Doubled Haploids for Faster and More Efficient Breeding in Cassava

Juan P Arciniegas, Francisco J Sánchez & Paul Chavarriaga

Alliance Bioversity-CIAT, Gene Editing Platform

Introduction:

In maize, haploids are routinely produced using crosses with Stock 6 to a rate of 2-3% maternal haploid seed when it is self-pollinated or outcrossed as a male (Coe et al., 1959), which is 10 to 20 times higher than spontaneous haploid induction rate (HIR) in maize (Zhao et al., 2013). The process is called gynogenesis, where the male gametes induce haploid embryo formation only from the female chromosomes (Jackson 2017). Breeding performed with Stock 6 derived inducers produces lines with elevated HIRs up to 20.42% (Cengiz et al., 2016). The genetic mechanism behind Stock 6 haploid induction was characterized by three groups and the gene implicated (GRMZM2G471240) was named NOT LIKE DAD (NLD) (Gilles et al., 2017), MATRILINEAL (MTL) (Kelliher et al., 2017) or ZmPHOSPHOLIPASE A1 (*ZmPLA1*) (Liu et al., 2017). This gene was found mutated in Stock 6 and the inducer line HKR, with a 4-bp insertion in the exon 4 that causes a frameshift that replaces the last 49 amino acids (aa) of the wild-type protein by an unrelated amino acid sequence of 20 aa followed by a premature STOP codon (Kelliher *et al.,* 2017).

CRISPR/Cas9-mediated genome editing was performed to knock out *ZmPLA1* (Liu *et al.,* 2017). Three lines with 1-bp insertion, 11-bp deletion and 1-bp deletion in the target region, which are putative knock outs alleles for *ZmPLA1*, were chosen for pollination assays. In self-pollinated knock out lines the HIR ranged from 3.7% to 6.67%. In the offspring of breeding male knock out lines with wild-type lines the chromosomes proceed only from the maternal genomes. The behavior of *ZmPLA1* knock out lines was similar to the gene mutation in Stock 6.

Target selection and genetic transformation:

The GRMZM2G471240 gene was aligned with *Manihot esculenta* v6.1 in Phytozome (https://phytozome.jgi.doe.gov/) to find possible candidate genes. This analysis revealed 6 candidate orthologs genes (Manes.05G171400, Manes.12G093400, Manes.18G033200,

Manes.13G126900, Manes.12G102200, Manes.18G033400). Multiple alignment with phylogenetic analysis with Maximum likelihood (Figure 1) determined Manes.12G093400 and Manes.12G102200 as the closest related to GRMZM2G471240 and the more suitable candidate genes to knock out through CRISPR/Cas9 for doubled-haploid induction.

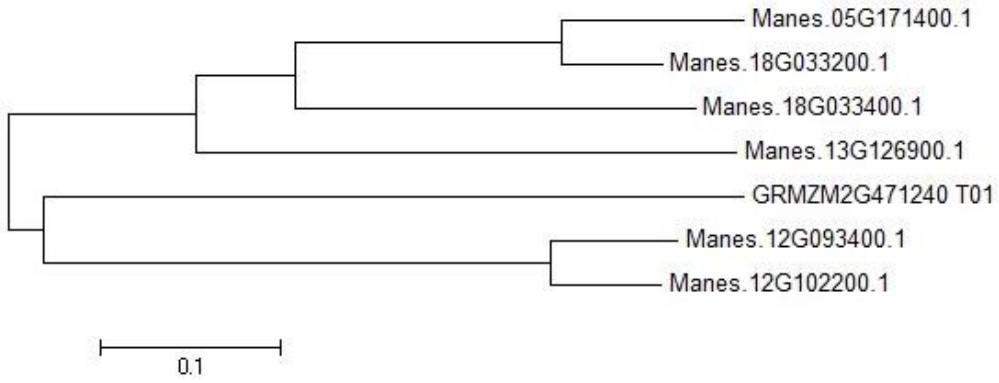


Figure 1. Maximum likelihood phylogeny tree of cassava candidate genes with GRMZM2G471240.

