



*The self-incompatibility system of *Theobroma cacao* L.: from genomics to diagnostic markers*

Claire Lanaud^{1†}, Olivier Fouet^{1†}, Thierry Legavre^{1†}, Uilson Lopes^{2†}, Olivier Sounigo^{3,4}, Marie Claire Eyango⁴, Benoit Mermaz¹, Marcos Ramos da Silva², Rey Gaston Loor Solorzano⁵, Xavier Argout¹, Gabor Gyapay⁶, Herman Ebai Ebaiarrey⁴, Kelly Colonges¹, Christine Sanier¹, Ronan Rivallan¹, Géraldine Mastin¹, Nicholas Cryer⁷, Michel Boccara¹, Jean-Luc Verdeil¹, Ives Bruno Efombagn Mousseni⁴, Karina Peres Gramacho², Didier Clement^{1,2}

1)-CIRAD-AGAP (France); 2) CEPEC-(CEPLAC (Brazil);3) CIRAD-UR Bioagresseurs (France); 4) IRAD (Cameroun); 5) INIAP (Ecuador); 6) CEA (France); 7) Mondelez International (UK)

Thanks to AGROPOLIS FONDATION and MONDELEZ International for their financial contribution

2017 International Symposium on cocoa research (ISCR), Lima, Peru, 13-17 novembre 2017

The several steps of cocoa self incompatibility studies

- Favorise genetic exchanges and recombination, increasing plant heterozygosity
 - Important factor to increase yield (young trees)
 - Important for clonal selection to ensure cocoa production,
- First Genetic studies in the fifties-sixties (Cope, 1939-62; Knight & Rogers, 1955; Bouharmont, 1960; Glendinning, 1967; etc... -> **multilocus-late acting self incompatibility system** –dominance relationships between alleles – results in **gamete non fusion and fruit drop**
 - Evidence of the non efficiency of incompatibility system for hybrid seed production in biclinal seed gardens (Lanaud et al.; 1987)
 - Genetic mapping (with molecular markers) of one locus (CH4) (Crouzillat et al., 1996; Royaert et al., 2010; Yamada et al., 2010) and recent cytological studies by Ford and Wilkinson (2012)

OBJECTIVES: Use new technology tools to increase our knowledge of the incompatibility system and produce diagnostic markers

Exploring cocoa genome sequence

→ new field of researches opened by the **availability of the genome sequences** (Argout et al, 2011; Motamayor et al, 2013) to refine the **genetic and molecular basis of SI**, and to define **diagnostic markers**, close to the genes, for genetic breeding

Genetic/genomic studies were based on several large populations :

- a F2 mapping population (850 ind.) located in Brazil used for fine mapping
- a diverse population (710 ind.) from Brazil, Ecuador, Cameroun, Trinidad evaluated for self-incompatibility used for association mapping and SC/SI prediction analyses

Fine mapping of two self-incompatibility loci

- F2 progeny (850 ind):**
- selfing self-incompatible TSH516 (ICS1 x SCA6), using mentor pollen. Fruits are maintained until maturity even in case of gametic non-fusions
 - Genotype selection could be observed in the progeny with markers – 2 skewed regions
 - TSH516 heterozygous at each locus with Amelonado (a)/SCA6(b) alleles: 3 genotypes are expected at each loci.

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
mSI_366	CH1	4053385	0	508	253	761
mSI_367	CH1	4054418	0	366	216	582
mSI_369	CH1	4057532	0	419	221	640
mSI_440	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

