Genetic Diversity of Puerto Rican Farmer-held Papaya (Carica papaya) Using SSR Markers

Dianiris Luciano-Rosario, Luis A. Cruz-Saavedra, and Dimuth Siritunga

Department of Biology, University of Puerto Rico Mayaguez Campus, Mayaguez, PR 00680

Abstract. Native to Central America, papaya (Carica papaya) is one of the most cultivated fruit crops in the tropical areas of the world. Genetic diversity analyses are an important aspect of conservation of plant genetic resources. In the island of Puerto Rico, where papaya has been consumed for centuries, knowledge on the genetic diversity of papaya is lacking. Therefore, 162 papaya accessions were evaluated using 23 simple sequence repeat (SSR) markers. Of these accessions, 139 were farmer-held samples from Puerto Rico, 13 were U.S. Department of Agriculture (USDA) repository samples, and 10 were commercial varieties. A total of 214 alleles were identified with a mean observed heterozygosity (H_o) of 0.219. Inbreeding coefficient (F) was 0.565, and when evaluating the population structure of these accessions, 2 groups (k = 2) were identified. Unweighted pair group method with arithmetic mean (UPGMA) dendrogram showed no geographical organization within the unknown Puerto Rican samples. This assessment provides an extensive record of the genetic diversity of papaya in Puerto Rico which can contribute to breeding strategies and to the conservation of papaya genetic resources in the Caribbean.

Conservation of plant genetic resources is important in addressing modern agricultural challenges. Population growth, food security, and climate change have resulted in the need to preserve the existing natural variation (Food and Agriculture Organization of the United Nations, 2010). Genetic diversity studies are an important aspect of conservation as they provide a record of the current variation which is the backbone of any breeding project. The Caribbean Islands are considered a zone of secondary diversification of crops leading to adaptation (Ocampo Perez et al., 2006a); therefore, its genetic diversity has always been an area of interest (Boza et al., 2013; Montero-Rojas et al., 2011, 2013; Muller et al., 2009; Rodriguez-Bonilla et al., 2014; Wendel et al., 1992).

Papaya is a tropical fruit crop belonging to the Caricaceae family. It is thought to be native to Central America and is cultivated in most of the world’s tropical areas (Arumuganathan and Earle, 1991; Food and Agriculture Organization of the United Nations, 2013; Ocampo, 2007; Teixeira da Silva et al., 2007). The commercial success of papaya is not only reliant on its beneficial nutritional properties but also on the commercial uses of papain, a proteolytic enzyme found in its latex (de Oliveira and Vitória, 2011). Papaya global production in 2013 was 12,420,485 t, of which 6.02% were produced in the Caribbean (Food and Agriculture Organization of the United Nations, 2013). With a production of 8,852 t in 2013, Puerto Rico is the third largest papaya producer in the Caribbean (Food and Agriculture Organization of the United Nations, 2013).

The genetic diversity of papaya has been studied both morphologically and at a molecular level (Aikopkpodion, 2012; Alonso et al., 2009; Asudi et al., 2013; Brown et al., 2012; Matos et al., 2013; Ocampo Perez et al., 2006a; Sengupta et al., 2013; Sudha et al., 2013). Morphologically, papaya has been shown to possess great diversity (Ocampo Perez et al., 2006a). Nevertheless, at a molecular level, it has been shown that commercial papaya offers a narrow genetic basis (Matos et al., 2013).

In the recent past, SSR markers have become an effective method for assessment of genetic diversity because of their high reproducibility, codominant inheritance, and genome-wide distribution (Idrees and Irshad, 2014; Wang et al., 2009). In papaya, several SSR marker libraries have been developed and its uses have led to success in assessing genetic diversity and molecular-assisted selection for breeding purposes (de Oliveira et al., 2010; Ocampo et al., 2004; Ocampo Perez et al., 2006a; Vidal et al., 2014). But, studies on papaya in the Caribbean and surrounding countries using SSR markers are limited and none have included papaya from Puerto Rico. Ocampo Perez (2007) studied 72 accessions from 13 locations in the Caribbean using 15 SSR markers and recorded a total of 99 alleles with the samples clustering according to their geographic region. Similarly, Brown et al. (2012) found heterozygosity deficiencies in natural populations of papaya in Costa Rica after studying 164 accessions from Costa Rica and 20 known cultivars from the USDA germplasm collection using 20 SSR markers. In our study, using 23 SSR markers, we assessed the genetic diversity within 139 farmer-held papaya accessions from different municipalities in Puerto Rico. For comparison purposes, 13 other accessions from USDA germplasm and 10 commercial varieties were also evaluated.

Materials and Methods

Plant material. A total of 162 samples were evaluated (Supplemental Table 1). Of these samples, 139 were unknown accessions from Puerto Rico, which were acquired voluntarily from Puerto Rico inhabitants’ personal gardens with a community of science approach. Thus, because of possible errors by the collectors, whom we have no control over, we did not collect any further morphological data including sexual reproductive system for each sample. We requested GPS coordinates but very few provided them. The rest of the samples in this study comprised 13 samples from the USDA germplasm repository in Hawaii and 10 samples acquired commercially. Leaf material was collected from the local farmers and frozen until further analysis. The USDA germplasm accessions and commercial varieties were planted from seeds in a greenhouse and leaves were collected 15 weeks after planting.

