Molecular-genetic analysis for validation of peach palm (*Bactris gasipaes* Kunt) landraces using RAPD markers

**Resumo**
Numerosas raças de pupunha (*Bactris gasipaes* Kunt) são conservadas em um banco de germoplasma do INPA. Utilizaram-se marcadores RAPD para analisar a diversidade genética de oito raças de pupunha e duas populações silvestres. Oito primers geraram 124 marcadores; 101 marcadores foram polimórficos (81.5 %). Há heterozigosidade observada foi de 0,38, e o polimorfismo foi de 93 %, ambos ligeiramente superiores aos estudos anteriores. As raças amazônicas apresentaram alta heterozigosidade (0,30) e uma porcentagem de polimorfismo (87,8 %) semelhante a raças da América Central (0,29 e 83,5 %, respectivamente), diferente de estudos anteriores, que descobriram que as raças da América Central têm valores de parâmetros mais baixos do que as raças da Amazônia. A estrutura do dendrograma com a distância genética de Nei foi semelhante à de estudos anteriores, com pequenas diferenças. As duas populações silvestres foram agrupadas longe das raças, sugerindo que não participaram da domesticação das populações cultivadas.

**Palavras-chave adicionais:** caracterização molecular, diversidade genética, fluxo gênico, relações genéticas,
the peach palm presents a complex landrace hierarchy, which was partially characterized morphologically and mapped (Mora Urpi & Clement, 1988; Mora Urpi et al., 1997) and partially validated genetically (Sousa et al., 2001; Rodrigues et al., 2004; Cristo-Araújo et al., 2010; Santos et al., 2011). Although peach palm was domesticated because of its fruit, the main demand today is for its palm heart. This is a profitable agribusiness in expansion (Bovi, 1997; Clement et al., 2004), which competes with heart-of-palm from *Euterpe oleracea* Mart. and *Euterpe edulis* Mart. (Batagin-Plotto et al., 2012).

The National Research Institute for Amazonia (INPA) created and maintains the Peach Palm Active Germplasm Bank (AGB) in collaboration with Embrapa Genetic Resources and Biotechnology. The germplasm bank has 375 accessions collected from wild and cultivated populations. It aims to make available high-quality material for the improvement of this species (Clement et al., 1996). The characterization and evaluation of accessions and populations, both phenotypically and genotypically, is important to understand the genetic diversity and structure of populations, and to evaluate characteristics linked to the production process. However, the morphometric characterization of the Peach Palm AGB is only partial (Clement et al., 1996; Clement et al., 2004), although the molecular genetic characterization has advanced considerably. Previous studies with molecular markers evaluated the genetic diversity of peach palm landraces, as well as gene flow among them, to determine the genetic validity of landraces defined with morphological criteria. Using RAPD markers, the existence of four landraces (Pará, Pampa Hermosa, Putumayo and Utilis) was validated, and that of three other landraces (Solimões, Tuíra and Guatuso) was not (Rodrigues et al., 2004). A morphometric analysis suggested that the Solimões landrace should be maintained, but with a smaller geographic distribution than originally proposed (Martel et al., 2003). Four hybrid populations (Belém, Manaus, Iquitos, Yurimaguas) were shown not to have higher genetic diversity than the landraces that surround them (Santos et al., 2011), contrary to the hypothesis of Mora Urpi and Clement (1988). Silva et al. (2018) determined that three populations around Yurimaguas, Peru, are not different from the Pampa Hermosa landrace (also adjacent to Yurimaguas), contrary to the ideas of Mora-Urpi and Clement (1988), who had suggested the existence of other landraces based on the morphometric variability observed in the Yurimaguas hybrid population. Using AFLP, Adin et al. (2004) compared the genetic diversity between domesticated populations along the Paraná and Cuiparillo Rivers (Yurimaguas, Peru), and observed little divergence and high gene flow, supporting the hypothesis that they are populations of the Pampa Hermosa landrace. Although the majority of the landrace and hybrid population accessions in the BAG have been examined with molecular markers, several cultivated and wild populations have not yet been studied.

This study furthered the validation process of peach palm landraces, with the following objectives: (1) characterize and evaluate the genetic variability of a sample of landraces in the INPA Peach Palm Active Germplasm Bank with RAPD markers; (2) validate the conclusions regarding Pampa Hermosa, Putumayo, Pará and Utilis landraces within the landrace complex; and (3) analyze the genetic relations among the Jurúá, Inírida, Vaupés and Cauca landraces, as well as the two wild populations included here.

