

Natural hybridization between a clonally propagated crop, cassava (*Manihot esculenta* Crantz) and a wild relative in French Guiana

ANNE DUPUTIÉ, PATRICE DAVID, CHANTAL DEBAIN and DOYLE MCKEY

Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175 CNRS, 1919 Route de Mende, 34293 Montpellier cedex 5, France

Abstract

Because domestication rarely leads to speciation, domesticated populations often hybridize with wild relatives when they occur in close proximity. Little work has focused on this question in clonally propagated crops. If selection on the capacity for sexual reproduction has been relaxed, these crops would not be expected to hybridize with their wild relatives as frequently as seed-propagated crops. Cassava is one of the most important clonally propagated plants in tropical agriculture. Gene flow between cassava and wild relatives has often been postulated, but never demonstrated in nature. We studied a population of a wild *Manihot* sp. in French Guiana, which was recently in contact with domesticated cassava, and characterized phenotypes (10 morphological traits) and genotypes (six microsatellite loci) of individuals in a transect parallel to the direction of hypothesized gene flow. Wild and domesticated populations were strongly differentiated at microsatellite loci. We identified many hybrids forming a continuum between these two populations, and phenotypic variation was strongly correlated with the degree of hybridization as determined by molecular markers. Analysis of linkage disequilibrium and of the diversity of hybrid pedigrees showed that hybridization has gone on for at least three generations and that no strong barrier prevents admixture of the populations. Hybrids were more heterozygous than either wild or domesticated individuals, and phenotypic comparisons suggested heterosis in vegetative traits. Our results also suggest that this situation is not uncommon, at least in French Guiana, and demonstrate the need for integrated management of wild and domesticated populations even in clonally propagated crops.

Keywords: biosafety, cassava, clonal propagation, morphological variation, natural hybridization, wild relative

Received 9 December 2006; revision received 28 February 2007; accepted 5 March 2007

Introduction

Hybridization is considered to be a major evolutionary process (Arnold & Hodges 1995; Ellstrand *et al.* 1999; Rieseberg *et al.* 2003). It is a primordial issue for many crop species because plant domestication has generally not led to total reproductive isolation of domesticated and wild populations (Ellstrand *et al.* 1999). Hybridization affects the functioning of both kinds of populations (Ellstrand

et al. 1999; Gepts & Papa 2003), and can lower or increase the degree of adaptation in each.

Because domestication often involved bottlenecks and loss of genetic diversity, gene flow from wild to domesticated taxa can be agronomically useful, favouring local adaptation and increasing genetic diversity, thereby lowering risks of crop failure due to variation in biotic or abiotic environmental factors (Jarvis & Hodgkin 1999). The crop's close relatives are thus considered important genetic resources, as a reservoir of diversity from which breeding programs can introgress traits into recipient cultivars.

On the other hand, introgression from domesticated plants may lower the fitness and jeopardize the persistence

Correspondence: Anne Duputié, Fax: +33 4 67 41 21 38; E-mail: anne.duputie@ens-lyon.org

of wild populations often already threatened by habitat loss (e.g. Ellstrand *et al.* 1999). Introgression sometimes results in genetic swamping of the wild taxon or even in its extinction by hybridization (Gepts & Papa 2003).

Whatever the direction of gene flow, hybrid fitness can differ substantially from that of parental taxa. Where parental populations show very divergent adaptations, reduced hybrid fitness can reinforce reproductive isolation (Barton & Hewitt 1985). Alternatively, hybrids can show heterosis, that is increased vigour, and sometimes become invasive weeds (reviewed by Ellstrand *et al.* 1999). They may also be fitter than their parents in some environments, which can ultimately lead to hybrid speciation (e.g. Rieseberg *et al.* 2003). Hybrid vigour depends on both genetic (Arnold & Hodges 1995) and environmental factors (Rieseberg *et al.* 2003).

Joint management of crops and their wild relatives thus appears to be a necessity in primary centres of diversity. Achieving it requires prior knowledge of their ability to hybridize in nature, and characterization of the environments in which introgression can occur. With the notable exceptions of potato (Celis *et al.* 2004) and yam (Scarcelli *et al.* 2006), relatively little attention appears to have been focused on this theme in clonally propagated crops. Yet, these may differ from sexually propagated crops because selection may have favoured traits that result in reduced gene flow from or towards their wild relatives. First, clonal propagation is expected to relax selection on traits involved in sexual reproduction (seeds and pollen) and could therefore decrease opportunities for pollen or seed flow. However, clonal propagation may not exclude the occasional or regular incorporation of individuals from seeds (and thus potential hybrids). The importance of sexual reproduction in populations of these plants has often been underestimated (e.g. Pujol *et al.* 2005a; Pujol & McKey 2006). Second, clonally propagated plants benefit from fixed favourable gene combinations. Hence, crop-wild hybrids might have lower agronomic interest, and thus lower fitness in domesticated populations, than established clones. There are thus many open questions about the extent to which clonally propagated crops hybridize with their wild relatives, and the effects of hybridization on vigour and fitness. Potato (Celis *et al.* 2004) and yam (Scarcelli *et al.* 2006) can hybridize in the field with several wild relatives, but neither F_1 vigour nor F_1 fertility have been studied.

In this study, we focused on cassava (*Manihot esculenta* Crantz), a crop propagated clonally by stem cuttings. Its domestication syndrome includes the thickening of stems, achieved in part through reduction in the extent of branching (Jennings 1995). As in all *Manihot* species, branching is strictly associated with flowering (e.g. Médard 1973); clonal propagation and sexual reproduction are thus in direct competition (Jennings 1995). Cassava therefore

typifies a crop where clonal propagation may have resulted in decreased opportunity for hybridization with wild relatives.

This view was recently challenged when we identified, on the basis of morphological traits, a putative hybrid zone between cassava and one of its wild relatives in French Guiana. Using data from six microsatellite markers, we assess the reality of hybridization, evaluate whether the phenomenon is frequent, whether it is ephemeral or long lasting, and discuss conservation issues arising from our results. The question of crop/wild gene flow acquires an additional interesting dimension as some field trials of transgenic cassava varieties have already been conducted, and others are planned.

Materials and methods

Cassava biology/study model

The genus *Manihot*, last revised over 30 years ago (Rogers & Appan 1973), includes 98 currently recognized species, although this number should probably be revised downward (Second *et al.* 1997). Domesticated cassava was already present in Panama 5000–7000 years ago (Piperno *et al.* 2000) but the plant was probably domesticated even earlier, as genetic data tend to show that the centre of origin of the crop lies along the southern rim of the Amazon basin (Olsen & Schaal 1999; Léotard 2003; Olsen 2004). It is still unclear whether cassava is a 'compilospecies' derived from multiple hybridization events (Rogers & Appan 1973; Nassar 2003), or has a single ancestor, *Manihot esculenta* ssp. *flabellifolia* (Pohl) Ciferri, the latter hypothesis being better supported by currently available molecular evidence (Olsen & Schaal 1999; Elias *et al.* 2000; Olsen 2004). The role of postdomestication gene flow in either direction is still unclear (Fregene *et al.* 1994).

Work by plant breeders has shown that cassava can be crossed with numerous wild congeners (Jennings 1959; Wanyera *et al.* 1994; Nassar 2003). There is no evidence for cytogenetic barriers, as all species studied so far share $2n = 36$ (Rogers & Appan 1973). Also, there is no evidence for strong phenological barriers. Growth, branching, and hence flowering, are all spread over several months, so that flowering periods of cassava and its wild relatives overlap broadly. Nevertheless, cassava shows a lower degree of flowering than its wild relatives, as expected for a clonally propagated crop.

Cassava has inherited from its wild ancestor(s) a set of adaptations to the periodically disturbed environments they typically inhabit (Pujol *et al.* 2005b). These adaptations continue to function in slash-and-burn agroecosystems. Seeds are buried by ants and remain dormant when a parcel is left to fallow; they germinate when the vegetation cover of the parcel is slashed and burned (Elias *et al.* 2001; Pujol

et al. 2005b). In traditional Amerindian farming systems, farmers regularly incorporate seedlings, injecting new products of sex into their stocks of clones (Elias *et al.* 2001; Pujol & McKey 2006).

