

Traditional Amerindian cultivators  
combine directional and ideotypic selection  
for sustainable management of cassava genetic diversity

Supplementary Information

Anne Duputié<sup>1,2</sup>, François Massol<sup>1,3</sup>, Patrice David<sup>1</sup>, Claudie Haxaire<sup>4</sup> & Doyle McKey<sup>1</sup>

**Contents**

<b>1</b>	<b>Supplementary Figures</b>	<b>2</b>
1.1	Supplementary Figure 1. . . . .	2
1.2	Supplementary Figure 2. . . . .	2
1.3	Supplementary Figure 3. . . . .	3
1.4	Supplementary Figure 4. . . . .	3
<b>2</b>	<b>R scripts</b>	<b>4</b>
2.1	Assessment of average relatedness within landraces and between products of sexual reproduction and the landrace they were assigned to. . . . .	4
2.2	AMOVA. . . . .	7
2.3	Benjamini & Hochberg's FDR test. . . . .	7
2.4	Saturation curve (Supplementary Figure 3). . . . .	8
<b>3</b>	<b>Supplementary Tables</b>	<b>8</b>

---

<sup>1</sup>CEFE UMR5175, 1919 Route de Mende, 34293 Montpellier CEDEX 5, France

<sup>2</sup>Corresponding author. Email: [anne.duputie@ens-lyon.org](mailto:anne.duputie@ens-lyon.org)

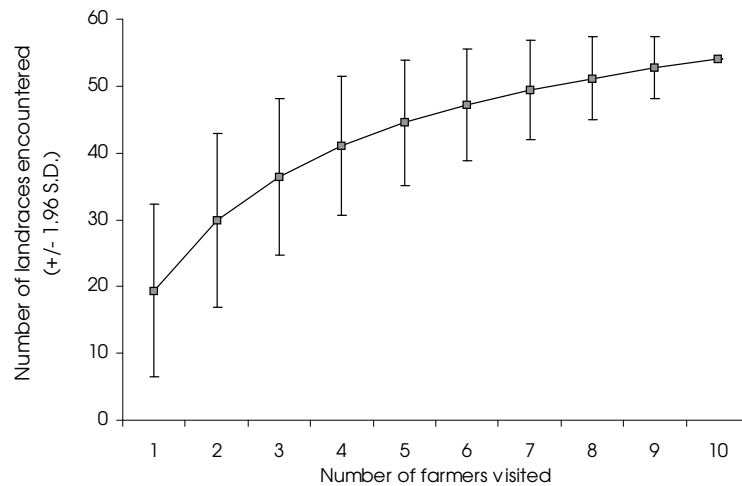
<sup>3</sup>Present address: UR HYAX, CEMAGREF, 3275 Route de Cézanne, Le Tholonet CS40061, 13182 Aix en Provence CEDEX 5, France

<sup>4</sup>Faculté de Médecine, UBO, 22 Avenue Camille Desmoulins, CS93837, 29238 Brest Cedex 3, France

# 1 Supplementary Figures

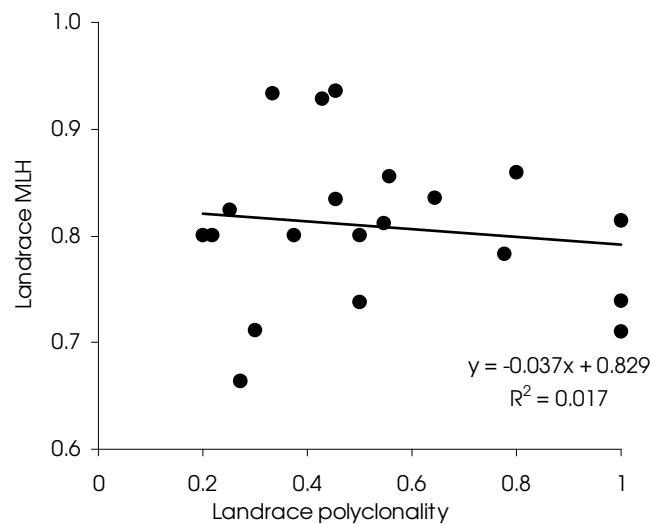
## 1.1 Supplementary Figure 1.

