Chromosome-scale pearl millet genomes reveal CLAMT1b as key determinant of strigolactone pattern and Striga susceptibility

Kuijer *et al.*

Supplementary Note 1. Comparison with existing assemblies

As of Feb 2024, there were 15 publicly available pearl millet genomes. The first pearl millet genome that was sequenced and published by Varshney et al [1] belongs to the reference genotype Tift $23D_2B_1$-P_1-P_5. The corresponding assembly was constructed using ten pair-end Illumina libraries and 13 mate-pair libraries and scaffolded using BAC libraries. In 2023, 14 new assemblies were published[2,3,4]. Yan et al published ten new assemblies, including one for Tift $23D_2B_1$-P_1-P_5 in 2023[2]. The assemblies of Yan et al. [2] were built using PacBio's HiFi reads and scaffolded based on Bionano optical maps. Final pseudo-molecules were constructed for some of the accessions using Hi-C data and, where Hi-C wasn't available, the genotype PI537069 was used as a reference[2]. In May 2023, Salson et al released yet another assembly for the reference genotype Tift $23D_2B_1$-P_1-P_5 using Oxford Nanopore sequencing and Bionano optical mapping for scaffolding[3]. The final pseudomolecules were built from the scaffolds using the Varshney et al assembly as a reference[1]. In September 2023, Ramu et al published three pearl millet genomes[4]. They re-sequenced, yet again, the reference genotype and two other new accessions[4]. This latest round of assemblies employed PacBio's HiFi reads. Scaffolding was done using Bionano optical mapping, and the final pseudomolecules were created and validated using Omni-C.

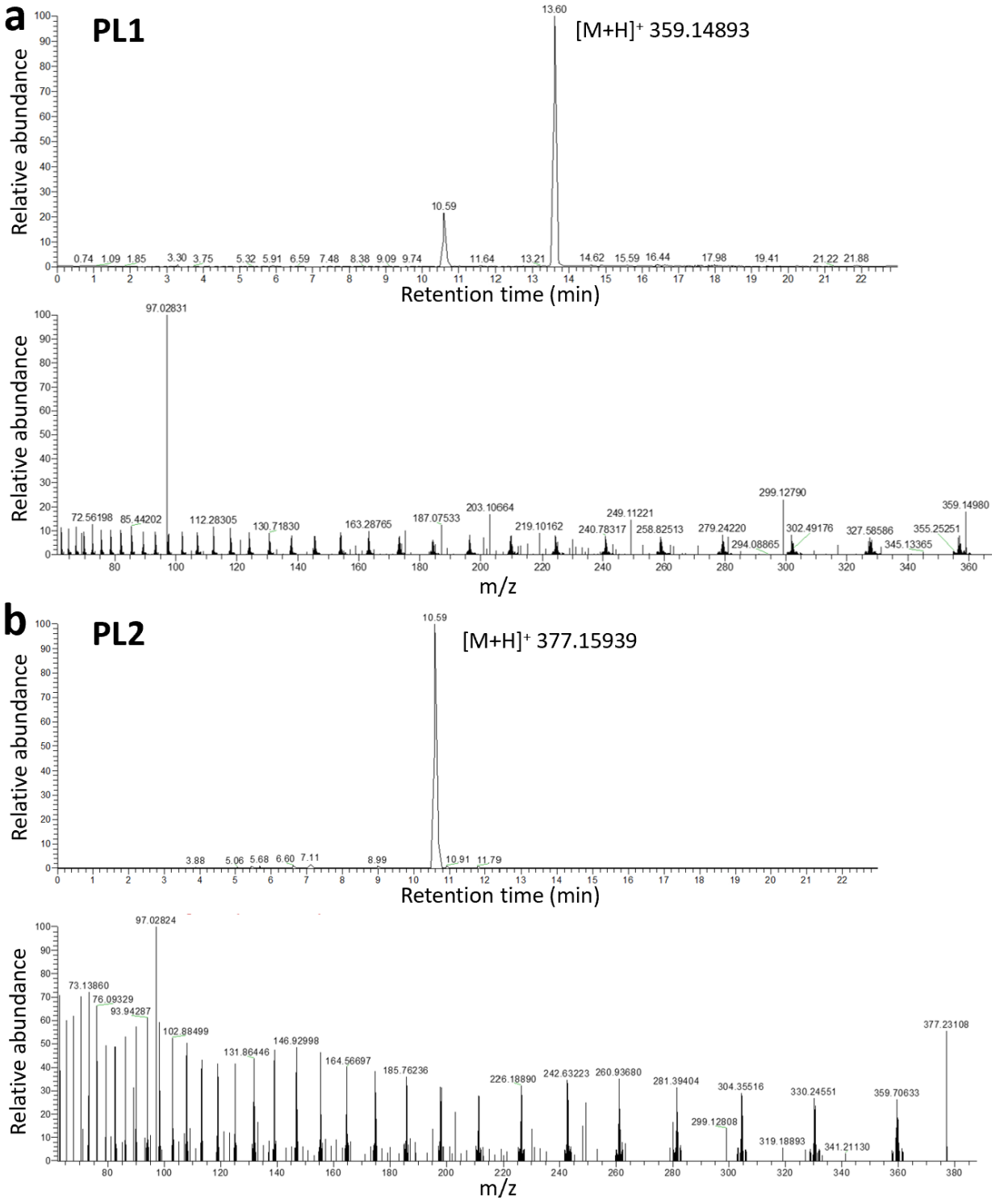
We compared our two pearl millet assemblies with the 14 most recent pearl millet assemblies and excluded the Varshney et al assembly (Supplementary Table 3)[1]. We used these metrics to demonstrate that our assemblies are of as high quality as what has been published, if not higher. Our assemblies are completely *de novo*, and do not rely on a reference for scaffolding. The initial HiFi assembly has produced a small number of contigs with four complete chromosomes for each of the two lines, which were all validated with Omni-C data (Supplementary Table 3). Using Omni-C, four of the remaining six chromosomes (both accessions together) were scaffolded out of two contigs and the remaining two out of three contigs.

