Supplementary Information File for

Population genetics of *Manihot esculenta* ssp. *flabellifolia* gives insight into past distribution of xeric vegetation in a postulated forest refugium area in northern Amazonia

Anne Duputié[1,2], Marc Delêtre[3], Jean-Jacques de Granville[4], Doyle McKey[1]

Contents

[1]CEFE UMR5175, 1919 Route de Mende, 34293 Montpellier Cedex 5, France
[2]Corresponding author. Email: anne.duputie@cefe.cnrs.fr
[3]Botany Department, School of Natural Sciences, Trinity College Dublin, Dublin 2, Ireland
[4]Unité 84 Biodival, Institut de Recherche pour le Développement, Herbier de Guyane, Route de Montabo, BP165, 97323 Cayenne Cedex, France

Supplementary Table 1.

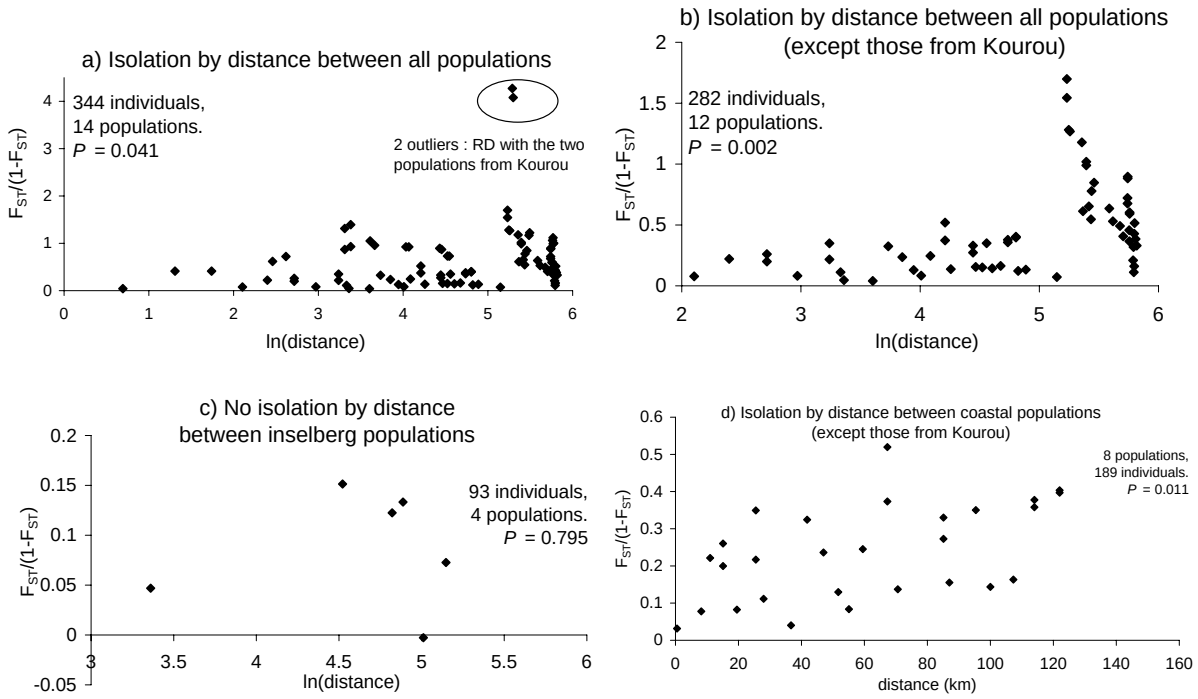
Table 1: Sampling locations for *Manihot esculenta* ssp. *flabellifolia* in French Guiana. Populations marked with an asterisk are the populations where hybridization with domesticated cassava was detected. Numbers in brackets indicate the numbers of hybrids found in each of these populations. In Savane Manuel, where hybridization was already studied (Duputié *et al.*, 2007), only wild individuals were included in the present sampling.

