white-fleshed accessions are underlined. Only bootstrap values of 50% and more are indicated for tree nodes after 10,000 bootstrappings. Accessions within the text box were supported by bootstrap values below 50% (see also Supplementary Figure S1).

tered mainly into groups of two or three accessions in the observed 22 clusters (Figure 2). In general, samples from all regions were found in the well sustained 22 clusters. It was interesting to observe that the white-fleshed accessions were clustered together with red-fleshed accessions in two of the clusters. It was also noted that the accessions within the well-supported 22 clusters grouped mainly according to their geographical origins, except in two of the groups.

The STRUCTURE analysis, however, did not reveal any genetic clusters based on both the *LnP(D)* and $\Delta K$ value (Evanno *et al.*, 2005) analyses (Figure 3 and Supplementary Figure S2). Consequently, the accessions were significantly admixed with any number of clusters, thereby pointing at the possibility of existence of only one genetic cluster. There was no

**Table 1.** Summary genetic variation statistics for 177 guava accessions collected from four regions of Kenya assessed with 13 simple sequence repeats.

| Locus | $N_a$ | $N_e$ | $H_o$ | $H_e$ | $F_{IS}$ |
|---|---|---|---|---|---|
| mPgCIR07 | 6 | 2.2 | 0.220 | 0.545 | 0.596 |
| mPgCIR08 | 4 | 2.1 | 0.203 | 0.518 | 0.608 |
| mPgCIR09 | 6 | 3.6 | 0.356 | 0.721 | 0.506 |
| mPgCIR10 | 13 | 6.4 | 0.497 | 0.843 | 0.410 |
| mPgCIR11 | 4 | 2.7 | 0.282 | 0.634 | 0.554 |
| mPgCIR13 | 4 | 2.0 | 0.192 | 0.507 | 0.621 |
| mPgCIR15 | 8 | 5.2 | 0.418 | 0.809 | 0.483 |
| mPgCIR17 | 5 | 2.3 | 0.232 | 0.565 | 0.590 |
| mPgCIR19 | 7 | 2.4 | 0.322 | 0.590 | 0.454 |
| mPgCIR20 | 9 | 2.4 | 0.311 | 0.583 | 0.467 |
| mPgCIR21 | 4 | 2.2 | 0.305 | 0.540 | 0.435 |
| mPgCIR22 | 6 | 2.4 | 0.328 | 0.592 | 0.446 |
| mPgCIR25 | 8 | 3.9 | 0.390 | 0.740 | 0.474 |
| Mean | 6.5 | 3.1 | 0.312 | 0.630 | 0.511 |

$N_a$ - number of alleles; $N_e$ - number of effective alleles; $H_o$ - observed heterozygosity; $H_e$ - expected heterozygosity; $F_{IS}$ - fixation index (see also Supplementary Table S2).

**Table 2.** AMOVA based on the region of collection of 177 guava accessions from four regions of Kenya.

| Source of variation | Degrees of freedom | Sum of squares | Mean square | Estimated variance | Variation (%) | Amova statistic | $P$* |
|---|---|---|---|---|---|---|---|
| Among regions | 3 | 243.8 | 81.3 | 1.7 | 13 | 0.131 | 0.001 |
| Within regions | 173 | 1,936.6 | 11.2 | 11.2 | 87 | | |
| Total | 176 | 2,180.4 | | 12.9 | 100 | | |

* Based on 1,000 random permutations.

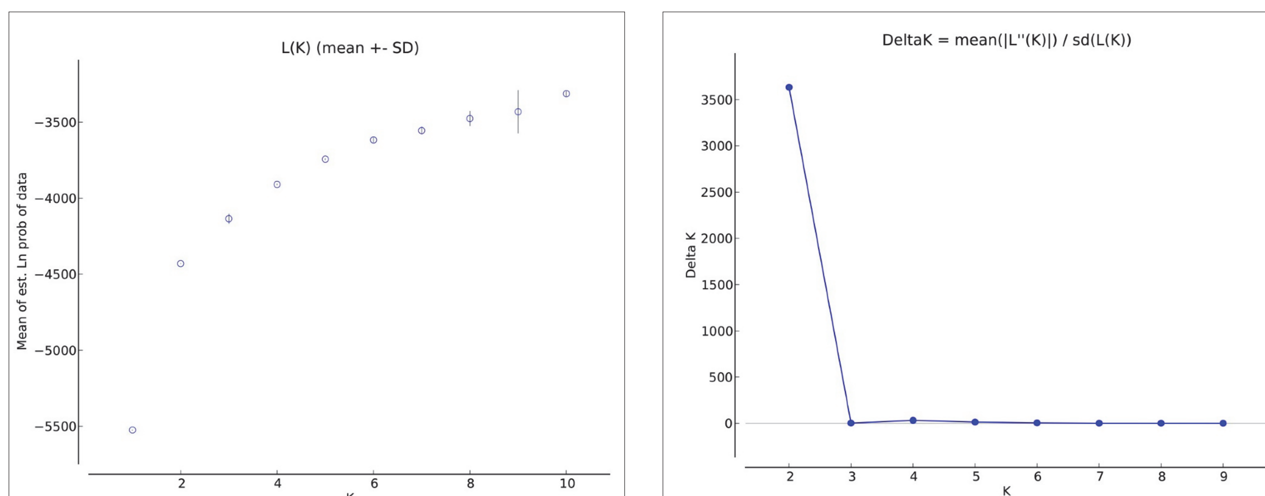



**Figure 3.** Estimate of probability of the data for a given *K*, *(LnP(D))* (Pritchard *et al.*, 2000) and plots for detecting the most probable number of K groups ($\Delta K$) (Evanno *et al.*, 2005) based on 13 SSR loci genotyped in 177 Kenyan guava accessions (see also Supplementary Figure S2).

