



**cirad**

LA RECHERCHE AGRONOMIQUE  
POUR LE DÉVELOPPEMENT

*Montpellier, mars 2019*

# **Analyse de la structure génétique d'une population de cacaoyers collectés en Guadeloupe**

Olivier Fouet - Claire Lanaud – **CIRAD**

## **Objectif :**

Analyse de la structure génétique de 21 cacaoyers cultivés en Guadeloupe, détermination de l'origine génétique par comparaison avec des cacaoyers témoins représentant la diversité de l'espèce *Theobroma cacao* et prédiction de l'auto-compatibilité.

## **I-Matériels**

### **I.2. Echantillons testés**

Les feuilles ont été récoltées en Guadeloupe et stockées sous vide directement dans un emballage plastique transparent respectant idéalement les consignes d'envoi (ANNEXE I). Les échantillons ont été réceptionnés au CIRAD (Montpellier) le 09/11/2018 sauf pour **Gp-75** (oublié dans un bagage) reçu le 11/01/2019. Les feuilles des 21 arbres sont arrivées dans un délai correct et étaient dans un parfait état sauf pour l'échantillon **Gp-75** qui montrait un début de dégradation.

Les 21 échantillons numérotés de 75 à 95 sur les échantillons ont été annotés **Gp-75 à Gp-95** dans l'analyse :

<b>Gp-1</b>	<b>Gp-2</b>	<b>Gp-3</b>	<b>Gp-4</b>	<b>Gp-5</b>	<b>Gp-6</b>	<b>Gp-7</b>
<b>Gp-8</b>	<b>Gp-9</b>	<b>GPp-10</b>	<b>Gp-11</b>	<b>Gp-12</b>	<b>Gp-13</b>	<b>Gp-14</b>
<b>Gp-15</b>	<b>Gp-16</b>	<b>Gp-17</b>	<b>Gp-18</b>	<b>Gp-19</b>	<b>Gp-20</b>	<b>Gp-21</b>

Les données notées sur les échantillons ont été recueillies dans un tableau (ANNEXE II).

### **I.2. Témoins**

La structure génétique des 21 cacaoyers collectés en Guadeloupe a été analysée avec 110 cacaoyers témoins représentant notamment les 10 groupes génétiques connus (Motamayor *et al.*, 2008) et 1 subdivision (Morona) se dissociant du groupe Nacional dans cette analyse. Des témoins spécifiques à la population étudiée (Trinitario) ont été intégrés à l'analyse pour une

meilleure comparaison tel que UF676 (Trinitario, hybride direct: Amelonado x Criollo), ICS95 (Trinitario «retrocroisé » : (Amelonado x Criollo) x Criollo) selon des informations données par la base ICGD (International Cocoa Germplasm Database).

## **II-Méthodes**

### **II.1. Extraction et purification de l'ADN**

L'ADN génomique a été extrait des feuilles à l'aide d'un tampon MATAB (Risterucci *et al.*, 2000) et purifié selon le protocole du kit Nucleobond® PC (Macherey-Nagel, n ° de catalogue 740.571.100) avec quelques modifications (Allegre *et al.*, 2011). La qualité a été visualisée par électrophorèse sur gel d'agarose (1%) par le dépôt d'un 1 µL d'ADN. La quantification de l'ADN a été réalisée à une absorbance de 260 nanomètres (nm) par le dépôt de 1 µL d'ADN sur un Infinite® 200 PRO *NanoQuant* (Tecan, Crailsheim, Germany).

### **II.2. Génotypage des arbres**

Dans cette analyse, 26 marqueurs moléculaires SSR (Simple Sequence Repeats) de type microsatellite (Risterucci *et al.*, 2000, Fouet *et al.*, 2011) ont été sélectionnés sur des critères de qualité et de polymorphisme pour étudier les dissimilarités entre les individus. Les amplifications PCR (Polymerase Chain Reaction) ont été réalisées avec un amorce M13 couplée à un colorant fluorescent 6-FAM™, NED™, VIC™ ou PET™ (Applied Biosystems, CA, USA) comme décrit dans des travaux précédents (Allegre *et al.*, 2011) et analysées en utilisant un séquenceur automatique de type ABI3500xL (Applied Biosystems). Les données brutes ont été analysées avec le logiciel Genemapper 4.0 (Applied Biosystems).