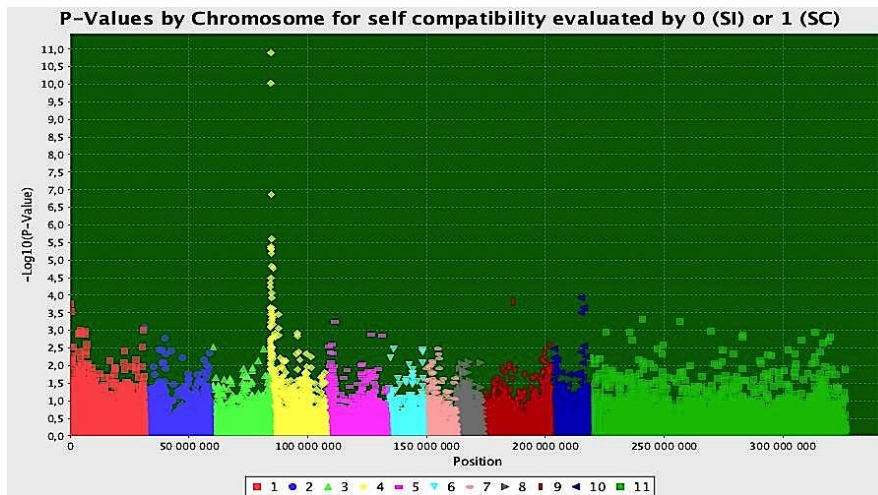
257 Kb

Identification of 2 regions involved in genotype selection, with a complete absence of recombinant plant. Incompatibility alleles closely linked to markers → candidate genes

Genome Wide Association Studies (GWAS)

Impact on fruit drop:

- GWAS conducted with 16480 SNP markers (GBS)
- on 570 cocoa trees from Brazil, Cameroun, Ecuador, Trinidad, evaluated for self incompatibility



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

189.4 kb * (GEX1)

1328.2 kb

69 Kb

Only one region located in CH4 was identified as involved in fruit drop, (evaluated by 0/1 or by the % fruit retention after selfing)

Only one significant marker, located in a **GEX (gamete expressed) gene**, included in the « non-recombinant » region of CH4 identified by fine mapping

Behavior of the CH1 locus

Is the CH1 functional when the cocoa tree is **self-compatible**, due to the presence of the *Sf* Amelonado allele, homozygous in the CH4 locus?

F1 : TSH516 (ICS1 x SCA6)

F2 self-compatible plants BR36 and BR59 ← F2 self-incompatible plants

CH1: 223 (*Amel*) / 225 (*Sca6*) + CH4: *Sf/Sf*

CH4: *Sf/Sx*
(*Amelonado/sca6*)

After selfing ↓

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
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

YES, it remains functional in case of self-compatibility

Does it result in selective gamete non-fusion?