Cassava has long been cultivated in French Guiana. In historical times, it has usually been grown in shifting-cultivation settings in forest clearings. In such settings, it is unlikely to be in proximity to its close wild relatives, which are found in savannas. Recently, however, small opportunistic plantings of a few 'sticks' of cassava are seen with increasing frequency in savanna settings, for example along drainage ditches or irrigation canals, where they sometimes are near populations of a wild relative.

This wild relative, widespread in the Guianas, is locally abundant in coastal savannas of French Guiana and also occurs in savanna islands ('savanes-roches') on inselbergs in the densely forested southern part of this overseas French department. Its taxonomic identity is not yet clear. Allem *et al.* (2001) considered it to belong to *M. esculenta* ssp. *flabellifolia* Pohl, but data on both microsatellite

markers and *G3pdh* haplotypes (Léotard 2003) place it closer to *Manihot pruinosa* Pohl. However, its morphology does not match that of specimens of *M. pruinosa* nor of *M. esculenta* ssp. *flabellifolia* that we have examined in several herbaria. Natural clonal propagation has not been reported for this species.

Study site

The study was conducted in Savane Manuel (05°21.804' N, 52°56.800' W), near the town of Sinnamary, French Guiana. This site was first noticed during a visit in 2001. At that time, several cassava plants were growing along a drainage ditch, at the northern edge of the area chosen for a study transect (Fig. 1), adjacent to a population of wild *Manihot*. Very near the drainage ditch, mixed with wild individuals, were a few spontaneously growing individuals of apparently intermediate phenotype (D. McKey, personal observation). Two years later, many more phenotypically intermediate individuals were present and a few

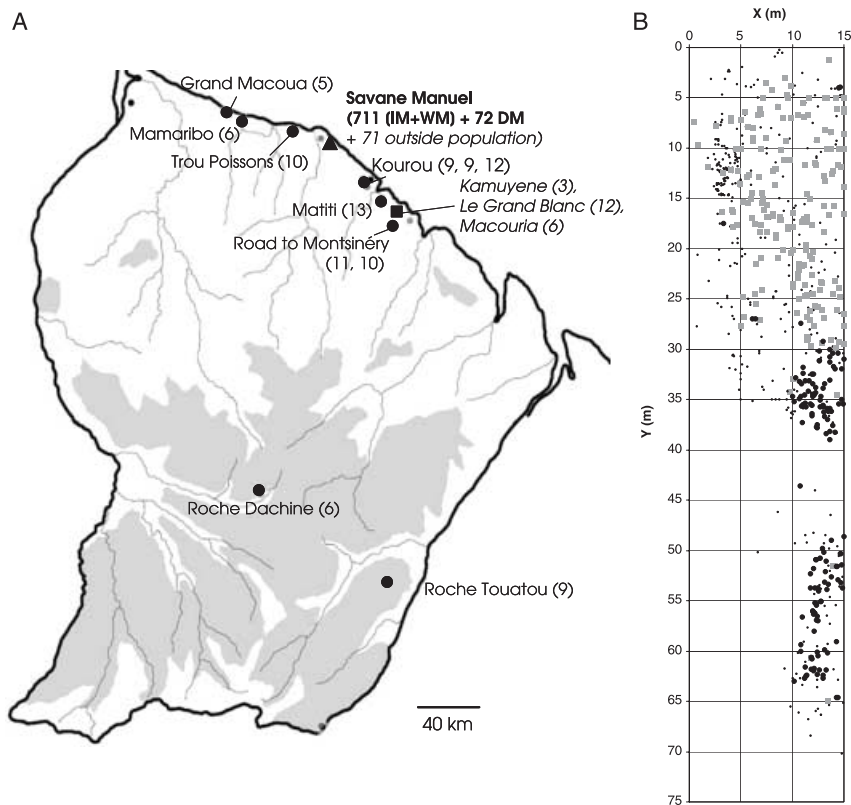


Fig. 1 Localization of the populations used in this study. (a) Map of French Guiana, showing the sites where allopatric wild individuals were sampled (black circles). Number of individuals at each location appears between parentheses (note that sometimes populations are very close to each other, e.g. the three populations close to Kourou). Savane Manuel is indicated by the triangle; the 72 domesticated individuals used in the analysis were sampled in the closest fields (within 3 km). The other small hybrid zones are indicated by the black square. (b) Map of the systematically sampled transect in Savane Manuel. Grey squares, IM individuals; black circles, WM individuals; dots, juveniles. DM individuals had been observed a few metres north (top of the map) of the sampled transect, but had been removed by the time of sampling.

domesticated plants still remained. Sequences of the single-copy nuclear gene *G3pdh* were obtained for a small number of domesticated, wild, and phenotypically intermediate individuals, and indicated introgression of domesticated cassava into the wild population was likely (Léotard 2003).

Sampling

Sampling was conducted in February 2004 along a 75-m long, 15-m wide transect parallel to the hypothesized direction of gene flow between domesticated and wild individuals. Within each 5 × 5-m quadrat, we mapped and collected a few young leaves from all sexually mature plants and 10 randomly chosen juveniles (Fig. 1). We defined sexually mature plants as those that had branched at least once (Médard 1973). Overall, 711 individuals (468 adults and 243 juveniles) were sampled in Savane Manuel.

Because the domesticated plants supposed to be at the origin of the hybridization events had been removed since February 2003, we sampled domesticated cassava plants in a few farmers' fields near Savane Manuel, in order to determine the alleles present in the local domesticated pool. We thus collected the greatest possible diversity of varieties (landraces). Each variety is phenotypically distinct and genetically differentiated from others, and corresponds to a clone or (more often) a set of clones (Elias *et al.* 2000, 2001). Farmers recognize varieties as different entities and manage them as such, purposely planting various assemblages of varieties within each field. A total of 69 cassava accessions, belonging to 27 varieties, were collected from the fields of four farmers.

To assess whether 'wild' phenotype individuals from Savane Manuel were truly wild, we included in the analysis 100 wild individuals from populations where domesticated cassava was absent and no hybridization was suspected. They belonged to 12 populations from the coastal area (from 90 km south to 60 km north of Sinnamary) and two inselberg populations. Four to 10 individuals were genotyped in each population.

Collection of morphological data

For each adult individual sampled, we recorded data for the following morphological traits: number of stems from the rootstock, basal diameter of each stem, total height of the plant, height at first branching, number of leaf lobes, length and width of the central lobe of the largest leaf of the plant, petiole colour (four classes: green, pink, red, purple), stipule colour (four classes: green, pink to red, brownish, and purple) and stipule persistence (coded in four classes from highly deciduous to highly persistent). In addition, all individuals were assigned to one of four morphotypes: domesticated (DM) for the plants found in cultivation, and

wild (WM), intermediate (IM) and juvenile (JM) for those from the transect, as judged by visual assessment of the overall phenotype ('Gestalt') in the field. Wild or domesticated status cannot be judged from morphology in juvenile individuals.

Morphological data were not collected for the wild individuals outside Savane Manuel.

Genotyping

All individuals were genotyped for six microsatellite loci: GA12, GA57, GA126 and GA127 (Chavarriaga-Aguirre *et al.* 1998), and SSR55 and SSR68 (Mba *et al.* 2001). All of these are on different linkage groups (Mba *et al.* 2001). The allopatric wild individuals were genotyped at all but one locus (GA127).

DNA was extracted using DNeasy 96 Plant kit (QIAGEN GmbH). Polymerase chain reaction (PCR) amplification was performed in a multiplex reaction, in a final volume of 10 µL, using 3 µL of diluted DNA in each reaction. Amplification was conducted over 30 amplification cycles with an annealing temperature of 55 °C.

Locus GA127 was amplified apart from the others, in a final volume of 12 µL, using 0.2 U *Taq* polymerase, 2.5 µM MgCl₂, 125 µM each nucleotide, 0.4 µM of each primer (all products from Eurogentec). Annealing took place at 45 °C.

All amplification reactions were performed in a PTC-100 thermocycler (MJ Research) and genotyping was conducted on two sequencers: an ABI 310 and an ABI 3100, after testing for calibration of allele length. Genotypes were then analysed using GENESCAN ANALYSIS version 3.1.2 and GENOTYPER 2.5 (all products, instruments and software from Applied Biosystems).