### **II.3. Analyse de la structure génétique**

Une première analyse a été réalisée sur un modèle de distance utilisant une matrice de dissimilarité calculée avec un index d'appariement simple avec la méthode d'assemblage par voisinage « neighbour-joining method » avec le logiciel DARWIN 6.0 (Perrier and Jacquemoud-Collet, 2006). Un dendrogramme été construit pour situer les arbres testés dans la phylogénie de *Theobroma cacao*.

Afin de définir la composition génétique des arbres testés, une deuxième méthode d'analyse a été effectuée sur un modèle de mise en œuvre bayésien (Pritchard *et al.*, 2000) avec le logiciel STRUCTURE 2.3.4 (Falush *et al.*, 2003), à un K=11 (nombre de groupes génétiques de l'analyse) avec « une burn-in period » de 100 000 itérations, 500 000 répétitions Markov Chain Monte Carlo (MCMC) et 10 répétitions.

### **II.3. Prédiction de l'auto-compatibilité**

Une analyse prédictive de l'auto-compatibilité des arbres a été réalisée à l'aide de 2 marqueurs moléculaires de type microsatellite récemment développés (Lanaud *et al.*, 2017).

## **III. Résultats**

### **III.1. Analyse phylogénétique**

Un arbre phylogénétique a été établi reflétant les distances génétiques calculées sur la base du pourcentage d'allèles partagés entre les différents arbres analysés par la méthode « Neighbor Joining » avec le logiciel DARWIN. La longueur des branches de l'arbre est proportionnelle à leur distance génétique. Les 21 arbres testés en provenance de Guadeloupe (**en rouge**) sont représentés dans le dendrogramme « radial » parmi les témoins spécifiques et les différents groupes génétiques (en noir).

Une représentation « hiérarchique-horizontale » du même arbre phylogénétique a été produit (ANNEXE III) pour avoir une autre visualisation des positions des arbres testés par rapport aux groupes génétiques et témoins référents.

Une analyse globale et une représentation graphique de l'ensemble des différents envois de cacaoyers issus de Guadeloupe ont été réalisées pour comparer la position des différents cacaoyers collectés sur un arbre phylogénétique (ANNEXE IV).