Once the candidate genes were selected, single-guided RNAs (sgRNAs) were designed with CRISPR P 2.0 (http://crispr.hzau.edu.cn/CRISPR2/). The selection was made based on affinity of sgRNAs with the gene, the position in the gene and the off-targets presents in the whole cassava genome (Figure 2A). Then, sgRNAs were cloned independently on pHSE401 vector and introduced on *Agrobacterium tumefaciens* LBA4404.

A

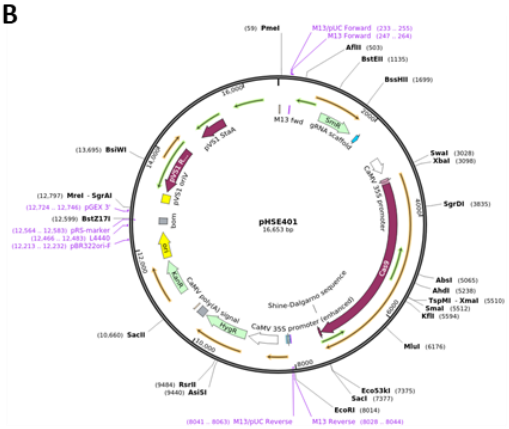
>Manes.12G093400 | Chromosome12:14822139..14823812 forward

ATGGCTACTGGTAATAACATTTCGAAACAATGGAAGAAGATAACAGTGTTGAGCATTGATGGAGGTGGGATCAG
AGGCATCATTCCTGCTACCATTCTTGCTTCTCTTGAATCCAAGCTTCAGGTACTTGAAAAGAAATTTTTCTTTT
TCTTTTTTTTTTTTTTTAAATTTCTTCATCTTAATTTGTTTTGCCTTTTTTTCCATGGTGTTAATCATTAGGAT
TTGGATGGACCAGATGCGAGAATTGCAGATTATTTTGATGTAATTGCAGGGACAAGTACTGGTGGGCTGGTAAC
CACCATGCTCACAGCTCCAAACAAGGACAAAAGACCTATGTTCCAAGCAAAAGACATCAAGAACTTCTATTTAG
AACACTGTCCCAAGATTTTCCCACAGAAAAGGTGAAATCTTTCAAGTTTTACTTAATTTCTTTCTTGTTTTTTT
ATTAAATGGAATTATAATAAATGGGTCTAATTTCATTTCTTCTATGGATTGAGCAGTCGAAATAACTTCATTAG
CTCGATAACAAATTTCTTTGGTGCGGTTATGGGACCAAAGTACGATGGAAAATACCTGAAAGGACTGACAAATG
AGCTGCTAGGGGACTTGACCATAAAAGAGACTTTGGCAGATGTGATTATACCTGCTTTTGATATCAAACTTCTT
CAGCCAGTCATCTTCTCAACTGATGATGTATGTATATATACATGCTAAAGCTGAGCTAGTAAAAATTTTTCCTT
GAATATGAGCTGAGATTTTTACAATTTTCTTGTCTTAACAGGCGAAAGTTGATGCTCTGAGAAATGCTAGACTA
GCTGATATCTGCATCAGCACCTCTGCAGCTCCAACTTTCTTTCCAGCACATTCCTTTACCATCAAGGATGATCA
AAGGACTCGCACTTTTGAACTCATTGATGGTGGGGTTGCAGCCAATAATCCTGTAAGCTACTCTTCTTCTGCAC
ATACACTCGTAGCAATAAAAGGCTTTTCATTGATAATTTGTTGCGCTTGTGACTGAAGACATTACTAGCAATAA