## Materials and methods

We sampled 182 plants of eight landraces (var. *gasipaes*) among the plants kept in the INPA Peach Palm Active Germplasm Bank (BR 174, km 38, Manaus, Amazonas state, Brazil): Putumayo (*n* = 30; 15 from the Solimões morphologically defined landrace distribution and 15 from the original Putumayo distribution), Pará (*n* = 30), Utilis (*n* = 30; 10 from the Tuíra morphologically defined landrace distribution, 10 from the Guatuso morphologically defined landrace, and 10 from the original Utilis distribution), Pampa Hermosa (*n* = 30), Jurúá (*n* = 30), Vaupés (*n* = 15), Cauca (*n* = 11) and Inírida (*n* = 6). We also sampled 18 plants from two populations of var. *chichagui* (from the Magdalena River, Colombia, var. *chichagui* type 2, *n* = 5) and from the Xingu River, Pará, Brazil [var. *chichagui* type 1 (Silva & Clement, 2005), *n* = 13]. Both var. *chichagui* populations were considered “out groups” in the analysis. We selected samples of the Putumayo and Utilis landraces to validate the analyses made by Rodrigues et al. (2004). We selected the accessions because they presented good geographical coverage of the landrace distributions. The plants of each accession were sampled at random.

The DNA was extracted using the DNAasy Plant Mini-Kit by Quiagen using 100 mg of the apical meristem of a lateral stem, as proposed by Clement et al. (1997) for enzyme extraction. The DNA was quantified in 0.8 % agarose gel stained with ethidium bromide. An average of 14 μg of DNA was obtained. In the absence of lateral stems, 200 mg of young leaf tissue was macerated using a pistil and liquid nitrogen (Weising et al., 1995). We obtained an average of 4 μg of DNA. In the case of the wild populations (var. *chichagui*), the DNA was extracted from seed embryos as they did not germinate fast enough to extract the DNA from their meristems. We obtained 6-10 μg of DNA.

The RAPDs were amplified according to Williams et al. (1990) with modifications (Rodrigues et al., 2004). Each amplification reaction, in a final volume of 30 μL, contained 10 ng of genomic DNA (5 ng μL⁻¹), 250 μM of dNTP (2.5 mM), 3 mM of MgCl₂ (25 mM), 50 ng of primer (10 ng μL⁻¹) and 1.5 U of Taq polymerase enzyme (CENBIOT/RS). The samples were amplified in a thermocycler (Perkin Elmer 9600) using two programs (according to the primer used, following Rodrigues et al., 2004). Program 1 (F-919-3):
1 step of 2 min at 94 °C, 40 cycles of 1 min at 92 °C, 1 min at 36 °C and 2 min at 72 °C, plus 1 step of 3 min at 72 °C. Program 2 (OPA-4, OPA-5, OPA-8, OPA-9, OPA-18, OPA-20, FC13): 2 cycles of 1 min at 94 °C, 1 min at 36 °C and 2 min at 72 °C, and 33 cycles of 10 sec at 94 °C, 20 sec at 40 °C and 2 min at 72 °C, plus 1 step of 5 min at 72 °C.

The amplified products were separated in agarose gel (1.5 %) in a TBE buffer at 120 V for 4 hours. Then, they were stained with 0.5 mg/mL ethidium bromide, visualized and photographed for further interpretation. One Juruá landrace plant was used as a control in the gels (with a replicate on the left side and one in the middle of the gel).

The bands were classified as present (1) or absent (0) based on the resolution and the degree of amplification by visual evaluation, as described by Ferreira and Grattapaglia (1998), to create a binary matrix. Based on this matrix, we estimated observed heterozygosity, assuming the absence of the marker as recessive (Weir, 1996), and the percentage of polymorphism for each population using the TFPGA program (Miller, 1997). We evaluated the relations among populations with Nei's genetic distance (1972) and a dendrogram was generated with UPGMA, using the TFPGA program. Differences in allele frequencies among populations were compared by the Exact Test (Raymond & Rousset, 1995) using the TFPGA program. We calculated gene flow [Nm = 0.5(1-Gst/Gsti)] among populations taking into account Gst estimates among populations (Slatkin and Barton, 1989) using the POPGEN program version 1.31 (Yeh et al., 1999).

Results and discussion

The eight primers generated 124 useful amplification products (defined as having good resolution), with an average of 15.5 bands per primer. Of these, 101 were polymorphic (81.5 %), lower than that obtained by Rodrigues et al. (2004) (89 %). There was variation in total numbers of polymorphic bands (Table 1), with a minimum for the OPA-9 primer (8 bands) and a maximum for the OPA-20 and OPA-4 primers (17 bands). There was also variation within landraces: only 5 bands for the Juruá landrace (OPA-9) and 17 bands for the Pará and Putumayo landraces (OPA-20). The Amazonian landraces presented high polymorphism, which was also observed by Rodrigues et al. (2004). The Amazonian landraces averaged 87.8 bands versus an average of 83.5 bands among the three Utilis populations in Central America.