To test completeness, we ran BUSCO[5] v5.1.2 for all assemblies (`busco -m genome --lineage_dataset poales_odb10`). BUSCO scores are listed in Supplementary Table 3. The scores are the highest for our assemblies even when compared with the most recent assemblies; albeit by a small margin. We acknowledge and highlight that BUSCO scores must be interpreted with caution, as absence of a gene when comparing different accessions could be a true biological reality rather than an assembly issue. It is difficult to reach a conclusive assessment without additional experimental validations. However, the multiple metrics demonstrate the high quality of our genome assemblies. Taken together, we believe that our assemblies and those published could serve as an excellent panel for a potential pangenome reference that can be used in future variant and other engineering endeavors.

For the purposes of this assessment, we used the following tools: 1) seqkit tool v2.2.0 to manipulate the fasta files and produce basic statistics. 2) The stats tool from the PAGIT package[6], which was used to produce assembly statistics e.g. Nx and Lx values. 3) BBTools's stats.sh tool to estimate the GC content and gaps (BBTools, https://sourceforge.net/projects/bbmap/). The linux Parallel tools was also used to parallelize data transfer between machines[7].

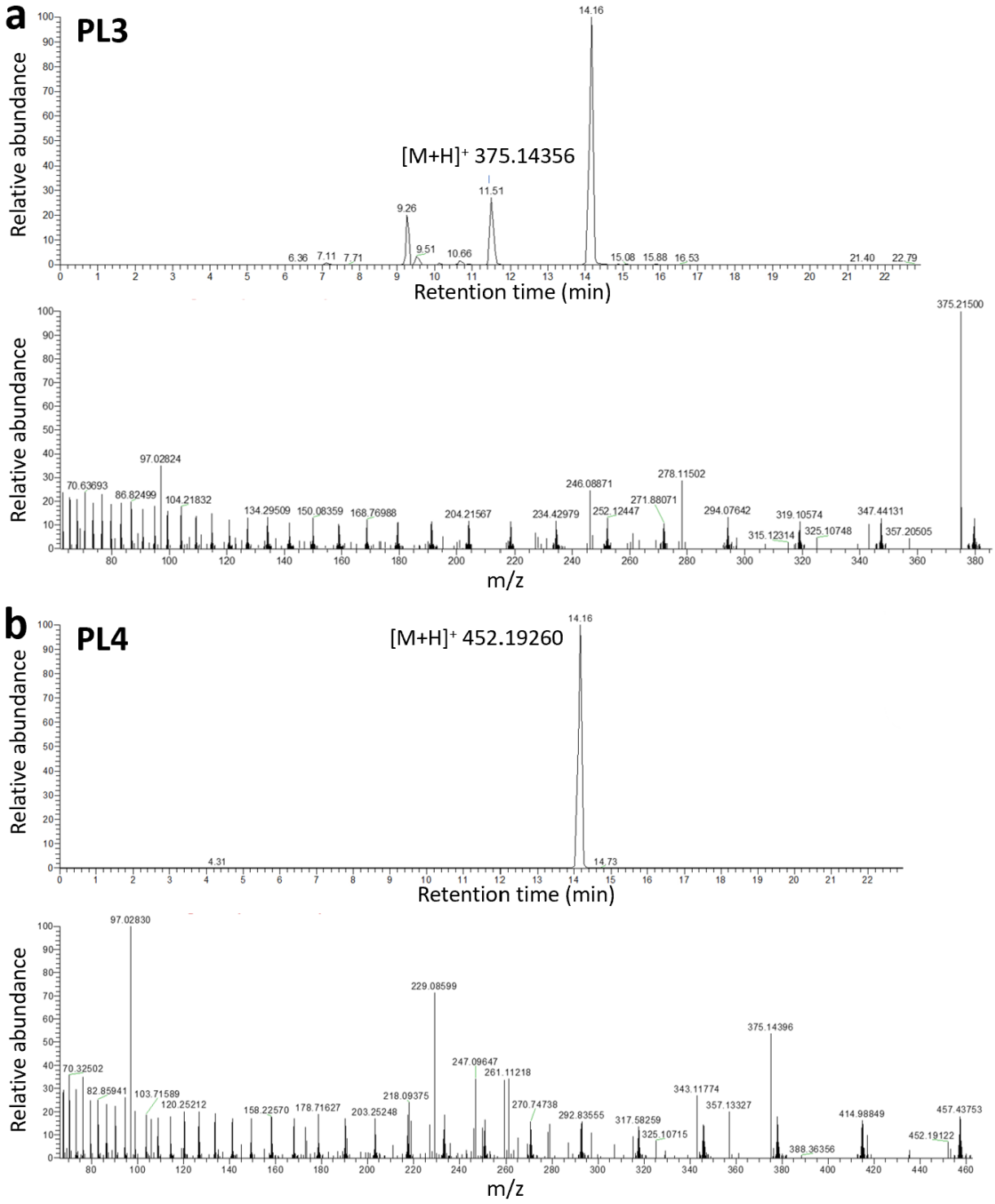