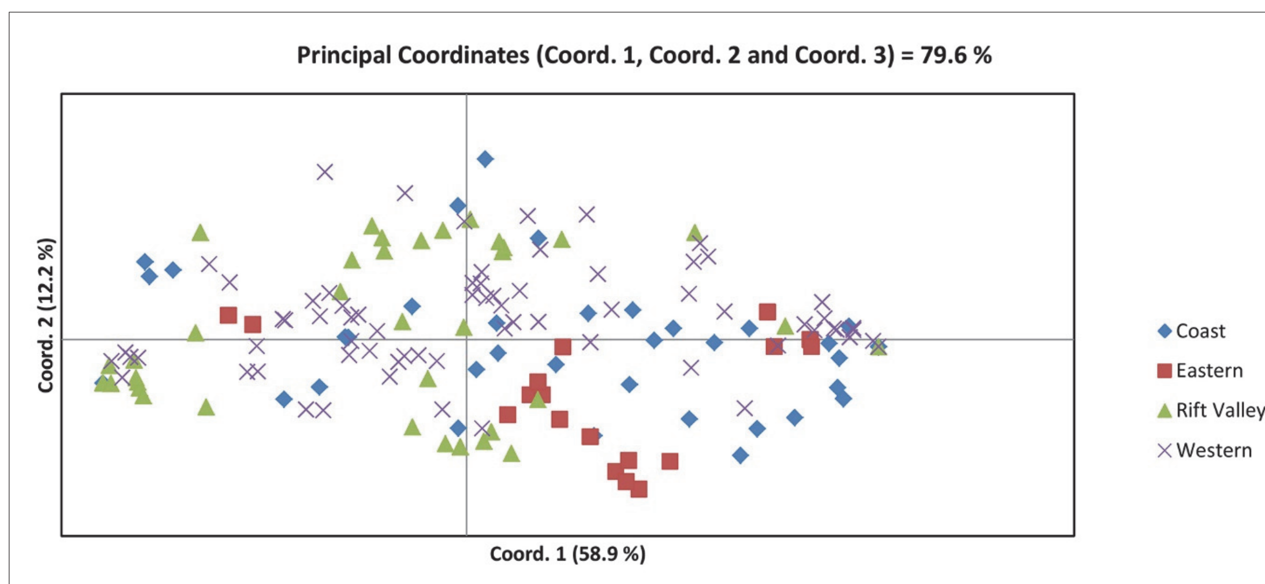



**FIGURE 4.** Principal coordinate analysis of 177 guava accessions collected from four regions of Kenya. The first three axes explained 79.6% of the total variation, with the first axis explaining 58.9%, the second 12.2%, and the third 8.5%.

preferential grouping of the accessions based on fruit flesh colour.

The PCoA confirmed the lack of strongly differentiated groups or clusters among accessions (Figure 4); this is similar to the results of the NJ clustering and Bayesian cluster analysis. Similarly, AMOVA performed in accordance with the region of collection of the accessions revealed that much of the genetic variation (87%) resided within accessions in a region (Table 2).

## Discussion

The comparison of the Kenyan guava germplasm multi-locus SSR genotypes revealed no identical or duplicate accessions; therefore, each accession was genetically distinct from the others. The low levels of observed heterozygosity (mean = 0.312) with respect to expected heterozygosity (mean = 0.630) likely indicates a high level of genetic differentiation between accessions within identified groups, including those that existed within the same geographical locations. Similar results were also reported using SSR markers by Sitther *et al.* (2014), where the observed and expected heterozygosities were 0.2 and 0.7 on the average, respectively, in the guava germplasm found in the United States. The expected heterozygosity was even much higher and varied between 0.392 and 0.961 with an average of 0.824 in the Indian guava based on the SSR genotypes (Kanupriya *et al.*, 2011), while a much lower expected heterozygosity ranging from 0.027 to 0.172 with an average of 0.085, was found in the Pakistan guava germplasm also based on the SSR markers (Mehmood *et al.*, 2015). Similarly, low to moderate levels of expected heterozygosity (0.057 to 0.568) were detected in the Cuban guava germplasm using microsatellites (Rodríguez *et al.*, 2007). The differences in the heterozygosity indices in the aforementioned studies were attributed to the high inbreeding and a possibility of cross incompatibility occurring in guava. The difference in the diversity among the mentioned studies, however, could be also due to the different microsatellite loci used (Pommer and Murakami, 2009), but it is more likely that they were accession or sample specific (Belaj *et al.*, 2003).

The average fixation index in our study was 0.511 (Table 1), implying a high genetic substructure within our guava accessions or a high inbreeding rate. The Myrtaceae flower has been reported to be hermaphrodite, which increases the possibility of selfing (Grattapaglia *et al.*, 2012). Nakasone and Paull (1998) estimated the outcrossing rate as only 35–40% in *Psidium guajava*, which is in agreement with our results. In contrast, very high inbreeding coefficients of 0.8 and 0.85 have been reported in the SSR studies by Sitther *et al.* (2014) and Mehmood *et al.* (2015), respectively. These very high values of an average inbreeding coefficient point at the possibility of cross-incompatibility, which may hinder the effectiveness of creating true hybrids and recombining favourable alleles from parental clones in guava as reported by Mehmood *et al.* (2015).

Based on the NJ phylogenetic tree, some of the Kenyan guava accessions mainly from one region were well supported by bootstrap values above 50% and grouped together in clusters of two or three individuals (Figure 2); although accessions from all the regions could be found together in different small genetic clusters when lower than 50% bootstrap values were considered (Supplementary Figure S1). A similar observation was also reported by Kareem *et al.* (2018) for 37 guava accessions in Pakistan. This could be indicative of parental material diversity due to a diverse ancestral breeding history. In the present study, the PCoA, however, depicted an overlap between these clusters and groups that was also supported by the observed genetically admixed individuals based on the Bayesian clustering implemented in the STRUCTURE software. This implies that some accessions are very similar and can form genetic groups, while others are genetically distinct and admixed irrespective of their existence within the same geographical environment as was also found by Kherwar *et al.* (2018) for 36 Indian guava varieties. This high intra-regional genetic heterogeneity was also supported by results of the AMOVA (Table 2). Population structure is a result of geographic adaptation and natural selection (Lehermeier *et al.*, 2015). Thus, it was expected that many individual guava accessions from one region rather than just two or three would cluster together and form distinct groups

based on their geographic origin, which was not the case in the present study. The lack of a robust sub-structuring in the Kenyan guava accessions could therefore be attributed to the high inbreeding as reported in other studies (Nakasone and Paull, 1998) and plant material (seeds and seedlings) transfer across different regions.

The white-fleshed guava accessions were found in groups together with the red-fleshed types in the NJ phylogenetic tree. Therefore, the expectation based on previous studies (Chen *et al*., 2007; Kanupriya *et al*., 2011) that all the white-fleshed accessions would be clustered in the same one group was not confirmed in our study. Moreover, grouping white- together with the red-fleshed types could be due to a shared ancestry of these accessions at some point in time, as supported by the Bayesian clustering in our study. However, the possibility of sympatric speciation cannot be ruled out in the Kenyan guava germplasm, especially when few accessions with similar flesh colour cluster together with higher bootstrap values. In related studies, Chen *et al.* (2007) and Alam *et al.* (2018) each identified two genetic groups based on RAPD markers in Taiwan and Bangladesh, respectively. The commercial and wild genotypes of guava were clustered separately in the Taiwan study, possibly depicting selection pressure on the traits of interest for the commercial group. The latter group included two subgroups, which roughly clustered white- and red-fleshed guavas separately. In the study of Bajpai *et al*. (2008), 22 guava genotypes were clustered according to their regions of origin based on the RAPD and directed amplification of minisatellite DNA (DAMD) markers. Additionally, molecular data allowed Coser *et al*. (2012) and Nogueira *et al*. (2014) to cluster most genotypes in accordance with their origins in Brazil. Notably, although the genotypes were registered as cultivars having been highly selected based on production quality traits, but they still clustered according to the initial parental material origin with minimal segregation (Coser *et al*., 2012). In the Indian guava germplasm, the pink flesh cultivars were reported to group separately from those with white flesh (Kanupriya *et al*., 2011), probably pointing at their distinct evolutionary pathways.