Region	Site	Label	Sampled individuals	Coordinates		Collector name	Date of collection
Coast	**279**						
	Savane Grand Macoua	GM	33	05°32'N	053°24'W	*Guillaume Léotard*	2003
	Savane Mammaribo	MB*	9 (3)	05°31'N	053°21'W	*Guillaume Léotard*	2003
	Savane Trou Poissons	TP	19	05°24'N	053°07'W	*Guillaume Léotard*	2003
	Savane Manuel	SM*	8	05°22'N	052°57'W	*Guillaume Léotard*	2003
	Kourou	CC	29	05°10'N	052°40'W	*Doyle McKey*	2003
		KP	33	05°10'N	052°40'W	*Doyle McKey*	2003
	Savane des Pères	SP	29	05°07'N	052°40'W	*Doyle McKey*	2004
	Savane Matiti	MT[1]	20	05°03'N	052°35'W	*Anne Duputié*	2006
	Tonate	T1*	40 (13)	04°59'N	052°28'W	*Benoît Pujol*	2002
		T2*	60 (13)	04°59'N	052°28'W	*Benoît Pujol*	2002
Inland	**93**						
	Marouini (4 inselbergs)	MA	13	02°36'N	054°00'N	*Jean-Jacques de Granville*	2002
	Roche Touatou (2 inselbergs)	RT	44	02°57'N	052°32'W	*Doyle McKey*	2002
	Wanapi (3 inselbergs)	WA	31	02°31'N	053°48'W	*Jean-Jacques de Granville*	2002
	Roche Dachine (1 inselberg)	RD	5	03°28'N	053°13'W	*Jean-Jacques de Granville*	2002

[1] Hybridization was detected (Duputié et al., in prep.), but only wild individuals were included in this study

Supplementary Figure 1.

Figure 1: Isolation by distance between French Guianan populations of *Manihot esculenta* ssp. *flabellifolia*.
Upper left panel: all populations included (significant IBD with $P = 0.041$).
Upper right panel: isolation by distance between all populations except those from Kourou (significant IBD with $P = 0.002$).
Lower left panel: no isolation by distance among inselberg populations only (no significant IBD, $P = 0.795$).
Lower right panel: isolation by distance between coastal populations (except those from Kourou). IBD is significant with $P = 0.011$.



Analysis of population structure, without the populations where introgressed individuals were found.

The populations included in these analyses are: MA, RD, RT, WA, GM, TP, CC, KP and SP and include 236 individuals.

A total of 32 alleles (3 - 7 per locus) were encountered. Overall, there was a strong deficit of heterozygotes, with f = 0.199 (95 % confidence interval: [0.091 - 0.315], with five of the nine populations showing significant values of F_{IS} (MA, RT, WA, GM, SP).

Population differentiation was high: θ = 0.357 (95 % confidence interval: [0.227 - 0.448]). Isolation by distance was not significant (regression of $F_{ST}/(1-F_{ST})$ with $ln(distance)$, Mantel test after 10,000 permutations, P = 0.064), but was significant once the two populations from Kourou were removed (regression of $F_{ST}/(1-F_{ST})$ with $ln(distance)$, Mantel test after 10,000 permutations, P = 0.040).

Bayesian clustering of the populations led to the formation of three clusters (not four, as when introgressed populations were included). The "missing" cluster is the one gathering the two populations from Tonate (not included in this sampling). Individual assignment to each cluster is presented on Supplementary Figure 2. Individuals from the inselbergs form a first cluster; a second one is formed by the populations from Kourou, a third one by the populations west of Kourou, and population SP is of mixed ancestry between those two last clusters.

Six of the 32 alleles were private to inselberg populations and six to coastal populations.