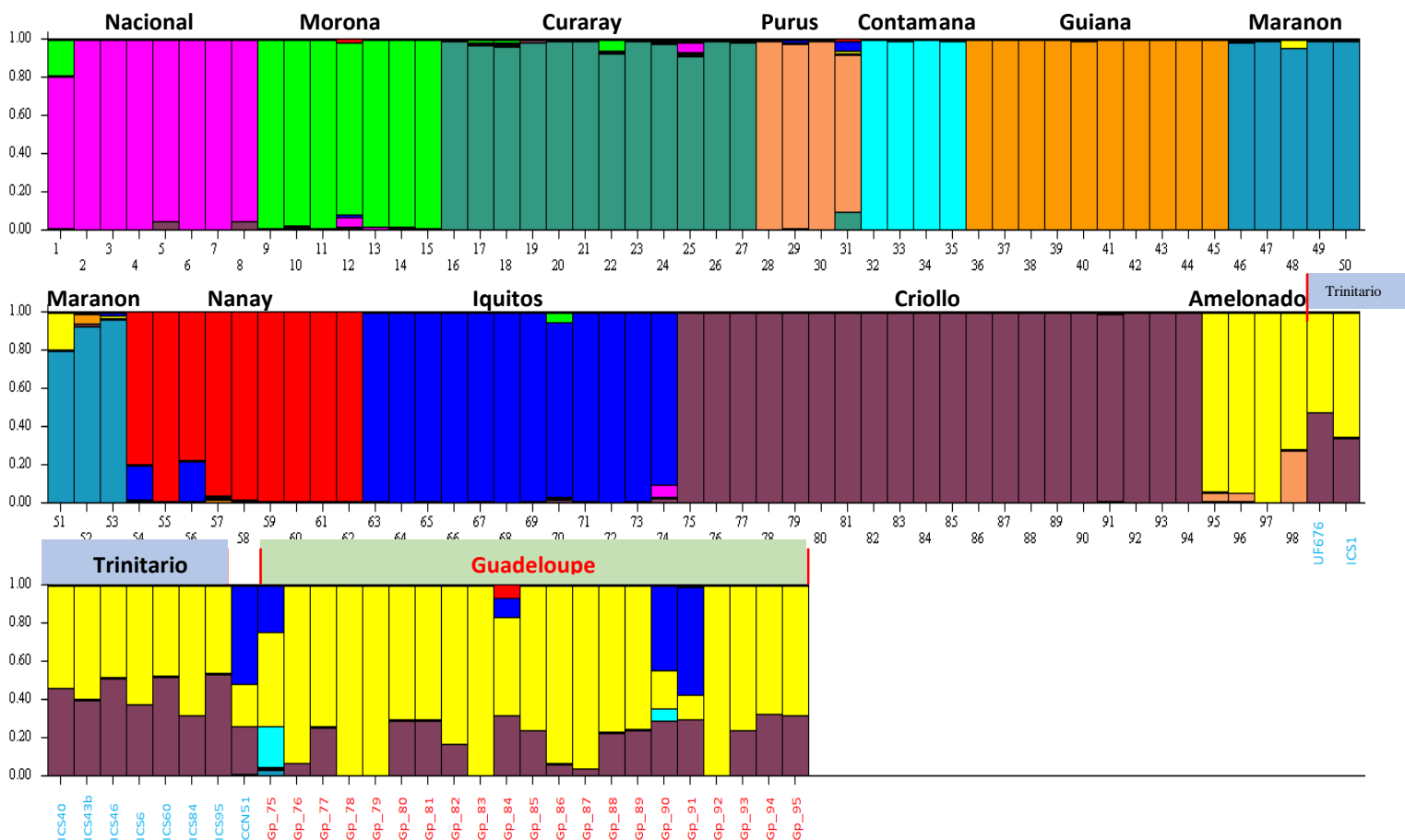


Cette première analyse montre clairement le regroupement de la grande majorité des arbres testés entre le groupe Amelonado et les témoins hybrides Trinitario (Criollo x Amelonado). Certains arbres comme Gp-78, Gp-86, Gp-87, Gp-92 se révèlent très proche des témoins Amelonado notamment de Matina1-6 (MAT1-6) pour Gp-79, Gp-83. Seuls deux arbres (Gp-90 et Gp-91) semblent plus distants de ces groupes génétiques (ANNEXE III-b).

La composition génétique des différents hybrides pourra être déterminée par l'analyse suivante (III.2).

### III.2. Structure et composition génétique des 21 arbres testés

L'analyse avec le logiciel STRUCTURE a permis d'identifier l'origine des arbres par la détermination de la contribution probable des 11 groupes génétiques témoins. Les résultats ont été représentés sur le graphique ci-dessous, où la composition génétique des 21 arbres testés a été révélée par la contribution des groupes génétiques de cacaoyer témoins représentés par différentes couleurs.