Supplementary Note 2. Hints at post-attachment resistance from comparing Aw and P10 homologs of genes identified in related cereals

We took the genes with significant association to Striga susceptibility from sorghum[8,9] and maize[10] and looked for the sequence differences between the homologous genes in P10 and Aw. Out of the 65 gene pairs analyzed, 63 have fewer than 10 amino acids difference between P10 and Aw, with a total average of 4.02 AA and a median of 2 AA difference (Supplementary data 1). Clear outliers *PgMAKR3* (PgP10c0401G048596.1 / PgAWc0401G047945 + PgAWc0401G047946) and *PgWRKY30* (PgP10c0701G036602.1 / PgAWc0701G036159) show 102 AA and 24 AA difference between P10 and Aw, respectively. Contrary to its annotation, *PgMAKR3* is a CC-NB-LRR resistance gene (Supplementary Figure 5), and most of the differences between the P10 and Aw gene are located in the variable LRR domain, which is involved in pathogen recognition. Pathogen recognition by CC-NB-LRR proteins is translated into an immune response by downstream WRKY transcription factors, a pathway that has been suggested for resistance to parasitic plants[11]. PgWRKY30 could be the WRKY TF paired with PgMAKR3 in the recognition of a pathogen, such as Striga. This finding could be a starting point towards discovering more about the differences in post-attachment resistance between Aw and P10.



