Diversity of Treegourd (Crescentia cujete) suggests introduction and prehistoric dispersal routes into Amazonia

Priscila Ambrósio Moreira, Xitlali Aguirre-Dugua, Cédric Mariac, Leila Zekraoui, Marie Couderc, Doriane Picango-Rodrigues, Alejandro Casas, Charles Roland Clément, Yves Vigouroux

To cite this version:


HAL Id: hal-03340603
https://hal.umontpellier.fr/hal-03340603
Submitted on 16 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Diversity of Treegourd (*Crescentia cujete*) Suggests Introduction and Prehistoric Dispersal Routes into Amazonia

Priscila A. Moreira¹*, Xitlali Aguirre-Dugua², Cédric Mariac³, Leila Zekraoui³, Marie Couderc³, Doriane P. Rodrigues⁴, Alejandro Casas³, Charles R. Clement⁵ and Yves Vigouroux³

¹ Post–Graduate Program in Botany, Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil, ² Centro de Investigaciones en Ecosistemas, Universidad Nacional Autónoma de México, Morelia, Mexico, ³ Institut de Recherche pour le Développement, Université de Montpellier, UMR DIADE, Montpellier, France, ⁴ Laboratório de Evolução Aplicada, Universidade Federal do Amazonas, Manaus, Brazil, ⁵ Coordenção de Tecnologia e Inovação, Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil

The use and dispersal of domesticated plants may reflect patterns of early human diffusion of technologies and lifestyles. Treegourd (*Crescentia cujete*) has fruits with ancient utilitarian and symbolic value in the Neotropics. We assessed diversity based on chloroplast (SNPs), nuclear (SSR) markers, and fruit shapes of cultivated treegourds and wild relatives across Amazonia and Mesoamerica in order to discuss hypothesis of dispersal routes and diversification of fruits along its distribution. The haplotype network showed three distinct groups: *Crescentia amazonica*, wild Mesoamerican *C. cujete*, and cultivated *C. cujete* from Brazilian Amazonia and Mexico. Mexico and Brazil shared two haplotypes, with slightly different distributions in Amazonia. The most divergent haplotype is well-represented in Eastern Amazonia. Nuclear differentiation between Mesoamerican wild and cultivated *C. cujete* is relatively low (\(F_{ST} = 0.35\)), compared with Amazonian cultivated (\(F_{ST} = 0.45–0.61\)). Differentiation is also higher between wild *C. amazonica* and cultivated *C. cujete* (\(F_{ST} = 0.57\)), but modest within cultivated *C. cujete* from Amazonia and Mexico (\(F_{ST} = 0.04\)), with higher genetic similarity in northwestern Amazonia. Mexico and Amazonia showed similar chloroplast nucleotide diversity (\(4.66 \times 10^{-2}\) and \(5.31 \times 10^{-2}\), respectively), although sample sizes are very different. Except in Northwestern and Eastern Amazonia, we found ample genetic homogeneity of cultivated *C. cujete* across Amazonia, but highest morphological diversity in the Northwest, with fruit shapes that are absent in Mexico. We conclude that treegourds introduced into the Amazon Basin and Mexico share a common ancestry with a currently unknown origin. The patterns of genetic diversity across Amazonia allow two hypotheses of the routes of introduction: a northwestern introduction into the Negro and Solimões Rivers, and an eastern introduction from the coastal Guianas into the Amazonas River. The dispersal into Amazonia followed previously proposed routes of human and plant migrations. The contrasting fruit shape diversity suggests different utilitarian demands and cultural preferences for treegourd fruits between Mexico and Amazonia.

Keywords: bignoniaceae, calabash, cuia, domestication, ethnobotany, historical ecology, phylogeography
INTRODUCTION

The use and dispersal of domesticated plants may reflect the patterns of diffusion of human technologies and lifestyles since prehistoric times (Bellwood, 2005; Blench, 2012). Humans greatly expand plants’ geographical distributions, which ultimately exerts different ecological and cultural pressures on the evolutionary pathways of plants (Rindos, 1984; Sodero Martins, 2005; Leclerc and Coppens’Eeckenbrugge, 2012; Meyer and Purugganan, 2013). Various centers of domestication have been proposed in the Americas (Meyer et al., 2012) from where people exchanged plants (Heiser, 1965; Schultes, 1984; Colunga-GarciaMarín and Zizumbo-Villarreal, 2004). Amazonia is one of them (Clement, 1999), and also encompasses great linguistic diversity (Blench, 2012), ceramic styles (Barreto et al., 2016), and landscape management strategies (Eriksen and Danielsen, 2014), whose geography and chronology are being disentangled (Mayle and Iriarte, 2014; Clement et al., 2015; Neves, 2016; Levis et al., 2017). The Amazonian routes of dispersal of plants and people have been associated with rivers and riparian environments (Schultes, 1984; Godoy et al., 1999; Guix, 2009). However, few studies have demonstrated the genetic signatures of the plants’ geographical dispersal mediated by humans in Amazonia (Clement et al., 2010; Shepard and Ramirez, 2011; Thomas et al., 2012; Freitas and Bustamante, 2013), even though they are persistent markers of the long-term use and management of resources (Hannote et al., 2002; Parker et al., 2010; Armstrong et al., 2017).