DNA extraction. DNA was extracted from papaya leaves using a modified protocol based on Doyle and Doyle (1991). About 0.5 g of leaf was ground with sterile sand using a pestle in a 2.0-mL tube and 800 μL of 3% CTAB buffer (20 mM EDTA, 0.1 M Tris- HCl pH 8.0, 1.4 M NaCl, 3% CTAB, 3% PVP, and 0.2% β-mercaptoethanol) was added and mixed by inversion. After incubating at 70 °C for 30 min, 500 μL chloroform:isoamyl alcohol (25:1) was added and gently mixed by inversion. The samples were centrifuged for 3 min at 12,200 rpm and a total of 500 μL of the supernatant were transferred to a new 2.0-mL tube. An equal amount of chloroform:isoamyl alcohol (25:1) and 200 μL of 3% CTAB buffer was then added. After mixing by inversion, samples were centrifuged for 3 min at 12,200 rpm. The supernatant was transferred to a 1.5-mL tube and 350 μL of cold isopropanol (−20 °C) was added. After gently mixing by inversion, samples...
The PCR cycle consisted of an initial denaturing step at 94°C for 4 min, followed by 30 s at 94°C for denaturing, and 1 min at 55°C for annealing and elongation. The PCR product was diluted 1:5 in LI-COR blue stop solution, denatured at 94°C for 5 min, and 150 mM 1X Promega colorless Gotaq Flexi buffer, rTaq DNA Polymerase (Bulldog Bio, Portsmouth, NH), and 40 ng of template DNA. 

**Table 1.** The list of 23 SSR markers used to assess the genetic diversity of papaya in Puerto Rico. Shown for each marker is their primer sequences, motif, location in the genome, allele sizes, observed heterozygosity ($H_o$), and expected heterozygosity ($H_t$) per locus.

| Locus    | Primer forward | Primer reverse | Motif         | Supercontig | Allele Size | $H_o$     | $H_t$     |
|----------|----------------|----------------|---------------|-------------|-------------|-----------|-----------|
| AJ81049  | GTCTATCACCTACCTACCA | GAGGTGTATCATATGCTACA | (TC)24        | 18          | 259–295     | 0.2308642 | 0.818682365 |
| AJ810490 | GAATCTCACCTACGGAATCT | ACTCTACCACGGGCA | (TC)14        | 19          | 202–236     | 0.39506173 | 0.687966773  |
| AJ810491 | AAGGCAAGAAGACAAACCA | ATGCTGGAAGTAAAGCA | (TC)10        | 11          | 239–253     | 0.222222222 | 0.7225800802 |
| AJ810492 | GCATTACTATCATGCTTCC | ACTATCTCTGCTGCTTCC | (CT)18        | 11          | 588–603     | 0.222222222 | 0.67162018  |
| AJ810493 | CCAAAAGGCAAAGAAACCA | ATCAAAGCCCTTCCTCAC | (TG)10(A)7(GA)10 | 1552 | 288–303 | 0.00617284 | 0.70064744 |
| AJ810494 | CCAACACATCATCCACCA | CGAAGCATACACGAGA | (TC)18        | 232         | 239–253     | 0.481841481 | 0.789037494 |
| AJ810495 | ATGACTGAAAGAAACACTTC | CTGAAATGCGCAATGCAAT | (CT)20...(AC)5 | 75 | 306–322 | 0.117283951 | 0.513850785 |
| AJ810505 | ATGCGTTATTTAAGGTTGGTATGC | TCAATGAGCCATAAAGCA | (CT)9...(CT)9 | 8 | 312–318 | 0.17912346 | 0.506649139 |
| CP27     | ATGCACGCGAAGATGGTACG | TCAAAAACCACTTCTCATGCTC | (GT)21 | 16 | 152–207 | 0.154320988 | 0.591210799 |
| CP31     | AAGGGAGCTTGCATGGAGACA | TCTGCGCGCTTTTATATCCTGCT | (AT)6(GT)10   | 27 | 167–178 | 0.265432099 | 0.87536988 |
| CP44     | TGCAACAGGAGCTTACCATCCTTA | CCTAGGTTTCTGGACTCTCTTAT | (AT)12       | 39          | 244–267     | 0.333333333 | 0.672210791 |
| CP49     | CCTGAAAGAACCAACCATTCTTA | TCGTGGAGGACCTGTAAGAAGA | (AT)12       | 78          | 210–222     | 0.364197531 | 0.665104405 |
| CPCIR2   | GGTCTTGGGATTGCTCAGGTTT | ATGATGGGACGGGTTT | (GA)12       | 82          | 252–294     | 0.265432099 | 0.861415943 |
| CPCIR3   | CGCATGGTATGACTCTAACT | ACCATGAGGCTGCT | (AT)10       | 103         | 203–234     | 0.172839506 | 0.709343088 |
| SP1      | GAGGACGAGGAGAGGGGCTG | GACGTGGGAGCGCTGGTG | (TTT)5(TTC)9 | 10 | 273–473 | 0.00617284 | 0.710099077 |
| SP2      | CACCAAGGGTGTTTGGGACTGGA | TGACATGCATGGTGTTG | (AC)9       | 4           | 648–700     | 0.080246914 | 0.623663662 |
| SP4      | TGCTCATATAAGGTAGATTGGTGGTGGT | ATGCAATTACATTTAAAACAAC | (AT)9       | 66          | 90–200      | 0.061728395 | 0.745141747 |
| SP5      | TGGGCTTACACATTGGGTGGTGGT | GCCGGTCTCTGGATCTGTAAT | (AC)9       | 6           | 242–267     | 0.037037037 | 0.868347356 |
| SP6      | CTTGACCGACACCACTAAAG | CATGAAACACACATGGTGCAAT | (AT)9       | 88          | 675–690     | 0.228395062 | 0.54913504 |
| SP7      | CAGATGAGGGATGGGATGGTGGT | ATCACAATACAGACCCCAT | (AA)T7       | 52          | 310–315     | 0.222222222 | 0.476870904 |
| SP8      | CAAATATGTGGATGGTGGTGTG | GCTCAGGGGCTTCTCTGAC | (AT)7       | 36          | 355–422     | 0.50617284 | 0.652853986 |
| SSP3     | CCAAGGAAACAAGCTACTCGGC | TCTCAGTTTTCAAGTTTGC | (AG)10      | 37          | 588–604     | 0.080246914 | 0.302716811 |
| SSP8     | TGCTCAGATATACCCCAA | ATGGCCTTTGGACACATGAC | (AT)12       | 37          | 588–604     | 0.080246914 | 0.302716811 |
| Mean     |                |                |              |             |            | 0.204777241 | 0.660707541 |
clusters ($k$) was evaluated from 1 to 10 with five iterations and the most probable $k$ was identified using Structure Harvester Web v0.6.94 (Earl and von Holdt, 2012) by the Evanno et al. (2005) method of delta $k$ ($\Delta k$). After identifying the most probable $k$, CLUMPP version 1.1.2 (Jakobsson and Rosenberg, 2007) was implemented and a bar plot was constructed with Distruct version 1.1 (Rosenberg, 2004).