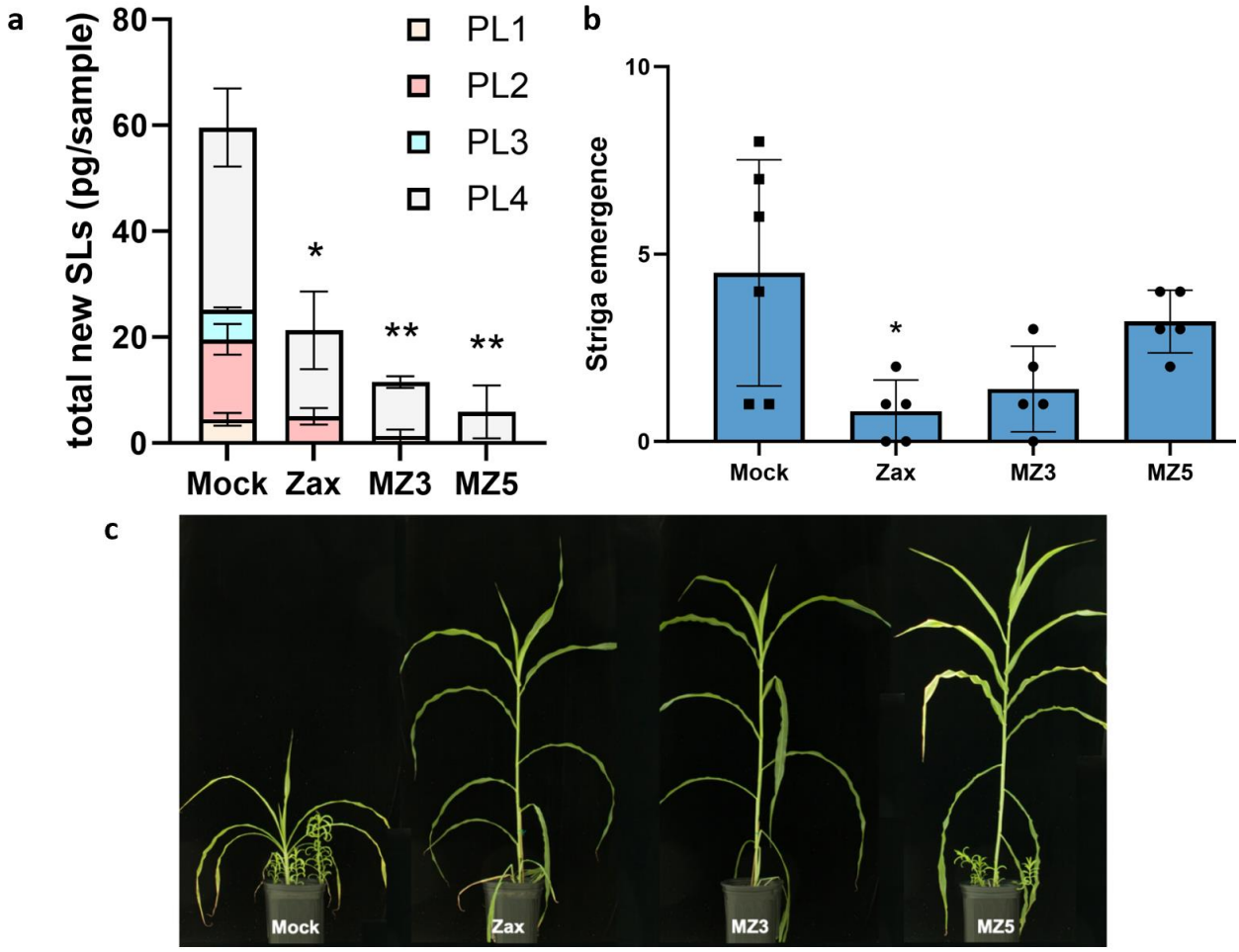
Supplementary Figure 1. LC-MS analysis of PL1 and PL2 identified in P10 root exudate.

a EIC chromatograms of PL1 (*m/z* 359.14893 $[M+H]^+$ in positive mode; retention time 13.60) and MS/MS fragmentation of PL1 with the characteristic D-ring fragment at 97.02831 *m/z*. **b** EIC chromatograms of PL2 (*m/z* 377.15939 $[M+H]^+$ in positive mode; retention time 10.59) and MS/MS fragmentation of PL2 with the characteristic D-ring at 97.02824 *m/z*.