Initial efforts to improve guava production in Kenya have only concentrated on the conservation of the available germplasm through collection based on morphological attributes such as leaf shape or fruit flesh colour, among others. A few of these genotypes collected from various regions of the country have been conserved at the Kenya Agricultural and Livestock Research Organization (KALRO). Therefore, guava farmers in Kenya rely on the genetic variation existing in the wild populations on their farms and probably on limited exchange of some genotypes they consider superior for production and quality traits between individual farmers.

## Conclusion

The SSR markers were able to distinguish among the Kenyan guava accessions. Much of the genetic variation resided within individual accessions found in different geographical locations of the country, and therefore, the hypothesis that the accessions would cluster according to their agroecological environments was rejected. In addition, the white-fleshed guava accessions clustered together with the red-fleshed guava accessions, thereby suggesting a shared ancestry. It is therefore recommended that sampling for conservation and improvement should include trees from different regions, covering the full ecological range of the species. This data also forms a basis for comparison of guava genetic diversity studies with other guava-producing countries, and therefore joint research aimed at improving guava production could be initiated.

## Acknowledgments

This research was funded by the International Fund for Agricultural Development (IFAD) and the European Commission under the World Agroforestry Centre (ICRAF) project "Tree crops development in Africa to benefit the poor". Thanks to Deutscher Akademischer Austauschdienst (DAAD) and the Kenyan government through National Commission for Science, Technology and Innovation (NACOSTI) for providing a Ph.D. scholarship to Mr. Chiveu Josiah; and to all farmers as well as field and lab staff at the ICRAF. All experimental analysis was done at the Department of Forest Genetics and Forest Tree Breeding, University of Göttingen, Germany. We appreciate Mrs. Alexandra Dolynska for the technical assistance with SSR fragment analysis.

## References

Alam, F., Islam, K.D., and Rahman, M.M. (2018). Variability among selective guava (*Psidium guajava* L.) varieties revealed by morphology and RAPD marker. J. Biol. Sci. *7*(2), 89–98. https://doi.org/10.3329/jujbs.v7i2.40750.

Bajpai, A., Chandra, R., Rajan, S., and Srivastava, N. (2008). RAPD and minisatellite markers for genetic diversity and relationship in guava varieties. Indian J. Genet. Plant Breed. *68*, 441.

Belaj, A., Satovic, Z., Cipriani, G., Baldoni, L., Testolin, R., Rallo, L., and Trujillo, I. (2003). Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in Olive. Theor. Appl. Genet. *107*, 736–744. https://doi.org/10.1007/s00122-003-1301-5.

Chen, T., Ng, C., Wang, C., and Shyu, Y. (2007). Molecular identification and analysis of *Psidium guajava* L. from indigenous tribes of Taiwan. J. Food Drug Anal. *15*, 82–88.

Coser, S.M., da Silva, F.M.F., Ferreira, A., Mitre, L.K., Carvalho, C.R., and Clarindo, W.R. (2012). Assessment of genetic diversity in *Psidium guajava* L. using different approaches. Sci. Hortic. *148*, 223–229. https://doi.org/10.1016/j.scienta.2012.09.030.

Dieringer, D., and Schlötterer, C. (2003). Microsatellite Analyser (MSA): A platform independent analysis tool for large microsatellite data sets. Mol. Ecol. Notes *3*, 167–169. https://doi.org/10.1046/j.1471-8286.2003.00351.x.

Earl, D.A., and von Holdt, B. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. *14*, 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Felsenstein, J. (1993). Phylogeny Inference Package. (Seattle, WA, USA: University of Washington, Department of Genetics).

Flores, G., Wu, S.-B., Negrin, A., and Kennelly, E.J. (2015). Chemical composition and antioxidant activity of seven cultivars of guava (*Psidium guajava*) fruits. Food Chem. *170*, 327–335. https://doi.org/10.1016/j.foodchem.2014.08.076.

Galli, J.A., Michelotto, M.D., Soares, M.B.B., Martins, A.L.M., Palharini, M.C.d.A., and Fischer, I.H. (2015). Characterization of guava plants belonging to a germplasm bank and cultivated in an organic system. Acta Hortic. *1137*, 213–218. https://doi.org/10.17660/ActaHortic.2016.1137.30.

Gautam, N.N., Singh, K., Singh, B., Seal, S., Goel, A., and Goel, V. (2010). Studies on clonal multiplication of guava (*Psidium guajava* L.) through cuttings under controlled conditions. Aust. J. Crop Sci. *4*, 666.

Grattapaglia, D., Vaillancourt, R.E., Shepherd, M., Thumma, B.R., Foley, W., Külheim, C., Potts, B.M., and Myburg, A.A. (2012). Progress in Myrtaceae genetics and genomics: eucalyptus as the pivotal genus. Tree Genet. Genomes *8*, 463–508. https://doi.org/10.1007/s11295-012-0491-x.

Gutiérrez, R.M.P., Mitchell, S., and Solis, R.V. (2008). *Psidium guajava*: A review of its traditional uses, phytochemistry and pharmacology. J. Ethnopharmacol. *117* (1), 1–27. https://doi.org/10.1016/j.jep.2008.01.025.

Horticultural Crops Directorate (2014). Horticulture validated report 2014. HCD, Nairobi, Kenya, pp. 68. http://www.agricultureauthority.go.ke/wp-content/uploads/2016/05/Horticulture-Validated-Report-2014-Final-copy.pdf (accessed November 30, 2017).

Jiménez-Escrig, A., Rincón, M., Pulido, R., and Saura-Calixto, F. (2001). Guava fruit (*Psidium guajava* L.) as a new source of antioxidant dietary fiber. J. Agric. Food Chem. *49*, 5489–5493. https://doi.org/10.1021/jf010147p.

Kanupriya, P., Madhavi, L.C., Aswath, L.R.B., Padmakar, C., Vasugi, and Dinesh, M.R. (2011). Cultivar identification and genetic fingerprinting of guava (*Psidium guajava*) using microsatellite markers. Int. J. Fruit Sci. *11*, 184–196. https://doi.org/10.1080/15538362.2011.578521.