Analysis of population genetic structure

Allelic richness, pairwise linkage disequilibria, R_{ST} , f and θ were calculated and tested using FSTAT version 2.9.3.2 (Goudet 1995). R_{ST} is a between-population differentiation coefficient, analogous to F_{ST} , which takes into account allelic sizes of microsatellites (Slatkin 1995). f and θ are unbiased estimates of F_{IS} and F_{ST} (respectively), derived by Weir & Cockerham (1984). Observed and expected heterozygosities were computed using GENEPOP version 1.2 (Raymond & Rousset 1995).

First, we determined whether WM individuals were representative of the wild taxon. We used the STRUCTURE software (version 2.1; Pritchard *et al.* 2000), which relies on a Bayesian method that clusters similar multilocus genotypes. We used the 'admixture' option, which infers, for each individual, the proportion of its genes coming from each cluster. Setting the number of clusters to two, this procedure is theoretically relevant to the study of hybridization, as it allows, without prior information, to identify two

Morphotype	Run 1		Run 2		Run 3		
	A	B	A	B	A	B	C
DM	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%	0.0%
WM	0.7%	66.9%	0.7%	64.8%	0.7%	81.7%	0.0%
Allopatric wild	2.0%	89.0%	—	—	—	—	—
IM	38.6%	5.8%	21.7%	3.7%	0.0%	3.2%	7.4%
JM	15.8%	33.3%	11.3%	30.6%	0.0%	25.3%	3.8%

Table 1 Proportion of individuals having more than 90% ancestry in one cluster, in three different STRUCTURE runs. Run 1, $K = 2$, no prior information, includes allopatric wild individuals (and thus only five loci). Run 2, $K = 2$, no prior information, allopatric wild individuals not included. Run 3, $K = 3$, prior population information for WM and DM individuals

sets of allele frequencies typical of parental populations, and the hybrid index within each individual. We first ran it including all individuals from Savane Manuel (IM, JM and WM), DM and the allopatric wild individuals (Table 1, run 1), using data for the five loci that had been sampled for all populations. As WM individuals clustered with the allopatric wild samples, meaning that WM individuals were representative of the wild taxon, we then dropped the allopatric samples from further analyses, in order to use data from all six loci. We then ran the program allowing for the formation of two clusters (Table 1, run 2). Finally, we ran the program allowing three clusters (Table 1, run 3), to evaluate the possibility that intermediate morphotypes are a completely different group, rather than being hybrids. Each run comprised a burn-in period of 200 000 iterations, followed by 10^6 MCMC steps, and was repeated two further times to confirm the repeatability of the result. We considered that an individual was 'pure X' when its inferred proportion of genes belonging to cluster X exceeded 90%. With $K = 2$, all domesticated individuals were inferred to be 'pure A' and most of the WM 'pure B' (see Results). Allelic frequencies were very different in groups A and B.

Population genetic differentiation was illustrated by a factorial correspondence analysis (FCA), which was conducted using GENETIX version 4.03 (Belkhir *et al.* 2002). This analysis discriminated wild and domesticated individuals (see Fig. 2). Alleles were clustered in two groups according to their PC1 coordinate. We defined two synthetic alleles: synthetic allele W, characteristic of population WM, and synthetic allele D, characteristic of DM individuals. Synthetic alleles D and W gathered all alleles with $PC1 > 0$ and $PC1 < 0$, respectively. They corresponded to the alleles inferred by STRUCTURE to be characteristic of clusters A and B, respectively. For each individual, hybrid index was defined as its proportion of W synthetic allele.

Synthetic alleles were further used in computing between-genome ($\kappa_{1,1}$, deviation from Hardy–Weinberg equilibrium) and within-genome ($\kappa_{0,2}$, linkage disequilibrium) pairwise associations, following the method described by Barton (2000), using MATHEMATICA 3.0 (Wolfram Research) add-ons provided by Barton (<http://helios.bto.ed.ac.uk/evolgen/barton/index.html>) and modified for six loci. This method estimates multilocus linkage disequilibrium

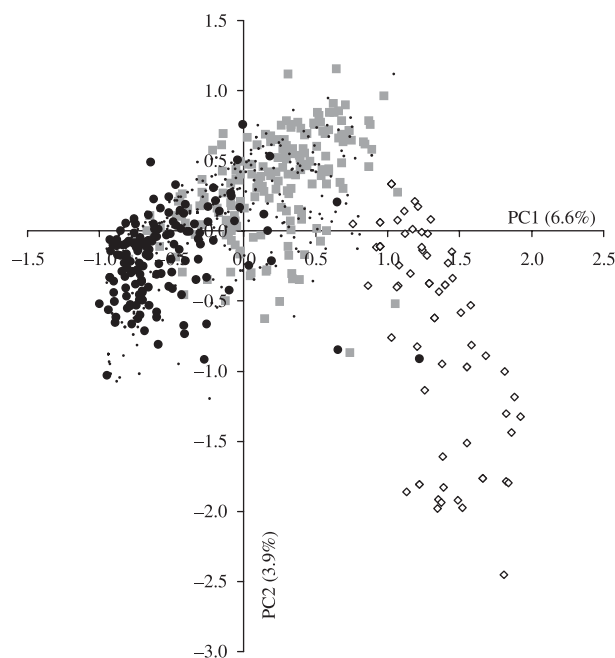


Fig. 2 Factorial correspondence analysis on the matrix of allele counts per morphotype: projection of the individuals in the principal plane. White diamonds, DM individuals; grey squares, IM individuals; black circles, WM individuals; dots, juveniles.

and Hardy–Weinberg disequilibrium in the whole sample. Different nested models were tested for their likelihood, as advised by Barton (2000). We first assumed a model with both Hardy–Weinberg and linkage equilibrium, then models allowing within-genome associations, both between-genome and within-genome associations, and finally also higher-order associations, assuming first equivalent, then different contributions of each locus to the multilocus association. The most likely model was then chosen by applying the likelihood ratio test, and a 95% confidence interval was computed carrying out a random walk algorithm with T initially set to 1, as recommended by Barton (2000). $\kappa_{1,1}$ and $\kappa_{0,2}$ were computed over all six loci and including all 668 individuals. The program was also run for the adult (IM + WM subsamples) and juvenile (JM) subsets.

Table 2 Summary statistics averaged over six loci, for the four morphotypes

	<i>N</i>	<i>A_R</i>	<i>H_O</i>	<i>H_E</i>	<i>f</i>	<i>P</i> value
DM	72	5.91	0.570	0.671	0.152	< 0.001
WM	148	5.71	0.417	0.561	0.253	< 0.001
IM	184	6.12	0.618	0.713	0.134	< 0.001
JM	265	6.11	0.453	0.681	0.335	< 0.001

N, sample size; *A_R*, number of alleles per locus; *H_O*, observed heterozygosity; *H_E*, expected heterozygosity (gene diversity); *f*, Weir & Cockerham's (1984) estimator of *F_{IS}*, the intrapopulation fixation index; *P* value associated with the exact test for deviations from Hardy–Weinberg proportions (computed using GENEPOP).

Analysis of morphological data

We analysed all morphological information, including both quantitative and qualitative variables, through an extended principal components analysis (PCA), as described by Hill & Smith (1976), performed with $\Delta E-4$ (Thioulouse *et al.* 1997). This analysis allows for equal weighting of the qualitative and quantitative variables, whatever the number of modalities used in coding the qualitative variables.

Results

Intrapopulation genetic structure

Allelic diversity was relatively high, with 47 alleles observed over the six loci for the plants sampled in Savane Manuel, and a total of 55 alleles when DM individuals were included. Table 2 summarizes statistics concerning genetic diversity across all six loci in the three morphotypes. Allelic richness was not significantly different among populations. Hardy–Weinberg expectations were not met, with a strong multilocus heterozygote deficit in all four morphological groups. More pairs of loci were in linkage disequilibrium among WM than in DM, IM or JM populations, probably because of a slight spatial structuring of the population (data not shown).

Interpopulation genetic structure

Basic statistics. Differentiation among morphotypes DM, IM and WM was high: $\theta = 0.183$ (0.130–0.262). *R_{ST}* was also high: *R_{ST}* = 0.217, meaning that these three morphotypes were genetically differentiated, not only in terms of allele frequencies, but also of allele sizes.