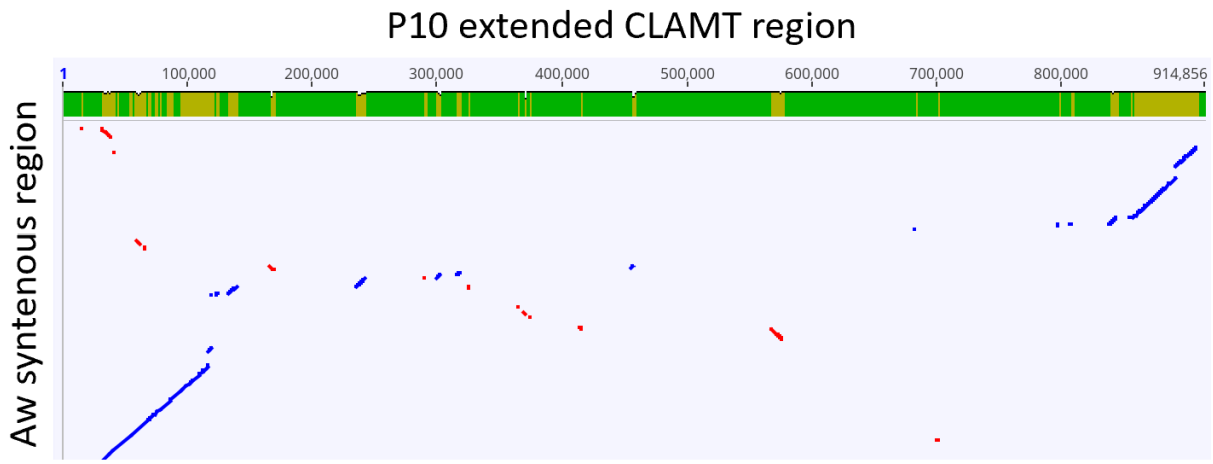
Supplementary Figure 2. LC-MS analysis of PL3 and PL4 identified in P10 root exudate.

a EIC chromatogram of PL3 (*m/z* 375.14356 $[M+H]^+$ in positive mode; retention time 11.51) and MS/MS fragmentation of PL3 with the characteristic D-ring at 97.02824 *m/z*. **b** EIC chromatogram of PL4 (*m/z* 452.19260 $[M+H]^+$ in positive mode; retention time 14.16) and MS/MS fragmentation of PL4 with the characteristic D-ring at 97.02830 *m/z*.



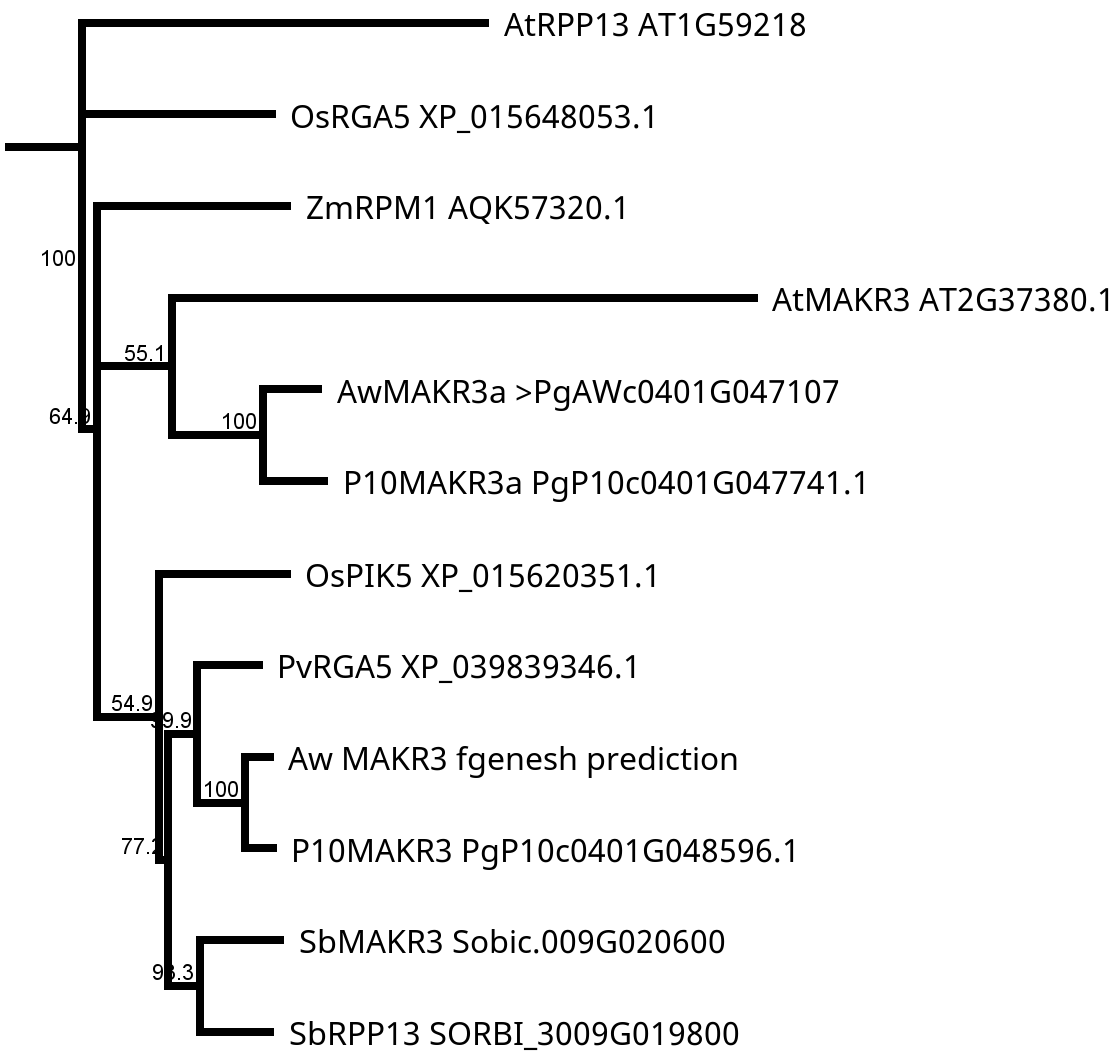
Supplementary Figure 3. Zaxinone and MiZax suppress the production of new SLs in P10, reducing susceptibility to Striga.