>Manes.12G102200 | Chromosome12:21375405..21377776 forward

ATGGCTGCTGATCATAATAATAATTTGAAGCCAGGAAAGAAGATAACAGTGCTGAGCATTGATGGAGGTGGGAT
TAGAGGTATCATTCCTGGTATCATTCTTGCTTCTCTTGAATCTAAACTTCAGGTAACAGAAATTTCTTTTTCT
TTTCTTTTATTTTAATTTCTCTTCATCTTAATTAATTTGTTGTGCTGTTCATGGTTGTAATCACAAGGATTTGG
ATGGACCAGATGCAAGAATTGCATATTATTTTGATATACTTGCAGGGACAAGCACTGGAGGGTTGTTAACCAACC
ATGCTTACAGCTCCTAACAAAGACAAAGACCTATGTTCCAAGCAAAAGACATGGAAAATTTCTATTTAGAACA
CTGTCCCAAGATTTTCCCACAGAAAAGGTGAAATCATCTCTGATTTTTTCAATTATTTTTGTTTTACATGGAAT
GATAATAAATGGGTTTTCAATTCTTCCATGGATTAAACAGTCGAAATAACTTCATTGACTCGATCACAAACTTT
TTTGGTGCTGTTTTGGGGCCAAAATATAATGGAAAATATCTGCATGAACTGACAAATCAGTTGTTAGGAGACTT
GACCATAAAGGAAACCTTGGCGAATGTGATTATACCTGCGTTTGACATAAAGCTTCTTCAGCCAGTCATCTTCT
CAACAGATGATGTATGTATATTACATATAGATTACCAAAATTAAGTTAAATTTTGCGTCCATGCATGCGCAAAA
ATATAAGATTTTTAAAAAGTTTATTGTCTTAACAGGCGAAAGTTGATGCTTTGAAAAATGCTAGGCTAGCTGAT
ATCTGCATCAGCACCTCTGCAGCTCCAACTTTCTTTCCACCACATTCCTTTACCACCAAGGATGATCAAAGGAC

B



C

Construct	5'-3' + PAM	Gene	Location
PHSE401-pL1a09	TTGCAGGGACAAGTACTGGTGGG	Manes.12G093400	Exon 2
PHSE401-pL1a10	GTTAGGAGCTGTAAGCATGGTGG	Manes.12G102200	Exon 2

Figure 2. (A) Coding sequences of candidate genes for doubled-haploid induction. Single-guided RNAs (sgRNAs) selected for Manes.12G093400 and Manes.12G102200. Introns not highlighted, exons in blue, Protospacer Adjacent Motif (PAM)in red, sgRNA sequence in green, primers for molecular evaluation in yellow (800-950 bp fragment) and purple (100-200 bp fragment). (B) pHSE401 Binary vector used as a destination vector for dicots. (C) Detail of sgRNA cloned sequences on pHSE401.

The CRISPR/Cas9 machinery was introduced into TMS60444 friable embryogenic callus (FEC) through *Agrobacterium tumefaciens* genetic transformation according to Taylor et al. (2012) with modifications.

Molecular analysis of regenerated lines:

Two genetic transformation essays were named T310718 and T100519. The following table summarize the results obtained:

