2017 International Symposium on Cocoa Research (ISCR), Lima, Peru, 13-17 November 2017

Exploration of the *T. cacao* genome sequence to decipher the incompatibility system of *Theobroma cacao* and to identify diagnostic markers

Claire Lanaud[1†], Olivier Fouet[1†], Thierry Legavre[1†], Uilson Lopes[2†], Olivier Sounigo[3], Marie Claire Eyango[4], Benoit Mermaz[1], Marcos Ramos da Silva[2], Rey Gaston Loor Solorzano[5], Xavier Argout[1], Gabor Gyapay[6], Herman Ebai Ebaiarrey[4], Kelly Colonges[1], Christine Sanier[1], Ronan Rivallan[1], Géraldine Mastin[1], Nicholas Cryer[7,] Michel Boccara[1], Ives Bruno Efombagn Mousseni[4], Karina Peres Gramacho[2], Didier Clement[1]

[†] These authors contributed equally to this study.

1- Centre de Cooperation Internationale en Recherche Agronomique pour le Developpement (CIRAD), UMR AGAP, F-34398 Montpellier, France.
AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France.
2-Centro de Pesquisas do Cacau (CEPEC), CEPLAC, Rod. Ilhéus-Itabuna, km 22, Ilhéus, BA 45605-350, Brazil
3-Centre de Cooperation Internationale en Recherche Agronomique pour le Developpement (CIRAD), UR Bioagresseurs, BP 2572 Elig-Essono, Yaoundé, Cameroun
4-Institut de Recherche Agricole pour le Developpement (IRAD) PO Box 2123, Yaoundé, Cameroun
5-Instituto Nacional de Investigaciones Agropecuarias (INIAP), EET-Pichilingue. CP 24 Km 5 vía Quevedo El Empalme, Quevedo, Los Ríos, Ecuador
6-Commissariat à l'Energie Antomique (CEA), Institut de Génomique (IG), Genoscope, Evry, France
7-Mondelez UK R&D Limited, PO Box 12, Bournville Place, Bournville Lane, Birmingham, B30 2LU, United Kingdom

Abstract

We explored the *Theobroma cacao* genome sequence to progress in the knowledge of the *T. cacao* incompatibility system. Cocoa self-compatibility is an important yield factor and has been described as controlled by a late gameto-sporophytic system involving several locus, and resulting in gametic non-fusion. In this work, we identified two different mechanisms controlling the *T. cacao* self-incompatibility system at two separate loci, located on chromosome one and four (CH1 and CH4). Both loci are responsible for gametic selection, but only one (the CH4 locus) is involved in the main fruit drop. The CH1 locus acts prior to gamete fusion and independently of CH4 locus. Fine mapping and genome wide association studies focused analyses of restricted regions without recombinant plants where several candidate genes were identified. Their expression analysis showed differential expression during incompatible or compatible reactions for some of them. Highly polymorphic SSR diagnostic markers, designed in the CH4 region identified by fine mapping, allowed the development of efficient diagnostic markers predicting self-compatibility and fruit setting according to allele or genotype presence. SSR alleles specific to self-compatible Amelonado and Criollo varieties were also identified allowing screening for self-compatible plants in cocoa populations.

Introduction

One important *T. cacao* yield factor is the self-compatibility status of cocoa trees. Self-compatible (SC) young trees would produce an average of 66% more fruits than the self-incompatible (SI) trees (Cope, 1939a). A higher proportion of self-compatible trees was also observed among higher producing trees (Lachenaud *et al*., 2005). Various mechanisms of self-incompatibility (SI), preventing the self-fertilisation of plants, have been described in angiosperms (de Nettancourt, 1997; Takayama *et al*, 2005; Rea *et al.*, 2008). These mechanisms act at the level of pollen growth inhibition for most species (gametophytic (GSI) and sporophytic (SSI) self-incompatibility systems) ((McClure, 2008; Suzuki et al. 1999), or at the ovary level (late acting self-incompatibility system (LSI), (Gibbs, 2014)), leading to an embryo development failure. The SI mechanism *in T. cacao* has been described as a LSI system: in SC and SI reactions the pollen can reach the ovary with a similar rate of pollen tube growth 4h after pollination (Bouharmont, 1960), but the double fertilization is completed after 24h in SC reactions, whereas in SI reactions male nuclei are released into the embryo sac, but fail to achieve gamete fusion in part of the ovules, resulting in floral abscission (Cheesman, 1927; Cope, 1939b, 1940, 1958, 1962; Knight and Rogers 1955; Posnette, 1940; Bouharmont, 1960; Ford *et al.,* 2012). The genetic control of *T. cacao* SI was studied by several authors (Cope, 1939b, 1940, 1958, 1962; Knight and Rogers,

1955; Glendinning, 1967), who hypothesized the existence of a S-locus and several alleles with dominance relationships between them, including an amorphous *Sf* allele, present in the SC Amelonado variety and leading to self-compatibility when homozygous. In addition, two other loci involved in *T. cacao* self-incompatibility were also hypothesized (Cope, 1962), based on cytological studies showing 25%, 50% or 100% of ovules without gamete fusion, after selfing of SI cocoa trees. It was concluded that the genetic system of SI in cacao has both aspects of sporophytic and gametophytic controls (Cope, 1958).
Two loci putatively involved in self-incompatibility observed by the % of fruit setting, were localised by QTL analysis at the top of chromosome 4 (Crouzillat *et al.*, 1996; Royaert et *al.*, 2010 ; Yamada *et al.*, 2010), and in chromosome 7 (Yamada *et al.*, 2010).
The gamete fusion failure, observed in part of the ovules and associated to fruit abscission, was observed only at the embryo sac level (Cope, 1939b; Bouharmont, 1960). However, fruits can reach maturity after self-pollination of a SI tree when compatible pollen is mixed with pollen from the SI trees ("mentor pollen" effect) (Opeke and Jacob, 1967 ; Bartley, 1969, 2005 ; Lanaud *et al.*, 1987 ; Glendinning, 1960), allowing the development of fertilized ovules. With the help of molecular markers, the observations, among these progenies, of skewed segregations potentially linked to SI is then possible.
Here, we report the analysis of a large F2 progeny produced from a SI tree using the "mentor pollen" effect. A first molecular analysis, performed on a subsample of this F2 progeny, revealed a skewed segregation at the level of CH1 and CH4 regions (Allegre *et al*., 2012). In this work, we carried out fine mapping of these genome regions where candidate genes were identified and their expression characterised during SC and SI reactions. In order to refine the existing predictive model for self-compatibility (Da Silva *et al*., 2016), multi-allelic diagnostic SSR markers were searched in the CH4 region identified by fine mapping.

Materials and Methods

Material

***Mapping population used for fine mapping*:** A large progeny was produced by CEPLAC, by selfing the clone TSH516, a hybrid of ICS 1 and Scavina 6 (SCA 6). Until 877 individuals from this progeny were analysed. Self seeds were produced with mixed pollinations using *Herrania* mariae pollen as mentor pollen, followed by pollination with TSH516 self-pollen.