La contribution génétique des 11 groupes témoins aux arbres analysés (en rouge) est reportée avec plus de précision dans le tableau suivant:

	Nom individu	Nanay	Morona	Iquitos	Amelonado	Nacional	Contamana	Guyana	Criollo	Purus	Maranon	Curaray
Témoins	UF676	0.001	0.001	0.001	0.52	0.001	0.001	0.001	0.473	0.001	0.001	0.001
	ICS1	0.001	0.001	0.001	0.654	0.001	0.001	0.001	0.338	0.001	0.001	0.001
	ICS40	0.001	0.001	0.001	0.535	0.001	0.001	0.001	0.458	0.001	0.001	0.001
	ICS43b	0.001	0.001	0.001	0.596	0.001	0.001	0.001	0.397	0.001	0.001	0.001
	ICS46	0.001	0.001	0.001	0.48	0.001	0.001	0.001	0.513	0.001	0.001	0.001
	ICS6	0.001	0.001	0.001	0.621	0.001	0.001	0.001	0.371	0.001	0.001	0.001
	ICS60	0.001	0.001	0.001	0.473	0.001	0.001	0.001	0.519	0.001	0.001	0.001
	ICS84	0.001	0.001	0.001	0.676	0.001	0.001	0.001	0.317	0.001	0.001	0.001
	ICS95	0.001	0.001	0.001	0.46	0.001	0.001	0.001	0.533	0.001	0.001	0.001
	CCN51	0.005	0.001	0.512	0.22	0.001	0.001	0.001	0.251	0.001	0.005	0.001
Guadeloupe	Gp_75	0.002	0.005	0.237	0.495	0.003	0.209	0.006	0.007	0.002	0.034	0.001
	Gp_76	0.001	0.002	0.003	0.923	0.002	0.001	0.001	0.064	0.001	0.001	0.001
	Gp_77	0.001	0.001	0.001	0.737	0.001	0.001	0.001	0.257	0.001	0.001	0.001
	Gp_78	0.001	0.001	0.001	0.993	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	Gp_79	0.001	0.001	0.001	0.993	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	Gp_80	0.001	0.001	0.001	0.704	0.001	0.001	0.001	0.289	0.001	0.001	0.001
	Gp_81	0.001	0.001	0.001	0.704	0.001	0.001	0.001	0.289	0.001	0.001	0.001
	Gp_82	0.001	0.001	0.001	0.827	0.001	0.001	0.001	0.166	0.001	0.001	0.001
	Gp_83	0.001	0.001	0.001	0.993	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	Gp_84	0.066	0.001	0.102	0.512	0.001	0.001	0.002	0.312	0.001	0.001	0.002
	Gp_85	0.001	0.001	0.001	0.756	0.001	0.001	0.001	0.238	0.001	0.001	0.001
	Gp_86	0.001	0.001	0.001	0.932	0.001	0.001	0.001	0.061	0.001	0.001	0.001
	Gp_87	0.001	0.001	0.001	0.955	0.001	0.001	0.001	0.037	0.001	0.001	0.001
	Gp_88	0.001	0.001	0.001	0.766	0.001	0.001	0.001	0.227	0.001	0.001	0.001
	Gp_89	0.001	0.001	0.001	0.753	0.001	0.001	0.001	0.24	0.001	0.001	0.001
	Gp_90	0.001	0.002	0.444	0.2	0.001	0.065	0.001	0.281	0.001	0.001	0.002
	Gp_91	0.001	0.01	0.563	0.127	0.001	0.001	0.001	0.29	0.001	0.004	0.002
	Gp_92	0.001	0.001	0.001	0.993	0.001	0.001	0.001	0.001	0.001	0.001	0.001
	Gp_93	0.001	0.001	0.001	0.757	0.001	0.001	0.001	0.236	0.001	0.001	0.001
	Gp_94	0.001	0.001	0.001	0.67	0.001	0.001	0.001	0.323	0.001	0.001	0.001
	Gp_95	0.001	0.001	0.001	0.677	0.001	0.001	0.001	0.315	0.001	0.001	0.001

Les arbres analysés se confirment très proches globalement des groupes Amelonado et Criollo. Le groupe Amelonado contribue majoritairement, avec des degrés variables, à la composition génétique des arbres testés.