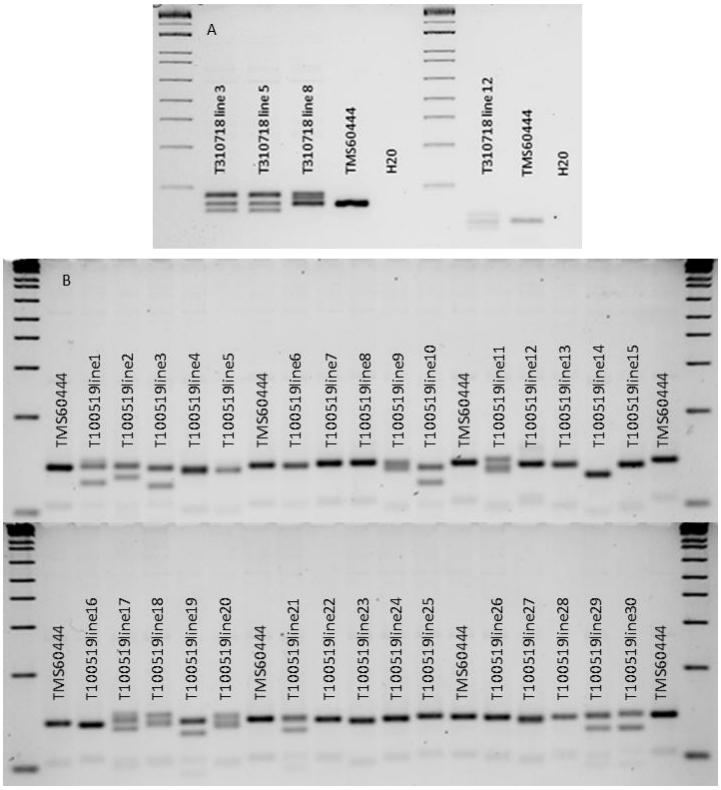
Table 1. Molecular analysis of regenerated lines from T310718 and T100519. Transformation line refers to a consecutive numbering during tissue culture stage, and Molecular line refers to a consecutive numbering for molecular tests. PCR amplified Hygromycin resistance gene, if positive the plant was considered transgenic. PCR results were confirmed by Southern blots, last column, indicating T-DNA copy number (1 to 5) and letters to match identical lines. N/A: not analyzed. T: transgenic lines. IL: independent transgenic lines.

Transformation ID	Genotype	Construct	Molecular line	PCR	Southern blot
T310718	TMS60444	PHSE401pL1a09	1	-	N/A
T310718	TMS60444	PHSE401pL1a09	2	-	N/A
T310718	TMS60444	PHSE401pL1a09	3	+	1a
T310718	TMS60444	PHSE401pL1a09	4	-	N/A
T310718	TMS60444	PHSE401pL1a09	5	+	1a
T310718	TMS60444	PHSE401pL1a09	6	+	1a
T310718	TMS60444	PHSE401pL1a09	7	N/A	N/A
T310718	TMS60444	PHSE401pL1a09	8	+	1b
T310718	TMS60444	PHSE401pL1a09	9	+	1b
T310718	TMS60444	PHSE401pL1a09	10	+	0
T310718	TMS60444	PHSE401pL1a10	11	-	N/A
T310718	TMS60444	PHSE401pL1a10	12	+	1c
T310718	TMS60444	PHSE401pL1a10	13	N/A	N/A
T310718	TMS60444	PHSE401pL1a10	14	+	1c
T310718	TMS60444	PHSE401pL1a10	15	+	1c
T310718	TMS60444	PHSE401pL1a10	16	+	1c
T310718	TMS60444	PHSE401pL1a10	17	-	N/A
T310718	TMS60444	PHSE401pL1a10	18	+	1c
T310718	TMS60444	PHSE401pL1a10	19	+	0
T310718	TMS60444	PHSE401pL1a10	20	+	1c

T310718	TMS60444	PHSE401pL1a10	21	-	N/A
T310718	TMS60444	PHSE401pL1a10	22	-	N/A
T310718	TMS60444	PHSE401pL1a10	23	+	1c
T310718	TMS60444	PHSE401pL1a10	24	-	N/A
T310718	TMS60444	PHSE401pL1a10	25	-	N/A
T310718	TMS60444	PHSE401pL1a10	26	+	1c
T100519	TMS60444	PHSE401pL1a10	1	+	2
T100519	TMS60444	PHSE401pL1a10	2	+	0
T100519	TMS60444	PHSE401pL1a10	3	+	N/A
T100519	TMS60444	PHSE401pL1a10	4	+	1
T100519	TMS60444	PHSE401pL1a10	5	+	1
T100519	TMS60444	PHSE401pL1a10	6	+	2
T100519	TMS60444	PHSE401pL1a10	7	+	3
T100519	TMS60444	PHSE401pL1a10	8	+	0
T100519	TMS60444	PHSE401pL1a10	9	+	1
T100519	TMS60444	PHSE401pL1a10	10	+	2
T100519	TMS60444	PHSE401pL1a10	11	+	5
T100519	TMS60444	PHSE401pL1a10	12	+	1
T100519	TMS60444	PHSE401pL1a10	13	+	1
T100519	TMS60444	PHSE401pL1a10	14	+	3
T100519	TMS60444	PHSE401pL1a10	15	+	2
T100519	TMS60444	PHSE401pL1a10	16	+	4
T100519	TMS60444	PHSE401pL1a10	17	+	2
T100519	TMS60444	PHSE401pL1a10	18	+	3
T100519	TMS60444	PHSE401pL1a10	19	+	1
T100519	TMS60444	PHSE401pL1a10	20	+	1
T100519	TMS60444	PHSE401pL1a10	21	+	4
T100519	TMS60444	PHSE401pL1a10	22	+	3
T100519	TMS60444	PHSE401pL1a10	23	+	1