Kareem, A., Jaskani, M.J., Mehmood, A., Khan, I.A., Awan, F.S., and Sajid, M.W. (2018). Morpho-genetic profiling and phylogenetic relationship of guava (*Psidium guajava* L.) as genetic resources in Pakistan. Rev. Bras. Frutic. *40*(4) (e-069). https://doi.org/10.1590/0100-29452018069.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., and Duran, C. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Kherwar, D., Usha, K., Mithra, S.V.A., and Singh, B. (2018). Microsatellite (SSR) marker assisted assessment of population structure and genetic diversity for morpho-physiological traits in guava (*Psidium guajava* L.). J. Plant. Biochem. and Biotechnol. *27*(3), 284–292. https://doi.org/10.1007/s13562-017-0438-2.

Lehermeier, C., Schön, C.-C., and de los Campos, G. (2015). Assessment of genetic heterogeneity in structured plant populations using multivariate whole-genome regression models. Genetics *201*, 323–337. https://doi.org/10.1534/genetics.115.177394.

Lukmanji, Z., Hertzmark, E., Mlingi, N., Assey, V., Ndossi, G., and Fawzi, W. (2008). Tanzania food composition tables (Dar es Salaam, Tanzania: MUHAS-TFNC, HSPH).

Mehmood, A., Jaskani, M.J., Ahmad, S., and Ahmad, R. (2013). Evaluation of genetic diversity in open pollinated guava by iPBS primers. Pak. J. Agric. Sci. *50*, 591–597.

Mehmood, A., Luo, S., Ahmad, N.M., Dong, C., Mahmood, T., Sajjad, Y., Jaskani, M.J., and Sharp, P. (2015). Molecular variability and phylogenetic relationships of guava (*Psidium guajava* L.) cultivars using inter-primer binding site (iPBS) and microsatellite (SSR) markers. Genet. Resour. Crop Evol. *63*, 1345–1361. https://doi.org/10.1007/s10722-015-0322-7.

Moon, P., Fu, Y., Bai, J., Plotto, A., Crane, J., and Chambers, A. (2018). Assessment of fruit aroma for twenty-seven guava (*Psidium guajava*) accessions through three fruit developmental stages. Sci. Hortic. *238*, 375–383. https://doi.org/10.1016/j.scienta.2018.04.067.

Nakasone, H.Y., and Paull, R.E. (1998). Tropical fruits (Wallingford, UK: Cab International), 461 pp.

Nogueira, A.M., Ferreira, M., Guilhen, J., and Ferreira, A. (2014). Multivariate analysis in a genetic divergence study of *Psidium guajava*. Genet. Mol. Res. *13*, 10657–10668. https://doi.org/10.4238/2014.December.18.8.

Peakall, R., and Smouse, P.E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics *28*, 2537–2539. https://doi.org/10.1093/bioinformatics/bts460.

Pommer, C.V., and Murakami, K.R.N. (2009). Breeding guava (*Psidium guajava* L.). In Breeding Plantation Tree Crops: Tropical Species (Springer), p. 83–120. https://doi.org/10.1007/978-0-387-71201-7_3.

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics *155*, 945–959.

Risterucci, A.M., Duval, M.F., Rohde, W., and Billotte, N. (2005). Isolation and characterization of microsatellite loci from *Psidium guajava* (L.). Mol. Ecol. Notes *5*, 745–748. https://doi.org/10.1111/j.1471-8286.2005.01050.x.

Rodríguez, N., Valdés-Infante, J., Becker, D., Velázquez, B., González, G., Sourd, D., Rodríguez, J., Billotte, N., Risterucci, A.-M., and Ritter, E. (2007). Characterization of guava accessions by SSR markers, extension of the molecular linkage map, and mapping of QTLs for vegetative and reproductive characters. Acta Hortic. *735*, 201–216. https://doi.org/10.17660/ActaHortic.2007.735.27.

Rodríguez, N.N., Valdés, J., Rodríguez, J.A., Velásquez, J.B., Rivero, D., Martínez, F., González, G., Sourd, D.G., González, L., and Cañizares, J. (2010). Genetic resources and breeding of guava (*Psidium guajava* L.) in Cuba. Biotecnología Aplicada *27*, 238–240.

Rousset, F. (2008). Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol. Ecol. Resour. *8*, 103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406–425.

Sharma, A., Sehrawat, S.K., Singhrot, R.S., and Ajinath, T. (2010). Morphological and chemical characterization of *Psidium* species. Notulae Botanicae Horti Agrobotanici Cluj-Napoca *38*, 28.

Sitther, V., Zhang, D., Harris, D.L., Yadav, A.K., Zee, F., Meinhardt, L.W., and Dhekney, S.A. (2014). Genetic characterization of guava (*Psidium guajava* L.) germplasm in the United States using microsatellite markers. Genet. Resour. Crop Evol. *61*, 829–839. https://doi.org/10.1007/s10722-014-0078-5.

Valdés-Infante, J., Rodríguez, N.N., Velasquez, B., Rivero, D., Martinez, F., Espinosa, G., Risterucci, A.M., Billotte, N., Becker, D., and Rohde, W. (2010). Simple sequence repeats (SSRs) for diversity characterization of guava (*Psidium guajava* L.). Acta Hortic. *849*, 155–162. https://doi.org/10.17660/ActaHortic.2010.849.17.

Valera-Montero, L., Muñoz-Rodríguez, P., Silos-Espino, H., and Flores-Benítez, S. (2016). Genetic diversity of guava (*Psidium guajava* L.) from Central Mexico revealed by morphological and RAPD markers. Int. J. Exp. Bot. *85*, 176–183.

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution *19*(3), 395–420. https://doi.org/10.1111/j.1558-5646.1965.tb01731.x.

Received: Apr. 22, 2018
Accepted: Jun. 2, 2019

**Supplemental Information Table S1.** Accession codes, region of collection, fruit flesh colour, altitude and geographic coordinates of the locations of the sampled Kenyan guava accessions. Related to Figure 1.