Treegourd (Crescentia cujete) is a good case study, since its trees produce fruits with ancient utilitarian and symbolic value widely dispersed across the Neotropics (Gentry, 1980; Arango-Ulloa et al., 2009; Meulenber, 2011; Aguirre-Dugua et al., 2013; Medeiros and Albuquerque, 2014; Moreira et al., 2017). It currently is one of the most common species in homegardens of the floodplains and adjacent communities of Amazonia (Santos, 1982; Lima and Saragoussi, 2000). Its fruits have different shapes and sizes that are used as bowls, vessels or bottles for drinking or transporting water, bags for provisions, utensils for cooking, and eating, bailing water from canoes, construction of fish traps, manufacture of body ornaments, and musical instruments (Steward, 1948; Patiño, 1967; Morton, 1968; Price, 1982; Bennett, 1992; Heiser, 1993; Meulenber, 2011). Medicinal uses are also similar across its distribution (Morton, 1968; Duke, 2009), which include neutralization of snake venom and intestinal parasites treatment (Otero et al., 2000; Volpato et al., 2009; Ramos, 2015; Paulo, 2016).

While the great phenotypic variability of cultivated treegourd is a distinctive feature among Crescentia species (Gentry, 1980), its wild populations from Mexican savannahs in the Yucatan Peninsula have smaller, elongated fruits with thinner exocarps (Aguirre-Dugua et al., 2012). The indehiscent and thicker exocarp of cultivated treegourd fruits makes the spontaneous dispersal of seeds impossible (Aguirre-Dugua et al., 2012). Its oldest remains found to date come from a Peruvian archaeological site dating to 5,000–3,800 years BP (Solis, 2006). This pattern contrasts to the bottle gourd, collected from a vine (Lagenaria siceraria), one of the ancient crops similarly used for technological purposes in the Americas (Heiser, 1993). Bottle gourd has been managed at least since the Late Pleistocene (Kistler et al., 2014) and was found in Colombian Amazon by 8,000 BP (Piperno, 2011). The wild progenitor of the cultivated Crescentia cujete remains elusive (Gentry, 1980; Arango-Ulloa et al., 2009; Aguirre-Dugua et al., 2012; Moreira et al., 2017). Gentry (1980) pointed out that C. cujete was certainly native to Mesoamerica, where putative wild populations are found in savannahs and semi-evergreen forests of southern Mexico and northern Central America (Figure 1). However, northern South America cannot be ruled out as part of the original distribution area of wild C. cujete, given the occurrence of apparently spontaneous C. cujete in grazed savannahs of Andean and Caribbean regions of Colombia (Arango-Ulloa et al., 2009). Historical anthropogenic fire management in savannahs (Pinter et al., 2011) may have been advantageous for its early dispersal (Bass, 2004) in these regions. Recently, the wild species native to Amazonian and Orinocan floodplains (Crescentia amazonica) was ruled out as the wild progenitor of cultivated C. cujete (Ducke, 1946; Moreira et al., 2017). Likewise, the wild C. cujete populations found in the southeastern Mexico are not the wild progenitor either (Aguirre-Dugua et al., under revision). In this study, we infer treegourd dispersal and diversification across two pivotal regions of the Neotropics: Amazonia and Mesoamerica. We (1) identify genetic relationships among Mesoamerican and Amazonian cultivated C. cujete; (2) infer routes of introduction into and dispersal within the Amazon Basin; and (3) identify centers of morphological and genetic diversity. We discuss whether this genetic/morphological diversity is linked to (1) introgression with local wild parents, (2) ecological diversification, or (3) cultural diversification, since all three of them are possible along the dispersal routes.

MATERIALS AND METHODS

Sampling

We performed molecular analyses using full chloroplast (SNPs) and nuclear (SSR) markers. We also analyzed fruit morphology along the major rivers of Brazilian Amazonia and in parts of Mesoamerica (Supplementary Table S1). We used a previously published genetic and morphological dataset (Moreira et al., 2017) of cultivated C. cujete (N = 372) distributed in 122 localities along the five major rivers of the Brazilian Amazon basin, as well as wild Brazilian treegourds (C. amazonica) (N = 20) distributed in three of the rivers mentioned (Figure 1).

From Mexico, we add new genetic data of cultivated C. cujete from the Yucatan Peninsula, Oaxaca and Chiapas, wild samples from the Yucatan savannahs and a putative wild sample from Costa Rica (Figure 1). We also integrate morphological data from Mesoamerican samples (N = 188), part of which (N = 124) was published previously (Aguirre-Dugua et al., 2013). All Mesoamerican wild samples were identified as C. cujete Linnaeus 1753. In order to depict the putative geographical distribution of wild C. cujete, we searched for individuals of C. cujete described as spontaneous in savannahs on herbarium descriptions found in GBIF (Global Biodiversity Information Facility) (Figure 1).
The geographical distribution of cultivated *C. cujete*, putative wild populations of *C. cujete* and wild *C. amazonica* in the Neotropics. Genetic or morphological analyses include samples from Mexico, Costa Rica and five rivers in Brazilian Amazonia (Negro, Branco, Solimões, Madeira, Amazonas). Their distributions were complemented with records from the Global Biodiversity Information Facility (GBIF) and plotted over the vegetation cover (Bartholomé and Belward, 2005). The wild *C. cujete* distribution was hypothesized based on apparently spontaneous individuals growing in a mosaic of shrub and grass cover, which does not rule out previous human dispersion, since areas might include abandoned or burned croplands.