La majorité (13 sur 21) des arbres (Gp-76, Gp-77, Gp-80, Gp-81, Gp-82, Gp-85, Gp-86, Gp-87, Gp-88, Gp-89, Gp-93, Gp-94 et Gp-95) apparaissent clairement hybrides entre Amelonado et Criollo. Les ratios Amelonado/Criollo pour ces arbres sont très variables de 95,5/3,7 (%) pour Gp-87 à 67,0/32,3 (%) pour Gp-94. L'arbre Gp-94 est celui qui intègre la part la plus importante de Criollo dans son génome parmi les 21 arbres analysés. Ces 13 arbres peuvent donc être considérés comme des Trinitario issus de différentes générations de croisements entre Amelonado et Criollo. Après vérification, allèle par allèle, Gp-80 et Gp-81 apparaissent génétiquement identiques sur les 26 loci SSR testés. Ces 2 échantillons ont été collectés sur le même lieu et ont les mêmes données GPS (ANNEXE II), ils sont donc peut-être issu du même arbre ou d'un arbre génétiquement très proche.

Certains arbres apparaissent comme Amelonado (**Gp-78**, **Gp-79**, **Gp-83** et **Gp-92**) avec une contribution génétique du groupe Amelonado estimée à 99,3%.

Après vérification des allèles sur les 26 loci SSR testés, l'arbre **Gp-83** apparaît totalement Amelonado avec 25 loci (sur 26) identiques à MAT1-6 et 1 locus (mTcCIR40) identique à Catongo.

L'arbre **Gp-79** est aussi de type Amelonado et très proche génétiquement de MAT1-6 avec seulement 3 loci (homozygotes) différents dont les allèles sont partagés avec Catongo (2 loci) et des témoins ICS (1 locus).

Les arbres **Gp-78** et **Gp-92** sont identiques sur l'ensemble des loci SSR testés, et sont proches de MAT1-6, avec seulement 4 loci (sur 26 testés) différents. Ces 4 loci sont homozygotes avec des allèles provenant d'autres témoins Amelonado et/ou d'autres groupes génétiques.

Seuls 4 arbres apparaissent comme des hybrides plus complexes constitués de plusieurs groupes génétiques. L'arbre **Gp-75** est constitué principalement des groupes Amelonado, Iquitos et Contamana dans un ratio respectivement de 49,5/23,7/20,9 (%) en intégrant une petite part de Criollo (7%).

L'arbre **Gp-84** est constitué majoritairement d'Amelonado (51,2%) et Criollo (31,2%) mais intègre aussi dans son génome une petite partie des groupes Iquitos (10,2%) et Nanay (6,6%).

Les arbres **Gp-90** et **Gp-91** sont constitués des groupes Iquitos/Criollo/Amelonado avec respectivement des ratios de 44,4/28,1/20 (%) pour **Gp-90** et 56,3/29,0/12,7 (%) pour **Gp-91**. Ce dernier est l'arbre qui contient la part d'Amelonado (12,7%) la plus faible des 21 arbres testés.

### **III.3. Hétérozygotie**

Le taux d'hétérozygotie des arbres testés et des témoins a été évalué sur 26 loci et reporté sur le tableau ci-après :



Nom groupe ou Population	Individu ou groupe	% Hétérozygotie
Contamana	Moyenne groupe	38.11
Nacional	Moyenne groupe	12.79
Morona	Moyenne groupe	45.19
Purus	Moyenne groupe	52.41
Curaray	Moyenne groupe	21.21
Iquitos	Moyenne groupe	66.36
Nanay	Moyenne groupe	35.95
Guyana	Moyenne groupe	9.57
Maranon	Moyenne groupe	40.50
Criollo	Moyenne groupe	2.07
Amelonado	Moyenne groupe	0.00
Trinitario	UF676	100.00
Trinitario	ICS1	73.08
Trinitario	ICS100	30.77
Trinitario	ICS40	100.00
Trinitario	ICS43b	84.00
Trinitario	ICS46	92.31
Trinitario	ICS6	57.69
Trinitario	ICS60	88.46
Trinitario	ICS84	69.23
Trinitario	ICS95	84.00
PopGuadeloupe	Gp_75	71.43
PopGuadeloupe	Gp_76	19.05
PopGuadeloupe	Gp_77	52.00
PopGuadeloupe	Gp_78	0.00
PopGuadeloupe	Gp_79	8.00
PopGuadeloupe	Gp_80	48.00
PopGuadeloupe	Gp_81	48.00
PopGuadeloupe	Gp_82	44.00
PopGuadeloupe	Gp_83	0.00
PopGuadeloupe	Gp_84	37.50
PopGuadeloupe	Gp_85	12.00
PopGuadeloupe	Gp_86	12.50
PopGuadeloupe	Gp_87	20.00
PopGuadeloupe	Gp_88	48.00
PopGuadeloupe	Gp_89	24.00
PopGuadeloupe	Gp_90	64.00
PopGuadeloupe	Gp_91	24.00
PopGuadeloupe	Gp_92	0.00
PopGuadeloupe	Gp_93	48.00
PopGuadeloupe	Gp_94	54.17
PopGuadeloupe	Gp_95	25.00