### Table 1 - Number of useful RAPD markers obtained from each primer for the eight peach palm landraces (B. gasipaes var. gasipaes) and the two wild populations (B. gasipaes var. chichagui) kept in the Peach Palm Active Germplasm Bank, Manaos, Amazonas, Brazil, and number of polymorphic markers for each landrace.

<table>
<thead>
<tr>
<th>Primer</th>
<th>T</th>
<th>M</th>
<th>Pará</th>
<th>Put</th>
<th>Utilis</th>
<th>Pampa</th>
<th>Vaupés</th>
<th>Cauca</th>
<th>Inirida</th>
<th>Juruá</th>
<th>Chi</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-919-3</td>
<td>17</td>
<td>12</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>OPA-20</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>16</td>
<td>15</td>
<td>16</td>
<td>12</td>
<td>16</td>
<td>16</td>
<td>-*</td>
</tr>
<tr>
<td>OPA-9</td>
<td>12</td>
<td>8</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>-*</td>
</tr>
<tr>
<td>FC-13</td>
<td>15</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>-*</td>
</tr>
<tr>
<td>OPA-5</td>
<td>16</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>10</td>
<td>12</td>
<td>-*</td>
</tr>
<tr>
<td>OPA-8</td>
<td>16</td>
<td>15</td>
<td>15</td>
<td>14</td>
<td>15</td>
<td>14</td>
<td>14</td>
<td>12</td>
<td>8</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>OPA-18</td>
<td>13</td>
<td>10</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>OPA-4</td>
<td>18</td>
<td>17</td>
<td>14</td>
<td>13</td>
<td>13</td>
<td>14</td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>14</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>124</td>
<td>101</td>
<td>93</td>
<td>91</td>
<td>89</td>
<td>91</td>
<td>87</td>
<td>78</td>
<td>75</td>
<td>90</td>
<td>50</td>
</tr>
</tbody>
</table>

T = number of markers. M = total number of polymorphic markers. Put = Putumayo, Pampa = Pampa Hermosa, chi = var. chichagui. * Only 4 primers were used (the most polymorphic, according to Rodrigues et al., 2004) for the analysis of the var. chichagui due to the small amount of DNA available for each individual.

The mean heterozygosity (H) estimated for this set of plants was 0.38, with 93 % of polymorphism (Table 2). Such values are higher than those obtained by Rodrigues et al. (2004) (0.31 and 89.4 %). The observed heterozygosities were always lower than or equal to the expected heterozygosities at the level of individual landraces and populations. However, it is worth remembering that heterozygosity estimates may be inflated because they are calculated based on an estimation of gene frequencies of the null homozygote, since RAPDs are dominant and do not distinguish between heterozygotes (Weir, 1996).

The average gene flow among all landraces was 4.35, somewhat higher than the value found by Rodrigues et al. (2004), who obtained a value of 2.7. Excluding the loci that did not present information for the two wild populations, gene flow was 2.43 (Table 3), a value closer to that of Rodrigues et al. (2004). In general, the closer the pair of landraces are geographically, the greater the gene flow. However, several pairs of landraces presented inconsistent estimates, such as the Juruá and Vaupés landraces, which are on different sides of the Amazon basin. Such inconsistent gene flows contribute to the inconsistent relations in genetic distances among landraces (Table 3).
The populations that represented the original Putumayo and Solimões landraces did not show differences in their allelic frequencies ($p = 0.08$) by the Exact test (Raymond & Rosset, 1995), as was also observed by Rodrigues et al. (2004) ($p = 0.11$). The Utilis, Darien and Guatuso populations of Central America also did not present differences. Rodrigues et al. (2004) found a $p = 0.97$ for the relation between Darien and Utilis, a value almost equal to the $p = 0.92$ we found in this study. Similarly, those authors found a $p = 0.09$ for the relation between Darien and Guatuso, while we found a $p = 0.07$, corroborating the hypothesis that these three populations are part of a same Utilis landrace. The Guatuso population is present around San Carlos de Alajuela in northern Costa Rica and is important because it contains many plants without spines on the stems and leaves, which makes them interesting for heart-of-palm production (Clement et al., 1996). The selective pressure for spineless plants could explain why the allele frequencies of Guatuso are different from those of the other two populations in the Utilis landrace. 