This research followed the International Society for Ethnobiology’s code of ethics (International Society of Ethnobiology, 2006) and was approved by the Committee for Ethics in Research with Human Beings of the National Research Institute for Amazonia (CEP INPA, proc. no. 408.611, 2013). Collection in Brazil was authorized by the Brazilian System for Authorization and Information in Biodiversity, Chico Mendes Institute for Biodiversity Conservation, proc. no. 25052–1, 2012, and transportation by the Brazilian Institute for the Environment and Renewable Natural Resources, proc. no. 14BR015576/DE, 2014. Collection in Mexico and Costa Rica was authorized by proc. no. SGPA/DGGFS/712/3691/10.

**Genetic Analysis**

We used a previously described protocol for genotyping nuclear microsatellites and the detection of single nucleotide polymorphisms along the entire sequence of the maternally inherited chloroplast genome (Moreira et al., 2016, 2017). In total, 250 samples were genotyped for eight nuclear microsatellites (SSR): 234 from Brazilian Amazonia (215 cultivated *C. cujete* and 19 wild *C. amazonica*), and 16 from Mesoamerica (7 cultivated *C. cujete* from Mexico, 8 wild *C. cujete* from Mexico, 1 wild *C. cujete* from Costa Rica). Data from the chloroplast genome was obtained from a total of 215 samples: 191 *C. cujete* and 16 *C. amazonica* from Amazonia, 5 cultivated *C. cujete* from Mexico, 2 wild *C. cujete* from Mexico, 1 wild *C. cujete* from Costa Rica. Among the total sample (*N* = 250), 80% were genotyped and sequenced for both kinds of markers.

The nuclear SSR dataset was used to assess population structure with a Bayesian approach (Structure 2.3, Pritchard et al., 2000). We applied the admixture model in order to identify ancestral population proportions for each individual and their probable populations of origin. Using total sampling and assuming independent allele frequencies in each population, which reduces the risk of overestimating the number of clusters (Pritchard et al., 2000), we assessed the number of clusters *K* varying from 1 to 20, with 100,000 burn-in, 100,000 iterations, and five different runs for each *K* value. To attempt to identify different genetic pools within the cultivated cluster, we performed an additional analysis on a subset including only cultivated *C. cujete* samples, whose membership probability was higher than 0.6 in the cultivated cluster (*N* = 200). Using the admixture model, we experimented with two allele frequency assumptions (Pritchard et al., 2000): the independent model as default; and the correlated (assuming lambda = 1), since it is likely that cultivated populations share ancestry due to migration and vegetative propagation. Evanno et al. (2005) Δ*K* was used to guide our choice of the most likely number of groups. Additionally, we performed a Principal Components Analysis (PCA) with stats R package (R Core Team, 2015) in order to uncover additional genetic structure in our data (Jombart et al., 2009). The PCA was non-centered, but scaled in order...
to compensate for differences in polymorphism and missing data among the loci analyzed. The spatial interpolation of the clusters obtained in Structure was analyzed using the kriging method in the fields R package (Nychka et al., 2015). Based on geostatistics and maximum likelihood, the krig function estimates the covariance in a grid (we used the scale parameter theta = 50) and infers the fitted surface between geographical coordinates and genetic relationship among samples (Nychka et al., 2015). Nuclear genetic diversity of C. cujete [allelic richness (A_r), private alleles (A_p), observed heterozygosity (H_o), expected heterozygosity (H_e)] was estimated for the five Amazonian rivers considered and the Mexican samples using hfrstat (Goudet, 2005) and poppr (Kamvar et al., 2014) R packages. Pairwise F_ST between regions were estimated and statistically evaluated using 1,000 bootstraps (Nei, 1987). A neighbor-joining dendrogram of regions was constructed based on Nei’s distance and 1,000 bootstraps (Saitou and Nei, 1987). The inbreeding coefficient F_IS for each region was estimated and its significance evaluated (considering a Bonferroni corrected p-value of 0.006) using pegas R package (Paradis, 2010).

For the identification of chloroplast SNPs, we used a bioinformatic pipeline previously validated for the sequencing of the entire chloroplast genome (Scarcelli et al., 2016). Briefly, SAMTOOLS 0.1.7 with option-B (Li et al., 2009) was used to generate an mpileup file. VARSCAN 2.3.7 (Koboldt et al., 2012) was used to call SNPs from this mpileup file. The variant call format file (VCF) generated was filtered following Scarcelli et al. (2016) and resulted in a total of 334 cpSNPs detected in our dataset. The final vcf file was exported as a fasta file using VCFtools 1.14 ( Danecek et al., 2011) and haplotypes identified with DNAsp 5.10.1 (Librado and Rozas, 2009). An haplotype network was constructed using the median joining algorithm (Bandelt et al., 1999) and samples with up to 6.5% of missing data using POPART 1.7 (Leigh and Bryant, 2015). The geographical distribution of the shared haplotypes of C. cujete samples was plotted using GenGIS 2.5 (Parks et al., 2009). The chloroplast diversity of C. cujete [total number of polymorphic sites (S), number of haplotypes (h), and nucleotide diversity (π)] were estimated according to Nei (1987) using DNAsp 5.10.1. The presence of singleton samples and their contribution with unique alleles were identified using VCFtools 1.14. Paired F_ST among the Amazonian rivers and Mexico were estimated using the distance method of Tajima and Nei (1984), and their significance was evaluated with 1,000 permutations at a significance level of 0.05 using Arlequin 3.5 (Excoffier and Lischer, 2010).