**Results**

Genetic diversity. A total of 214 alleles were observed across the 23 SSR markers. Allele per locus ranged between 2 and 18 with a mean of 9.304 (Supplemental Table 1). All evaluated loci were polymorphic with its PIC ranging from 0.292 (locus AJ810493) to 0.863 (CPCIR2) with a mean of 0.626 (Supplemental Table 2). A low $H_e$ for all individually evaluated loci was obtained with a mean of 0.2047. Values ($H_e$) per locus ranged from 0.00617 (AJ810493 and SP1) to 0.506 (SP8) (Table 1). The mean $H_e$ for all the samples is 0.219, mean expected heterozygosity ($H_o$) is 0.559, and the mean F-index is 0.565 (Table 2). When evaluating Puerto Rican unknown samples by different geographical regions, the $H_e$ ranged from 0.196 in the Northwest area to 0.284 in the Northeast area, contrasting with a higher $H_o$ which ranged from 0.492 in the Central area to 0.536 in the Southwest area. The $H_e$ for the USDA germplasm samples was 0.138 (Table 2). The distribution of the private alleles considering the 23 assessed loci per group is shown in Table 3. The USDA germplasm group and known commercial varieties group showed to have the greatest amount of private alleles with 22 and 20, respectively, whereas a total of 26 private alleles were found among the Puerto Rico’s unknown samples. Linkage disequilibrium analysis for the assessed SSR marker pairs showed that of the 253 analyzed SSR marker combinations, 146 locus pairs are statistically linked, showing a $P$ value <0.05 (Supplemental Table 3).

Population structure. A genetic distance method–based dendrogram was constructed using Euclidean distance and UPGMA (Fig. 1). Two main clusters were identified and no geographical grouping among the samples was identified. Samples from the USDA germplasm and commercial accessions grouped within the same cluster but not exclusively because samples ‘Mona’, ‘130’ (SE), and ‘67’ (SW) grouped with them. Cluster 1 is composed of 7 samples from the Puerto Rico’s unknown group (from the Northeast (2), Northwest (2), Southeast (1), and Southwest areas (2)), whereas cluster 2 included the rest of the unknown Puerto Rico samples, all of the commercial varieties and USDA germplasm samples.

A total of 2 clusters ($k = 2$) were also identified after using the method of delta $k$ with Structure Harvester software (Supplemental Fig. 2). All the Puerto Rico’s unknown samples but one, the ‘Mona’ sample, were found to belong to cluster 1 (Fig. 2). Cluster 2 comprised the USDA germplasm samples and the known commercial samples. Limited admixture was observed among the two identified clusters with few samples showing the following gene frequencies: sample 37s showed 0.409 inferred ancestry level of cluster 2 and 0.150 of cluster 1, and sample 79s showed 0.409 inferred ancestry level of cluster 2 and 0.592 ancestry level of cluster 1.

Discussion

Genetic diversity estimators revealed similar results to other genetic diversity studies using SSR markers (Asudi et al., 2013; Brown et al., 2012; Matos et al., 2013). When evaluating $H_e$ for the USDA germplasm samples, we found similar results as Brown et al. (2012) which evaluated a total of 20 samples from the USDA germplasm repository and reported an $H_e$ of 0.14 (similar to 0.138 in our analysis). We also found low levels of heterozygosity in Puerto Rico’s unknown samples with an $H_e$ ranging from 0.196 to 0.284 which was comparable with Costa Rica’s natural populations that ranged from 0.31 to 0.45 (Brown et al., 2012). Likewise, Matos et al. (2013) also reported low levels of $H_e$ when evaluating a Brazilian germplasm that resulted in a mean of 0.20.