T100519	TMS60444	PHSE401pL1a10	24	+	3
T100519	TMS60444	PHSE401pL1a10	25	+	2
T100519	TMS60444	PHSE401pL1a10	26	+	2
T100519	TMS60444	PHSE401pL1a10	27	+	1
T100519	TMS60444	PHSE401pL1a10	28	+	4
T100519	TMS60444	PHSE401pL1a10	29	+	2
T100519	TMS60444	PHSE401pL1a10	30	+	N/A
	Total		56 lines	T:45	IL:29

Nested PCR was performed on Southern-blot-confirmed transgenic lines to detect signs of gene editing. PCR products were run on a Metaphor agarose 3% electrophoresis gel. Bands different than control (TMS60444) indicated gene editing events. Further confirmation through sequencing INDELS is undergoing.



Figures 3 (above) & 4 (two below). Metaphor agarose 3% electrophoresis of transgenic putative doubled-haploid inducer lines (A) Metaphor agarose 3% electrophoresis of T310718 lines. Lines 3 and 5 appear to have the same mutation patterns. (B) Metaphor agarose 3% electrophoresis of T100519

Pollination essays with putative haploid-inducer lines.

Phenotypic differences between transgenic/edited lines and TMS60444 were observed. The size of anthers and pollen of transgenic/edited lines was smaller (Figure 5). Crosses were made between T310718 pollen and TMS60444 female flowers. Seeds product of crosses with putative haploid inducer and control did not show physical differences (Figure 6) and embryos were rescued from them. For embryo rescue, seeds were disinfected with ethanol 70% with TWEEN20 drop wash for 1 minute, followed by 3 washes with sterile water, sodium hypochlorite 0.425% with TWEEN20 drop for 5 minutes, and 3 washes with sterile water. To isolate immature embryos, it was cut transversally below the stigma to expose the carpels and obtain the embryos.

Regeneration of rescued embryos is a challenging process, the optimal requirements and culture medium composition are highly variable during the development of embryos, but efforts has been made on this field (Lentini et al., 2020). A first essay consisted in rescuing embryos from seeds that fell from plants 8-10 days after pollination. The embryos were cultured in ME056.2 (MS Micro and Macroelements 4.3 g/L, MS vitamins 1X, sucrose 2%, BAP 0.45 mg/L, agar 4.5 g/L, pH 6.11-6.14). Significant tissue development was evidenced after 25 days (Figure 7). Embryos were subcultured in a new medium to promote organogenesis as described by Li et al. (1998): MS Micro and Macroelements 4.3 g/L, MS vitamins 1X, sucrose 2%, BAP 1 mg/L, IBA 0.5 mg/L, $CuSO_4$ 200 µM, agar 4.5 g/L, pH 6.11-6.14. Nevertheless, the embryos did not develop further.