**RESULTS**

**Geographic Patterns of Nuclear Diversity**

Evanno et al. (2005) ΔK suggested that two clusters are the most likely structure in the dataset (K = 2, Figure 2A, Supplementary Figure S1). At K = 2, a clear distinction among wild and cultivated samples was observed (clusters shown in blue and red in Figure 2A, respectively), regardless of their geographical origin. Mexican cultivated C. cujete samples showed an admixed pattern (membership probability to wild cluster from 0.16 to 0.87), as did some of the cultivated C. cujete from the Amazon Basin (membership probability to wild cluster from 0.01 to 0.98). The wild admixture within cultivated C. cujete in the Amazon Basin had higher proportions along the Amazonas River, decreasing values along the Solimões, Madeira, and Negro rivers, and was absent along the Branco River (Figure 2A). The wild Costa Rican sample displayed a membership probability of 0.25 to the cultivated cluster, a larger proportion than the membership shown by the Mexican wild samples (0.01–0.02). In the Principal Component Analysis (PCA), the first two principal components explained 16.7% of the total variance found in the dataset (Figure 2B). Principal component one separated wild from cultivated samples, while principal component two separated the Brazilian wild C. amazonica from the Mesoamerican wild C. cujete samples. The wild sample from Costa Rica was intermediate between wild and cultivated Mexican samples, which agree with its ancestry pattern observed in the clustering analysis performed by Structure. One Brazilian sample from Amazonas River was relatively closer to the Costa Rican sample (Figure 2B). To assess to what extent the intermediate ancestry of cultivated Mexican samples between wild Mesoamerican and Brazilian cultivated samples (Figure 2B) was associated with hybridization or divergence, we performed a Structure analysis among only Mesoamerican samples. This analysis clearly differentiates two groups of wild and cultivated Mesoamerican C. cujete (Supplementary Figure S2). However, we still observed the Costa Rican sample as having intermediate ancestry among these Mesoamerican samples (Supplementary Figure S2). Consequently, the intermediate ancestry detected in cultivated Mesoamerican samples may reflect divergence rather than hybridization. The differentiation between wild C. cujete and cultivated samples was also evident in the neighbor-joining
The level of differentiation between Mesoamerican wild and cultivated *C. cujete* was relatively low ($F_{ST} = 0.35, I_{G95\%} = 0.13–0.60$). The differentiation between Mesoamerican wild and Amazonian cultivated *C. cujete* samples was lowest with the Negro River ($F_{ST} = 0.45, I_{G95\%} = 0.27–0.65$), followed by the Amazonas ($F_{ST} = 0.50, I_{G95\%} = 0.37–0.64$), the Solimões ($F_{ST} = 0.52, I_{G95\%} = 0.40–0.68$), Madeira ($F_{ST} = 0.57, I_{G95\%} = 0.39–0.81$), and Branco ($F_{ST} = 0.61, I_{G95\%} = 0.45–0.82$). The wild *C. amazonica* samples showed high differentiation compared with cultivated *C. cujete* samples ($F_{ST} = 0.57, I_{G95\%} = 0.37–0.64$).

We performed another Structure analysis with the cultivated samples, using only plants whose membership probability was higher than 0.6 in the cultivated cluster (Figure 2A). The two allele frequency models showed similar patterns, with better defined clusters using the correlated model (Supplementary Figures S3, S4). Again, Evanno et al. (2005) $\Delta K$ suggested that two clusters are the most likely structure ($K = 2$, Supplementary Figure S3); these distinguished Mexican from Brazilian samples, with considerable admixture widely distributed in the Amazon Basin (Figure 3A). Evanno et al.’s $\Delta K$ suggested decreasing likelihood of structure up to four clusters ($K = 3$ and $K = 4$), although the fourth cluster did not show a pattern that was clearly different from $K = 3$. At $K = 3$, Mexican and Brazilian samples showed strong admixture (green and yellow clusters in Figure 3A). The green cluster membership was found in Mexico, but was higher along the Negro River and upper sections of the Branco River, with decreasing membership along the Solimões, Amazonas, and Madeira rivers (Figure 3A). In contrast, the third yellow cluster, also found in Mexico, was predominant along the Amazonas and Madeira rivers, scattered along the Solimões, but also high in the middle Negro River (Figure 3A). The neighbor-joining tree differentiated two groups within Amazonia that are both genetically different from Mexico (Figure 3B). However, the differentiation between Amazonia and Mexico is modest ($F_{ST} = 0.04, I_{C95\%} = 0.006–0.08$). Spatial interpolation of the Structure clusters highlights that, although the admixture between Mexico and Amazonia (Figure 3A), genetic similarity is higher between Mexican samples and northwestern Amazonia (Figure 3C). The spatial interpolation also reveals the wide genetic homogeneity of cultivated *C. cujete* across Amazonia, except for the genetic differentiation in the Northwest and East, which is free from local wild-admixture effect in this data set (Figure 3C). The Northwestern and Eastern regions are relatively similar (Figure 3C), which agrees with the distribution of the Eastern yellow cluster up to the middle Negro River (Figure 3A). As expected, the Structure clusters in Amazonia without the Mexican samples show similar spatial interpolation pattern (Supplementary Figure S5).