When evaluating $H_o$, we found that the values are higher than the $H_e$. This trend is observed in other studies. For example, Ocampo Perez (2007) reported $H_o$ values that ranged from 0.37 to 0.69 for different populations in the Caribbean region and Brown et al. (2012) reported a $H_o$ of 0.64 for the assessed known cultivars, similar to our results with a $H_o$ of 0.58 for the USDA germplasm samples. Regarding the total allele number, our study showed a higher number of alleles (214) when compared with other studies (Asudi et al., 2013; Ocampo Perez, 2007; Matos et al., 2013) but similar to Brown et al. (2012). Nevertheless, allele per locus in our study which ranged from 2 to 18 with a mean of 9.304 is comparable with other studies such as Asudi et al. (2013) that reported 8 to 18 alleles per locus with a mean of 11.93 and Brown et al. (2012) that reported 6–25 allele per locus with a mean of 11.6. A possible explanation for high allelic abundance may be the history of papaya in Puerto Rico. Although not well documented, we infer that historically multiple introductions have been made because of the islands’ geographic location and political history. It is thought that papaya was first introduced to Puerto Rico around 1525 because of the proximity and likewise history of the Dominican

Table 2. Genetic diversity estimators for the 162 assessed samples and groups from Puerto Rico. Number of alleles (Na), observed heterozygosity ($H_o$), expected heterozygosity ($H_e$), and inbreeding coefficient ($F$).

| Samples                     | Na     | $H_o$  | $H_e$  | $F$   |
|-----------------------------|--------|--------|--------|-------|
| Puerto Rico unknown         | Mean   | 7.174  | 0.231  | 0.559 | 0.576 |
|                             | SE     | 0.628  | 0.031  | 0.031 | 0.050 |
| Puerto Rico unknown (Groups)|        |        |        |       |
| Puerto Rico unknown (NW)    | 4.174  | 0.196  | 0.494  | 0.579 |
|                             | SE     | 0.331  | 0.031  | 0.031 | 0.063 |
| Puerto Rico unknown (NE)    | 4.043  | 0.284  | 0.532  | 0.431 |
|                             | SE     | 0.336  | 0.034  | 0.033 | 0.074 |
| Puerto Rico unknown (SE)    | 4.348  | 0.257  | 0.516  | 0.502 |
|                             | SE     | 0.353  | 0.038  | 0.031 | 0.064 |
| Puerto Rico unknown (Center)| 3.739  | 0.238  | 0.492  | 0.527 |
|                             | SE     | 0.334  | 0.033  | 0.029 | 0.065 |
| Puerto Rico unknown (SW)    | 3.826  | 0.209  | 0.536  | 0.614 |
|                             | SE     | 0.272  | 0.038  | 0.036 | 0.060 |
| USDA germplasm              | Mean   | 4.087  | 0.138  | 0.582 | 0.787 |
|                             | SE     | 0.371  | 0.045  | 0.033 | 0.064 |
| Commercial varieties        | Mean   | 4.043  | 0.288  | 0.549 | 0.510 |
|                             | SE     | 0.347  | 0.063  | 0.042 | 0.105 |
| Mean over loci and groups   | Mean   | 5.101  | 0.219  | 0.563 | 0.565 |
|                             | SE     | 0.319  | 0.028  | 0.021 | 0.046 |

Table 3. Genetic diversity estimators for the 162 assessed samples and groups from Puerto Rico. Number of private alleles per group.

| Population                  | Number of private alleles | Percentage (%) of private alleles per population | Percentage (%) of private alleles (total) |
|-----------------------------|---------------------------|-------------------------------------------------|-----------------------------------------|
| Puerto Rico unknown NW      | 5                         | 3.13                                            | 1.40                                    |
| Puerto Rico unknown NE      | 7                         | 7.29                                            | 3.27                                    |
| Puerto Rico unknown SE      | 4                         | 4.00                                            | 1.87                                    |
| Puerto Rico unknown Center  | 8                         | 9.30                                            | 3.74                                    |
| Puerto Rico unknown SW      | 4                         | 4.55                                            | 1.87                                    |
| USDA                        | 22                        | 21.36                                           | 10.28                                   |
| Commercial                  | 20                        | 26.32                                           | 9.35                                    |

USDA = U.S. Department of Agriculture.
Republic, where the introduction of papaya has been documented to have occurred around 1525 (Teixeira da Silva et al., 2007). Because Puerto Rico does not produce enough papaya to meet the local demand, papaya fruits are regularly imported from the Dominican Republic, Costa Rica, and the United States (Junta de Planificación, 2016; Morton, 1987; Zambrana-Echevarría et al., 2016). This undoubtedly contributes to more allelic diversity across the island as papaya grows from seeds and is easily cultivated for personal consumption by residents of Puerto Rico. For example, during 2015, Puerto Rico imported 512,861 kg of papaya from Costa Rica and interestingly exported 50,072 kg to the United States (Junta de Planificación, 2016). Another possible introduction event of papaya to Puerto Rico was during 1978, when a new economic development strategy was implemented by establishing agriculture as one of the pillars for an export-based economy. This led to different approaches with one of them being converting the southern coastal area of Puerto Rico as an intensive fruit and vegetable farming area for local consumption and winter exportation (Carro-Figueroa, 2002). After linkage disequilibrium analysis, we found that 57% of the SSR used in this study are statistically linked, although physically only 2 SSR marker pairs are in the same supercontig. We propose two possible explanations: 1) the assessed alleles may be associated as a result of domestication (Matos et al., 2013) or 2) the SSR markers are physically linked but poor genomic annotation does not provide enough resolution to confirm this. Similar studies in the future should take this into account and increase the number of SSR markers that will be used to reduce redundancy.