At each of the six loci, allelic frequencies were very different in each morphotype (Fig. 3). Most alleles were diagnostic or quasi-diagnostic between wild and domesticated

morphotypes (61% of them were present at frequencies < 1% in one or the other of these two groups). Allopatric wild individuals exhibited allelic frequencies different from WM individuals, but the alleles encountered were the same, with the exceptions of allele 193 at locus GA126, being exclusively present in WM and IM individuals from Savane Manuel, and of allele 286 at locus SSR68, encountered only in allopatric wild populations. Some alleles (e.g. allele 140 at locus SSR55 and allele 246 at locus SSR68) are shared between DM and WM populations, but were also present in allopatric populations. Several alleles were present in DM individuals, and in WM at low frequencies, but absent from allopatric wild samples (e.g. alleles 144 and 146 at locus GA12; see Fig. 3). IM individuals bore a mixture of the alleles present in DM and WM, mostly with intermediate frequencies. They did not display a subset of the alleles present in DM, ruling out the hypothesis that they could be escapees from cultivation.

Characterization of WM individuals. Further analysis using STRUCTURE software and including all individuals and all five loci genotyped for the allopatric wild individuals, with no prior population information and allowing the formation of two clusters, led to a clear clustering of DM individuals in cluster A (100% of them were 'pure A') and of allopatric wild individuals, together with WM individuals, in cluster B (89.0% and 66.9% of them, respectively, were 'pure B') (Table 1, run 1). Average posterior probability of belonging to cluster B was 85.9% for WM individuals. Of the juveniles, 33.6% also were 'pure B', but only 5.8% of IM were.

These two analyses showed that WM individuals truly belong to the wild taxon. Given the allelic frequencies of each morphotype, it is likely that a small proportion of individuals classed as WM were introgressed. Ideally, one would expect the same proportion of WM individuals to be classified as 'pure B' as of allopatric individuals. Our results indicate that approximately 20% of individuals classed as WM may have in fact been introgressed. Nevertheless, in further analyses, we excluded individuals from the allopatric wild populations, to use only WM individuals as reference wild individuals. As members of the wild population actually involved in hybridization, they should be a better reference for assessing hybridization in Savane Manuel than are samples from allopatric wild populations. Furthermore, exclusion of the latter populations allowed us to make use of all six loci typed for WM, IM, JM and DM individuals. In any case, using as a reference WM individuals, including some that are likely introgressed, means that our estimations of the extent of hybridization are conservative.

Genetic assignment using STRUCTURE 2.1. Results of further analyses, conducted for the 668 individuals typed for all six

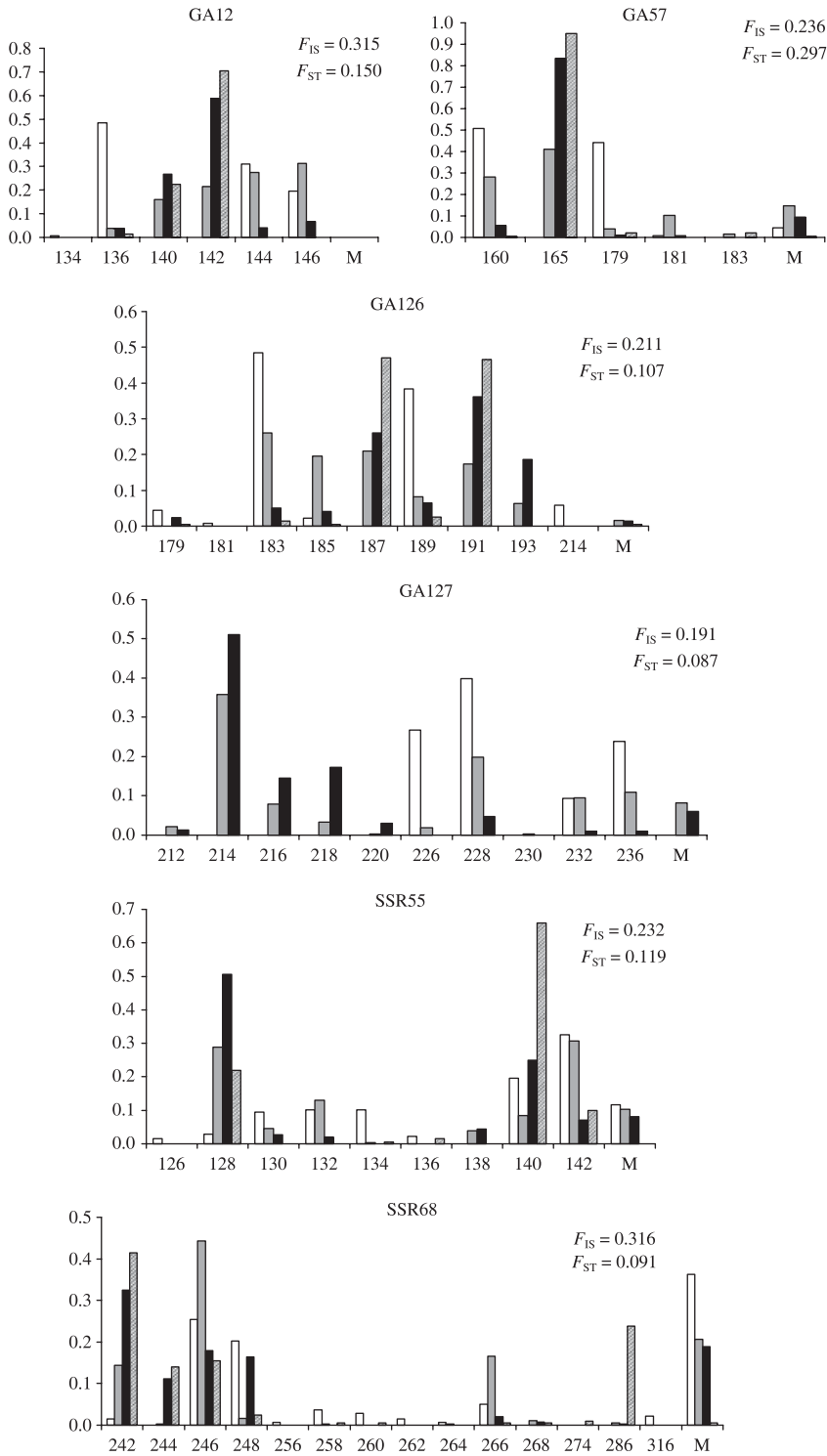


Fig. 3 Allelic frequencies as a function of individual morphotype. White, domesticated individuals; grey, intermediate individuals; black, wild-type individuals; hatched, allopatric wild individuals. Alleles are designated by their size. M, missing data. Note that allopatric wild individuals were not genotyped at locus GA127.

loci (72 DM, 189 IM, 142 WM and 265 JM), are summarized in Table 1. We first conducted the analysis without prior information on morphotype and assuming two clusters, A and B ($K = 2$). All domesticated plants were classified as 'pure A'. Of WM individuals, 64.8% were classified as

'pure B'. Nevertheless, the average proportion classified as belonging to cluster B was 87.1%. The presence of a few shared alleles probably explains why some WM individuals are not classified as 'pure B'. The great majority (74.6%) of IM individuals had mixed ancestry (between

10% and 90%), and 21.7% were classed as 'pure A'. Juveniles either had mixed ancestry or belonged to cluster A or B (Table 1, run 2).

These results suggest that IM individuals are hybrids. To test this hypothesis, STRUCTURE was allowed to form three clusters, using prior population information for the DM and WM populations. If hybrids formed a third genetic entity, distinct from DM and WM populations, they should cluster in the third group. All DM (respectively, most of the WM) individuals were inferred in cluster A (respectively B) but only 10.6% of the IM individuals had clear ancestry in any of the inferred clusters, and 7.4% in cluster C (Table 1, run 3). This clearly proves they do not form a third entity, but rather are hybrids. Juveniles mostly had a mixed or 'pure B' ancestry (Table 1, subrun 2).

Factorial correspondence analysis. Figure 2 presents a projection of the individual coordinates on the principal plane of the FCA. DM, WM and IM morphotypes had very different means on the first factorial axis ($P < 0.001$ for all pairwise t -tests). IM individuals formed a continuum between the two other groups, consistent with their being hybrids.