### Geographic Patterns of Chloroplast Diversity

The haplotype network showed three distinct groups: *C. amazonica*, wild Mesoamerican *C. cujete*, and cultivated *C. cujete* from Brazil and Mexico (Figure 4A). The wild Mexican *C. cujete* lineage is more distant from cultivated *C. cujete* (55 substitutions + 12 substitutions) than is wild *C. amazonica* (39 + 12 substitutions). In the cultivated *C. cujete* group, five common haplotypes were identified, among which four are very close to each other (1 and 2 substitutions) at the core of the cultivated haplogroup (H1, H2, H3, H4). Haplotype H5 is differentiated by at least four substitutions from the core of the network. Divergent cultivated *C. cujete* samples from the Amazon basin were arranged in the extreme branches of the *C. cujete* group in the haplotype network (Figure 4A); the highest number of substitutions (36 and 75) was comparable to the differentiation between the wild and cultivated groups.

The most common haplotype in the Amazon basin (H1) was widely dispersed, but not found in Mexico. Mexico and...
Brazil shared haplotypes H2 and H3, which, although different by only one substitution, showed slightly different distributions in the Amazon Basin (Figure 4B). Haplotype H2, the most common in Mexico, is restricted to the western half of Brazilian Amazonia, with higher frequency in the Northwest. Haplotype H3 is unevenly distributed in the Amazon Basin, but absent in the Northwest. Haplotype H4 is widely distributed, whereas haplotype H5, the most divergent haplotype (Figure 4A), is less abundant and found at low frequencies along the middle Negro River, but is well-represented in Eastern Amazonia. The most divergent rare haplotypes (H6, H10, H11) agree with the geographical distribution of the haplotype H5. The other rare haplotypes (H7, H8, H9) were sparsely distributed along the Solimões and Madeira rivers, except the haplotype H12 shared between Madeira and Branco River and the haplotype H13, restricted to the upper sections of Negro and Solimões Rivers (Figure 4B). None of the Amazonian rivers were significantly divergent from Mexico (Table 1), certainly because of the small sample size from Mexico. Within the Amazon Basin, the Amazonas River is the most differentiated from all other rivers (Table 1).

**Genetic Diversity in Cultivated C. cujete**

Based on 8 nSSR of cultivated samples, there were 31 alleles in Mexico and 55 in the Amazon Basin (Table 2), although the sample sizes of the two regions are very different. The number of private alleles among cultivated samples showed that seven alleles were only found in cultivated Mexican samples and 31 alleles in cultivated Amazonian samples (Table 2), among which six are also found in wild Mesoamerican samples. Among Amazonian samples, the Amazonas River concentrated private alleles (5) not found in local wild C. amazonica. The Negro, Solimões and Madeira rivers had fewer private alleles, while none was found in the Branco River (Table 2). Mexico presented the highest expected heterozygosity ($H_e$). In the Amazon Basin, heterozygosity was highest along the Negro River, followed by the Solimões, Amazonas, Madeira rivers, and was lowest along the Branco River (Table 2). Mexico presented significant inbreeding, while in the Amazon Basin inbreeding was significant along the Branco and Madeira rivers (Table 2).

Among the 334 SNPs found in chloroplast sequences, 206 were found in cultivated C. cujete. Mexico and the Amazon Basin showed similar nucleotide diversity ($\pi$), $3.78 \times 10^{-2}$ and $3.83 \times 10^{-2}$, respectively, although sample sizes are very different and the Amazon Basin harbors highly divergent samples (Figure 4A). Among cultivated C. cujete, 15 samples produced 119 unique SNP alleles, of which 66 % were from only two samples collected along the Amazonas River, which thus produced an extremely high nucleotide diversity estimate for this river ($\pi = 9.31 \times 10^{-2}$). When these 15 singleton samples were discarded, there were 93 SNPs and nucleotide diversity in Mexico was still similar to the Amazon Basin (Table 2). The highest nucleotide diversity was still along the Amazonas River, with decreasing values along the Solimões, Madeira, Negro, and Branco rivers (Table 2).

**Morphological Diversity of Cultivated C. cujete**

We identified a total of eight fruit shapes in the Amazon Basin and five in Mexico (Figure 5A). Fruit shapes shared among these regions were spherical, flattened, oblong, elongated, and cuneate, with higher frequencies of spherical, flattened, and...
oblong shapes in both regions. Three types (globular, rounded-drop, and oblong-drop) were only recorded in the Amazon Basin. The kidney-shaped fruit found in Colombia was not found in Mexico or Brazilian Amazonia. The absence of drop-shaped fruits in Mexico, which are types clearly distinguished from the others, indicates higher morphological diversity along Amazonian rivers than in Mexico. The Solimões River harbors all the eight fruit shapes described (Figure 5B). The spherical shape, the most frequent in Mexico, is relatively rare in the Amazon Basin, with a higher frequency along the Amazonas River (Figure 5B). The fruit types absent in Mexico were rare in the Amazon Basin as well, except the rounded-drop shape. This fruit type showed relatively high frequency along the Negro River, more than the more common flattened and oblong shapes (Figure 5B). The fruit shape diversity index was higher along the Negro River, with decreasing values along the Solimões and Amazonas rivers, followed by Mexico, and lowest along the Madeira and Branco rivers (Table 2). The fruit shape diversity index was not
correlated with any of the genetic estimators (p > 0.05). The fruit diameters showed the lowest average along the Negro River and in Mexico, and the highest along the Madeira River (Table 2). Mexico and the Negro River also showed the extremes of size variation, with Mexico least variable and the Negro most variable (Table 2).