Number of landraces sampled as a function of the number of farmers sampled. Only C plants included (55 landraces as a total). No plateau is reached, suggesting that visiting more farms would probably have led to an even higher number of named landraces.



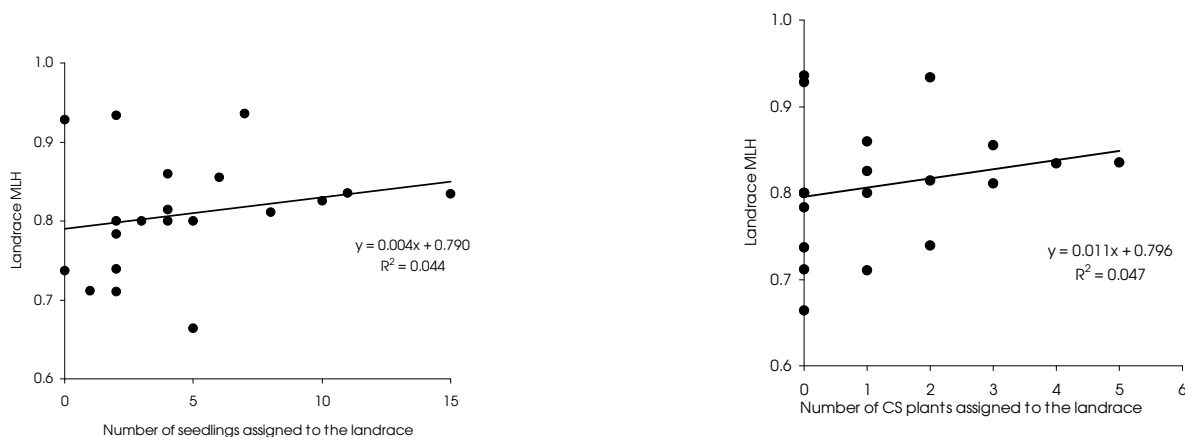
## 1.2 Supplementary Figure 2.

Average multilocus heterozygosity of landraces (MLH) as a function of their polyclonality (number of genotypes/number of plants). Landraces incorporating numerous plants issued from sex are expected to have high polyclonality, together with reduced MLH. No such trend is observed.



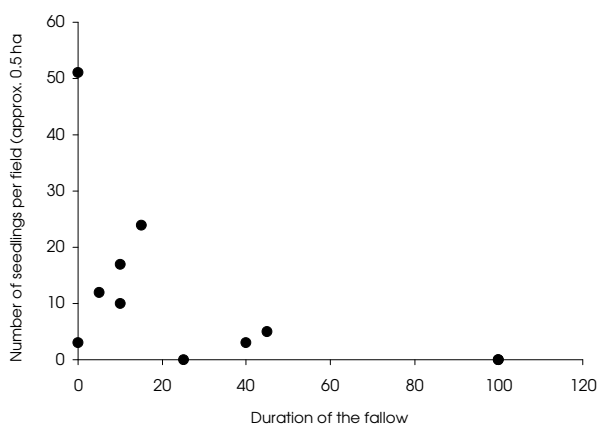
### 1.3 Supplementary Figure 3.

Left panel: Relation between the number of seedlings assigned to each landrace (with three or more C individuals) and the average multilocus heterozygosity (MLH) of the landrace. Right panel: Average MLH of landraces as a function of the number of CS plants assigned to them. As MLH of S and CS plants is lower than that of C plants, landraces incorporating numerous CS plants (which may be the same as those to which numerous S plants are assigned) would be expected to have lower heterozygosity than those incorporating few products of sexual reproduction. However, no such negative correlation is observed.