Analyses of potential skewed segregations in 2 other progenies

• A progeny of 550 individuals was created at IRAD (Barombi-kang, Cameroon) from IMC 60 (SI) using a mixture of pollens of IMC 60 and of Catongo as mentor pollen. Self progenies were identified using molecular markers.

• A progeny of 96 plants from the cross UF 676 x ICS 95, planted in French Guyana, was analysed to observe the segregations in the CH1 locus.

Cocoa populations used for genome wide association study (GWAS) and to analyse prediction level of diagnostic markers

A population of 710 individuals evaluated for self-incompatibility, and from different origins (a farm and a breeding populations from IRAD (Cameroun) breeding populations from CEPLAC (Brazil) and INIAP (Ecuador), collections from INIAP, CRC (Trinidad and Tobago) and CIRAD (France)) was used to assess the efficiency of prediction based on SSR genotypes and alleles. A subset of 570 individuals from these populations was analysed by GWAS using SNP markers revealed by GBS (genotyping by sequencing).

Samples for gene expression analyses

Pollinations were carried out on the SCA 6 SI clone, present in the greenhouse of CIRAD/Montpellier, using pollen from ICS 1 (SC) or from SCA 6 (SI), and ovaries with their pistil (named "ovaries" in the text) were collected at different times after pollination. QPCR experiments were conducted on three biological repetitions of each RNA extract from these samples to analyse gene expression.

Methods

Genotyping

F2 progeny: new markers were defined in the CH1 and CH4 regions:
SSR markers were defined using the whole sequence of the Criollo genome V1 (http://cocoagendb.cirad.fr/gbrowse/cgi-bin/gbrowse/theobroma/) (Argout et *al.*, 2011), and the tool « Search for SSR » integrated in the ESTTIK database (http://esttik.cirad.fr/cgi-bin/SSR_server.cgi). Deletions/insertions were identified in the CH4 region between the Criollo genome and the Amelonado genome (http://www.cacaogenomedb.org/) after alignment of both sequence fragments.

GWAS population: The GWAS population was genotyped by sequencing (GBS) using the DArTseq technology after DNA restriction with *Pst*I and *Mse*I (Killian *et al.,* 2012).

***Genome wide association studies*:**

GWAS was conducted using Tassel 5.2.31 software (Glaubitz *et al.,* 2014) on 570 individuals assessed for SC/SI coded as 0 (SI) and 1 (SC) and on a subsample of 388 individuals, using the % of fruit setting 14 days after self-pollination, taken as a quantitative variable to assess SC/SI status. These individuals were genotyped using 16480 SNPs (GBS) stored in the TropGENE-DB (http://tropgenedb.cirad.fr/tropgene/) and with a minor allele frequency > 0,05. The structure of the population was determined with a subset of 150 SNPs distributed over all chromosomes, and a bayesian clustering method implemented in the STRUCTURE software (Pritchard *et al.,* 2000), with a burning period of 100,000 iterations, 500,000 Markov Chain Monte Carlo repetitions and ten independent runs.

Search for candidate genes and gene expression analyses

The expression of candidate genes, identified in the CH1 and CH4 regions using the Criollo cocoa genome sequence, version V1, available at genome (http://cocoagendb.cirad.fr/tools.html) (Argout *et al.*, 2011), was analysed during SC and SI reactions carried out on the SCA 6 accession. Q-PCR were conducted using a Roche LightCycler 480 Real-time PCR System and a SYBR Green dye included in the supermix to detect dsDNA amplification products, and with two reference genes used for normalization: Tc04_g000050 (Isocitrate dehydrogenase) and Tc08_g003640 (Tubulin beta-6 chain).

Immunolocalization of Tc01_g007270 and Tc01_g007290 proteins

Rabbit polyclonal antibodies, anti-Tc01_g007270 and anti-Tc01_g007290, were produced by Eurogentec (Anti-peptide Speedy 28-Day https://secure.eurogentec.com/speedy.html) using sythetized peptides as antigen. Peptides sequences LGNDKTVRIWTQENE, corresponded to residues 310-324 of Tc01_g007270 protein and RSVDKSNDESESQVS corresponded to residues 478-492 of Tc01_g007290 protein. Immunolocalization was revealed with a Alexa Fluor 488 dye conjugated goat anti-rabbit antiboddies (Interchim, France, Montluçon). The microscope imaging was performed in Montpellier RIO Imaging Platform (http://www/mri/cnrs.fr) with a confocal microscope (LSM510, Meta; Carl Zeiss Micro Imaging).

Identification of diagnostic markers

Eleven SSR markers (Table 1), identified in the CH1 and CH4 regions, were used to genotype a population of 710 individuals, and to establish predictions for SC/SI according to the genotype or allele, considered alone or in combination with one or two other markers.

Table 1: SSR or INDEL markers defined in the CH1 and CH4 regions and used to establish predictions of SC or SI cocoa plants

Markers	Chromosome	Position	type	5'-3' forward primer	Tm	5'-3' reverse primer	Tm	PCR product size (bp)
mSI_103	CH1	4021267	(TG)6(TA)7	CAGGCTGCCATTTTCTC	55,1	TCAAGGACTGCTCCAAAA	55,2	209
mSI_107	CH1	4130575	(AT)10	GAAAATACTCCGTAACAACCA	54,4	ACCTTACCAACACCACACA	54,6	221
mSI_7	CH4	20673	(AGA)8	TTTCATGGAGGTTGGGA	55,5	GTTGCACAAAGGATGGG	55,7	183
mSI_35	CH4	33618	(AG)14	TCCCGATAGCCTCAACA	56,0	ACAAATTCCTTCATCCCCT	55,9	122
mSI_2	CH4	43494	(TA)9	CATCGAAAGTCAAGAAAAGG	55,1	ATTGAAATGGTGGTTTGGT	55,1	268
mSI_303	CH4	119995	(AT)11	CAAGTCGTTGGGAGGG	55,7	AAAGTTTCAATCCCATTTCC	55,6	255
mSI_458	CH4	136890	(TA)11	GACACGAGATGTATCCTGACCA	59,3	TGCAACCGTGAGCATTTTGT	59,3	284
mSI_460	CH4	139590	(TC)8	TGAGAACAAAGCCAAAGAAAGGA	58,7	CCGAGACAAAGCCCAGAAG	58,2	117
mSI_315	CH4	233706	(TC)6	CAAGGGGTCTTGGGTTT	55,7	AATGATGGCGATGGAGA	55,7	206
mSI_408	CH4	236686	INDEL	TGCAGAGGCCATGCGAGTAT	61,7	TGCACTTGAAAGAGGGGGAA	59,2	244
mSI_411	CH4	258684	INDEL	CGCCAGGCATTCTTACTCTT	58,0	ATACTGGACATCTTGTTGAATGAC	57,0	274

Allelic frequencies and identification of alleles specific to Amelonado and Criollo SC varieties were search in a subset of 108 *T. cacao* genotypes, capturing the diversity of the *T. cacao* genetic groups. Reference SSR profiles allowing cocoa breeders to characterise clones for potential SC/SI status were established for these SSR (Table 2).