**DISCUSSION**

Cultivated *C. cujete* are quite similar from Mexico to Brazil, suggesting a common genetic origin. But these cultivated types are strongly differentiated from wild types, both from Mexico and Amazonia, suggesting these wild populations are not the direct ancestors of cultivated *C. cujete*. The geographical origin of the domestication of this species is still uncertain. However, the high diversity of cultivated *C. cujete* from Mexico, compared to Amazonia, suggests that its origin may be in Central America. Diversity analyses allowed discussion of the different routes of introduction into Amazonia and subsequent dispersal. More than one route may have been used: a northwestern introduction into the Negro and Solimões Rivers; and an eastern introduction from the coastal Guianas into the Amazonas River. Finally, fruit shape diversity suggests distinct selection pressures across the crop’s distribution.

**Relationships among Mesoamerican and Amazonian Treegourd Populations**

The wild samples from Mexico (taxonomically identified as *C. cujete*) and the Amazon Basin (identified as *C. amazonica*) were strongly differentiated from the cultivated samples, given their FST values based on nuclear SSR and number of substitutions in the chloroplast genome. The high number of substitutions in the chloroplast sequences between these wild taxa suggests ancient divergence. The differentiation between wild and cultivated in Mexico (Aguirre-Dugua et al., 2012; under revision) and between wild and cultivated in Amazonia was already noted Moreira et al., 2017). These results suggest that neither of these wild relatives are the direct ancestor of cultivated *C. cujete*, although Mexican wild samples present clear morphological identification as *C. cujete* based on Gentry (1980) description.

The Costa Rican sample showed an intermediate admixed nuclear pattern, but high chloroplast differentiation from the cultivated samples (Figures 2, 4A). Consequently, it could be a wild individual pollinated by cultivated *C. cujete*. However, because ancestry could also reflect divergence, increased sampling in Central America is of interest. Although our results rule out the possibility that cultivated *C. cujete* was derived from the wild samples from the Yucatan Peninsula, we cannot rule out an origin somewhere between Central America and northern South America, where other potentially wild *C. cujete* populations occur in savannahs (Figure 1). Nevertheless, our results provide evidence that introduction of domesticated *C. cujete* in Mexico and Amazonia originated from the same source, given the Mexican relationship with Amazonian samples (Figure 3A, yellow and green clusters) and occurrence of wild Mesoamerican alleles in cultivated Amazonian *C. cujete* samples.

**Table 1** | Paired FST distance matrix between cultivated *C. cujete* chloroplast sequences (N = 181) based on 93 SNPs from Mexico and five Amazonian rivers.

<table>
<thead>
<tr>
<th>Regions</th>
<th>Mexico</th>
<th>Negro</th>
<th>Branco</th>
<th>Solimões</th>
<th>Amazonas</th>
<th>Madeira</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.09</td>
<td>0.06</td>
<td>0.32</td>
<td>0.18</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>Negro</td>
<td>0.11</td>
<td>0.14</td>
<td>0.41</td>
<td>0.05</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>Branco</td>
<td>0.14</td>
<td>0.03</td>
<td>0.87</td>
<td>0.00</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>Solimões</td>
<td>−0.02</td>
<td>0.00</td>
<td>−0.01</td>
<td>0.00</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>Amazonas</td>
<td>0.07</td>
<td>0.05</td>
<td>0.11</td>
<td>0.08</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Madeira</td>
<td>0.03</td>
<td>0.02</td>
<td>−0.01</td>
<td>0.00</td>
<td>0.14</td>
<td></td>
</tr>
</tbody>
</table>

The FST-values are below the diagonal; in italics above the diagonal the significance evaluated using 1,000 bootstraps at p ≤ 0.05. Significant FST are indicated with bold script. Samples with singletons (see text) were not included.

**Table 2** | Genetic diversity of cultivated *Crescentia cujete* in Mexico and along major rivers of the Brazilian Amazonia, based on 8 nuclear SSR, 93 chloroplast SNPs and eight fruit shapes.

<table>
<thead>
<tr>
<th>Regions</th>
<th>Nuclear diversity</th>
<th>Chloroplast diversity#</th>
<th>Fruit morphology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>A_r</td>
<td>A_r</td>
</tr>
<tr>
<td>Mexico</td>
<td>7</td>
<td>31</td>
<td>3.84</td>
</tr>
<tr>
<td>Amazon Basin</td>
<td>215</td>
<td>55</td>
<td>6.75</td>
</tr>
<tr>
<td>Negro</td>
<td>24</td>
<td>28</td>
<td>3.43</td>
</tr>
<tr>
<td>Branco</td>
<td>17</td>
<td>17</td>
<td>2.05</td>
</tr>
<tr>
<td>Solimões</td>
<td>79</td>
<td>41</td>
<td>5.07</td>
</tr>
<tr>
<td>Amazonas</td>
<td>48</td>
<td>41</td>
<td>5.02</td>
</tr>
<tr>
<td>Madeira</td>
<td>47</td>
<td>32</td>
<td>3.89</td>
</tr>
<tr>
<td>Overall</td>
<td>222</td>
<td>62</td>
<td>7.6</td>
</tr>
<tr>
<td></td>
<td>117</td>
<td>0.49</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.65</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>118</td>
<td>0.56</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>0.47</td>
<td>30</td>
</tr>
</tbody>
</table>

N, number of samples; A_r, total number of alleles; A_p, rarefied allele counts; A_r, observed heterozygosity; H_o, expected gene diversity, mean FST (*significant at p < 0.05 at least at 50 % of loci); S, number of polymorphic sites; h, number of haplotypes; π, nucleotide diversity; H’shape, Shannon index of fruit shape diversity estimated for each region; and D, fruit diameter (average ± SD).