Based on de Evanno method for $k$ size selection, STRUCTURE analysis suggests a total of two distinct groups. One of the groups contains 138 unknown Puerto Rico accessions.
samples with the exception of a sample from Mona Island, an uninhabited island belonging to the archipelago of Puerto Rico. A possible explanation for this sample being more genetically similar to the known commercial and USDA germplasm is that it may actually be another known cultivar that has remained isolated and self-fertilized, therefore, more similar to known samples. Mass human migration from the Dominican Republic to Puerto Rico is documented since 1961, when political events such as the fall of Trujillo regime and consequent events led to Puerto Rico being a preferred destination due to its proximity, similar history, geography, culture, and language (Duany, 2005). This could possibly explain the arrival of ‘Mona’ sample to this uninhabited island because it is known that migrants attempt to cross the Mona passage to access Puerto Rico (United States Coast Guard, 2016).

The fact that the UPGMA dendrogram does not perfectly match the STRUCTURE analysis may be due to the difference in clustering methods; UPGMA method is distance based, whereas the STRUCTURE method uses Bayesian inference (Evanno et al., 2005). Nevertheless, these analyses are similar. The UPGMA dendrogram has several clusters, one of which contains the known commercial samples and the USDA germplasm samples but also containing samples from the SE (130) and SW (67). We believe human transportation of papaya seeds as the reason for the lack of geographical clustering within the unknown samples of Puerto Rico. It is known that plant distribution and diversity are influenced by human behavior due to their mobility (Antrop, 2004; Niggemann et al., 2009). We had duplicates of the samples ‘Known You’ and ‘Red Lady’ acquired differently. One sample each of these varieties was acquired with the USDA repository samples, whereas the other was acquired commercially. Interestingly, the duplicate samples were not a match for either ‘Known You’ or ‘Red Lady’ varieties.

In general, we suggest that Puerto Rico is an allele reservoir for papaya that should be further studied for possible breeding applications, given the allelic abundance in the island. We believe that the abundance is due to historical reasons, specifically due to the geographic location of Puerto Rico which is central and accessible to the entire American continent (Zambrana-Echevarria et al., 2016). We suggest future studies evaluating the genetic diversity at a morphological level taking in to consideration the allele abundance in our assessed samples. This study provides the first exhaustive record of the genetic diversity of papaya in Puerto Rico that can be used in conservation or future breeding programs.

**Literature Cited**

Akolkpodion, P.O. 2012. Assessment of genetic diversity in horticultural and morphological traits among papaya (Carica papaya) accessions in Nigeria. Fruits 67:173–187.

Alonso, M., B. Alor, O. García, Q. Moreno, S. Teyor, and L. Felipe. 2009. Caracterización de accesiones de papaya (Carica papaya L.) a través de marcadores AFLP en Cuba. Caracterising Cuban papaya accessions (Carica papaya L.) by AFLP markers. Rev. Colomb. Biotecnologa 2:31–39.

Antrop, M. 2004. Landscape change and the urbanization process in Europe. Landsc. Urban Plan. 67:9–26.

Arumuganathan, K. and E. Earle. 1991. Nuclear DNA content of some important plant species. Plant Mol. Biol. Rpt. 9:208–218.

Asudi, G., F.K. Ombwara, F.K. Rimberia, A.B. Nyende, E.M. Ateka, and L.S. Wamocho. 2013. Evaluating diversity among Kenyan papaya germplasm using simple sequence repeat markers. Afr. J. Food Agr. Nutr. Dev. 13:7307–7324.

Boza, E.J., B.M. Irish, A.W. Meereor, C.L. Tondo, O.A. Rodriguez, M. Ventura-López, J.A. Gómez, J.M. Moore, D. Zhang, J.C. Matamayor, and R.J. Schnell. 2013. Genetic diversity, conservation, and utilization of Theobroma cacao L.: Genetic resources in the Dominican Republic. Genet. Resources Crop Evol. 60:605–620.

Brown, J.E., J.M. Bauman, J.F. Lawrie, O.J. Rocha, and R.C. Moore. 2012. The structure of morphological and genetic diversity in natural populations of Carica papaya (Caricaceae) in Costa Rica. Biotropica 44:179–188.

Carro-Figueroa, V. 2002. Agricultural decline and food import dependency in Puerto Rico: A historical perspective on the outcomes of postwar farm and food policies. Caribb. Stud. 30:77–107.

de Oliveira, E.J., V.B.O. Amorim, E.L.S. Matos, J.L. Costa, M. daSilva Castellen, J.G. Pádua, and J.L.L. Dantas. 2010. Polymorphism of microsatellite markers in papaya (Carica papaya L.). Plant Mol. Biol. Rpt. 28:519–530.

de Oliveira, J.G. and A.P. Vitória. 2011. Papaya: Nutritional and pharmacological characterization, and quality loss due to physiological disorders. Food Res. Intl. 44:1306–1313.