### 1.4 Supplementary Figure 4.

Number of seedlings found in each field, in relation to the duration of the fallow. On average, less seedlings were encountered in fields left in fallow for a long time than on fields left into fallow for a short time, or not at all. Duration of the fallow was estimated according to the farmer's memory, and confirmed by observation of the diameters of the trunks still present in the fields. In some cases, the farmers had settled places which were not remembered to have been occupied. The forest looked like primary forest; however, in all of these fields, we found pieces of ceramics, testifying for past occupancy of the site. Arbitrarily, we attributed to all of these fields the age of 100 years. No seedling was encountered in any of these very old fields. A trend towards a decreasing number of seedlings in older fallows is observed.



## 2 R scripts

### 2.1 Assessment of average relatedness within landraces and between products of sexual reproduction and the landrace they were assigned to.

```

#load packages 'Hmisc' and 'FactoMiner'

#PREPARATION OF DATA.

#Read tables:
#Table var (1 column, 436 rows): names of landraces, for all individuals (S, CS and C).
#Table sex (1 column, 436 rows): indicates for all individuals whether it's a S, CS or C. S
plants: rows 1-129; CS: rows 130-167; C: rows 168-436.
#Table data (437 columns, 437 rows): contains the labels of each individual and the table
of pairwise relatedness.
var<-read.table("var.txt")
sex<-read.table("sex.txt")
data<-read.table("DataWang.txt",header=TRUE,sep="\t")

#disj.var and disj.sex are disjunctive tables of landrace names and status (S, CS or C).
disj.var<-tab.disjonctif(var)
disj.sex<-tab.disjonctif(sex)

#revar: recoding of landraces with numbers (here, landraces are numbered 1-63).
#parvar: subset of table disj.var, including only C plants.
revar<-recodevar(var[,1],levels(var[,1]),1:63)
parvar<-function(v){subset(1:436,as.logical(disj.var[,v]*disj.sex[,3]))}

#AVERAGE RELATEDNESSES OF S AND CS PLANTS WITH C PLANTS.

#Function mean.sex used for computing average relatedness between each plant issued
from sex and each landrace
mean.var<-function(v,i){
if(sum(disj.var[,v]*disj.sex[,3])>2) mean(data[parvar(v),i+1]) else NA}
interm<-function(i){sapply(1:63,mean.var,i)}
mean.sex<-sapply(1:167,interm)

#Mean relatedness of each S and CS plant with the landrace it was assigned to:
mean.mylandrace
interm2.S<-function(i){mean.sex[revar[i],i]}
mean.mylandrace<-sapply(1:167,interm2.S)
write.table(mean.mylandrace,'mean.mylandrace.txt',sep='\t')