% of aborted ovules easily observable 7 days after pollination

25 % of aborted ovules expected in case of selective gamete non fusion

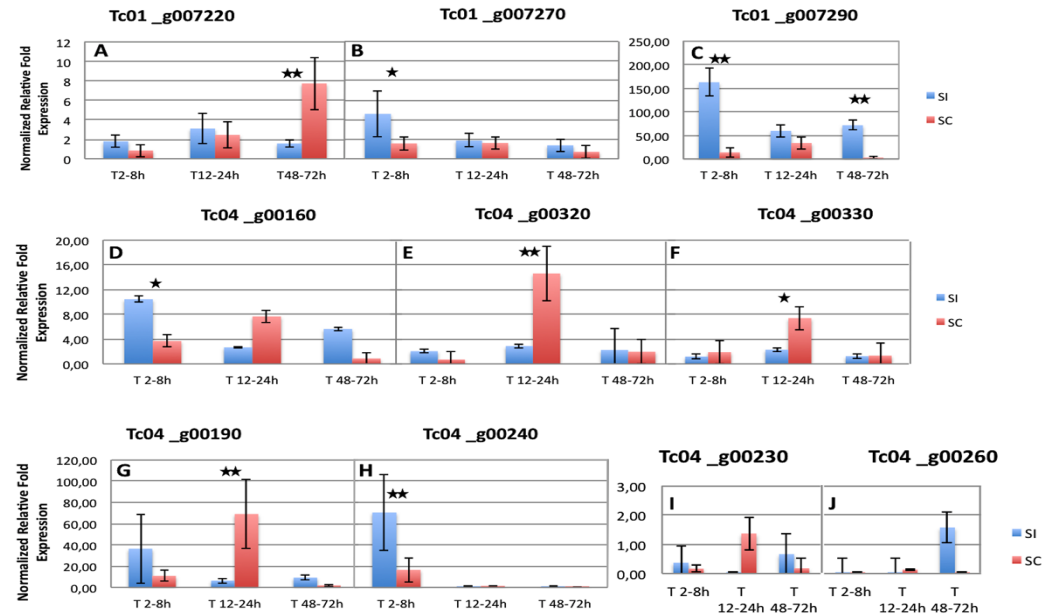
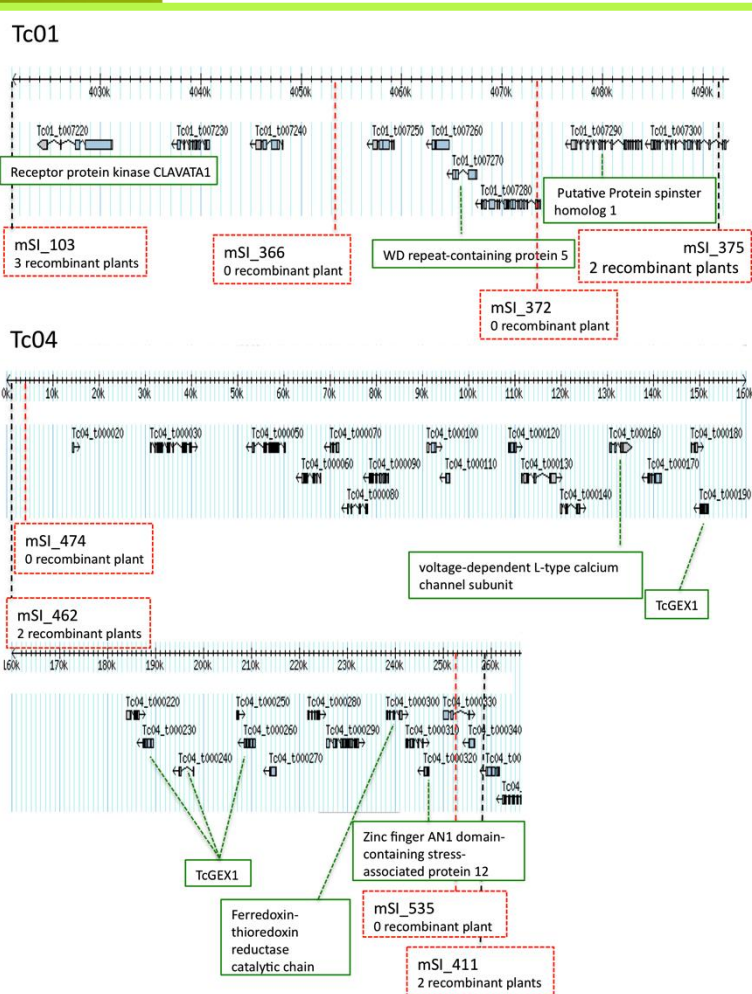
11 % and 9% of aborted ovules respectively observed after self or cross pollination (CCN51)

Genotypic selection happened before gamete non-fusion stage and after pollen germination