Doyle, J. 1991. DNA protocols for plants. In: G.M. Hewitt, A.W.B. Johnston, and J.P.W. Young (eds.). Molecular techniques in taxonomy. NATO ASI Series (Series II: Cell Biology). Vol. 57. Springer, Berlin, Heidelberg.

Duany, J. 2005. Dominican migration to Puerto Rico: A transnational perspective. Centro J. 17:243–268.

Earl, D.A. and B. VonHoldt. 2012. STRUCTURE HARVESTER. A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resources 4:359–361.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611–2620.

Food and Agriculture Organization of the United Nations. 2010. The Second Report on The State of the World’s Plant Genetic Resources for Food and Agriculture. pp. 1–16.
Food and Agriculture Organization of the United Nations. 2013. FAOSTAT database. 12 Dec. 2017. <http://faostat.fao.org/>.

Idrees, M. and M. Irshad. 2014. Molecular markers in plants for analysis of genetic diversity: A review. Eur. Acad. Res. 2:1513–1540.

Jakobsson, M. and N.A. Rosenberg. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806.

Junta de Planificació/C19. 2016. Puerto Rico external trade data. 12 Dec. 2017. <http://www.jspobierno.pr/>.

Letunic, I. and P. Bork. 2007. Interactive Tree of Life (iTOL): An online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128.

Liu, J., S. Muse, and W. Bruce. 2005. PowerMarker V. 3.25: Integrated analysis environment for genetic marker data. Bioinformatics 21:2128–2129.

Matos, E.L.S., E.J. Oliveira, O.N. Jesus, and J.L.L. Dantas. 2013. Microsatellite markers of genetic diversity and population structure of Carica papaya. Ann. Appl. Biol. 163:298–310.

Montero-Rojas, M., A.M. Correa, and D. Siritunga. 2011. Molecular differentiation and diversity of cassava (Manihot esculenta) assessed from 162 locations across Puerto Rico and assessed with microsatellite markers. Annals of Botany Plants https://doi.org/10.1093/aobpla/plt010.

Montero-Rojas, M., M. Ortiz, J. Beaver, and D. Siritunga. 2013. Genetic, morphological and cytoagen content evaluation of a new collection of Caribbean Lima bean (Phaseolus lunatus L.) landraces. Genet. Resources Crop Evol. 60:2241–2252.

Morton, J. 1987. Papaya, p. 336–346. In: J. Morton (ed.). Fruits of warm climates. Echo Point Books & Media, Brattleboro, VT.

Muller, F., M. Voccia, A. Bé, and J.M. Bouvet. 2009. Genetic diversity and gene flow in a Caribbean tree Pterocarpus officinalis Jacq.: A study based on chloroplast and nuclear microsatellites. Genetica 135:185–198.

Niggemann, M., J. Jetzkowitz, S. Brunzel, M.C. Wichmann, and R. Bialozyt. 2009. Distribution patterns of plants explained by human movement behavior. Ecol. Modell. 220:1339–1346.

Ocampo, J.A., D. Dambier, P. Ollitrault, G. Coppons d’Eeckenbrugge, P. Brottier, and A. Risterucci. 2004. Development of microsatellite markers in papaya: Isolation, characterization and cross amplification in mountain papayas. Proc. Interam. Soc. For Trop. Hort. 48:90–93.

Ocampo, Perez, J. 2007. Papaya genetic diversity assessed with microsatellite markers in germplasm from the Caribbean region. Acta Hort. 740:93–102.

Ocampo Perez, J., D. Dambier, P. Ollitrault, G. Coppons d’Eeckenbrugge, P. Brottier, Y. Froelicher, and A. Risterucci. 2006a. Microsatellite markers in Carica papaya L.: Isolation, characterization and transferability to vasconcellea species. Mol. Ecol. Notes 6:212–217.

Ocampo Perez, J., G.C. D’Eeckenbrugge, S. Bruyere, L. Bellaire, and P. Ollitrault. 2006b. Organization of morphological and genetic diversity of Caribbean and Venezuelan papaya germplasm. Fruits 61:25–37.

Peakall, R. and P.E. Smouse. 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539.

Rodriguez-Bonilla, L., H.E. Cuevas, M. Montero-Rojas, F. Bird-Pico, D. Luciano-Rosario, and D. Siritunga. 2014. Assessment of Genetic Diversity of Sweet Potato in Puerto Rico. PLOS One 9(12):e116184.

Rosenberg, N.A. 2004. DISTRUCT: A program for genetic marker data. Bioinformatics 20:1247–1248.

Sengupta, S., B. Das, M. Prasad, P. Acharaya, and T.K. Ghose. 2013. A comparative survey of genetic diversity among a set of Caricaceae accessions using microsatellite markers. Springerplus 2:345–355.

Sudha, R., D.R. Singh, M. Sankaran, S. Singh, V. Damodaran, and P. Simachalam. 2013. Genetic diversity analysis of papaya (Carica papaya L.) genotypes in Andaman Islands using morphological and molecular markers. Afr. J. Agr. Res. 8:5187–5192.

Teixeira da Silva, J.A., Z. Rashid, D. Tan, N. Dharini, A. Gera, M. Teixeira, and P.F. Tennant. 2007. Papaya (Carica papaya L.) biology and biotechnology, p. 47–73. Tree and forestry science and biotechnology, Global Science Books, London.