Prediction analyses were carried out according to genotype or allele presence, using SAS software modules (SAS Institute Inc., 2004) using Fisher´s Exact Test and PROC FREQ. Probabilities.

Table 2: Newly references SSR profiles were established for a collection of diverse *T. cacao* clones available from international germplasm collections, and which could be used as standards for prediction of self-compatibility status of cocoa trees.

	Chromosome	CH4	CH4	CH4	CH4	CH4	CH4	CH4	CH4	CH4	CH1	CH1
	mk position	20673	33618	43494	119995	136890	139590	233706	236686	258684	4024677	4130575
Genetic group	marker / clone	mSi7	mSI_35	mSi_2	mSI_303	mSI_458	mSI_460	mSi_315	mSi_408	mSi_411	mSI_103	mSI_107
Amelonado	CATONGO	200/200	119/119	283/283	227/227	292/292	161/161	247/247	272/272	286/286	223/223	245/245
Amelonado	Matina 1-6	200/200	119/119	283/283	227/227	292/292	161/161	247/247	272/272	286/286	223/223	245/245
Contamana	Scavina 12	178/194	129/139	293/293	239/241	302/322	140/146	247/249	260/266	284/284	225/225	243/257
Contamana	Scavina 6	178/194	129/139	291/291	221/221	298/306	140/149	245/249	266/272	284/290	225/225	241/243
Curaray	LCTEEN_188	178/194	139/139	287/287	215/223	300/306	132/159	247/249	260/260	284/292	225/228	250/265
Curaray	LCTEEN_255	194/200	119/129	281/287	225/225	292/308	142/159	247/247	272/272	284/284	228/228	233/233
Curaray	LCTeen_32	194/194	139/146	287/287	215/233	300/304	146/159	249/249	260/260	292/292	225/231	235/235
Curaray	LCTEEN_327	194/194	139/139	287/287	215/215	300/300	159/159	249/249	260/260	292/292	225/225	235/235
Curaray	LCTEEN_36	194/197	142/142	275/275	209/209	307/320	132/142	247/249	272/272	286/290	225/225	235/235
Curaray	LCTEEN_37	194/197	119/137	283/288	211/211	294/304	140/146	245/247	260/260	284/284	225/225	235/243
Curaray	LCTEEN_403	197/197	119/142	283/283	209/235	290/290	132/146	247/249	260/272	284/290	225/231	235/235
Curaray	LCTEEN_189	178/194	119/139	287/287	215/223	300/306	132/159	247/249	260/272	284/292	225/228	250/265
Guiana	GU114_P	197/197	119/151	283/283	231/243	304/304	145/151	245/247	272/272	284/286	225/228	245/245
Guiana	GU151_F	197/197	119/119	283/283	243/243	304/304	151/151	247/247	272/272	284/284	225/225	245/245
Guiana	GU195_P	194/197	119/119	283/283	233/233	304/304	145/151	247/249	272/272	284/292	228/228	245/245
Guiana	GU219_P	197/197	119/119	283/283	243/243	304/304	151/151	247/247	272/272	284/290	228/228	245/245
Guiana	GU241_P	194/197	119/119	283/283	233/243	304/304	145/151	247/249	272/272	286/292	228/228	245/245
Guiana	GU261_P	197/197	119/119	283/283	243/243	304/304	151/151	247/247	272/272	284/284	225/228	245/245
Guiana	GU277_G	197/197	119/119	283/283	243/243	304/304	151/151	247/247	272/272	284/284	225/228	245/245
Guiana	GU286	197/200	119/151	283/283	231/243	304/304	151/151	245/247	272/272	286/286	225/228	245/245
Guiana	GU310_P	194/197	119/119	283/283	233/243	304/304	145/151	247/249	272/272	284/292	223/228	245/245
Guiana	GU335_P	194/197	119/119	283/283	233/243	304/304	145/151	247/249	272/272	284/292	225/228	245/245
Iquitos	IMC105	194/200	119/135	283/283	225/233	302/304	133/145	245/249	260/272	286/292	228/228	235/243
Iquitos	IMC107	178/194	135/142	278/283	233/233	304/304	133/140	245/245	266/272	284/292	223/228	243/245
Iquitos	IMC2	178/178	135/142	278/278	227/227	304/306	140/151	245/249	266/272	284/290	228/228	235/243
Iquitos	IMC48	194/200	119/135	283/283	225/233	302/304	130/145	245/249	260/272	286/292	223/228	243/245
Iquitos	IMC50	178/194	135/142	278/283	233/233	304/304	130/137	245/245	266/272	284/292	228/228	235/243
Iquitos	IMC55	178/194	135/142	278/283	233/233	304/304	130/137	245/245	266/272	284/292	223/228	243/245
Iquitos	IMC60	194/194	135/151	283/283	233/233	304/306	133/145	247/247	272/272	284/286	223/228	233/245
Iquitos	IMC76	194/200	119/135	283/283	225/233	302/304	133/145	245/249	260/272	286/292	228/228	235/243
Iquitos	IMC98	178/178	142/151	278/278	237/237	304/306	140/145	245/247	266/272	286/286	228/228	235/243
Marañón	PA141	194/200	119/135	283/283	231/231	312/312	132/146	247/247	272/272	286/286	228/228	245/245
Marañón	PA151	194/200	119/135	283/283	231/231	312/312	132/146	247/247	272/272	286/286	228/228	245/245
Marañón	PA16	194/194	135/135	283/283	231/231	312/312	132/146	247/247	272/272	286/286	225/228	245/250
Marañón	PA30	194/194	135/135	283/283	231/231	312/312	132/146	247/247	272/272	286/286	228/228	245/245
Marañón	PA32	194/194	135/135	283/283	231/231	304/304	133/139	247/247	272/272	286/286	228/228	243/243
Marañón	PA39	194/200	119/135	283/283	231/231	312/312	132/145	247/249	272/272	286/286	223/228	245/245
Marañón	PA7a	194/194	129/135	283/283	223/231	318/318	129/145	247/249	266/272	286/286	228/228	245/245
Nacional	MO109	200/200	119/119	283/283	227/227	292/292	161/161	247/247	272/272	286/286	223/223	245/245
Nacional	MO96	178/194	137/142	278/278	211/211	304/304	145/146	247/249	260/260	284/284	233/233	245/247
Nanay	NA30	200/200	119/119	283/283	227/227	292/292	161/161	247/247	272/272	286/286	223/228	245/245
Nanay	NA32	194/194	135/151	283/283	233/233	304/306	130/145	247/247	272/272	284/286	223/223	245/245
Nanay	NA34	194/194	135/151	283/283	233/233	304/306	130/145	247/247	272/272	284/286	223/228	233/245
Nanay	NA84	194/194	135/135	283/283	233/233	304/304	133/133	247/247	272/272	284/284	228/228	243/243
Nanay	P10-C	194/194	119/135	283/283	233/233	304/304	133/133	247/247	272/272	286/286	228/228	243/243
Nanay	P25-A	194/194	135/135	283/283	233/233	304/304	133/133	247/247	272/272	284/284	228/228	243/243
Nanay	P26	194/197	119/135	283/283	233/235	304/304	133/139	247/247	272/272	292/292	228/228	243/243
Nanay	P32-A	194/197	119/135	283/283	233/235	304/304	133/139	247/247	272/272	284/292	228/228	243/243
Purus	LCTEEN_362	178/194	137/139	297/313	241/241	300/320	135/149	247/249	260/272	292/292	228/228	257/261
Trinitario	GS29	194/200	119/137	280/283	227/237	292/300	134/161	247/247	260/272	286/292	223/228	239/245
Trinitario	GS77	200/200	119/119	283/283	227/227	292/292	161/161	247/247	272/272	286/286	223/228	239/245
Trinitario	ICS1	200/200	119/119	283/283	227/227	292/292	161/161	247/259	272/272	286/286	223/228	239/245
Trinitario	ICS100	194/200	119/137	280/283	227/227	292/292	134/161	247/247	260/272	286/292	228/228	239/239
Trinitario	ICS15	194/200	119/137	280/283	227/239	292/300	134/161	247/247	260/272	286/292	223/228	239/245
Trinitario	ICS24	200/200	119/119	283/283	227/227	292/292	161/161	247/247	272/272	286/286	223/223	245/245
Trinitario	ICS27	200/200	119/119	283/283	227/227	292/292	161/161	247/247	272/272	286/286	223/228	239/239
Trinitario	ICS40	194/194	135/151	283/283	231/237	306/306	145/145	247/247	272/272	286/286	225/228	235/243
Trinitario	ICS46	194/194	135/151	283/283	231/237	306/306	145/145	247/247	272/272	286/286	225/228	235/243
Trinitario	ICS52	194/200	119/137	280/283	227/233	292/300	134/161	247/247	260/272	286/292	223/228	239/245
Trinitario	ICS53	197/197	119/151	283/283	235/239	302/306	139/145	247/247	272/272	286/292	223/228	243/245
Trinitario	ICS61	178/200	119/139	283/283	227/239	292/322	140/161	247/249	266/272	284/286	225/228	241/241
Trinitario	ICS62	197/200	119/119	283/283	225/235	302/302	145/145	247/249	260/272	286/292	225/228	235/241
Trinitario	ICS67	178/197	119/139	283/283	235/239	302/302	140/140	247/249	266/272	284/292	223/225	241/245
Trinitario	ICS73	178/194	135/139	283/283	233/239	304/304	140/140	245/249	266/272	284/292	223/228	245/245
Trinitario	ICS76	194/200	119/137	280/283	227/239	292/300	134/161	247/247	260/272	286/286	228/228	239/239
Trinitario	ICS77	178/200	119/139	283/283	227/239	292/322	140/161	247/249	266/272	284/286	223/225	241/245
Trinitario	ICS8	200/200	119/119	283/283	227/227	292/292	161/161	247/249	272/272	286/286	223/228	239/245
Trinitario	ICS83	194/200	119/137	280/283	227/239	292/300	134/161	247/247	260/272	286/292	223/223	245/245
Trinitario	UF676	194/200	119/137	283/285	217/227	292/300	134/161	247/247	260/272	286/292	223/228	239/245
hybrid	EET_103	194/200	119/139	283/287	215/227	292/300	138/161	247/249	260/272	286/292	223/233	245/245
hybrid	EET_399	194/197	119/133	283/285	223/235	302/302	140/140	247/247	266/272	286/292	228/231	235/243