Ces valeurs montrent un taux d'hétérozygotie très variable selon les cacaoyers testés. Certains arbres ont montré un taux d'hétérozygotie assez élevé notamment **Gp-75** avec 71.43 % de loci hétérozygotes. D'autres, en revanche, ont été révélés faiblement hétérozygotes comme **Gp-79** (8%) ou entièrement homozygotes sur les 26 loci analysés comme les arbres **Gp-78**, **Gp-83** et **Gp-92** (0%).

Une analyse de l'hétérozygotie regroupant l'ensemble des arbres (différents envois) collectés en Guadeloupe est présentée en ANNEXE V.

### III.4. Prédiction de l'auto-compatibilité des arbres

Une analyse prédictive de l'auto-compatibilité des arbres a été réalisée à l'aide de marqueurs moléculaires de type microsatellites (mSi303 et mSi460) récemment développés (Lanaud *et al.*, 2017). La prédiction a été réalisée par l'identification des allèles de type "Amelonado" et de type "Criollo" qui confèrent l'auto-compatibilité (SC) aux arbres à l'état homozygote, et inversement l'auto-incompatibilité (SI) à l'état hétérozygote.

Name	mSi303	mSi303	mSi460	mSi460	SI/SC	Type SC
Criollo	217	217	132	132	SC	SC type "Criollo"
Amelonado	227	227	159	159	SC	SC type "Amelonado"
Gp_75	221	239	132	146	-	
Gp_76	217	227	132	159	SI	
Gp_77	227	227	159	159	SC	SC type "Amelonado"
Gp_78	227	227	159	159	SC	SC type "Amelonado"
Gp_79	227	227	159	159	SC	SC type "Amelonado"
Gp_80	227	237	132	159	SI	
Gp_81	227	237	132	159	SI	
Gp_82	217	227	132	159	SI	
Gp_83	227	227	159	159	SC	SC type "Amelonado"
Gp_84	233	237	132	132	SC	SC type "Criollo"
Gp_85	227	227	159	159	SC	SC type "Amelonado"
Gp_86	217	227	132	159	SI	
Gp_87	217	227	132	159	SI	
Gp_88	217	227	132	159	SI	
Gp_89	217	227	132	159	SI	
Gp_90	237	239	138	142	-	
Gp_91	237	237	142	142	-	
Gp_92	227	227	159	159	SC	SC type "Amelonado"
Gp_93	227	227	159	159	SC	SC type "Amelonado"
Gp_94	227	227	159	159	SC	SC type "Amelonado"
Gp_95	227	227	159	159	SC	SC type "Amelonado"

Il a donc été possible d'établir une prédiction sur 18 arbres (10 SC et 8 SI), 3 arbres n'ont pas pu être déterminés.

Les arbres auto-compatibles (SC) et de même type, croisés entre eux, donneront des plantes auto-compatibles (par exemple **Gp-77** x **Gp-78**). Un croisement entre 2 plantes SC de type différent (par exemple **Gp-77** x **Gp-84**) donnera des plantes auto-incompatibles.