In this analysis, the low heterozygosities and the percentages of polymorphism of the wild populations (var. chichagui of Magdalena and Xingu rivers) are curious: approximately one third of the values of the landraces. This may be due to the small number of samples and markers used (Table 1). Many markers
were monomorphic, especially with an absence of bands, suggesting that there was a mutation in the sequence of the primer that does not allow the generation of fragments (Ferreira & Grattapaglia, 1998).

The UPGMA dendrogram using Nei’s genetic distances (1972) and the same four landraces analyzed by Rodrigues et al. (2004) generated two groups (Figure 1-A). The first group contained the Putumayo, Útilis and Pampa Hermosa landraces and the second group contained the Pará landrace. Not surprisingly, the structure of the dendrogram was similar to that of Rodrigues et al. (2004), except that the Útilis landrace grouped with the Putumayo landrace before this pair grouped with the Pampa Hermosa landrace. The consistencies of these groups were similar to those found by Rodrigues et al. (2004) using a different set of plants.

The Juruá is the only landrace in the new set of populations that had enough plants (n = 30) to estimate its genetic relationships with other landraces (Figure 1-B). Its position, together with the three western landraces, makes geographical sense and suggests that the genetic similarities are due to a common origin of this set of landraces. This relationship of the Juruá landrace with the other western landraces was also observed with nSSR (Cristo-Araújo et al., 2013; Clement et al., 2017).

The other three landraces have small samples (n = 6-15), which does not allow much confidence in genetic relationships when compared to the better sampled landraces. The problematic groupings in this dendrogram (Figure 2) are between the Vaupés and Juruá landraces, because these two races are geographically distant and have remarkably different fruit shapes and sizes. The Vaupés landrace is classified as macrocarpa because it has large fruits (average 70 g, maximum 250 g), while the Juruá is classified as microcarpa because it has fruits with an average weight of 20 g (Mora Urpi & Clement, 1988). They would only be genetically related if they were geographically close, which is not the case since they are on different sides of the Amazon basin, so this relationship may be due to the small sample size of the Vaupés landrace or to gene bank errors, which have been found in other validation studies (Sousa et al., 2001; Rodrigues et al., 2004; Cristo-Araújo et al., 2010; Santos et al., 2011). The relation between the Cauca and Inirida landraces is also problematic, since they are geographically separated by the Andes. The Inirida landrace is geographically close to the Vaupés landrace and has a similar fruit shape (flattened). A priori, a relation between Inirida and Vaupés landraces was expected, but the non-expected result can be attributed to one or both of the problems just mentioned.

Figure 1 – UPGMA dendograms based on Nei’s genetic distances (1972) and 30 plants per landrace. A. Four peach palm landraces (Bactris gasipaes var. gasipaes) from Amazonia and Central America previously analyzed by Rodrigues et al. (2004). B. Five peach palm landraces from Amazonia and Central America. The consistency of the nodes is expressed as the percentage of markers that support each node.
The two wild populations (var. chichagui type 2 from the Magdalena River and type 1 from the Xingu River) are very different from the cultivated landraces. This suggests that these two wild populations were not involved in the origin of cultivated peach palm, as also suggested by Couvreur et al. (2007) for the var. chichagui type 2 with chloroplast sequences. The distance from the Xingu River also suggests that only the western populations of var. chichagui type 1 were involved in the origin of cultivated peach palm (Clement et al., 2017).

In spite of these inconsistencies, the separation between the Pará landrace and all the other landraces agrees with all previous analyses. This suggests that all landraces and populations of the western Amazon and northwestern South America originated from the same region in southwestern Amazonia, as suggested by Rodrigues et al. (2004) and confirmed with nSSR and chloroplast sequences by Cristo-Araújo et al. (2013) and Clement et al. (2017).

Conclusions

The peach palm landraces we analyzed in this study are closely related, although individually they presented a high degrees of polymorphism and heterozygosity. The analysis validated the four landraces analyzed by Rodrigues et al. (2004), especially the separation of the Pará landrace from the others. The Juruá landrace was related to the three western landraces (Putumayo, Utilis and Pampa Hermosa), which makes geographic sense. The other landraces (Vaupés, Cauca and Inirida) were represented by very small numbers of individuals, not allowing a more precise conclusion. The two wild populations (var. chichagui type 2 of the Magdalena River and type 1 of the Xingu River) were quite different from the cultivated landraces. This suggests that these two wild populations did not participate in the origin of the cultivated peach palm.

Acknowledgements

The authors are grateful to the Banco da Amazônia SA for financing the project "Molecular markers (RAPDs) for the discrimination of peach palm landraces (Bactris gasipaes) kept in the Active Germplasm Bank", coordinated by Clement and Astolfi Filho, and to the Foundation Djalma Batista for administrative and logistical support.
References