#(samples with singletons were not included [N = 15]).
The patterns of treegourd genetic diversity across the Amazon Basin allow two, not mutually exclusive, hypotheses of introduction: a Northwestern route and an Eastern route. A Northwestern route into the upper Negro River is supported by the relatively high levels of heterozygosity and fruit shape diversity (Table 2), higher proportions of Mexican ancestry (Figure 3A, green cluster) and higher frequency of the most common haplotype in Mexico (Figure 4B, haplotype H2). This route into Negro River is possible from the Orinoco River, given the fluvial connections via de Cassiquiare canal. This route was part of an extensive social trading network (Hornborg, 2005), based at least in part on the Arawak network (Eriksen and Danielsen, 2014). This route has also been suggested for various crop dispersals (Schultes, 1984), such as cocona (Solanum sessiliflorum), whose populations were domesticated in the upper Orinoco River (Volpato et al., 2004) and which was widely cultivated in Northwestern Amazonia (Schultes, 1957). Similarly, people from the upper Negro River reported intentional collection of treegourd propagules from the Cassiquiare, where treegourd is considered a spontaneous tree in the floodplains, while along the Negro River cultivation demands more effort (P.A.M., personal observation).

A possible Western route into the upper Solimões River is partially supported by heterozygosity and fruit diversity (Table 2); the presence of all fruit shapes described enhances the possibility (Figure 5B). Moderately high nucleotide diversity with the highest number of haplotypes are the strongest evidence (Table 2), especially because hybridization with wild populations was not reported (Moreira et al., 2017), suggesting that this is C. cujete diversity. This route might reflect introduction from the Pacific coast and crossing of the Andes mountains via the Napo and Putumayo rivers (Schultes, 1984), as might be the case of cacao (Theobroma cacao) (Thomas et al., 2012) and peach palm (Bactris gasipaes) (Rodrigues et al., 2005) demonstrated by molecular evidence. However, it is also possible that this is a continuation of the Negro River route across interfluvial areas, as suggested by the distribution of abundant haplotype H2 and the rare haplotype H13 (Figure 4B).

The Eastern route into the Amazonas River is supported by high heterozygosity and fruit diversity (Table 2), with high Mexican ancestry not found in Western Amazonia (Figure 3A, yellow cluster). The highest levels of nucleotide diversity (Table 2) and the particular distribution of haplotypes not found in Western Amazonia (Figure 4B, haplotype H5), which include one of the Mexican haplotypes (Figure 4B, H3), agree with the nuclear pattern. This route is linked to the coastal Guianas, an ancient area of exchange of Amazonian crops with Mesoamerica (Schultes, 1984). Molecular data of early maize (Zea mays) introduction into South America support dispersal from Mesoamerica through the Caribbean, spreading along the lowlands of the northeastern coast of South America to finally reach Amazon Basin through river systems (Freitas et al., 2003; Bedoya et al., 2017), although the oldest archaeological remains of maize are western (Bush et al., 2016). This route also agrees with pineapple dispersal from the Guianas, where it was domesticated and introduced into Mexico (Coppens D’Eeckenbrugge and Duval, 2009).

The extremely high chloroplast nucleotide diversity along the Amazonas River, almost twice that along the Solimões River (Table 2), is an unexpected result. Such high diversity was also observed with nuclear markers, given the relatively higher number of exclusive cultivated alleles along the Amazonas River (Table 2), which might not be related to local hybridization,
since they were not found in C. amazonica (Moreira et al., 2017). While nuclear information is limited by the small number of loci analyzed, the chloroplast pattern is robust and they are in agreement. Therefore, we do not rule out that diversity along the Amazonas River might have been promoted by interspecific hybridization between Mesoamerica and northern South America, where most diversity of Crescentia species is found (Gentry, 1980) and hybrid samples might have been introduced into Amazonia. Another process that is complementary and also deserves future investigations is the role of seed cultivation to deal with high flooding described along the Amazonas River (Moreira et al., 2017), since seeds might show diversity not found among cuttings as usually practiced (Arago-Ulloa et al., 2009; Aguirre-Dugua et al., 2012; Moreira et al., 2017). This hypothesis follows that of manioc (Manihot esculenta), where cuttings are usually practiced, but seed propagation is important to maintain diversity (Peroni and Sodero Martins, 2000; Elias et al., 2001; Duputié et al., 2009; McKey et al., 2010).