```

```

#For each CS and S plant, ranking of landraces according to their maximum relatedness
to the focal plant: rangmax.S.results and rangmax.CS.results (containing the ranks of ALL
landraces for EACH S/CS plant).
max.S<-max.sex[,1:129]
max.CS<-max.sex[,130:167]
maxopp.S<--max.S
maxopp.CS<--max.CS
rangmaxfun<-function(i,kruk){rank(kruk[,i], na.last = "keep", ties.method =
"min")}
rangmaxsexS<-function(krok){
as.data.frame(sapply(1:129,rangmaxfun,kruk=krok))}
rangmax.S.results<-rangmaxsexS(maxopp.S)
write.table(rangmax.S.results,"rangmaxS.txt",sep=" ")
rangmaxsexCS<-function(krok2){
as.data.frame(sapply(1:38,rangmaxfun,kruk=krok2))}
rangmax.CS.results<-rangmaxsexCS(maxopp.CS)
write.table(rangmax.CS.results,"rangmaxCS.txt",sep=" ")

#Extraction of the rank of the landrace each S/CS plant was assigned to:
rangmaxindSresults and rangmaxindCSresults.
rag1S<-function(i,krak){rangmaxsexS(krak)[revar[i],i]}
rag1CS<-function(i,krak2){rangmaxsexCS(krak2)[revar[i+129],i]}
rangmaxindS<-function(krik){as.matrix(sapply(1:129,rag1S,krak=krik))}
rangmaxindCS<-function(krik2){as.matrix(sapply(1:38,rag1CS,krak2=krik2))}
rangmaxindSresults<-rangmaxindS(maxopp.S)
rangmaxindCSresults<-rangmaxindCS(maxopp.CS)
write.table(rangmaxindSresults,"rangmaxindS.txt",sep="\t")
write.table(rangmaxindCSresults,"rangmaxindCS.txt",sep="\t")

#TESTS FOR DISTRIBUTION OF RANKINGS OF AVERAGE AND MAXIMUM
RELATEDNESSES WITH THE LANDRACE THEY WERE ASSIGNED TO.

#Kolmogorov Smirnov test
ks.test(rangmaxindSresults,rangmaxindCSresults,alternative = "two.sided")
ks.test(rangmoyindSresults,rangmoyindCSresults,alternative = "two.sided")

#Test of the proportions of individuals assigned to the best possible landrace (rank 1)
among S and CS plants. Vector rk1moySCS: (number of seedlings assigned to the best
variety (when criterion is average belonging), number of CS assigned to the best possible
variety (same criterion). Vector rk1maxSCS is the same, but with criterion "maximum" (ie,
the genetically closest plant of the focal plant is found in the landrace the focal plant actually
was assigned to.) Vector effSCS is the number of (S,CS) plants assigned to landraces with
N(Cplants)>2.

```

```

rk1moySCS<-c(16,7)
effSCS<-c(108,29)
pairwise.prop.test(rk1moySCS, effSCS)
rk1maxSCS<-c(13,12)
pairwise.prop.test(rk1maxSCS, effSCS)
#COMPUTATION OF AVERAGE, VARIANCE, AND QUANTILES OF INTRA-
LANDRACE RELATEDNESSES (ONLY C PLANTS CONSIDERED)

#Variances of relatednesses between C plants of the same landrace (only landraces with
>2 C plants sampled). Output: variancevar.
variance<-function(v){
if(sum(disj.var[,v]*disj.sex[,3])>2)
var(as.dist(data[parvar(v),parvar(v)+1])
[1:((length(parvar(v))*(length(parvar(v))-1))/2)], na.rm = TRUE) else NA}
variancevar<-sapply(1:63,variance)
write.table(variancevar,"variancevar.txt",sep=" ")
write.table(as.data.frame(levels(var[,1])), "ordrevar.txt",sep=" ")

#Averages of relatednesses between C plants of the same landrace (only landraces with >2
C plants sampled). Output: moyennevar.
moyenne<-function(v){
if(sum(disj.var[,v]*disj.sex[,3])>2)
mean(as.dist(data[parvar(v),parvar(v)+1])
[1:((length(parvar(v))*(length(parvar(v))-1))/2)], na.rm = TRUE) else NA}
moyennevar<-sapply(1:63,moyenne)
write.table(moyennevar,"moyennevar.txt",sep=" ")

#Quantiles (2.5, 5, 25, 50, 75, 90, 95, 97.5%) of the distribution of relatednesses between
C plants of the same landrace(only landraces with >2 C plants sampled). Output: quantvar.
p<-c(2.5,5,25,50,75,90,95,97.5)/100
quant<-function(v){
if(sum(disj.var[,v]*disj.sex[,3])>2)
quantile(as.dist(data[parvar(v),parvar(v)+1])
[1:((length(parvar(v))*(length(parvar(v))-1))/2)], p, na.rm = TRUE) else NA}
quantvar<-sapply(1:63,quant)
quantvar<-as.matrix(quantvar)
write.table(quantvar,"quantvar.txt",sep=" ")