Dynamics of this contact zone. Of the 142 genotyped WM plants from Savane Manuel, 43% possessed only W alleles at each locus, and the remainder had at least one D allele. This is mostly due to the presence of three shared alleles with high frequencies, which were also present in the allopatric wild populations. This suggests that these alleles either represent shared ancestral polymorphism or are homoplasious. Of the 189 IM plants from Savane Manuel, 0.5% possessed only W alleles, 2.7% possessed, at each locus, one D allele and one W allele (as expected for F_1 hybrids), 52.2% presented a multilocus genotype consistent with being post- F_1 hybrids and the others had genotypes consistent with being backcrosses towards either wild or domesticated individuals. Among the juveniles from Savane Manuel, 15.5% had only W alleles; 0.4% had the multilocus genotype expected for F_1 hybrids, and 53.6% had a genotype expected for post- F_1 hybrids. The remainder had genotypes consistent with being backcrosses (mostly towards wild individuals).

Multilocus linkage disequilibrium ($\kappa_{0,2}$) and Hardy-Weinberg disequilibrium ($\kappa_{1,1}$) were then assessed using MULTILOCUS (Barton 2000), on the individuals from Savane Manuel. For the whole population as well as for the subsample of sexually mature plants, the chosen model was the most complex one, but for the JM sample, the model assuming equal δp among loci and no higher-order association coefficients was chosen. Values of $\kappa_{1,1}$ and $\kappa_{0,2}$ are shown in Table 3.

In the entire population, $\kappa_{1,1}$ was significant [$\kappa_{1,1} = 4.94\%$ (3.6–5.0%)] but far from its maximum possible value

Table 3 Multilocus intra- ($\kappa_{0,2}$) and inter- ($\kappa_{1,1}$) genome associations, computed with MULTILOCUS MATHEMATICA add-ons (Barton 2000). Maximum values of $\kappa_{0,2}$ and $\kappa_{1,1}$ are 13.6% in each case

	$\kappa_{0,2}$	$\kappa_{1,1}$
WM + IM + JM (whole population)	4.13% (4.0–5.0%)	4.94% (3.6–5.0%)
WM + IM (mature only)	5.34% (4.9–7.0%)	3.97% (3.6–5.1%)
JM (juveniles only)	1.94% (1.5–2.4%)	3.24% (2.6–3.5%)

(13.6%, with the allelic frequencies present in DM and WM), suggesting the existence of a barrier to hybridization (incomplete mixing of the populations), due either to genetic or reproductive incompatibility or to a spatial structure, or both. Spatial structure is the more likely explanation, as IM and WM individuals were not located randomly (see Fig. 1).

This result was corroborated by the high, but also far from maximum, value of $\kappa_{0,2}$ [4.13% (4.0–5.0%)], the estimator of multilocus linkage disequilibrium. If there is no admixture of the populations, $\kappa_{0,2}$ should remain at its highest value (13.6%, given the allele frequencies found in Savane Manuel), whereas it should decrease by half at each generation if complete mixing occurs. Our results show that at least the equivalent of two to three generations of free recombination must have occurred in Savane Manuel since the first hybridization event. Comparison of $\kappa_{0,2}$ values in adult vs. juvenile subsamples [5.34% (4.9–7.0%) vs. 1.94% (1.5–2.4%)] suggests that linkage disequilibrium is currently decreasing, implying that recombination of wild and domesticated genomes is continuing, even if not necessarily completely unconstrained.

A classical representation of the admixture phenomenon is given in Fig. 4, showing that most of the IM individuals were genetically intermediate between the two parental populations, and some were apparently strongly introgressed in one or the other direction, at these particular microsatellite loci.

Morphological differentiation between wild and domesticated populations

Six of the seven size variables (i.e. total height, height at first branching, leaf length and width, stem diameter, and leaf lobe number; stem number was the sole exception) decreased with hybrid index (t -tests, $P < 0.001$ for all tests). Most IM plants (62%) were larger than any wild-type plant studied.

Figure 5a shows the first factorial plane of the Hill & Smith analysis on all 10 morphological variables, including only adult plants with no missing morphological data (396 individuals). The first axis (PC1) explained 23.6% of the

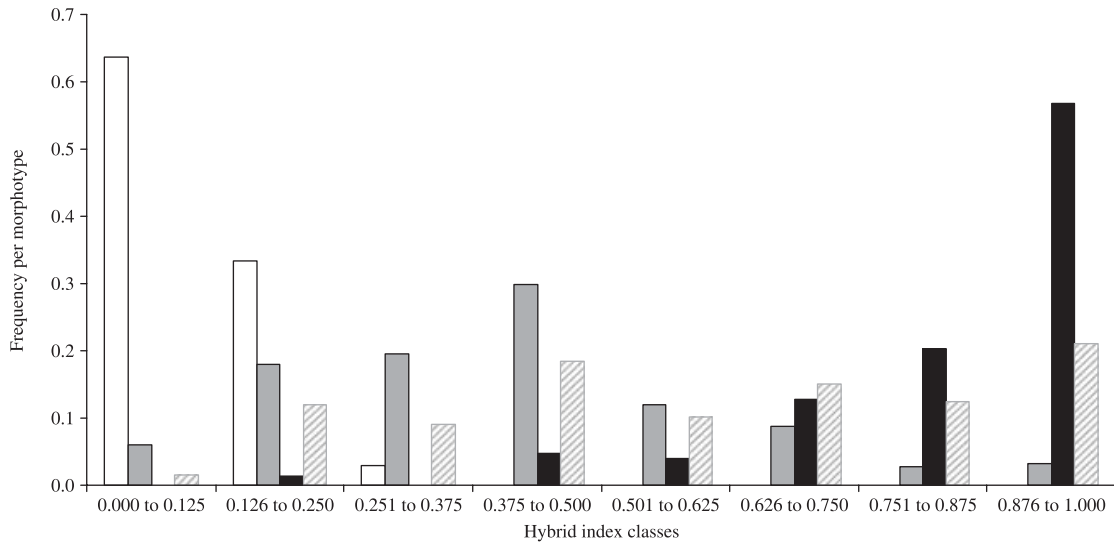


Fig. 4 Distribution of hybrid index per morphotype. DM individuals figure in white (69 individuals); IM in grey (184 individuals); WM in black (148 individuals) and juveniles in hatched grey (265 individuals).

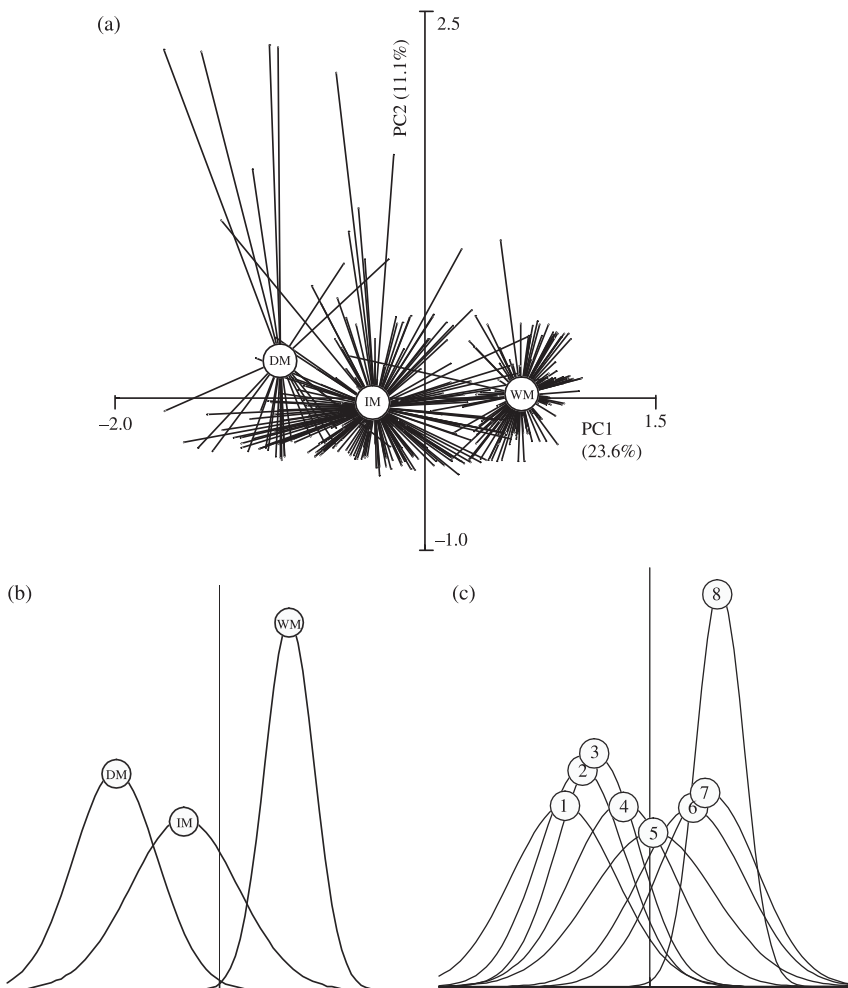


Fig. 5 Hill and Smith analysis on 10 morphological variables. (a) Projection of the individuals on the principal plane. Individuals were grouped *a posteriori* according to their morphotype. (b) Normalized distribution of PC1 of individuals from different morphotypes. Each curve represents the distribution of PC1 for one morphotype. Abbreviations: see Materials and methods. (c) Normalized distribution of PC1 of the same individuals, grouped by genotype composition in W allele (hybrid index HI). 1 : $0.000 \leq HI \leq 0.125$; 2 : $0.125 < HI \leq 0.250$... 8 : $0.875 \leq HI \leq 1.000$.