## **IV Conclusion**

Les arbres collectés en Guadeloupe testés dans cette analyse ont montré globalement un taux d'hétérozygotie très variable pouvant être très élevé dans certain cas. Trois arbres (Gp-78, Gp-83 et Gp-92) ont été détectés entièrement homozygote sur l'ensemble des 26 loci SSR testés.

La majorité (13 sur 21) des arbres (Gp-76, Gp-77, Gp-80, Gp-81, Gp-82, Gp-85, Gp-86, Gp-87, Gp-88, Gp-89, Gp-93, Gp-94 et Gp-95) apparaissent comme hybrides entre les groupes Amelonado et Criollo, et peuvent donc être considérés comme des Trinitario issus de différentes générations de croisements entre Amélonado et Criollo.

Les arbres GP-79 et GP-83 sont de type Amelonado et très proche de MAT1-6, sur les 26 loci testés.

Les arbres GP-78 et GP-92 sont identiques entre eux et proches de MAT1-6, avec seulement 4 loci différents (sur 26 testés), ces 4 loci sont homozygotes avec des allèles provenant d'autres individus Amelonado ou d'autres groupes génétiques.

Les arbres (Gp-75, Gp-84, Gp-90 et Gp-91) apparaissent comme des hybrides plus complexes constitués d'au moins 3 groupes génétiques.

Il a été possible d'établir une prédiction de l'auto-compatibilité (SC) ou l'auto-incompatibilité (SI) sur 18 arbres, 9 arbres SC type "Amelonado" (Gp-77, Gp-78, Gp-79, Gp-83, Gp-85, Gp-92, Gp-93, Gp-94 et Gp-95), 1 arbre SC type "Criollo" (Gp-84), 8 arbres SI (Gp-76, Gp-80, Gp-81, Gp-82, Gp-86, Gp-87, Gp-88 et Gp-89). Trois arbres n'ont pas pu être déterminés (Gp-75, Gp-90 et Gp-91).

## **Référence**

Allegre, M., Argout, X., Boccara, M., Fouet, O., Roguet, Y., Bérard, A., Thévenin, J.M., Chauveau, A., Rivallan, R., Clement, D., 2011. **Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of *Theobroma cacao* L.** DNA Res. 19, 23–35.

Falush, D., Stephens, M., Pritchard, J.K., 2003. **Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies.** Genetics 164, 1567–1587.

Fouet, O., Allegre, M., Argout, X., Jeanneau, M., Lemainque, A., Pavék, S., Boland, A., Risterucci, A.M., Loor, G., Tahi, M., 2011. **Structural characterization and mapping of functional EST-SSR markers in *Theobroma cacao*.** Tree Genet. Genomes 7, 799–817.

Lanaud, C., Fouet, O., Legavre, T., Lopes, U., Sounigo, O., Eyango, M.C., Mermaz, B., Silva, D., Ramos, M., Solorzano, L., Gaston, R., Argout, X., Gyapay, G., Ebaiarrey, H.E., Colonges, K., Sanier, C., Rivallan, R., Mastin, G., Cryer, N., Boccara, M., Verdeil, J.-L., Mousseni, E., Bruno, I., Peres Gramacho, K., Clément, D., Wilson, Z., 2017. **Deciphering the *Theobroma cacao* self-incompatibility system: from genomics to diagnostic markers for self-compatibility.** J. Exp. Bot. 68, 4775–4790.

Motamayor, J.C., Lachenaud, P., Mota, J.W. da S. e, Loor, R., Kuhn, D.N., Brown, J.S., Schnell, R.J., 2008. **Geographic and Genetic Population Differentiation of the Amazonian Chocolate Tree (*Theobroma cacao* L).** PLOS ONE 3, e3311.

Perrier, X., Jacquemoud-Collet, J.P., 2006. **DARwin software.**

Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Risterucci, A.M., Grivet, L., N’Goran, J.A.K., Pieretti, I., Flament, M.H., Lanaud, C., 2000. **A high-density linkage map of *Theobroma cacao* L.** Theor. Appl. Genet. 101, 948–955.