```

## 2.2 AMOVA.

```

#Load packages 'ADE4' and 'FactoMiner'.

#Table rawgeno contains the genotypes of each clone (each clone (row) is coded by 20
columns containing each the length of one allele (2 columns per locus)).
#Table genosamples.txt contains landraces (populations) as columns, clones as lines and
clone abundance by population as entries.
#Table genocode contains genotypes as rows, landraces as columns and is filled with the
counts of genotypes belonging to each of these landraces.

#READ DATA
rawgeno<-read.table("rawgenotypeadults.txt", sep="\t",header = TRUE)
disj.geno<-tab.disjonctif(rawgeno[,3:22])
write.table(disj.geno, "disj.geno.txt", sep = "\t")
genocode<-read.table("genocode.txt", sep = "\t", header = FALSE)

#Table rawgenodisj.txt has to be manually remanied (regrouping of data corresponding to
the same allele) in order to get the full disjunctive table of clone composition (where rows
are genotypes, columns are alleles, and the table contains 0, 1 and 2 as values, indicating
the genetic composition of each genotype).
genodisj<-read.table("rawgenodisj.txt", sep="\t", header = TRUE)

#COMPUTATION OF DISTANCE MATRIX
distgeno<-dist(genodisj[,2:53])
distances.euclide<-as.matrix(distgeno)
write.table(distances.euclide,"disteucl.txt",sep="\t")

#AMOVA
genocode1<-genocode[,which(apply(genocode,2,sum)>4)]
amovares<-amova(genocode1,distgeno)
randtest(amovares, nrepet = 1000)

```

## 2.3 Benjamini & Hochberg's FDR test.

```

#Input table 'pvalues.txt' contains a single columns of p-vales, arranged by ascending values.
Here we have 20 landraces so  $20 * (20 - 1) / 2 = 190$  tests.
pval<-read.table("pvalues.txt",sep="\t",header=FALSE)
pvalvect<-as.vector(pval[,1])
adjustedpval<-p.adjust(pvalvect, method = "BH", n = 190)
write.table(adjustedpval,"adjustedpval.txt",sep="\t")

```

## 2.4 Saturation curve (Supplementary Figure 3).

```
#load package 'vegan'  
data<-read.table("data.txt")  
data<-t(data)  
specaccum.farm<-specaccum(data,method="random",permutations = 10000)  
plot(specaccum.farm)  
summary(specaccum.farm)  
write.table(specaccum.farm$richness,"specaccum.richness.txt",sep="\t")  
write.table(specaccum.farm$sd,"specaccum.sd.txt",sep="\t")
```

## 3 Supplementary Tables



**Supplementary Table 1.** Identity of the landraces cultivated by each farmer. Farmers are indicated in columns and landraces in rows. Each blue cell indicates the landrace was found in at least one of the farmer's fields.

The total number of landraces cultivated by each farmer is indicated in the lower-most row, and the number of farmers (out of 10) cultivating each landrace figures in the last column.

	ALE	ALI	ADR	ARM	CHT	GTN	GHS	GLB	MUN	YUI	
akā panini			█				█				1
akawa			█				█				2
akusiway	█		█			█	█		█		5
alaku	█	█	█	█		█		█		█	8
aleō	█		█							█	3
alime	█		█		█	█			█		7
apā			█								1
apetanemu			█								1
ātā		█	█				█				3
ho'wi'i			█							█	1
iyī	█	█		█	█			█	█	█	7
kalaku					█						1
kalai				█	█						1
kapelulu			█								1
kapopī	█	█				█					3
kasili pupu	█			█	█	█		█	█	█	7
kasilisiililia		█									2
kulupa'i	█	█	█		█		█			█	3
kwasi ulu	█	█	█		█	█	█		█	█	7
malay kuluka			█		█	█			█		6
manuepe		█	█		█						2
matawale	█	█			█		█		█		3
mulutuku	█	█									2
nikolo		█	█						█		3
panali									█	█	2
pepitā	█	█				█	█				5
pilepupu	█	█	█								1
piyū		█	█		█	█	█	█	█	█	8
pōpō						█	█		█		4
pu'aii	█						█		█		1
si	█	█			█	█	█		█		6
sili	█	█	█	█	█						5
silipu			█						█	█	3
simo	█	█	█	█	█	█		█	█	█	8
takwāy milā	█	█	█					█		█	5
tapak			█								1
tawa	█		█	█	█		█	█		█	7
te			█							█	4
tēti makāsi						█	█				2
te'u	█					█					3
tīpī'o	█	█			█	█	█		█	█	7
toakulu											1
u	█	█					█				3
wakoloke	█					█	█	█			3
wala			█								1
wasey					█				█		2
wat e'e	█				█						2
weluī	█	█	█			█	█		█		5
wīla	█	█				█	█				4
wīlapapoā	█	█	█								2
wipe'i	█		█			█	█		█	█	5
yau	█		█	█	█	█	█	█	█	█	9
yeneya										█	1
yīkīī	█	█	█		█						3
unknown name	█	█	█		█		█				3
	26	23	34	9	18	21	19	10	19	17	