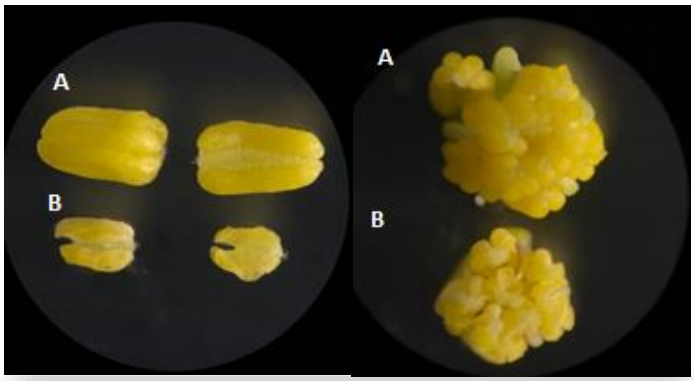


Figure 5. Anthers and pollen of TMS60444 (A) and line T310718 pL1a09 (B).

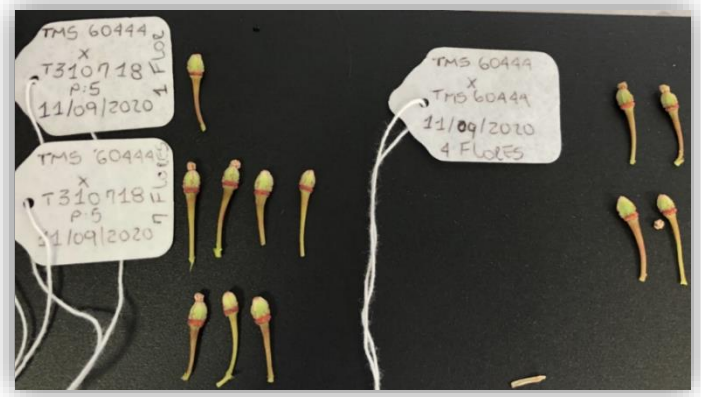


Figure 6. Immature seeds obtained after 7 days of crossing T310718 PHSE401p5 (corresponds to well #3 on gel A, Figure 3, labelled T310718line5) **X** TMS60444 (left) and TMS60444 **X** TMS60444.

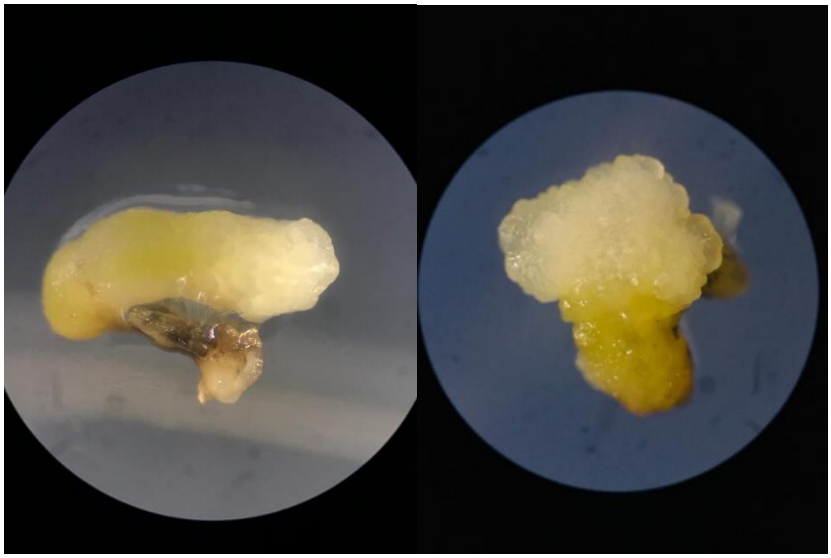


Figure 7. Rescued seeds from PHSE401pL1a09-5 x TMS60444 crossing in ME056.2 medium after 25 days of culture.