| Sample number | Accession code | Region | Fruit flesh colour | Latitude (N°/S°) | Longitude (E°) | Altitude (m) |
|---|---|---|---|---|---|---|
| 1 | KIL001 | Coast | White | 03.69568 °S | 039.72340 °E | 208 |
| 2 | KIL002 | Coast | White | 03.69580 °S | 039.72343 °E | 199 |
| 3 | KIL003 | Coast | Red | 03.69679 °S | 039.72604 °E | 202 |
| 4 | KIL004 | Coast | Red | 03.69518 °S | 039.72219 °E | 200 |
| 5 | KIL009 | Coast | White | 03.92239 °S | 039.74352 °E | 23 |
| 6 | KIL010 | Coast | Red | 03.92240 °S | 039.74314 °E | 25 |
| 7 | KIL011 | Coast | White | 03.92226 °S | 039.74282 °E | 22 |
| 8 | KIL013 | Coast | Red | 03.91339 °S | 039.74015 °E | 18 |
| 9 | KIL014 | Coast | Red | 03.91348 °S | 039.74015 °E | 17 |
| 10 | KIL015 | Coast | Red | 03.91338 °S | 039.73997 °E | 18 |
| 11 | KIL016 | Coast | Red | 03.91332 °S | 039.73999 °E | 21 |
| 12 | KIL017 | Coast | White | 03.91347 °S | 039.73988 °E | 20 |
| 13 | KWA001 | Coast | Red | 04.16923 °S | 039.59783 °E | 23 |
| 14 | KWA002 | Coast | Red | 04.16853 °S | 039.59749 °E | 19 |
| 15 | KWA003 | Coast | White | 04.16856 °S | 039.59748 °E | 19 |
| 16 | KWA004 | Coast | White | 04.16854 °S | 039.59750 °E | 19 |
| 17 | KWA005 | Coast | White | 04.16494 °S | 039.57737 °E | 104 |
| 18 | KWA006 | Coast | Red | 04.16495 °S | 039.57743 °E | 97 |
| 19 | KWA007 | Coast | White | 04.16496 °S | 039.57764 °E | 119 |
| 20 | KWA008 | Coast | Red | 04.16782 °S | 039.56780 °E | 108 |
| 21 | KWA009 | Coast | Red | 04.16837 °S | 039.56796 °E | 92 |
| 22 | KWA010 | Coast | Red | 04.16860 °S | 039.56822 °E | 94 |
| 23 | KWA011 | Coast | Red | 04.34928 °S | 039.53458 °E | 22 |
| 24 | KWA012 | Coast | White | 04.34926 °S | 039.53447 °E | 23 |
| 25 | KWA013 | Coast | Red | 04.34938 °S | 039.53400 °E | 26 |
| 26 | KWA014 | Coast | Red | 04.34318 °S | 039.51459 °E | 35 |
| 27 | KWA015 | Coast | Red | 04.33752 °S | 039.44971 °E | 117 |
| 28 | KWA016 | Coast | White | 04.33753 °S | 039.44975 °E | 118 |
| 29 | KWA017 | Coast | White | 04.49746 °S | 039.25124 °E | 39 |
| 30 | KWA018 | Coast | White | 04.49765 °S | 039.25125 °E | 45 |
| 31 | KWA019 | Coast | White | 04.49763 °S | 039.25131 °E | 41 |
| 32 | KWA021 | Coast | White | 04.49715 °S | 039.25139 °E | 45 |
| 33 | KWA022 | Coast | Red | 04.60348 °S | 039.18504 °E | 25 |
| 34 | KWA023 | Coast | Red | 04.60352 °S | 039.18509 °E | 20 |
| 35 | KWA024 | Coast | White | 04.60323 °S | 039.18452 °E | 21 |
| 36 | MOM006 | Coast | White | 03.96482 °S | 039.73122 °E | 15 |
| 37 | MOM007 | Coast | Red | 03.96493 °S | 039.73089 °E | 14 |
| 38 | MOM008 | Coast | Red | 03.96229 °S | 039.73233 °E | 16 |
| 39 | MER001 | Eastern | Red | 00.17234 °S | 037.64283 °E | 1564 |
| 40 | MER002 | Eastern | Red | 00.17239 °S | 037.64275 °E | 1545 |
| 41 | MER003 | Eastern | Red | 00.16647 °S | 037.65030 °E | 1457 |
| 42 | MER004 | Eastern | Red | 00.16708 °S | 037.65543 °E | 1449 |
| 43 | MER005 | Eastern | Red | 00.17249 °S | 037.65120 °E | 1479 |
| 44 | MER006 | Eastern | Red | 00.17247 °S | 037.65128 °E | 1481 |
| 45 | MER007 | Eastern | Red | 00.17251 °S | 037.63130 °E | 1481 |
| 46 | MER008 | Eastern | Red | 00.19338 °S | 037.66548 °E | 1429 |
| 47 | MER009 | Eastern | Red | 00.08721 °S | 037.66675 °E | 1455 |
| 48 | MER010 | Eastern | Red | 00.08726 °S | 037.66695 °E | 1452 |
| 49 | MER011 | Eastern | Red | 00.08583 °S | 037.66500 °E | 1474 |
| 50 | MER012 | Eastern | Red | 00.08564 °S | 037.66451 °E | 1478 |
| 51 | MER013 | Eastern | Red | 00.08536 °S | 037.66438 °E | 1481 |
| 52 | MER014 | Eastern | White | 00.11461 °S | 037.69637 °E | 1384 |
| 53 | MER015 | Eastern | White | 00.11443 °S | 037.69638 °E | 1380 |
| 54 | MER016 | Eastern | Red | 00.18701 °S | 037.69572 °E | 1290 |