Results

Study of skewed segregations linked to the incompatibility system

***Fine mapping of restricted genome regions including the skewed segregations in the F2 progeny*:**

• In the skewed CH1 and CH4 regions, respectively 22 and 26 new markers segregating in the F2 progeny could be defined. The expected a (homozygous)/h (heterozygous)/b (homozygous) genotype segregations for this F2 was 1/2/1, whereas only two genotypes were predominant in thess regions, with the “a” Amelonado genotype totally absent in a small region of 70,3 Kb of CH1, and the scavina 6 allele (b) totally absent in a region of 257,2 Kb of CH4 (Fig 1).

Markers	CH	Position	Number of individuals			
			a	h	b	Total
mSI_26	CH1	3377732	5	228	118	351
mSI_88	CH1	3499444	5	224	111	340
mSI_89	CH1	3525756	5	233	114	352
mSI_32	CH1	3649333	4	236	120	361
mTcCIR15	CH1	3711664	4	227	117	348
mSI_73	CH1	3790637	4	291	138	429
mSI_101	CH1	3935902	3	243	119	365
mSI_102	CH1	3966163	2	249	119	370
mSI_140	CH1	3988656	2	217	125	344
mSI_141	CH1	4010921	2	308	138	448
mSI_103	CH1	4024677	3	387	228	618
[illegible]56	CH1	4053385	0	508	253	761
[illegible]67	CH1	4054418	0	366	216	582
[illegible]69	CH1	4057532	0	419	221	640
[illegible]40	CH1	4066036	0	476	253	729
mSI_370	CH1	4070474	0	466	266	732
mSI_372	CH1	4073585	0	297	170	467
mSI_375	CH1	4091577	2	429	228	659
mSI_107	CH1	4130575	4	374	212	590
mTcCIR356	CH1	4149062	6	232	118	354
mSI_112	CH1	4233257	6	339	163	502
mSI_113	CH1	4252975	6	284	172	456

70,3 Kb

Markers	CH	Position	Number of individuals			
			a	h	b	Total
mSI_462	CH4	1414	222	425	2	649
mSI_466	CH4	4737	191	385	0	576
mSI_474	CH4	10127	198	380	0	578
mSI_7	CH4	20673	259	488	0	747
mSI_34	CH4	28166	238	487	0	725
mSI_8	CH4	28166	142	270	0	412
mTcCir312	CH4	32259	277	516	0	793
mSI_35	CH4	33618	183	348	0	531
mSI_2	CH4	43494	191	335	0	526
mSI_542	CH4	63388	33	68	0	101
mSI_303	CH4	119995	299	578	0	877
mSI_458	CH4	136890	208	399	0	607
mSI_460	CH4	139590	198	406	0	604
mSI_308	CH4	139780	211	408	0	619
mSI_309	CH4	141679	205	411	0	616
mSI_310	CH4	142517	193	383	0	576
mSI_315	CH4	233706	301	563	0	864
mSI_402	CH4	246098	259	506	0	765
mSI_535	CH4	252815	110	232	0	342
mSI_411	CH4	258684	125	242	1	368
mS_413	CH4	270916	118	253	1	372
mSI_39	CH4	278179	277	421	1	699
mSI_42	CH4	343424	109	233	2	345
mSI_46	CH4	428250	119	230	2	352
mSI_54	CH4	751986	55	127	4	521
mSI_294	CH4	1686245	109	248	16	373

257 Kb

Fig 1: Segregations observed in the F2 progeny at the level of the CH1 and CH4 genome regions.

Skewed segregations observed in other progenies (Table 3).