A second essay was performed with culture medium MS3 reported by Lentini et al. 2020: MS Micro and Macroelements 4.3 g/L, $CuSO_4$ 0.8 mg/L, nicotinic acid 2.5 mg/L, pyridoxine 1.2 mg/L, thiamine 10 mg/L, glycine 4 mg/L, biotin 0.2 mg/L, Ca-pantothenate 0.2. ascorbic acid 0.2 mg/L, riboflavin 0.4 mg/L, L-proline 200 mg/L, L-glutamine 400 mg/L, casein hydrolysate 150 mg/L, 2,4-D 2mg/L, BAP 2 mg/L, GA3 1 mg/L, sucrose 80 mg/L, gelrite 4 g/L, pH 6.11-6.14. Another medium, named MS3-BAP (MS3 with BAP 0.45 mg/L) was tested in parallel. Different seed development stages were also tested: 4, 5 and 7 days after pollination.

Results with MS3 with BAP were no different. Tissues turned brown and no development of new callus was observed (Figure 8A). On the other hand, seeds cultured in MS3 independently of the seed development stage generated different types of calli (Figure 8B, C & D). Nevertheless, at this point, the origin of the new tissues is unclear, i.e., if they come from haploid, unpollinated cells or just diploid somatic cells.

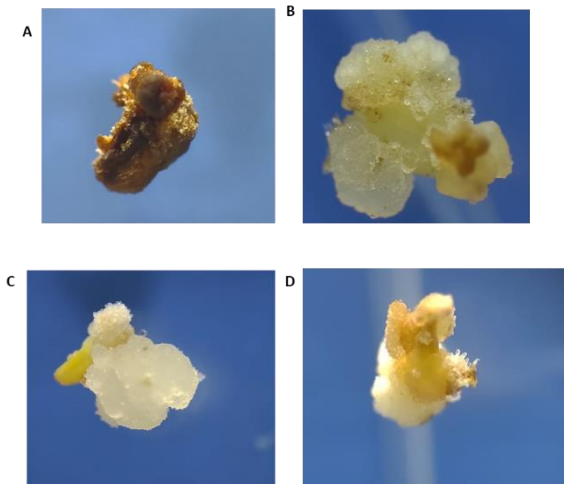


Figure 8. Rescued embryos from PHSE401pL1a09 x TMS60444 crossing in MS3 and MSE-BAP medium after a month of culture. (A) Rescued embryos on MS3-BAP. (B) Rescued embryos after 4 days of pollination in MS3. (C) Rescued embryos after 5 days of pollination in MS3. (D) Rescued embryos after 7 days of pollination in MS3.

New Haploid Inducer Lines in Field.

A new set of T100519 lines are now established in the field and started flowering in November-December/2021 (Table 2 at the end of document), so new pollination assays just began, but this time crosses are being made with non-TMS60444 genotypes, with early flowering and flowering during long period, i.e., SM3559-11 (DM), AM1521-3 (BC), and C243 (MD), which were provide by the cassava breeder, Dr Xiaofei Zhang. T100519 lines carry the PHSE401pL1a10 construct to mutate the gene Manes.12G102200, to test another mutation, different that the one on T310718 lines. A recent checkup (December 16/2022) of male flowers from these plants showed that they do produce pollen (Figure 9), although pollen viability and germination will be tested at the beginning of 2022.

Figure 9. The new set of haploid inducers, T100519 lines, planted in the field (left) and started flowering producing female and male flowers (center), the llaters with pollen grains (right).

Perspectives.

To produce haploids the use of inducer lines increases the frequency of success. This is the most practical and fastest way to generate homozygous parentals breeding lines to create true F1s where hybrid vigor can be clearly exploited for breeding in cassava. For corn this is the most common route (Jackson 2017), and the research described above points towards using the same technology for cassava.

However, the frequency of haploid production in corn using inducer lines is 3.59% in the best case (Gilles et al 2017), which is a moderately low frequency, but still useful for breeding, for a plant that produces hundreds of seeds per fruit. For cassava, which only produces three seeds per fruit, the effort to obtain haploids will be greater. Therefore, our future effort will focus exclusively on obtaining mature seeds from crosses with putative haploid inducer cassava lines. Seed thus produced must contain mature and

differentiated embryos, which can be germinated *in vitro* directly, without going through intermediate callus stages that complicate genotyping. From *in vitro* seedlings, DNA will be extracted to perform Genotyping by Sequencing-GBS and determine the heterozygosity of genes with respect to parents, for example, for genes involved in agronomic performance of cassava (Hu et al 2021).