United States Coast Guard. 2016. Alien migrant interdiction statistics. 13 Dec. 2017. <https://www.uscg.mil/hq/cg5/cg531/AMIO/amio.asp>.

Vidal, N.M., A.L. Graziotin, H.C. Ramos, M.G. Pereira, and T.M. Venancio. 2014. Development of a gene-centered SSR Atlas as a resource for Papaya (Carica papaya) marker-assisted selection and population genetic studies. PLOS One. 9:e12654.

Wang, M.L., N.A. Barkley, and T.M. Jenkins. 2009. Microsatellite markers in plants and insects. Part 1: Applications of biotechnology. Genes. Genomes Genomics 3:54–67.

Wendel, J.F., C.L. Brubaker, and P.E. Percival. 1992. Genetic diversity in Gossypium hirsutum and the origin of upland cotton. Amer. J. Bot. 79:1291–1310.

Zambrano-Echevarria, C., L. de Jesús-Kim, R. Márquez-Karry, D. Jenkins, and D. Siritunga. 2016. Diversity of Papaya ringspot virus isolates in Puerto Rico. HortScience 51:362–369.
Supplemental Fig. 1. Number of alleles per simple sequence repeat (SSR) marker for 162 farmer-held papaya samples consisting of 139 unknown accessions from Puerto Rico, 13 accessions from U.S. Department of Agriculture germplasm collection and 10 known commercial varieties. Allele number ranged from 2 to 18 alleles per locus with a mean of 9.304.

Supplemental Fig. 2. Delta $k$ value by number of $k$ (groups) calculated using Structure Harvester software.
### Supplemental Table 1. List of evaluated samples. NW, northwest; NE, northeast; C center; SW, southwest; USDA, USDA germplasm repository samples, commercial, commercial varieties.

| Number | Provenance | Area |
|--------|------------|------|
| 1      | Aguadilla  | NW   |
| 3      | Aguadilla  | NW   |
| 11     | Aguadilla  | NW   |
| 14     | Aguadilla  | NW   |
| 18     | Arecibo    | NW   |
| 19     | Manati     | NW   |
| 29     | Aguadilla  | NW   |
| 33     | Aguada     | NW   |
| 36     | Aguada     | NW   |
| 38     | Isabel     | NW   |
| 45     | Moca       | NW   |
| 48     | Barceloneta| NW   |
| 49     | Camuy      | NW   |
| 50     | Quebradillas| NW |
| 55     | Rincon     | NW   |
| 56     | Aguada     | NW   |
| 60     | Arecibo    | NW   |
| 65     | Manati     | NW   |
| 66     | Manati     | NW   |
| 70     | Hatillo    | NW   |
| 73     | Arecibo    | NW   |
| 78     | San Sebastian| NW |
| 82     | Aguadilla  | NW   |
| 85     | Camuy      | NW   |
| 86     | Moca       | NW   |
| 89     | Aguada     | NW   |
| 91     | San Sebastian| NW |
| 94     | Aguadilla  | NW   |
| 96     | Florida    | NW   |
| 97     | Arecibo    | NW   |
| 101    | Aguadilla  | NW   |
| 102    | Aguada     | NW   |
| 104    | Aguada     | NW   |
| 105    | Barceloneta| NW   |
| 106    | Barceloneta| NW   |
| 107    | Florida    | NW   |
| 108    | Florida    | NW   |
| 6      | Rio Grande | NE   |
| 8      | Carolina   | NE   |
| 9      | Guayanabo  | NE   |
| 16     | Rio Piedras| NE  |
| 17     | Vega Alta  | NE   |
| 20     | Vega Baja  | NE   |
| 21     | Vega Baja  | NE   |
| 27     | Dorado     | NE   |
| 30     | Dorado     | NE   |
| 42     | Carolina   | NE   |
| 51     | Vega Baja  | NE   |
| 53     | Rio Piedras| NE  |
| 72     | Rio Piedras| NE  |
| 90     | Trujillo Alto| NE |
| 113    | Vega Baja  | NE   |
| 114    | Vega Baja  | NE   |
| 127    | Trujillo Alto| NE |
| 128    | Toa Baja   | NE   |
| 134    | Vega Baja  | NE   |
| 137    | Naguabo    | NE   |
| 146    | Toa Alta   | NE   |
| 154    | Toa Baja   | NE   |
| 2      | Arroyo     | SE   |
| 35     | Aibonito   | SE   |
| 37     | Salinas    | SE   |
| 68     | Las Piedras| SE  |
| 79     | Cayey      | SE   |
| 88     | Coamo      | SE   |
| 116    | Coamo      | SE   |
| 118    | Guayama    | SE   |
| 119    | Salinas    | SE   |
| 120    | Santa Isabel| SE |
| 121    | Guayama    | SE   |
| 122    | Arroyo     | SE   |
| 124    | Maunabo    | SE   |
| 125    | Yabucoa    | SE   |

(Continued on next page)
Supplemental Table 1. (Continued) List of evaluated samples. NW, northwest; NE, northeast; C center; SW, southwest; USDA, USDA germplasm repository samples, commercial, commercial varieties.