Search for candidate genes

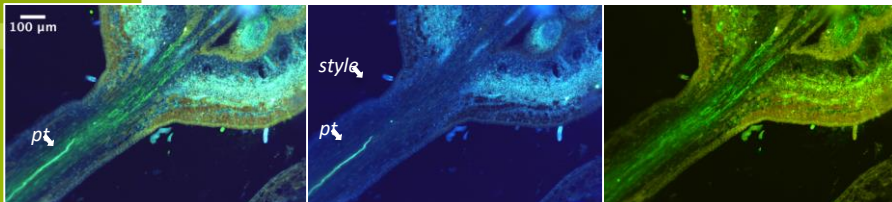
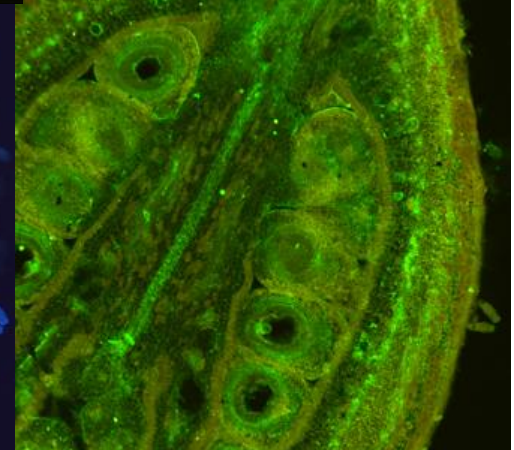
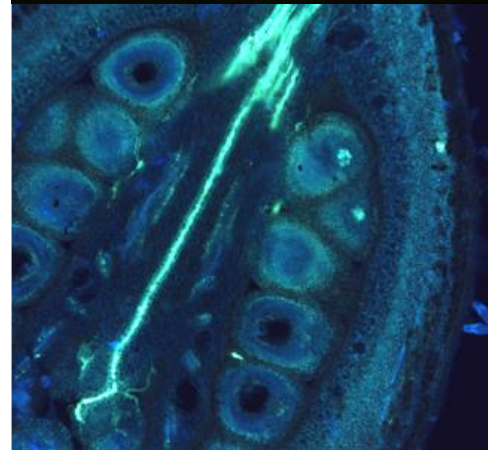
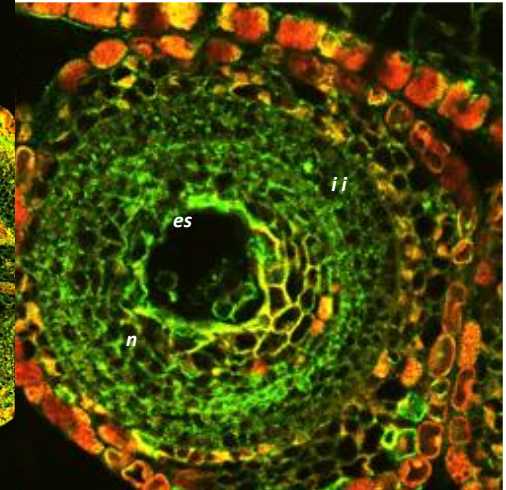
Search for candidate genes, orthologous to genes known to interact with incompatibility system, in the non-recombinant regions of CH1 and CH4, and analyses of their expression during compatible and incompatible reactions

CH1: 3 candidate genes differentially expressed, and among them a transporter and a gene containing a site of protein-protein interactions



CH4: 7 candidate genes differentially expressed, and among them 2 GEX1 orthologs known to be involved in early embryo and gametophyte development. They interact also with HAPLESS genes essential for pollen tube guidance and fertilization

Localization of CH1 candidate genes proteins by immunolocalization



Tc01_g007290

Search for diagnostic markers

Predictions established

- from a population of 710 cocoa trees from Brazil, Cameroun, Trinidad and Ecuador, covering a large diversity
- with highly multi-allelic SSR markers (until 24 alleles) located in the CH4 restricted region (based on a strong linkage between SSR alleles and incompatibility alleles)

Identification of SSR alleles specific to self-compatible varieties: Amelonado (MSI_460-161) and Criollo (MSI_303-217)

Some examples of predictions

Other *T. cacao* incompatibility allele combinations could result in SC or SI plants.

Multi-allelic status of SSR markers allowed a better discrimination of S-allele interactions and result of improved predictions.

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

Conclusions

- Identification of 2 loci involved in the *T. cacao* incompatibility system by 2 different and independent mechanisms.
 - only one (on CH4) seems to be linked to fruit drop, but the CH1 locus generates a genotypic selection, even in case of self-compatibility of the plant, and acts prior to the gamete fusion stage.
-
- A fine mapping of these 2 genomic regions, made thanks to the available *T. cacao* genome sequences, allowed us to identify several candidate genes potentially involved in the self incompatibility system, and some of them are differentially expressed between SC/SI reactions.
 - Immunolocalization of CH1 candidate genes show that these genes are expressed in the ovule tissues surrounding the embryo sac and in the style during pollen tube progression.
-
- Diagnostic SSR markers, highly polymorphic, were designed in the CH4 region identified by fine mapping, linked by a strong linkage disequilibrium with incompatibility alleles.
 - SSR alleles specific to self-compatible Amelonado and Criollo varieties were identified, as well as some genotypic combinations, allowing to screen efficiently self compatible plants in cocoa populations.



Thank
you for
your
attention