**Supplemental Information Table S1.** Continued.

| Sample number | Accession code | Region | Fruit flesh colour | Latitude (N°/S°) | Longitude (E°) | Altitude (m) |
|---|---|---|---|---|---|---|
| 55 | MER017 | Eastern | Red | 00.18693 °S | 037.69600 °E | 1288 |
| 56 | MER018 | Eastern | Red | 00.12048 °S | 037.72087 °E | 1393 |
| 57 | MER019 | Eastern | Red | 00.12024 °S | 037.72074 °E | 1385 |
| 58 | ELG001 | Rift Valley | White | 00.64776 °N | 035.51977 °E | 2089 |
| 59 | ELG002 | Rift Valley | Red | 00.64203 °N | 035.52221 °E | 2064 |
| 60 | ELG003 | Rift Valley | Red | 00.64265 °N | 035.52145 °E | 2077 |
| 61 | ELG004 | Rift Valley | Red | 00.64264 °N | 035.52150 °E | 2071 |
| 62 | ELG005 | Rift Valley | Red | 00.67029 °N | 035.51809 °E | 2214 |
| 63 | ELG006 | Rift Valley | Red | 00.67030 °N | 035.51812 °E | 2209 |
| 64 | ELG007 | Rift Valley | Red | 00.64350 °N | 035.51839 °E | 2104 |
| 65 | ELG008 | Rift Valley | Red | 00.64349 °N | 035.51843 °E | 2104 |
| 66 | ELG009 | Rift Valley | Red | 00.64338 °N | 035.51852 °E | 2102 |
| 67 | ELG010 | Rift Valley | Red | 00.64505 °N | 035.51627 °E | 2132 |
| 68 | ELG012 | Rift Valley | Red | 00.63469 °N | 035.52243 °E | 2031 |
| 69 | ELG013 | Rift Valley | Red | 00.63185 °N | 035.52095 °E | 2024 |
| 70 | ELG018 | Rift Valley | Red | 00.58769 °N | 035.46060 °E | 2325 |
| 71 | ELG022 | Rift Valley | Red | 00.63766 °N | 035.51977 °E | 2079 |
| 72 | ELG023 | Rift Valley | Red | 00.64214 °N | 035.52221 °E | 2056 |
| 73 | ELG041 | Rift Valley | Red | 00.63469 °N | 035.52043 °E | 2021 |
| 74 | ELG046 | Rift Valley | Red | 00.63187 °N | 035.52195 °E | 2024 |
| 75 | ELG047 | Rift Valley | Red | 00.66551 °N | 035.53129 °E | 1972 |
| 76 | ELG048 | Rift Valley | Red | 00.66582 °N | 035.53104 °E | 1985 |
| 77 | ELG049 | Rift Valley | Red | 00.56152 °N | 035.30367 °E | 2142 |
| 78 | ELG050 | Rift Valley | Red | 00.58151 °N | 035.30357 °E | 2150 |
| 79 | UAG014 | Rift Valley | Red | 00.57152 °N | 035.30377 °E | 2142 |
| 80 | UAG015 | Rift Valley | Red | 00.57151 °N | 035.30377 °E | 2150 |
| 81 | UAG016 | Rift Valley | Red | 00.58574 °N | 035.46054 °E | 2317 |
| 82 | UAG017 | Rift Valley | Red | 00.57162 °N | 035.30367 °E | 2142 |
| 83 | UAG019 | Rift Valley | Red | 00.58788 °N | 035.46055 °E | 2322 |
| 84 | UAG020 | Rift Valley | Red | 00.66651 °N | 035.53149 °E | 1972 |
| 85 | UAG021 | Rift Valley | Red | 00.66682 °N | 035.53004 °E | 1985 |
| 86 | UAG024 | Rift Valley | * | 00.64256 °N | 035.52145 °E | 2067 |
| 87 | UAG025 | Rift Valley | * | 00.64264 °N | 035.52150 °E | 2076 |
| 88 | UAG026 | Rift Valley | * | 00.67019 °N | 035.51809 °E | 2267 |
| 89 | UAG027 | Rift Valley | * | 00.67028 °N | 035.51812 °E | 2210 |
| 90 | UAG028 | Rift Valley | * | 00.64352 °N | 035.51839 °E | 2114 |
| 91 | UAG029 | Rift Valley | * | 00.64356 °N | 035.51843 °E | 2106 |
| 92 | UAG030 | Rift Valley | * | 00.64109 °N | 035.51783 °E | 2119 |
| 93 | UAG031 | Rift Valley | * | 00.64348 °N | 035.51852 °E | 2112 |
| 94 | UAG032 | Rift Valley | * | 00.64509 °N | 035.51627 °E | 2125 |
| 95 | UAG033 | Rift Valley | * | 00.64109 °N | 035.51783 °E | 2120 |
| 96 | UAG034 | Rift Valley | * | 00.63469 °N | 035.52243 °E | 2041 |
| 97 | UAG035 | Rift Valley | * | 00.63185 °N | 035.52095 °E | 2021 |
| 98 | UAG036 | Rift Valley | * | 00.66651 °N | 035.53149 °E | 1972 |
| 99 | UAG037 | Rift Valley | * | 00.66682 °N | 035.53004 °E | 1985 |
| 100 | UAG039 | Rift Valley | * | 00.57152 °N | 035.30377 °E | 2142 |
| 101 | UAG040 | Rift Valley | * | 00.64348 °N | 035.51852 °E | 2102 |
| 102 | UAG042 | Rift Valley | * | 00.64438 °N | 035.51752 °E | 2102 |
| 103 | UAG043 | Rift Valley | * | 00.64507 °N | 035.51632 °E | 2142 |
| 104 | UAG044 | Rift Valley | * | 00.64505 °N | 035.51627 °E | 2142 |
| 105 | UAG045 | Rift Valley | * | 00.64129 °N | 035.51783 °E | 2112 |
| 106 | HOM001 | Western | Red | 00.59582 °N | 034.57717 °E | 1308 |
| 107 | HOM002 | Western | White | 00.59580 °N | 034.57707 °E | 1302 |
| 108 | HOM003 | Western | Red | 00.59585 °N | 034.57596 °E | 1307 |
| 109 | HOM004 | Western | Red | 00.59594 °N | 034.57690 °E | 1306 |