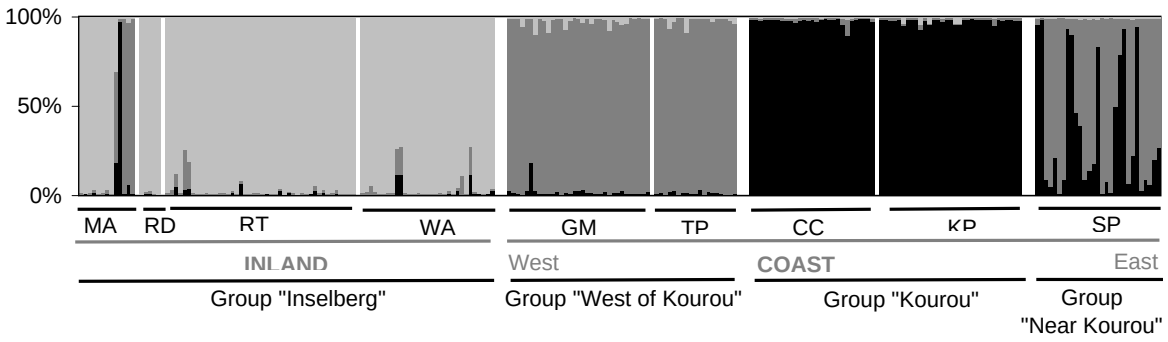


Figure 2: Proportion of the genome of each individual assigned to each of the three clusters. Each individual is represented by a vertical bar.

Conclusions Removing introgressed populations does not change the main conclusions of the paper:

- coastal populations are strongly differentiated from inselberg populations
- inselberg populations are not highly differentiated
- coastal populations form different genetic groups, supporting founder effects through bottlenecks.

Analysis of population structure, without the small populations (N $<$ 19).

The populations included in these analyses are: CC, GM, KP, MT, SP, T1, T2, RT, TP, WA and include 312 individuals.

All 36 alleles documented in the main text were present. Overall, there was a strong deficit of heterozygotes, with f = 0.167 (95 % confidence interval: [0.091 - 0.271], with five of the ten populations showing significant values of F_{IS} (RT, WA, GM, SP, T2).

Population differentiation was high: θ = 0.373 (95 % confidence interval: [0.277 - 0.441]). Isolation by distance was significant at the 5 % level (regression of $F_{ST}/(1-F_{ST})$ with $ln(distance)$, Mantel test after 10,000 permutations, P = 0.048), and even more significant when removing the two populations from Kourou (regression of $F_{ST}/(1-F_{ST})$ with $ln(distance)$, Mantel test after 10,000 permutations, P = =0.010).

Bayesian clustering of the populations led to the formation of four clusters, the same as described in the main text of the manuscript. Individual assignment to each cluster is presented on Supplementary Figure 3. As in the main text, individuals from SP and MT were found to be of admixed ancestry between the three clusters of individuals from the coast.

Five of the 32 alleles were private to inselberg populations and eight to coastal populations.

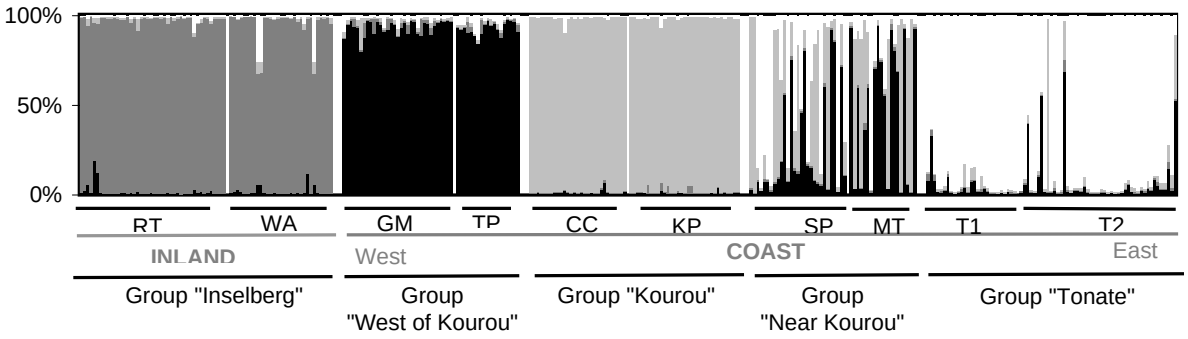


Figure 3: Proportion of the genome of each individual assigned to each of the four clusters. Small populations (N $<$ 19) were removed. Each individual is represented by a vertical bar.

Conclusions Removing the small populations does not change our conclusions either.

Null allele quantification.

Because the primers for the microsatellites we used were designed for cassava, and not for its wild relative, null alleles may be encountered.