Self-progeny from IMC 60: among the 39 self-fertilized seeds identified from the 550 individuals, skewed segregations, with a complete absence of one homozygous genotype, were only observed in the CH4 region and not in the CH1 region.

Segregations observed in UF 676 x ICS 95 involving the Amelonado CH1 allele: Each Trinitario parent of this progeny, hybrid between Amelonado and Criollo genotypes, is heterozygous for the mSI_103 (223) Amelonado allele and the mSI_103 (228) allele originated from Criollo. No skewed segregation was observed in this progeny.

Segregations observed in the Self-compatible F2 plants, homozygous for the CH4 loci (Sf/Sf) and heterozygous for the CH1 loci, after self pollinations: Two F2 plants, BR36 and BR59, homozygous *Sf/Sf* for the CH4 locus and heterozygous for the CH1 locus, were self-pollinated. It was observed in both cases, as in the F2 progeny, a total absence of the "a" genotype corresponding to plants homozygous for the Amelonado allele (mSI_103-223) for the CH1 locus. These results show clearly that even if the plants are SC due to the *Sf/Sf* CH4 genotype, the CH1 locus is still functional, and a genotypic selection could still be observed.

Table 3: segregations observed at the level of CH1 and CH4 loci for other progenies.

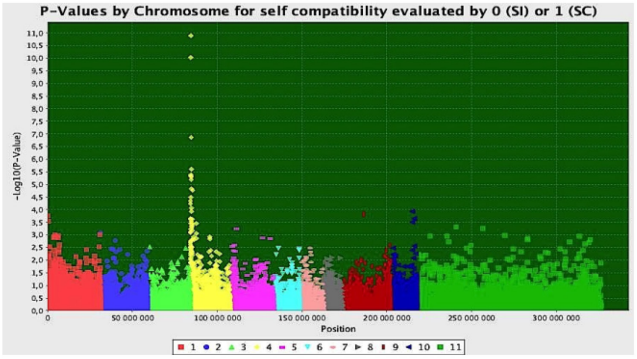
Progeny	marker	CH	parental genotypes	number of seeds analysed	expected genotypes	number
IMC60 x IMC60 self-incompatible	mSI_460	CH4	133-145 x 133-145	39	145-145	0
					133-145	22
					133-133	17
	mSI_103	CH1	223-228 x 223-228	38	228-228	12
					223-228	19
					223-223	7
	mSI_107	CH1	233-245 x 233-245	38	233/233	12
					233/245	19
					245/245	7
BR36 x BR36 self-compatible	mSI_460	CH4	161-161 x 161-161	34	161-161	34
	mSI_103	CH1	223-225 x 223-225	34	223-223	0
					223-225	20
					225-225	14
	mSI_107	CH1	241-245 x 241-245	33	245-245	0
					241-245	21
					241-241	12
BR59 x BR59 self-compatible	mSI_460	CH4	161-161 x 161-161	109	161-161	109
	mSI_103	CH1	223-225 x 223-225	109	223-223	0
					223-225	77
					225-225	32
	mSI_107	CH1	241-245 x 241-245	108	245-245	0
					241-245	76
					241-241	32
UF676 x ICS95 cross-compatible	mSI_460	CH4	134-161 x 134-134	87	134-134	43
					134-158	44
	mSI_103	CH1	223-228 x 223-228	87	223-223	26
					223-228	39
					228-228	22
	mSI_107	CH1	239-245 x 239-245	87	239-239	23
					239-245	38
					245-245	26
Reference genotypes						
Amelonado Mat1-6	mSI_460	CH4	161-161			
	mSI_103	CH1	223-223			
	mSI_107	CH1	245-245			
Criollo B97-61	mSI_460	CH4	134-134			
	mSI_103	CH1	228-228			
	mSI_107	CH1	239-239			
Scavina 6	mSI_460	CH4	140-149			
	mSI_103	CH1	225-225			
	mSI_107	CH1	241-243			

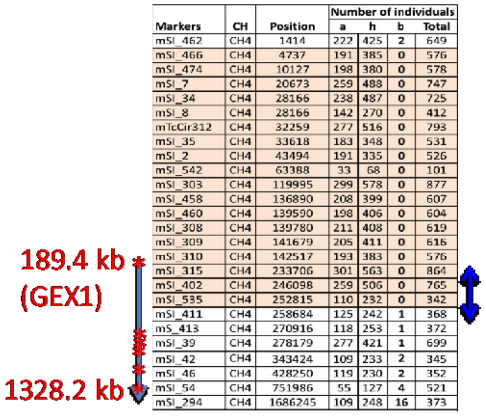
Origin of genotypic selection in self pollinated young fruits of BR59

In order to check if the genotypic selection provided by the CH1 locus could result from gamete non fusion, the proportion of fertilized ovules were assessed in young fruits of BR59 seven days after self-pollination. If the skewed segregations observed in the CH1 region reflect gamete non fusion events, a proportion of 25% aborted ovules is expected. The proportion of aborted ovules was observed similar between self pollinated ovaries and ovaries cross-pollinated with a compatible pollen from a different origin (CCN51), taken as control (respectively 11 % and 9.2 %). This shows that in that case, the SI reaction did not result in gamete non fusions and consequently aborted ovules.

Association studies for the fruit setting linked to self-incompatibility

Marker/incompatibility trait association studies, conducted on the 570 individuals evaluated by 0/1 for SC/SI trait using 16480 SNP, revealed positive associations for 11 SNP markers, all located at the top of CH4 from the positions 189447 bp to 1328172 bp (Fig. 2). Only one marker, the 3673000|F|0-37:G>A located at position 189447 bp is included in the restricted region identified by fine mapping which contained 36 other SNP markers subjected to GWAS. This marker is located inside the Tc04_g000230 gene, an ortholog to the *GEX1* gene of *Arabidopsis thaliana.* No association was detected in CH1. These observations revealed that only the CH4 locus is associated to fruit drop, contrary to the CH1 locus.





Markers	CH	Position	Number of individuals a	h	b	Total
mSI_462	CH4	1414	222	425	2	649
mSI_466	CH4	4737	191	385	0	576
mSI_474	CH4	10127	198	380	0	578
mSI_7	CH4	20673	259	488	0	747
mSI_34	CH4	28166	238	487	0	725
mSI_8	CH4	28166	142	270	0	412
mTcCir312	CH4	32259	277	516	0	793
mSI_35	CH4	33618	183	348	0	531
mSI_2	CH4	43494	191	335	0	526
mSI_542	CH4	63388	33	68	0	101
mSI_303	CH4	119995	299	578	0	877
mSI_458	CH4	136890	208	399	0	607
mSI_460	CH4	139590	198	406	0	604
mSI_308	CH4	139780	211	408	0	619
mSI_309	CH4	141679	205	411	0	616
mSI_310	CH4	142517	193	383	0	576
mSI_315	CH4	233706	301	563	0	864
mSI_402	CH4	246098	259	506	0	765
mSI_585	CH4	252815	110	232	0	342
mSI_411	CH4	258684	125	242	1	368
mS_413	CH4	270916	118	253	1	372
mSI_39	CH4	278179	277	421	1	699
mSI_42	CH4	343424	109	233	2	345
mSI_46	CH4	428250	119	230	2	352
mSI_54	CH4	751986	55	127	4	521
mSI_294	CH4	1686245	109	248	16	373

Fig 2: Genome-wide association analysis of self-incompatibility traits and localisation of the positive markers in the CH4 region. Only one marker is inside the genomic region defined by fine mapping and is located in a GEX1 orthologous gene.