total inertia, and mainly included contributions of five quantitative variables (total height, height at first branching, stem diameter, leaf length and number of leaf lobes). PC1 clearly separated WM and DM morphotypes, with IM forming a continuum, as expected (*t*-tests, $P < 0.001$ for all pairs). Figure 5(b) shows the normalized distribution of PC1 of the different morphotypes. WM individuals were more homogeneous for PC1 than were DM individuals (Fisher test, $P < 0.01$), which were in turn less variable than IM individuals (Fisher test, $P = 0.05$).

PC1 of the Hill & Smith analysis was also highly correlated with hybrid index (i.e. the frequency of the *W* synthetic allele within an individual, $r^2 = 0.548$, $P < 0.001$; Fig. 5c). Morphological variance depended on the hybrid index, being relatively large in intermediate and extreme D classes, and low in extreme *W* classes.

The geographical coordinate *Y* (position along the length of the transect) was also correlated with both PC1 of the Hill & Smith analysis ($r^2 = 0.502$, $n = 291$ adults; $P < 0.001$) and PC1 of the genetic FCA ($r^2 = 0.435$, $n = 711$ mature and juvenile plants; $P < 0.001$).

Identification of other small hybrid zones

Three small hybrid zones were identified about 100 km away from Savane Manuel [Kamuyene (04°58.2' N, 052°26.3' W); a field near the southern limit of Macouria (04°58.2' N, 052°26.4' W) and Le Grand Blanc (04°58.8' N, 052°26.6' W)]. Respectively, 2 out of 3, 2 out of 6 and 5 out of 12 individuals appeared to be hybrids (the others being wild; see the FCA analysis presented as Supplementary material). Furthermore, 71 individuals were sampled on the roadside from Savane Manuel to Sinnamary (3 km). In this environment, wild *Manihot* grows spontaneously (but is periodically mowed if close to the road) and some domesticated plants had been planted in nearby gardens or along the roadside. This was indeed a favourable context for hybridization between the two taxa, and 27 individuals (38%) were hybrids (four were wild and 40 were domesticated).

Discussion

Hybridization within the genus *Manihot*

Species boundaries in *Manihot* are 'not sharply delimited' (Rogers & Appan 1973), and it has been argued that all taxa of the genus are interfertile (Jennings 1995). Despite these claims, however, natural hybrids have rarely been documented, and never characterized on a molecular basis. Although Lefèvre & Charrier (1993) noted that tree cassava (*Manihot glaziovii*) naturally hybridized with cassava in western Africa, they did not assess the fertility of the hybrids, and the identity of the wild taxon is unclear,

as tree cassava is stated elsewhere to be itself an artificial hybrid between *M. glaziovii* and cassava (Wanyera *et al.* 1994).

Here, for the first time, we have demonstrated, based on molecular evidence, that a clonally propagated crop has hybridized in nature with a wild relative and that hybrids were fertile. Thus, selection associated with clonal propagation has not prevented cassava from retaining the capacity for sexual reproduction and from intercrossing with wild relatives. Detection of hybridization in this site was greatly facilitated by the high degree of morphological and genetic differentiation presented by domesticated and wild populations in French Guiana. Indeed, cassava shares alleles (for both microsatellites and three nuclear genes) only with wild populations from the southern part of the Amazon basin, where domestication apparently took place (Olsen & Schaal 1999; Olsen 2004). In contrast, the wild *Manihot* taxon we studied in French Guiana is strongly differentiated from domesticated cassava, showing greater genetic affinity with populations of wild *Manihot* (*Manihot pruinosa*) along the southeastern rim of Amazonia (G. Léotard, unpublished data).

Frequency of hybridization events

This study focused on only one hybrid zone, in which the hybrids possessed up to five alleles per locus private to WM populations and five exclusive to DM populations. This implies that at least three wild and three domesticated individuals were involved in the first hybridization events. Moreover, the hybrids found between Savane Manuel and the town of Sinnamary possessed domesticated alleles absent from Savane Manuel. Their formation thus involved still other domesticated parents. Furthermore, three very small hybrid zones were identified about 100 km southeast of Savane Manuel. This suggests that the extensive hybridization observed in Savane Manuel has also occurred in several other places.

Hybridization between cassava and this wild relative is thus not an isolated event in coastal French Guiana. It seems to occur easily whenever cassava is planted in proximity to wild populations in coastal savannas.

Heterosis and the genetic and morphological diversity of hybrids

Cassava is known to be a highly heterozygous crop (Chavarriaga-Aguirre *et al.* 1998), a characteristic allowed by clonal reproduction. In our sample, the average heterozygosity of domesticated individuals was 57%, similar to that found by Fregene *et al.* (2003). Both farmer selection (Pujol *et al.* 2005a) and natural selection (Pujol & McKey 2006) help maintain a high rate of heterozygosity in planted clones. In contrast, wild individuals presented a

substantially lower heterozygosity (41.7%). IM individuals, however, were even more heterozygous (61.8%) than were domesticated individuals, because of the mixing of D and W alleles. On the whole, genetic diversity was increased in hybrids, with no loss of rare alleles.

In contrast to wild individuals, hybrids exhibited huge morphological variation (Fisher-Snedecor test, $P < 0.01$), consistent with their high genetic diversity and with the diversity of hybrid pedigrees (F_1 , F_2 , and probably backcrosses with the wild relative). Hybrids are expected to present high morphological variation, as they express a mosaic of parental phenotypes (Allendorf *et al.* 2001).

Increased heterozygosity and genetic diversity in F_1 hybrids – and to a lesser extent, in subsequent generations – is invoked in all cases of high hybrid vigour. Heterosis, the higher vigour of hybrids, has been observed in numerous situations, mainly involving plants (e.g. Arnold & Hodges 1995), and can result from overdominance, masking of deleterious recessive alleles, increased mutation rate, and new allele combinations (Jarvis & Hodgkin 1999; Rieseberg *et al.* 2003).

In this study, we did not evaluate the fitness of the hybrids. However, we did observe that hybrids, although likely younger, were much larger on average than were wild individuals. They can thus outcompete the wild parents in competition for light. Nevertheless, heterosis is often restricted to vegetative characters, and hybrid fitness is not always correlated with hybrid vigour; Dobzhansky (1952) described as 'luxuriance' the case of interspecific hybrids that display some heterotic traits together with a reduced reproductive fitness. Therefore, our observation of heterosis for size does not mean that *Manihot* hybrids are necessarily fitter than their wild parents. Moreover, we observed abnormal phenotypes, with a very low frequency of branching (and consequently of flowering) in eight hybrid individuals, genotypes of which were inconsistent with their being F_1 s, but compatible with F_2 status. Low fitness often appears in the F_2 as a consequence of the rupture of co-adapted gene complexes (Barton & Hewitt 1985; Rieseberg *et al.* 2003).