| Number | Provenance | Area |
|--------|------------|------|
| 130    | Vieques    | SE   |
| 136    | Yabucoa    | SE   |
| 139    | San Lorenzo| SE   |
| 142    | Coamo      | SE   |
| 143    | Salinas    | SE   |
| 148    | Cidra      | SE   |
| 150    | San Lorenzo| SE   |
| 151    | Cidra      | SE   |
| 152    | Aibonito   | SE   |
| 153    | Salinas    | SE   |
| 23     | Utuado     | C    |
| 26     | Lares      | C    |
| 58     | Naranjito  | C    |
| 64     | Caguas     | C    |
| 69     | Ciales     | C    |
| 93     | Lares      | C    |
| 95     | Morovis    | C    |
| 109    | Ciales     | C    |
| 110    | Ciales     | C    |
| 111    | Morovis    | C    |
| 112    | Morovis    | C    |
| 115    | Jayuya     | C    |
| 117    | Barranquitas| C  |
| 126    | Caguas     | C    |
| 129    | Corozal    | C    |
| 131    | Lares      | C    |
| 132    | Adjuntas   | C    |
| 133    | Morovis    | C    |
| 138    | Gurabo     | C    |
| 140    | Lares      | C    |
| 141    | Adjuntas   | C    |
| 145    | Ciales     | C    |
| 147    | Utuado     | C    |
| 149    | Utuado     | C    |
| 13     | Mayaguez   | SW   |
| 15     | Anasco     | SW   |
| 22     | Hormigueros| SW   |
| 24     | Peñuelas   | SW   |
| 25     | Mayaguez   | SW   |
| 34     | Yauco      | SW   |
| 39     | Mayaguez   | SW   |
| 41     | Yauco      | SW   |
| 47     | Mayaguez   | SW   |
| 52     | Cabo Rojo  | SW   |
| 54     | Ponce      | SW   |
| 57     | Mayaguez   | SW   |
| 59     | Guayanilla | SW   |
| 62     | Hormigueros| SW   |
| 63     | Cabo Rojo  | SW   |
| 67     | Mayaguez   | SW   |
| 71     | Cabo Rojo  | SW   |
| 74     | Mayaguez   | SW   |
| 75     | Cabo Rojo  | SW   |
| 76     | Juana Diaz | SW   |
| 77     | Mayaguez   | SW   |
| 80     | Yauco      | SW   |
| 81     | Mayaguez   | SW   |
| 83     | Las Marias | SW   |
| 84     | Mayaguez   | SW   |
| 87     | Las Marias | SW   |
| 92     | San German | SW   |
| 98     | Guanica    | SW   |
| 99     | Yauco      | SW   |
| 100    | Las Marias | SW   |
| 103    | Sabana Grande| SW  |
| Mona   | USDA       |
| k14    | Panama (Brash, *Carica papaya*)| USDA |
| k17    | Northern Mariana (Saipan Red, *Carica papaya*)| USDA |
| k20    | Thailand (Khag Naun, *Carica papaya*)| USDA |
| k164   | United States (Hawaii, *Carica papaya*)| USDA |
| k207   | Taiwan (Tainung No. 5, *Carica papaya*)| USDA |
| k217   | Puerto Rico| USDA |
| k309   | ? (Kaek Dum) | USDA |
### Supplemental Table 2

List of analyzed locus with the number of alleles recorded (Na), number of effective alleles (Ne), and polymorphic index content (PIC) for each. Values calculated by analysis of 162 papaya samples with 23 SSR markers using PowerMarker software.

| Locus            | Na     | Ne     | PIC               |
|------------------|--------|--------|-------------------|
| AJ810489a        | 13.000 | 4.568  | 0.796199846       |
| AJ810490a        | 10.000 | 2.740  | 0.648929532       |
| AJ810491a        | 13.000 | 2.713  | 0.60508896        |
| AJ810492a        | 6.000  | 2.084  | 0.609898418       |
| AJ810493a        | 7.000  | 1.239  | 0.291694153       |
| AJ810494a        | 8.000  | 2.928  | 0.689352364       |
| AJ810495a        | 8.000  | 2.601  | 0.640798891       |
| AJ810505a        | 6.000  | 2.596  | 0.658747159       |
| CP21b            | 17.000 | 4.210  | 0.773336617       |
| CP31b            | 6.000  | 1.846  | 0.486742983       |
| CP44b            | 10.000 | 1.819  | 0.49177939        |
| CP49b            | 6.000  | 2.080  | 0.558827407       |
| CPCIR2a          | 13.000 | 7.025  | 0.862844808       |
| CPCIR3a          | 13.000 | 2.778  | 0.640764542       |
| SP1c             | 8.000  | 2.474  | 0.624536215       |
| SP3c             | 18.000 | 6.221  | 0.848939533       |
| SP4c             | 10.000 | 2.985  | 0.708552437       |
| SP5c             | 10.000 | 3.037  | 0.673751893       |
| SP6c             | 8.000  | 2.622  | 0.638296699       |
| SP7c             | 2.000  | 1.882  | 0.465980788       |
| SP8c             | 7.000  | 1.637  | 0.444810713       |
| SSPA3c           | 5.000  | 3.109  | 0.654240109       |
| SSPA8c           | 10.000 | 2.029  | 0.598436895       |
| Mean             | 9.304  | 2.923  | 0.626632624       |

### Supplemental Table 3

Linkage disequilibrium (LD) analysis within assessed SSR marker pairs using GENEPOP V 4.2 software.

| Number of pairwise comparisons | P value <05 | P value >05 |
|--------------------------------|-------------|-------------|
| 153                            | 146         | 107         |
|                                | 57.70%      | 42.30%      |