Search for candidate genes potentially involved in the *T. cacao* self-incompatibility system in the CH1 and CH4 genome regions and analysis of their expression.

Candidate genes potentially involved in cocoa self-incompatibility were first searched in the CH1 and CH4 regions without recombinant plants identified by fine mapping and the expression of twelve candidate genes was analysed during SC and SI reactions.

CH1 Region: In this region, only nine genes have been annotated in the whole genome sequence V1 of Criollo (Argout *et al.*, 2011), from Tc01_g007220 to Tc01_g007300. Among them three genes are considered as possible candidate genes:

• **Tc01_g007220** is an ortholog of *Arabidopsis thaliana* BAM1 which encodes CLAVATA1-related Leucine rich repeat receptor like kinases having an important role in early anther development and development of both male and female gametophyte (Hord *et al.,* 2006; DeYoung *et al.* , 2006). Tc01_g007220 is differentially expressed between SC and SI reactions at later stages with a higher expression during SC reaction.

• **Tc01_g007270** is homologous to the COMPASS-like H3K4 histone methylase component WDR5a from *Arabidopsis thaliana* (At3g49660). It is a transducin WD-40 repeat-containing protein acting as a site of protein-protein interactions playing central roles in biological processes (Stirniman *et al.*, 2010)

• **Tc01_g007290**, is a putative transmembrane transporter, part of the Major facilitator superfamily (MFS-1), and homologous to *A. thaliana* At5g65687 gene, a probable sphingolipid transporter spinster homolog 1. This gene is also predicted to interact with several genes involved in ubiquination, an important process during self-incompatibility reactions.

Tc01_g007270 and Tc01_g007290 are differentially expressed between SC/SI reactions at early stages (2h to 8h) with a much higher expression observed for Tc01_g007290 during SI reactions.

Immunolocalization experiments conducted with Tc01_g007270 and Tc01_g007290 showed that these genes are expressed in the cell layers of the style where are progressing the pollen tubes and in those surrounding the embryo sac (Fig. 3)

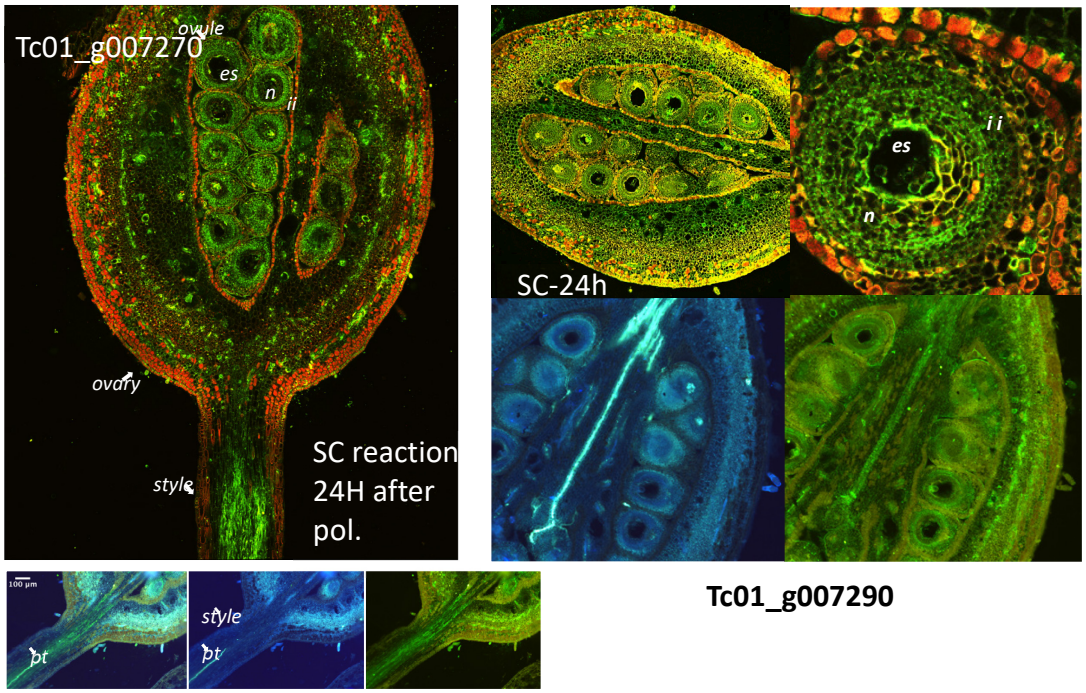


Fig 3: Immunolocalization of Tc01_g007270 and Tc01_g007290 proteins. Observations (confocal microscope) of sections of ovules, ovary and style of Scavina 6 accession 24H after pollinations by a compatible pollen (SC) and probed with antibodies of Tc01_g007270 and Tc01_g007290 proteins (green points).

***CH4 region*:** The region of 257.270 Kb identified by fine mapping includes 30 genes annotated in the Criollo genome V1. Among them, nine genes were identified as candidates potentially involved in the SI reactions.

• **Tc04_g000160** is homologous to the Voltage-dependent L-type calcium channel and could have a role in Ca^{2+} influx into stigma papilla cells mediating the SI signalling as was shown in the *Brassicaceae*

• **Tc04_g000170** is ortholog to a Kanadi transcription factor, potentially interacting with an Auxin factor response 3

• **Tc04_g000190, Tc04_g000230, Tc04_g000240** and **Tc04_g000260** are homologous to *GEX1* (Gamete Expressed Protein - At5g55490) *A. thaliana* genes potentially involved in early embryo development (Alandete-Saez *et al*., 2011). They have a dual function during male and female gametophyte development and early embryogenesis and are required for correct pollen maturation. Using STRING network view, it was shown that in *A. thaliana*, *GEX1* is interacting with the *HAPLESS2* gene involved in male fertility, essential for pollen tube guidance, successful gamete attachment and fertilization. Among them only Tc04_g000190 and Tc04_g000240 displayed a differential expression between SC/SI reactions compared to un-pollinated ovules.

• **Tc04_g000300** is ortholog to a ferredoxin thioredoxin reductase, but not differentially expressed between SC/SI reactions.

• **Tc04_g000320** is an ortholog of a Zinc finger AN1 domain-containing stress-associated protein 12 (PMZ) gene from *A. thaliana*. PMZ could interact with several genes associated to protein degradation through the ubiquination complex**.** Tc04_g000320 is significantly more expressed at the stage 12-24h during the SC reaction.