Another potential difference between hybrids and wild *Manihot* lies in ecological preferences. It has often been suggested that, although hybrids may have lower fitness than their parents in their respective niches, they may outcompete parental forms in some new ecological niche [the bounded hybrid superiority hypothesis (Arnold 1995 and references therein)]. In the present study, even though hybrids belonged to several generations, they are clustered on a relatively small area of talus formed by the soil removed when the ditch was dug. Wild parents were found in a less disturbed part of the sampled area. Hybrids might be fitter than their wild parent only in well-drained disturbed microenvironments such as this talus. Importantly, the persistence of a strong spatial structure and

ecological segregation of hybrids at a small scale, several generations after primary hybridization, suggests either strong differential selection among environments or very low gene flow between the hybrid and wild compartments, or both. In our case, pollen is thought to be dispersed by stingless bees, and seeds are dispersed first ballistically (through explosive fruit dehiscence), and then by ants. Seed dispersal in cassava thus probably occurs at a more local scale than pollen dispersal. The clustering of the hybrids suggests that pollen flow from the hybrid to the wild subpopulation is limited. Reduced male fertility in hybrids is a possible cause. Planting of wild and hybrid individuals in different controlled environments and assessment of their fertility would help in understanding genetic and environmental contributions to their relative vigour.

Fate of the hybrid zone: barriers to hybridization, or gene swamping?

Despite some evidence for partial hybrid breakdown (see above), there seem to be no strong barriers to hybridization in Savane Manuel. Hybrid index was highly variable (Fig. 4), with some intermediate individuals presenting an F_1 pattern, but the great majority of them (96.7%) probably being post- F_1 or backcross. Some F_1 s may be masked by the presence of a few shared alleles, but the very low value of $\kappa_{0,2}$ (an estimator of linkage disequilibrium between alleles issuing from wild populations on the one hand and domesticated plants on the other) shows that hybridization has in any case long since passed the F_1 stage. Furthermore, dramatic reduction of $\kappa_{0,2}$ in juveniles compared to adults shows that genetic mixing is still going on. One potential barrier is spatial (spatial clustering of the wild and hybrid gene pools), rather than biological. The currently restricted dispersal and gene flow should therefore not be considered a permanent feature of this hybrid zone.

Ultimately, the fate of this hybrid zone will depend on the relative fitness of wild and hybrid individuals. If hybrids are as fit as wild individuals both on the talus and in the remainder of the savanna, domesticated genes may introgress into the wild population and the hybrid zone will become a uniform hybrid swarm. If, on the other hand, hybrids are fitter than wild individuals only on the better-drained talus, the swarm may remain confined there (Allendorf *et al.* 2001). However, in this very small contact zone, gene flow will oppose local adaptation. Hybrid zones can be stabilized in the long term only when selection is strong relative to gene flow, and affects many loci throughout the genome, or when there is a permanent input of parental forms from external populations (Barton & Hewitt 1985). Otherwise, any locus not tightly linked with a selected locus is expected to homogenize rapidly

throughout the zone. This may well happen in Savane Manuel, given the small size and relative isolation of the contact zone, the apparent weakness of intrinsic barriers, and the evidence for effective recombination among markers (i.e. low linkage disequilibrium).

Conservation concerns

In Savane Manuel, gene flow was strongly asymmetrical, from domesticated to wild, probably because domesticated plants were removed some years ago. The mostly clonal propagation of domesticated cassava limits the incorporation of hybrids into the domesticated pool. Although Amerindian farmers regularly incorporate plants from self-sown seeds into their stocks of clones (Elias *et al.* 2001), this practice seems to be much less frequent among the Creoles, Haitians and Surinamese who plant cassava in the coastal savannas of French Guiana (D. McKey, personal observation). Further analysis is needed to determine whether gene flow is asymmetric in all contact zones.

Crop-to-wild gene flow could lead either to the appearance of weeds, as has been documented in bean (Papa & Gepts 2003) and beet (Viard *et al.* 2004), or to the introduction of domesticated genes into the wild genome, which could eventually lead to the extinction of wild populations by complete admixture (Ellstrand *et al.* 1999; Allendorf *et al.* 2001). This is most likely to happen when the wild populations are rather small (Ellstrand *et al.* 1999), as is often the case for wild *Manihot* species.

Gene flow from domesticated to wild taxa also raises the concern of transgene flow. Impacts of transgene flow are unclear, but could include decreased variability of wild taxa or cultivated landraces (Gepts & Papa 2003) or enhancement of the weediness of the feral taxon, as shown for several crops (Ellstrand *et al.* 1999). The spread of a transgene will depend on its fitness effect. The fitness effect of a transgene depends both on the crop in which it is introduced and on the transgene itself, stressing the need for case-by-case studies (Den Nijs *et al.* 2004). Indeed, some transgenes impose fitness costs in the absence of the selective pressure (e.g. herbicide) for which they are designed, which would prevent their spread into nontarget environments (e.g. Bartsch *et al.* 2003). Others are neutral and are expected to diffuse neutrally in nontarget environments (e.g. Burke & Rieseberg 2003), and others still can confer a selective advantage in a wide range of environments and are expected to spread widely into both domesticated plants and taxa intercrossing with them (e.g. Snow *et al.* 2003).

With the exception of potato (Celis *et al.* 2004), clonally propagated crop plants appear to have been neglected in discussions of biosafety concerns. Our study demonstrates the pertinence of this theme for cassava, on which

development of transgenic plants is an active focus of current research (Taylor *et al.* 2004; Ihemere *et al.* 2006), and suggests the interest of conducting similar studies with other clonally propagated crops. Indeed, at least for cassava, selection for traits enhancing clonal propagation, which is in direct competition with sexual reproduction, has not led to reproductive isolation of the crop from wild relatives, and crop-wild hybridization is neither an isolated phenomenon, nor an ephemeral process.

In the case of cassava, a hybrid bridge could allow gene flow to wild taxa not in direct contact with populations of the domesticated taxon. This bridge could be provided by a wild taxon such as the one we encountered in Savane Manuel. Hybrid zones such as those we have described need to be monitored over longer periods, in order to understand their dynamics and the evolutionary consequences of hybridization.

Acknowledgements

This study was funded by grants to D. McKey from various agencies of the French government (programme 'Impact des Biotechnologies dans les Agroécosystèmes' of the Ministry of Research, programme 'Ecosystèmes Tropicaux' of the Ministry of Ecology and Sustainable Development, the 'Bureau des Ressources Génétiques', and the 'Contrat Plan Etat-Région Guyane'). A. Duputié was supported by a stipend from the Ecole Normale Supérieure, Lyon. We thank the Office of the Mayor, Sinnamary, for permission to study in the site, as well as the farmers who allowed us to sample and measure the cassava varieties growing in their fields. Genotyping was done in the 'Service Commun de Marqueurs Génétiques en Ecologie' of the CEFÉ. We thank G. Léotard, M. Delêtre and B. Pujol for permission to cite their unpublished results on *G3pdh* and microsatellite loci, and F. Massol for adapting the MULTILOCUS add-ons for 6 loci. We are grateful to J. D. Lebreton for invaluable comments on statistical analysis and for bringing to our attention the study by Hill and Smith. We also appreciated the constructive comments provided by two anonymous reviewers.

References

- Allem AC, Mendes RA, Salomão AN, Burle ML (2001) The primary gene pool of cassava (*Manihot esculenta* Crantz subspecies *esculenta*, Euphorbiaceae). *Euphytica*, **120**, 127–132.
- Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution*, **16**, 613–622.
- Arnold ML, Hodges SA (1995) Are natural hybrids fit or unfit relative to their parents? *Trends in Ecology & Evolution*, **10**, 67–71.
- Barton NH (2000) Estimating multilocus linkage disequilibria. *Heredity*, **84**, 373–389.
- Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.
- Bartsch D, Cuguen J, Biancardi E, Sweet J (2003) Environmental implications of gene flow from sugar beet to wild beet – current status and future research needs. *Environmental Biosafety Research*, **2**, 105–115.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2002) *GENETIX 4.03, Logiciel Sous Windows™ Pour la Génétique Des*