**Supplemental Information Table S1.** Continued.

| Sample number | Accession code | Region | Fruit flesh colour | Latitude (N°/S°) | Longitude (E°) | Altitude (m) |
|---|---|---|---|---|---|---|
| 110 | HOM005 | Western | Red | 00.59596 °N | 034.57690 °E | 1306 |
| 111 | HOM006 | Western | Red | 00.59593 °N | 034.57688 °E | 1303 |
| 112 | HOM007 | Western | Red | 00.59593 °N | 034.57692 °E | 1307 |
| 113 | HOM008 | Western | Red | 00.59596 °N | 034.57689 °E | 1307 |
| 114 | HOM009 | Western | White | 00.59600 °N | 034.57698 °E | 1305 |
| 115 | HOM010 | Western | Red | 00.59596 °N | 034.57703 °E | 1307 |
| 116 | HOM011 | Western | Red | 00.59603 °N | 034.57717 °E | 1302 |
| 117 | HOM012 | Western | Red | 00.60963 °N | 034.58897 °E | 1329 |
| 118 | HOM013 | Western | Red | 00.60974 °N | 034.58366 °E | 1335 |
| 119 | HOM014 | Western | Red | 00.60961 °N | 034.58369 °E | 1339 |
| 120 | HOM015 | Western | White | 00.60961 °N | 034.58374 °E | 1337 |
| 121 | HOM016 | Western | Red | 00.60984 °N | 034.58377 °E | 1336 |
| 122 | HOM017 | Western | Red | 00.60610 °N | 034.63214 °E | 1463 |
| 123 | HOM018 | Western | White | 00.60611 °N | 034.63223 °E | 1456 |
| 124 | HOM019 | Western | Red | 00.61762 °N | 034.64497 °E | 1498 |
| 125 | HOM020 | Western | Red | 00.61760 °N | 034.64495 °E | 1502 |
| 126 | HOM021 | Western | White | 00.61766 °N | 034.64488 °E | 1800 |
| 127 | HOM022 | Western | Red | 00.53904 °N | 034.50943 °E | 1242 |
| 128 | HOM023 | Western | White | 00.53907 °N | 034.50946 °E | 1238 |
| 129 | HOM024 | Western | White | 00.53907 °N | 034.50945 °E | 1240 |
| 130 | HOM025 | Western | Red | 00.53907 °N | 034.50941 °E | 1237 |
| 131 | HOM026 | Western | Red | 00.53908 °N | 034.50942 °E | 1238 |
| 132 | HOM027 | Western | White | 00.53906 °N | 034.50946 °E | 1242 |
| 133 | HOM028 | Western | Red | 00.53905 °N | 034.50951 °E | 1239 |
| 134 | HOM029 | Western | Red | 00.53893 °N | 034.50956 °E | 1240 |
| 135 | HOM030 | Western | White | 00.53880 °N | 034.50989 °E | 1239 |
| 136 | HOM031 | Western | Red | 00.53887 °N | 034.51012 °E | 1238 |
| 137 | HOM032 | Western | Red | 00.53987 °N | 034.50855 °E | 1246 |
| 138 | HOM033 | Western | Red | 00.72493 °N | 034.45583 °E | 1289 |
| 139 | HOM034 | Western | Red | 00.72484 °N | 034.45608 °E | 1292 |
| 140 | HOM035 | Western | Red | 00.72481 °N | 034.45610 °E | 1289 |
| 141 | HOM036 | Western | Red | 00.72479 °N | 034.45597 °E | 1290 |
| 142 | HOM037 | Western | Red | 00.72493 °N | 034.45608 °E | 1293 |
| 143 | HOM038 | Western | Red | 00.72485 °N | 034.45566 °E | 1285 |
| 144 | HOM039 | Western | Red | 00.72471 °N | 034.45581 °E | 1292 |
| 145 | HOM040 | Western | Red | 00.72468 °N | 034.45585 °E | 1289 |
| 146 | HOM041 | Western | Red | 00.72472 °N | 034.45564 °E | 1287 |
| 147 | HOM042 | Western | Red | 00.72455 °N | 034.45533 °E | 1283 |
| 148 | HOM043 | Western | Red | 00.72442 °N | 034.45531 °E | 1283 |
| 149 | HOM044 | Western | Red | 00.72436 °N | 034.45530 °E | 1285 |
| 150 | HOM046 | Western | White | 00.72439 °N | 034.45518 °E | 1283 |
| 151 | HOM047 | Western | White | 00.72412 °N | 034.45534 °E | 1265 |
| 152 | HOM048 | Western | White | 00.72412 °N | 034.45539 °E | 1275 |
| 153 | KAK001 | Western | Red | 00.27951 °N | 034.67358 °E | 1419 |
| 154 | KAK002 | Western | Red | 00.27863 °N | 034.67363 °E | 1409 |
| 155 | KAK003 | Western | Red | 00.27861 °N | 034.67367 °E | 1420 |
| 156 | KAK004 | Western | Red | 00.27791 °N | 034.69564 °E | 1447 |
| 157 | KAK005 | Western | Red | 00.27700 °N | 034.69589 °E | 1441 |
| 158 | KAK006 | Western | Red | 00.27777 °N | 034.69579 °E | 1443 |
| 159 | KAK007 | Western | Red | 00.24446 °N | 034.82470 °E | 1571 |
| 160 | KAK008 | Western | Red | 00.24442 °N | 034.82479 °E | 1572 |
| 161 | SIA001 | Western | Red | 00.19481 °N | 034.34081 °E | 1297 |
| 162 | SIA002 | Western | Red | 00.19376 °N | 034.33390 °E | 1286 |
| 163 | SIA003 | Western | Red | 00.19423 °N | 034.33385 °E | 1280 |
| 164 | SIA004 | Western | Red | 00.13007 °N | 034.42597 °E | 1358 |

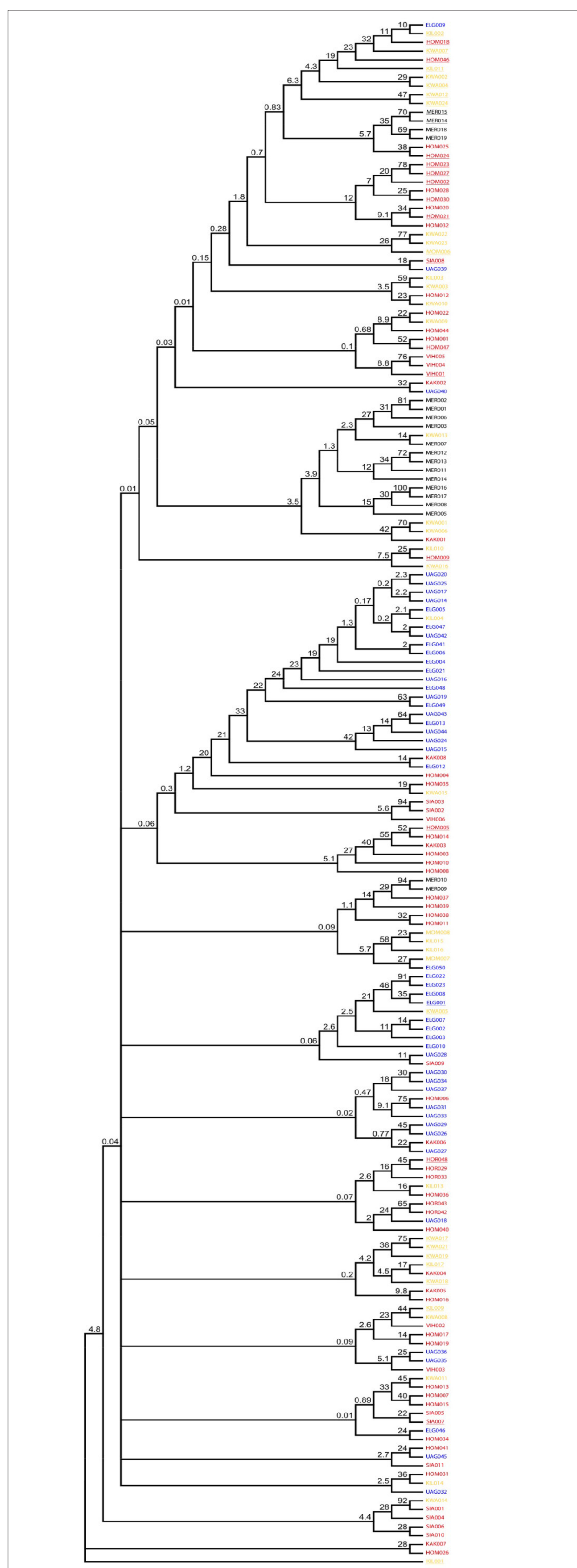
**Supplemental Information Table S1.** Continued.