• **Tc04_g000330** is an ortholog of the Arm repeat-Containing protein ARC1 gene from *Brassica napa* (AGP76183.1). The ARC1 protein is involved in SI signalling in *Brassica* and targets proteins for degradation during SI response. Tc04_g000330 is differentially expressed, at the stage 12h-24h, where a higher expression is observed for SC reaction, which is not in agreement with what is expected during the SI reaction.

Self-compatibility/incompatibility predictions

***Allele specificity*:** Specific Amelonado and Criollo alleles were identified among the CH4 markers, predicting SC trees homozygous for these alleles:

The 161 bp allele of mSI_460 is specific to Amelonado. Its low frequency in the Nanay and Nacional groups probably reflects introgression of Amelonado in some individuals of these 2 groups. The allele 217 of mSI_303 is specific to Criollo. These 2 alleles could be used to screen for SC plants, homozygous for these alleles.
***Genotype analysis*:** Other *T. cacao* allele combinations than those identified in Amelonado and Criollo could result in SC plants. Dominance relationships exist between cocoa S-alleles and the multi-allelic status of SSR markers allowed a better discrimination of S-haplotype interactions. When only one genotype was considered, 24 genotypes from 7 markers were significantly associated to cocoa self-incompatibility, by Fisher´s Exact test (q-value < 0.05). Among those genotypes, 16, when present, resulted in higher frequency of SC plants and 10 resulted in higher frequency of SI plants. When two genotypes were considered, 178 combinations of genotypes were significantly associated to SI in cacao, with 134 combinations resulting in more SC clones and 44 in more SI clones.
Allele effect: Among the 128 alleles tested, 43 were significantly associated with SI status by the Fisher´s Exact test.
Examples of predicions are reported in Table 4

Table 4: examples of predictions. Probabilities of an individual to be self-incompatible (prob-SI) or self-compatible (Prob-SC) according to its genotype

marker1	marker2	genotype		S_0	S_1	Prob-SI	Prob-SC
mSI_303		227/227		17	57	0,23	0,77
mSI_303	mSI_7	227/227	200/200	0	50	0,00	**1,00**
mSI_303	mSI_411	227/227	286/286	1	53	0,02	**0,98**
mSI_303		225/227		1	14	0,07	**0,93**
mSI_35		119/119		28	86	0,25	0,75
mSI_35	mSI_303	119/119	227/227	1	53	0,02	**0,98**
mSI_460		161/161		0	44	0,00	**1,00**
mSI_107	mSI_35	239/245	119/119	0	27	0,00	**1,00**
mSI_107	mSI_458	239/245	292/292	1	26	0,04	**0,96**
mSI_2		280/283		35	6	**0,85**	0,15
mSI_303		231/233		23	0	**1,00**	0,00
mSI_303		231/231		24	0	**1,00**	0,00
mSI_458		306/306		26	2	**0,93**	0,07
mSI_460		155/161		22	1	**0,96**	0,04
mSI_460		145/146		24	2	**0,92**	0,08
mSI_7		194/197		22	1	**0,96**	0,04

Discussion- conclusion

The *T. cacao* self–incompatibility system has already been described as a late incompatibility system having both gametophytic and sporophytic aspects and dominance relationships between alleles (Cope, 1958, 1962; Glendinning, 1960). Our results confirmed its gametophytic and sporophytic features, and led to the identifcation of two independent loci involved in the SI system through two different late acting mechanisms. The fine mapping and GWAS analyses lead to restricted chromosome regions where several candidate genes could be identified.

the CH1 locus, identifed for the first time in this study, acts prior gamete fusion, but after pollen tube germination. Indeed, all ovules were fertilized which reflects a gametic selection that happened prior to the gamete fusion step. This selection could involve a lack of penetration or migration of the sperm nuclei in the embryo sac as was already observed in SCA 24 (closely related to SCA 6 used as pro-genitor in this study) (Ford and Wilkinson, 2012). Two main candidate genes were identified in the 70.3 kb CH1 region without recombinants plants, identified by fine mapping (*Tc01_g0007270* and *Tc01_g0007290*). Their proteins did not seem to be present in the pollen tubes themselves, observed at 8 or 24 h after pollination, but were apparent in the style cells where there were growing pollen tubes, and in several embryo sac cell layers.

The CH4 locus is involved in the SI reaction through a different mechanism, strongly associated with fruit drop. All positive associations identified by the GWAS were determined in the CH4 region where a gametic selection (probably linked to the observations on gamete non-fusions made by several authors: Cheesman 1927; Cope 1939*b*, 1940, 1958, 1962; Posnette, 1940; Knight and Rogers, 1955; Bouharmont, 1960; Ford and Wilkinson,

2012) was also observed in the F_2 progeny under study, as in the self-progeny of IMC 60, another SI clone. In these two progenies, only the use of mentor pollen, which prevented the fruit from dropping, had allowed the seed development in ovules where gametic fusions were effective.

In the region of 257.270 kb, without recombination identified by fine mapping in the CH4, several candidate genes were identified, potentially involved mainly in ubiquination steps or in early embryo or gamete development and guidance, as the GEX1 (gamete expressed protein) orthologous genes. Their potential role in the T. cacao self-incompatibility system was enhanced by the results of GWAS for which the only association identified in the region without recombinant identified by fine mapping was located in a GEX1 orthologous gene. Recently, it has been observed that the non-fusion of gametes, as observed after pollination of *T. cacao* using incompatible pollen, could result from two different and independant mechanisms: an incomplete migration of the sperm nucleus, which fails to reach the female nuclei, or a successful sperm nucleus migration and reaching of the female nuclei, but followed by a non-fusion of gametes (Ford and Wilkinson, 2012). The role of *GEX1* genes, interacting with the *HAPLESS2* gene that is known to be essential for pollen tube guidance, successful gamete attachment and fertilization (von Besser *et al.*, 2006; Mori *et al.*, 2006), could be determinant in this late manifestation of SI in *T. cacao* . However, *Tc04_ g000160*, *Tc04_g000320*, and *Tc04_g000330*, orthologs to genes involved in SI signalling in *Brassica* and protein degradation during the SI response, cannot be ruled out and could also participate in the pathway leading to self-incompatibility.

The practical output from this study is the identification of markers that could help breeders to select for SC plants in cocoa populations at an early stage in the breeding process. Indeed, the identification of highly polymorphic SSR markers in the CH4 region gives the potential to have a strong linkage disequilibrium between SSR alleles and incompatibility alleles, leading to a good ability of SSR to predict SI or SC genotypes. Specific Amelonado and Criollo alleles, identified in this study, will be particularly effcient in predicting SC varieties.

Acknowledgements

This project was supported by Agropolis Fondation under the reference ID 1403-046 through the "Investissements d'avenir" programme (Labex Agro: ANR-10-LABX-0001-01). We also thank Mondelēz International for their financial contribution to this project. We thank Marc Lartaud and Geneviève Conéjéro for their helpful advices in immunolocalization experiments and imaging. We thank Neba Ngwa AKONGNWI and Semi Melliti for their help in managing field pollinations in Cameroun and greenhouse in Montpellier.