**Supplementary Table 3.** Pairwise  $F_{ST}$  values between landraces (lower-left matrix), and their significance at 5% level after performing Benjamini & Hochberg's FDR test (upper-right matrix). NS: non-significant population differentiation; empty cells correspond to significant tests (5% level). Darker shades stand for higher values of  $F_{ST}$ . Numbers of C plants for each landrace are indicated in the left-most column.

		akusiway	alaku	alime	iyi	kasili pupu	kwakī tawa	kwasi ulu	malay kuluke	pepitā	piyū	si	sili	silipu	simo	takwāy miā	te	tīpī'o	welu'i	wipe'i	yau
5	akusiway			NS							NS						NS			NS	
10	alaku	0.23																			
9	alime	0.02	0.15									NS					NS				
8	iyi	0.21	0.36	0.25																	
12	kasili pupu	0.07	0.16	0.07	0.24							NS					NS				
11	kwakī tawa	0.13	0.21	0.04	0.36	0.15															
10	kwasi ulu	0.25	0.27	0.24	0.49	0.24	0.29														
8	malay kuluka	0.12	0.26	0.13	0.27	0.09	0.18	0.31													
9	pepitā	0.11	0.13	0.06	0.27	0.03	0.11	0.27	0.11							NS	NS				
11	piyū	0.09	0.09	0.08	0.29	0.08	0.16	0.22	0.20	0.10											
7	si	0.05	0.23	0.07	0.29	0.07	0.12	0.19	0.07	0.08	0.12	NS			NS		NS				
7	sili	0.06	0.20	0.03	0.30	0.05	0.11	0.25	0.10	0.04	0.12	0.01					NS				
5	silipu	0.14	0.37	0.15	0.41	0.23	0.24	0.37	0.19	0.28	0.27	0.19	0.19								
14	simo	0.13	0.25	0.14	0.32	0.10	0.20	0.21	0.16	0.10	0.15	0.04	0.10	0.28							
7	takwāy miā	0.13	0.25	0.12	0.33	0.08	0.21	0.30	0.16	0.03	0.18	0.08	0.05	0.32	0.10		NS				
5	te	0.09	0.14	0.08	0.25	0.04	0.17	0.26	0.14	###	0.10	0.07	0.05	0.30	0.09	0.05				NS	NS
11	tīpī'o	0.21	0.31	0.19	0.26	0.20	0.25	0.42	0.26	0.18	0.19	0.18	0.16	0.33	0.23	0.24	0.14				
11	welu'i	0.13	0.23	0.15	0.32	0.11	0.23	0.25	0.21	0.09	0.18	0.11	0.07	0.34	0.10	0.08	0.06	0.26			
6	wipe'i	0.09	0.17	0.12	0.26	0.19	0.17	0.31	0.25	0.15	0.13	0.18	0.21	0.28	0.22	0.23	0.11	0.22	0.23		
23	yau	0.23	0.29	0.21	0.41	0.22	0.22	0.22	0.25	0.22	0.26	0.20	0.19	0.32	0.23	0.22	0.23	0.34	0.24	0.31	