We'll continue making crosses with several mutant lines of the two candidate NLD genes of cassava, Manes.12G093400 and Manes.12G102200, to increase the chances of obtaining and detecting an early haploid or doubled-haploid *in vitro*.

References

Cengiz R (2016). Development of double haploid maize lines by using in vivo haploid technique. Graduate Theses and Dissertations. Master thesis, Namık Kemal Üniversitesi (Published), Tekirdağ

Coe, EH. 1959. A line of maize with high haploid frequency. Am. Nat. 93, 381-382.

Gilles LM, Khaled A, Laffaire JB, Chaignon S, Gendrot G, Laplaige J, Bergès H, Beydon G, Bayle V, Barret P, Comadran J, Martinant JP, Rogowsky PM, Widiez T. 2017. Loss of pollen-specific phospholipase NOT LIKE DAD triggers gynogenesis in maize. The EMBO Journal, 36(6):707-717.

Hu, W., Ji, C., Liang, Z. et al. Resequencing of 388 cassava accessions identifies valuable loci and selection for variation in heterozygosity. Genome Biol 22, 316 (2021). https://doi.org/10.1186/s13059-021-02524-7

Jackson. 2017. No sex please, we're (in)breeding. EMBO J., 36(6):703-704

Kelliher T, Starr D, Richbourg L, Chintamanani S, Delzer B, Nuccio ML, Green J, Chen Z, McCuiston J, Wang W, Liebler T, Bullock P, Martin B. 2017. MATRILINEAL, a sperm-specific phospholipase, triggers maize haploid induction. Nature, 542(7639):105-109.

Zhao X, Xu X, Xie H, Chen S, Jin W. 2013. Fertilization and uniparental chromosome elimination during crosses with maize haploid inducers. Plant Physiol 163: 721 – 731.

Liu C, Li X, Meng D, Zhong Y, Chen C, Dong X, Chen B, Li W, Li L, Tian X, Zhao H Song W, Lou H, Zhang Q, Lai K, Jin W, Yan J, Chen S. A 4-bp Insertion at ZmPLA1 Encoding a Putative Phospholipase A Generates Haploid Induction in Maize. Mol Plant, 10(3):520-522.

Lentini, Z. Restrepo, G. Buitrago, M. Tabares, E. (2020). Protocol for rescuing young *cassava* embryos. Frontiers in plant science 11: 1-13.

H.Q. Li · J.-Y. Guo · Y.-W. Huang · C.-Y. Liang H.-X. Liu · I. Potrykus · J. Puonti-Kaerlas. (1998). Regeneration of cassava plants via shoot organogenesis. Plant Cell Reports 17: 410–414

Table 2. New set of haploid-inducer lines (Transformation ID: T100519) established in the field during the second semester of 2021.

Transformation ID	Genotype	Construct ID	Line ID from Tissue Culture	Line ID from Molecular Analysis	Field Planting Date
T310718	TMS 60444	PHSE401PL1a09	A-1	1	9-May-19
			A-4	4	
			B-1	5	
			B-3	7	
			B-4	8	
			B-6	10	
			A-4	14	
			A-5	15	
			A-6	16	
			A-7	17	
			A-8	18	
			A-10	20	
			A-11	21	
			A-13	23	
			A-15	25	
			A-16	26	
T100519		pL1a10	2	2	10-Dec-20
			3	2	
			4	3	
			5	5	
			6	2	
			7	4	
			8	3	
			9	3	
			10	5	
			11	4	
			12	4	
			13	2	
			14	5	
			15	4	
			16	2	
			17	3	
			18	2	
			19	5	
			20	3	
			21	4	
			22	2	
			23	3	
			24	5	
			25	4	
			26	5	
			27	5	
			28	5	
			29	5	
			30	3	
		60444	Control	1	