| Sample number | Accession code | Region | Fruit flesh colour | Latitude (N°/S°) | Longitude (E°) | Altitude (m) |
|---|---|---|---|---|---|---|
| 165 | SIA005 | Western | Red | 00.13003 °N | 034.42687 °E | 1357 |
| 166 | SIA006 | Western | Red | 00.12687 °N | 034.42089 °E | 1340 |
| 167 | SIA007 | Western | White | 00.12680 °N | 034.42102 °E | 1342 |
| 168 | SIA008 | Western | White | 00.12804 °N | 034.42337 °E | 1347 |
| 169 | SIA009 | Western | Red | 00.12810 °N | 034.42309 °E | 1347 |
| 170 | SIA010 | Western | Red | 00.13046 °N | 034.42354 °E | 1348 |
| 171 | SIA011 | Western | Red | 00.13008 °N | 034.42255 °E | 1349 |
| 172 | VIH001 | Western | White | 00.08540 °N | 034.79936 °E | 1680 |
| 173 | VIH002 | Western | Red | 00.08539 °N | 034.79936 °E | 1679 |
| 174 | VIH003 | Western | Red | 00.08532 °N | 034.79938 °E | 1682 |
| 175 | VIH004 | Western | Red | 00.84470 °N | 034.79931 °E | 1683 |
| 176 | VIH005 | Western | Red | 00.84360 °N | 034.79930 °E | 1684 |
| 177 | VIH006 | Western | Red | 00.08413 °N | 034.79875 °E | 1688 |

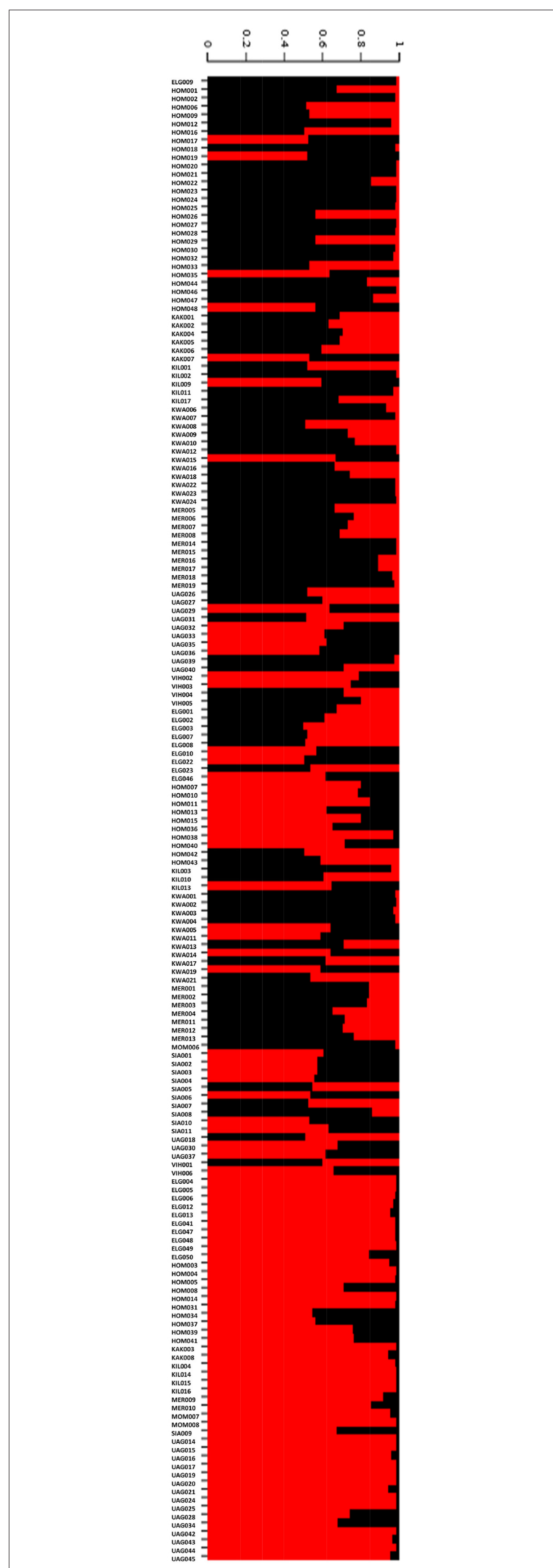
* There were no fruits on the trees at the time of sampling, hence fruit flesh colour was not determined.

**Supplemental Information Table S2.** PCR primer sequences and pools used for the PCR multiplexing in guava (*Psidium guajava* L.) DNA fragment analysis and size ranges of alleles amplified. Related to Table 1.

| Primer Multiplex pool | Name | Forward | Reverse | Fluorescent dye | Allele size range, bp |
|---|---|---|---|---|---|
| 1 | mPgCIR11 | TGAAAGACAACAAACGAG | TTACACCCACCTAAATAAGA | HEX | 301-316 |
| | mPgCIR15 | TCTAATCCCCTGAGTTTC | CCGATCATCTCTTTCTTT | HEX | 146-166 |
| | mPgCIR17 | CCTTTCGTCATATTCACTT | CATTGGATGGTTGACAT | HEX | 225-243 |
| | mPgCIR19 | AAAATCCTGAAGACGAAC | TATCAGAGGCTTGCATTA | HEX | 255-280 |
| 2 | mPgCIR07 | ATGGAGGTAGGTTGATG | CGTAGTAATCGAAGAAATG | HEX | 143-158 |
| | mPgCIR09 | GCGTGTCGTATTGTTTC | ATTTTCTTCTGCCTTGTC | FAM | 155-175 |
| | mPgCIR10 | GTTGGCTCTTATTTTGGT | GCCCCATATCTAGGAAG | FAM | 260-326 |
| | mPgCIR13 | CCTTTTTCCCGACCATTACA | TCGCACTGAGATTTTGTGCT | FAM | 246-258 |
| 3 | mPgCIR08 | ACTTTCGGTCTCAACAAG | AGGCTTCCTACAAAAGTG | HEX | 214-224 |
| | mPgCIR20 | TATACCACACGCTGAAAC | TTCCCCATAAACATCTCT | FAM | 265-296 |
| | mPgCIR21 | TGCCCTTCTAAGTATAACAG | AGCTACAAACCTTCCTAAA | HEX | 147-162 |
| | mPgCIR22 | CATAAGGACATTTGAGGAA | AATAAGAAAGCGAGCAGA | HEX | 237-253 |
| | mPgCIR25 | GACAATCCAATCTCACTTT | TGTGTCAAGCATACCTTC | FAM | 99-131 |



**Supplemental Information Figure S1.** The neighbour-joining phylogenetic tree of 177 guava accessions collected from four regions of Kenya. The accession colour codes depict the region of collection (Red = Western, Blue = Rift Valley, Gold = Coast, Black = Eastern). The white-fleshed accessions are underlined. The bootstrap values are indicated at the tree nodes after 10,000 bootstrappings. Related to Figure 2.




**Supplemental Information Figure S2.** Bayesian analysis cluster plot of 177 guava accessions from four regions of Kenya. The most probable number of clusters (K = 2) is represented by colours, which however, depict the accessions as having admixed genotypes. Related to Figure 3.