Hypothesis of Fruit Dispersal and Diversification

Domesticated varieties often present greater fruit shape diversity than their wild relatives, as observed in bottle gourd (L. siceraria), whose fruits have similar technological uses (Heiser, 1993; Morimoto et al., 2005). Across its distribution, the pattern of treegourd fruit shape diversity (Figure 5) suggests different cultural preferences affecting diversification. The highest shape diversity was found along the Negro and Solimões rivers (Figure 5B, Table 2). Similar high diversity was also observed in the Orinoco and Caribbean regions of Colombia (Arago-Ulloa et al., 2009), suggesting northwestern South America is an area of treegourd diversification. This pattern of diversity agrees with Amazonian ethnographies that underscore the cultural value of morphotype diversity cultivated for its own sake, such as in manioc (Rival and McKey, 2008) and pequi (Caryocar brasiliense) (Smith and Fausto, 2016). Nevertheless, the greater local frequency of the spherical type in Mexico and rounded-drop shape along the Negro River (Figure 5) suggests distinct selection pressures, as also described for popcorn in Peru (Groban et al., 2012) and the differential selection of bitter and sweet manioc between Amazonia and the Atlantic Forest in Brazil (Emperaire and Peroni, 2007). Modern Maya people in Mexico and Guatemala have a long history of strong selection of spherical fruits of C. cucute for bowls (jicas) to use with traditional beverages in rituals and also daily life situations (Ventura, 1996; Aguirre-Dugua et al., 2012, 2013). In Amazonia, the spherical and drop-shaped fruits of C. cucute have different symbolic importance and are recognized with distinct names by Tukano Oriental speakers (Pieter van der Veld, pers. communication), a linguistic family found in Northwestern Amazonia. The spherical fruit is called wawatöwë, and is used as bowls to prepare ipadu powder (Erythroxylum coca var. ipadu) in rituals. In contrast, the rounded-drop, called naašâwa, is common in daily life as a spoon and cup for collective food consumption (xibë, a meal of water and manioc flour, and açai, the juice from Euterpe precatoria). Local people along the upper Negro River reported that the spherical type was also used as an ashtray by healers (pajë) in blessing rituals with tobacco smoke. Ethnographies also reported different treegourd fruits for each type of use, such as cuia-de-tapioca and cuia-de-ipadu (Ribeiro, 1995), although shape differences were not mentioned. In Northwestern South America, these bowls are cultural markers for the traditional use of coca introduced from the Andeanfoothills (Plowman, 1984). Interestingly, the spherical fruit type selected in Mexico was the same as the one used in special rituals in Negro River Basin. This suggests that the wide dispersal of plants between South America and Mesoamerica in pre-Columbian times was motivated not essentially by food consumption, as would be expected for agrarian societies, but mainly for recreational and religious purposes (Neves, 2016). Indeed, archaeological remains of C. cucute in Central America and the Antilles were found in ritualistic contexts, such as offerings in funerary rituals (Beaubien, 1993; Conrad et al., 2001). This hypothesis of recreational and religious exchanges is also supported by the ancient dispersal of maize (Zea spp.) for beer preparation and tobacco (Nicotiana spp.) for magic and therapeutic uses, both widely exchanged between these continents (Heiser, 1965; Smalley and Blake, 2003), possibly as sacred gifts (Norton, 2008).

The relatively high morphological diversity found along the Solimões and Amazonas rivers, where most of rare fruit shapes were found (Figure 5B), suggests different demands for fruit shapes since pre-historic times, as expected among plants with technological uses (Blench, 2012). The upper Solimões River and middle Amazonas River were ancient treegourd handicraft centers that were regarded by both Europeans and Native Amazonians as one of the best expressions of their arts and an important article of trade (Rodrigues-Ferreira, 1933; Métraux, 1948). During the colonial period, villages along the Amazonas River produced 5,000–6,000 bowls a year that were exchanged for food (Rodrigues-Ferreira, 1933). This handicraft tradition extends until today, especially for the production of tacaca bowls (a kind of soup), which are made with the rounded (spherical and flattened) fruits (Moreira et al., 2017).

Although there is similarly high biological and cultural complexity in Mesoamerica and Amazonia (Blench, 2012; Clement et al., 2015; Casas et al., 2017), these two plant domestication centers contrast in terms of the morphological diversity of cultivated C. cucute fruits. Curiously, although Mexico pre-history is especially rich in complex societies, such as the Maya (Willey, 1956), morphological fruit diversity is lower and particular fruit shapes are absent, which also reinforces different cultural selection pressures between these regions. It follows that, although the introduction of the cultivated germplasm into both Mexico and Amazonia should lead to a bottleneck (i.e., through founder effect), it might be less severe in Amazonia due to a more diverse array of usages. Moreover, although the spread of a phenotype during dispersal might also be influenced by wild introgression/hybridization (Meyer and Purugganan, 2013), this effect was remarkable only on treegourd fruit size and not on shape diversity in Amazonia (Moreira et al., 2017). Within Mexico, elongated and smaller shapes spontaneously grown in homegardens, resulted possibly from gene flow with wild populations, are not appreciated in Yucatan.
Peninsula (Aguirre-Dugua et al., 2012), but are selected in the Pacific Coast as spoons (X.A.D, personal observation), although at low frequencies (Figure 5A). Therefore, cultural selection influences the bottlenecks during introduction and afterwards the management of hybridization with local wild congeners. Whereas, distribution of shape diversity reflects different culture preferences, size is more influenced by local wild introgression effects.

CONCLUSIONS

We demonstrated with molecular evidence that C. cujete introduced into the Amazon Basin and Mexico shares a common ancestry with a currently unknown origin. The dispersal followed previously proposed routes of human and plant migrations into Amazonia. The patterns of genetic diversity across Amazonia allow two, not mutually exclusive, hypotheses of the routes of introduction: a Northwestern introduction into the Negro and Solimões rivers, and an Eastern introduction from the coastal Guianas into the Amazonas River. The fruit shape diversity reveals different ancient utilitarian demands for the fruits. Mesoamerica and Amazonia have contrasting fruit morphological diversity, which suggests different cultural preferences along treegourd’s dispersal routes. More comparative studies of its different uses, with a broader genetic and phenotypic distribution, would be useful to better understand the dispersal and diversification of C. cujete in the Americas.

REFERENCES