- Populations. Laboratoire Génome, Populations, Interactions, Université Montpellier II, Montpellier, France.
- Burke JM, Rieseberg LH (2003) Fitness effects of transgenic disease resistance in sunflowers. *Science*, **300**, 1250.
- Celis C, Scurrah M, Cowgill S *et al.* (2004) Environmental biosafety and transgenic potato in a centre of diversity for this crop. *Nature*, **432**, 222–225.
- Chavarriga-Aguirre P, Maya MM, Bonierbale MW *et al.* (1998) Microsatellites in cassava (*Manihot esculenta* Crantz): discovery, inheritance and variability. *Theoretical and Applied Genetics*, **97**, 493–501.
- Den Nijs H, Bartsch D, Sweet J (2004) *Introgression from Genetically Modified Plants into Wild Relatives and its Consequences*. CAB International, Wallingford, UK.
- Dobzhansky TG (1952) Nature and origin of heterosis. In: *Heterosis* (ed. Gowen JW), pp. 218–223. Iowa State College Press, New York.
- Elias M, Panaud O, Robert T (2000) Assessment of genetic variability in a traditional cassava (*Manihot esculenta* Crantz) farming system, using AFLP markers. *Heredity*, **85**, 219–230.
- Elias M, Penet L, Vindry P, McKey D, Panaud O, Robert T (2001) Unmanaged sexual reproduction and the dynamics of genetic diversity of a vegetatively propagated crop plant, cassava (*Manihot esculenta* Crantz), in a traditional farming system. *Molecular Ecology*, **10**, 1895–1907.
- Ellstrand NC, Prentice HC, Hancock JF (1999) Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics*, **30**, 539–563.
- Fregene MA, Vargas J, Ikea J *et al.* (1994) Variability of chloroplast DNA and nuclear ribosomal DNA in cassava (*Manihot esculenta* Crantz) and its wild relatives. *Theoretical and Applied Genetics*, **89**, 719–727.
- Fregene MA, Suarez M, Mkumbira J *et al.* (2003) Simple sequence repeat marker diversity in cassava landraces: genetic diversity and differentiation in an asexually propagated crop. *Theoretical and Applied Genetics*, **107**, 1083–1093.
- Gepts P, Papa R (2003) Possible effects of (trans) gene flow from crops on the genetic diversity from landraces and wild relatives. *Environmental Biosafety Research*, **2**, 89–103.
- Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485–486.
- Hill MO, Smith AJE (1976) Principal component analysis of taxonomic data with multi-state discrete characters. *Taxon*, **25**, 249–255.
- Ihemere U, Arias-Garzon D, Lawrence S, Sayre R (2006) Genetic modification of cassava for enhanced starch production. *Plant Biotechnology Journal*, **4**, 453–465.
- Jarvis DI, Hodgkin T (1999) Wild relatives and crop cultivars: detecting natural introgression and farmer selection of new genetic combinations in agroecosystems. *Molecular Ecology*, **8**, S159–S173.
- Jennings DL (1959) *Manihot melanobasis* Müll. Arg. — a useful parent for cassava breeding. *Euphytica*, **8**, 157–162.
- Jennings DL (1995) Cassava. In: *Evolution of Crop Plants*, 2nd edn (eds Smartt J, Simmonds NW), pp. 128–132. Longman, London.
- Lefèvre F, Charrier A (1993) Isozyme diversity within African cassava germplasm. *Euphytica*, **66**, 73–80.
- Léotard G (2003) *Phylogéographie et origine de la domestication du manioc (Manihot esculenta Crantz, Euphorbiaceae): les apports d'un échantillon élargi à l'écotone nord de l'Amazonie*. DEA Biologie de l'Evolution et Ecologie, Université Montpellier II, France.
- Mba REC, Stephenson P, Edwards K *et al.* (2001) Simple sequence repeat (SSR) markers survey of the cassava (*Manihot esculenta* Crantz) genome: towards an SSR-based molecular genetic map of cassava. *Theoretical and Applied Genetics*, **102**, 21–31.
- Médard R (1973) Morphogénèse du manioc, *Manihot esculenta* Crantz (Euphorbiacées-Crotonoidées): étude descriptive. *Adansonia, Sér.*, **2** (13), 483–494.
- Nassar NMA (2003) Gene flow between cassava, *Manihot esculenta* Crantz, and wild relatives. *Genetics and Molecular Research*, **2**, 334–347.
- Olsen KM (2004) SNPs, SSRs and inferences on cassava's origin. *Plant Molecular Biology*, **56**, 517–526.
- Olsen KM, Schaal BA (1999) Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. *Proceedings of the National Academy of Sciences, USA*, **96**, 5586–5591.
- Papa R, Gepts P (2003) Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theoretical and Applied Genetics*, **106**, 239–250.
- Piperno DR, Ranere AJ, Holst I, Hansell P (2000) Starch grains reveal early root crop horticulture in the Panamanian tropical forest. *Nature*, **407**, 894–897.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Pujol B, McKey D (2006) Size asymmetry in intraspecific competition and the density-dependence of inbreeding depression in a natural plant population: a case study in cassava (*Manihot esculenta* Crantz, Euphorbiaceae). *Journal of Evolutionary Biology*, **19**, 85–96.
- Pujol B, David P, McKey D (2005a) Microevolution in agricultural environments: how a traditional Amerindian farming practice favours heterozygosity in cassava (*Manihot esculenta* Crantz, Euphorbiaceae). *Ecology Letters*, **8**, 138–147.
- Pujol B, Mühlén GS, Garwood N, Horoszowski Y, Douzery EJP, McKey D (2005b) Evolution under domestication: contrasting functional morphology of seedlings in domesticated cassava and its closest wild relatives. *New Phytologist*, **166**, 305–318.
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenism. *Journal of Heredity*, **86**, 248–249.
- Rieseberg LH, Raymond O, Rosenthal DM *et al.* (2003) Major ecological transitions in wild sunflowers facilitated by hybridisation. *Science*, **301**, 1211–1216.
- Rogers DJ, Appan SG (1973) *Manihot, Manihotoides* (Euphorbiaceae). Flora Neotropica. Monograph 13. Hafner, New York.
- Scarcelli N, Tostain S, Vigouroux Y *et al.* (2006) Farmers' use of wild relative and sexual reproduction in a vegetatively propagated crop. The case of yam in Benin. *Molecular Ecology*, **15**, 2421–2431.
- Second G, Allem ACA, Mendes RA *et al.* (1997) Molecular marker (AFLP)-based *Manihot* and cassava numerical taxonomy and genetic structure analysis in progress: implications for their dynamic conservation and genetic mapping. *African Journal of Root and Tuber Crops*, **2**, 140–147.
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.
- Snow A, Pilsen D, Rieseberg L *et al.* (2003) A Bt transgene reduces herbivory and enhances fecundity in wild sunflowers. *Ecological Applications*, **13**, 279–286.
- Taylor N, Chavarriga P, Raemakers K, Siritunga D, Zhang P (2004) Development and application of transgenic technologies in cassava. *Plant Molecular Biology*, **56**, 671–688.

- Thioulouse J, Chessel D, Dolédec S, Olivier J-M (1997) ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing*, **7**, 75–83.
- Viard F, Arnaud JF, Delescluse M, Cuguen J (2004) Tracing back seed and pollen flow within the crop-wild *Beta vulgaris* complex: genetic distinctiveness vs. hot spots of hybridisation over a regional scale. *Molecular Ecology*, **13**, 1357–1364.
- Wanyera NMW, Hahn SK, Aken'Ova SE (1994) Introgression of ceara rubber (*Manihot glaziovii* Muell-Arg) into cassava (*Manihot esculenta* Crantz): a morphological and electrophoretic evidence. *Proceedings of 5th Symposium ISTRC-AB*, 125–130.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

This study is part of Anne Duputié's PhD thesis, a work aimed at understanding the ecology of the domestication of cassava, and the evolutionary relationships between cassava and its wild relatives. Patrice David is a population geneticist and evolutionary biologist working on various organisms and evolutionary issues. Chantal Debain manages the CEFÉ's molecular markers service. Doyle McKey is an evolutionary ecologist interested in plant-animal mutualism and plant domestication.

Supplementary material

The following supplementary material is available for this article:

Fig. S1 Factorial correspondence analysis on the matrix of allele counts per morphotype: projection of the individuals from four small hybrid zones, treated as supplementary individuals, on the principal plane. The analysis was conducted on the same individuals as those presented in Fig. 2. The ellipses include 95% of the individuals from each morphotype. Black squares, Kamuyene (04°58.2' N, 052°26.3' W); grey diamonds, a field near the southern limit of Macouria (04°58.2' N, 052°26.4' W); black triangles, Le Grand Blanc (04°58.8' N, 052°26.6' W); white squares, individuals located between Savane Manuel and the town of Sinnamary, along the road.

This material is available as part of the online article from:

<http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2007.03340.x>

(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