All details about markers, primers and gene expession values are reported in *Journal of Experimental Botany*, Volume 68, Issue 17, 13 October 2017, Pages 4775-4790, https://doi.org/10.1093/jxb/erx293

References

Alandete-Saez M, Ron M, Leiboff S, McCormick S. 2011. Arabidopsis thaliana GEX1 has dual functions in gametophyte development and early embryogenesis. The Plant Journal **68,** 620–632.

Allegre M, Argout X, Boccara M, *et al.* 2012. 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 Research **19,** 23–35.

Argout X, Salse J, Aury JM, *et al.* 2011. The genome of *Theobroma cacao* . Nature Genetics **43,** 101–8.

Bartley BGD. 1969. Selfing of self-incompatible trees. Annual Report on Cacao Research (1968) Trinidad, 22–23.

Bartley BGD. 2005. The genetic diversity of cacao and its utilization. CABI Publishing, Wallingford, 341.

Bouharmont J. 1960. Recherches cytologiques sur la fructification et l'incompatibilité chez *Theobroma cacao* L. Publication de l'Institut National pour l'Etude Agronomique du Congo, No. 633.74 B6.

Cheesman EE. 1927. Fertilisation and embryogeny in *Theobroma cacao* L. Annals of Botany **41,** 107–126.

Cope FW. 1939a. Some factors controlling the yield of young cacao. 8th Annual Report on Cacao Research, Trinidad, 4-15.

Cope FW. 1939b. Studies in the mechanism of self-incompatibility in cacao I. 8th Annual Report on Cocoa Research, Trinidad, 20–21

Cope FW. 1940. Studies in the mechanism of self-incompatibility in cacao II. 9th Annual Report of Cocoa Research, Trinidad, 19–23.

Cope FW. 1958. Incompatibility in *Theobroma cacao* . Nature **181,** 279–279.

Cope FW. 1962. The mechanism of pollen incompatibility in *Theobroma cacao* L. Heredity **17,** 157–182.

Crouzillat D, Lerceteau E, Petiard V, *et al.* 1996. *Theobroma cacao* L.: a genetic linkage map and quantitative trait loci analysis. Theoretical and Applied Genetics **93**, 205–214.

Da Silva MR, Clément D, Gramacho KP, Monteiro WR, Argout X, Lanaud C, Lopes UV 2016. Genome-wide association mapping of sexual incompatibility genes in cacao (*Theobroma cacao* L.). Tree Genetics and Genomes, **12,** 62.

De Nettancourt D. 1997, Incompatibility in angiosperms. Sexual Plant Reproduction **10,** 185–199.

DeYoung BJ, Bickle KL, Schrage KJ, Muskett P, Patel K, Clark SE. 2006. The CLAVATA1-related BAM1, BAM2 and BAM3 receptor kinase-like proteins are required for meristem function in Arabidopsis. The Plant Journal for Cell and Molecular Biology **45**, 1–16.

Ford CS, Wilkinson JW. 2012. Confocal observations of late-acting self-incompatibility in *Theobroma cacao* L. Sexual Plant Reproduction **25,** 169–183.

Gibbs PE. 2014. Late-acting self-incompatibility – the pariah breeding system in flowering plants. New Phytologist **203,** 717–734.

Glaubitz JC, Casstevens TM, Lu F. 2014. TASSEL-GBS: high capacity genotyping by sequencing analysis pipeline. PLoS One **9,** e90346.

Glendinning DR. 1960. Selfing of self-incompatible cocoa. Nature **187,** 170–170.

Glendinning DR. 1967. Incompatibility alleles of cocoa. Nature **213**, 306.

Hord CLH, Chen C, Deyoung BJ, Clark SE, Ma H. 2006. The BAM1/BAM2 receptor-like kinases are important regulators of Arabidopsis early anther development. Plant Cell **18**,1667-1680.

Killian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, *et al.* 2012. Diversity arrays technology: a generic genome profiling technology on open platforms. Data Production and Analysis in Population Genomics: Methods and Protocols, 67-89.

Knight R, Rogers HH. 1955. Incompatibility in *Theobroma cacao* . Heredity **9,** 69–77.

Kotturi MF, Carlow DA, Lee JC, Ziltener HJ, Jefferies WA. 2003. Identification and functional characterization of voltage-dependent calcium channels in T lymphocytes. Journal of Biological. Chemistry **278,** 46949-46960.

Lachenaud P, Sounigo O, Clément D. 2005. The compatibililty - yield efficiency relationship. Ingenic newsletter **10,** 13-16.

Lanaud C, Sounigo O, Amefia YK, Paulin D, Lachenaud P, Clement D. 1987. Nouvelles données sur le fonctionnement du système d'incompatibilité du cacaoyer et ses conséquences pour la sélection. Café Cacao Thé **31,** 267–277.

McClure B. 2008. Comparing models for S-RNase-based self-incompatibility. In:Franklin-Tong V, editor. Self-incompatibility, in flowering plants: evolution, diversity, and mechanisms. Berlin: Springer; 217-236.

Mori T, Kuroiwa H, Higashiyama T, Kuroiwa T. 2006. GENERATIVE CELL SPECIFIC 1 is essential for angiosperm fertilization. Nature Cell Biology **8,** 64–71.

Opeke LK, Jacob VJ. 1967. Studies on methods of overcoming self-incompatibility in *Theobroma cacao* Linn. 2e Conférence Internationale sur les Recherches Cacaoyères, 356–359.

Posnette AF. 1940. Self-incompatibility in cocoa (*Theobroma* spp.). Tropical Agriculture, Trinidad and

Tobago **17,** 67-71.
Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics **155,** 945–959.
Rea A, Nasrallah JB. 2008. Self-incompatibility systems: barriers to self-fertilization in flowering plants. International Journal of Developmental Biology **52,** 627–636.
Royaert S, Phillips-Mora W, Arciniegas Leal AM, *et al.* 2010. Identification of marker-trait associations for self-compatibility in a segregating mapping population of *Theobroma cacao* L. Tree Genetics & Genomes **7,** 1159–1168.
SAS Institute Inc. 2004. SAS/STAT ® 9.1 User's Guide. Cary, NC: SAS Institute Inc.
Stirnimann CU, Petsalaki E, Russell RB, Muller CW. 2010. WD40 proteins propel cellular networks. Trends in Biochemical Sciences **35,** 565-574.
Suzuki G, Kai N, Hirose T, Fukui K, Nishio T *et al*. 1999. Genomic organization of the S locus: Identification and characterization of genes in SLG/SRK region of S9 haplotype of *Brassica campestris* (syn. rapa). Genetics **153**, 391–400.
Takayama S, Isogai A. 2005. Self-incompatibility in plants. Annual Review of Plant Biology **56,** 467–489.
von Besser K, Frank AC, Johnson MA, Preuss D. 2006. *Arabidopsis HAP2* (*GCS1*) is a sperm-specific gene required for pollen tube guidance and fertilization. Development **133,** 4761–4769.
Yamada MM, Faleiro FG, Clement D, Lopes UV, Pires JL, Pires Melo GR. 2010. Relationship between molecular markers and incompatibility in *Theobroma cacao* L. Agrotropica **22,** 71-74.