Examination of the control wells in the PCR plates led to the conclusion that, if no discrepancy between two amplifications of the same sample were observed, some loci often showed a lack of amplification in one of the trials. Therefore, a number of the observed double nulls are, in fact, individuals for which unconspicuous peaks were observed: they were not truly double nulls, but suffered a technical problem for amplification. This lack of amplification was observed only in the locus showing the longest alleles: SSR68. We removed the individuals showing a double null genotype at this locus (24 individuals) and computed the expected frequency of null alleles in the remaining individuals, using the algorithm of Dempster *et al.* (1977), as implemented in FREENA (Chapuis & Estoup, 2007). Unfortunately, this method, like the other methods dedicated to estimating null allele frequencies, is based on the hypothesis that the populations are at Hardy-Weinberg equilibrium, which is false in our case.

Average null allele frequency was estimated to 5.2 %, with the highest frequency of null alleles found at locus GA21 (9.7 %). Null allele frequency ranged between 3.0 and 5.2 % for all other loci (Table 2). F_{IT} was very high for all loci (0.36 - 0.54), but was not highest for locus GA21 (Table 2).

Table 2: Estimation of F_{IT} and null allele frequencies for each locus.

locus	F_{IT}	estimated frequency of null alleles
GA12	0,360	0.052
GA126	0.401	0.049
GA21	0.544	0.097
SSR169	0.487	0.030
SSR55	0.537	0.045
SSR68	0.448	0.039

When locus GA21 was removed from the analysis, there was a strong deficit of heterozygotes, with $f = 0.183$ (95 % confidence interval: [0.106 - 0.296].Population differentiation was high: $\theta = 0.382$ (95 % confidence interval: [0.290 - 0.437]).

Isolation by distance was significant at the 5 % level (regression of $F_{ST}/(1 - F_{ST})$ with $ln(distance)$, Mantel test after 10,000 permutations, $P = 0.031$), and even more significant when removing the two populations from Kourou (regression of $F_{ST}/(1 - F_{ST})$ with $ln(distance)$, Mantel test after 10,000 permutations, $P < 0.001$).

Bayesian clustering of the populations led to the formation of the four clusters described in the main text. Individual assignment to each cluster is presented on Supplementary Figure 4. As in the main text, individuals from SP and MT were found to be of admixed ancestry between the three clusters of individuals from the coast.

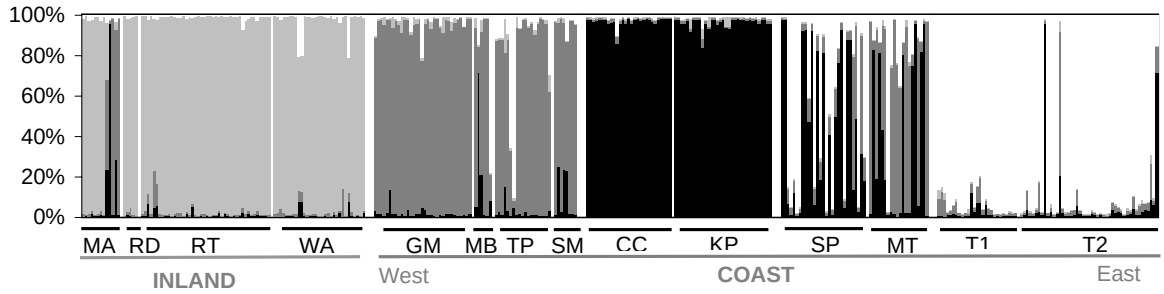


Figure 4: Proportion of the genome of each individual assigned to each of the four clusters. Each individual is represented by a vertical bar.

Conclusions Removing the locus exhibiting the highest frequency of null alleles does not modify the conclusions of the manuscript.

References

Chapuis M, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621–631.

Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B*, **39**, 1–38.

Duputié A, David P, Debain C, McKey D (2007) Natural hybridization between a clonally propagated crop, cassava (*Manihot esculenta* Crantz) and a wild relative in French Guiana. *Molecular Ecology*, **16**, 3025–3038.