a Decrease in the amount of the SLs PL1-4 released by P10 roots upon treatment with zaxinone (Zax), MiZax3 (MZ3) or MiZax5 (MZ5). **b** Treatment with zaxinone (0.25 µM) also significantly decreased Striga emergence, while treatment with MiZax3 and MiZax5 followed the same trend, but did not reach significance in this experiment. **c** Picture of treated plants, showing the decreased susceptibility of P10 upon the application of zaxinone, MiZax3 and MiZax5. Significant differences to mock treatment were tested using ROUT (Q=5%) to discard outliers followed by a two-tailed *t*-test (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$). Source data are provided as a Source Data file.



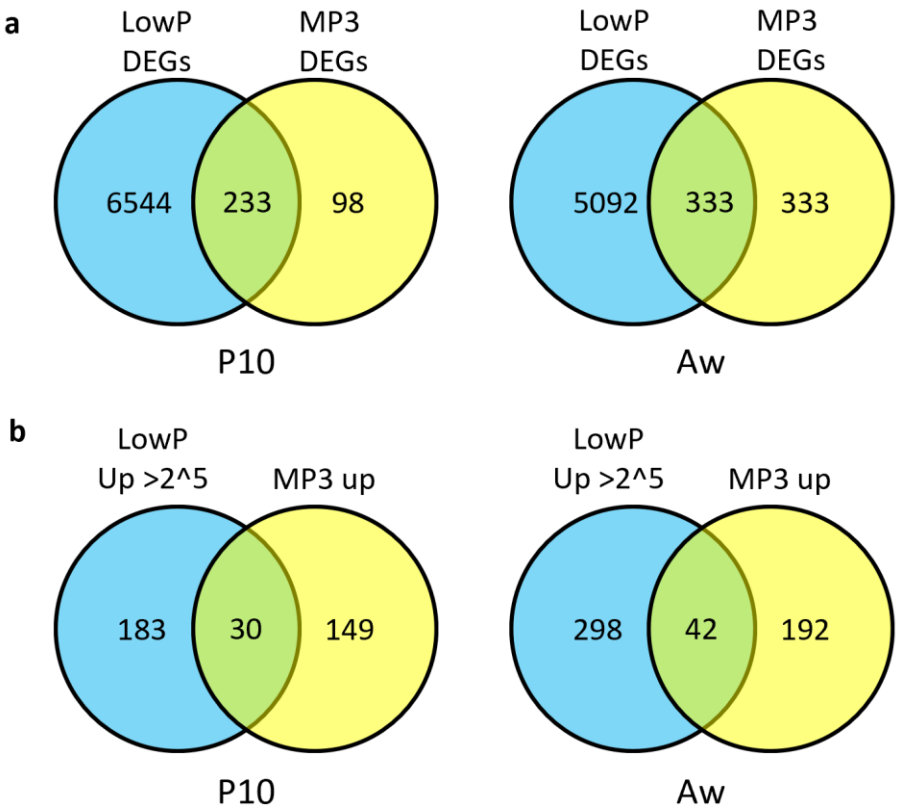
Supplementary Figure 4. **Alignment of the *CLAMT* region and the flanking sequences on chromosome 2 of P10 and Aw**.

The *CLAMT* region on P10 is 0.7Mbp and contains four HC genes, while the equivalent region in Aw is only 0.1Mbp and does not contain predicted genes. Figure created with LASTZ 7.0.3 plugin for Geneious Prime.



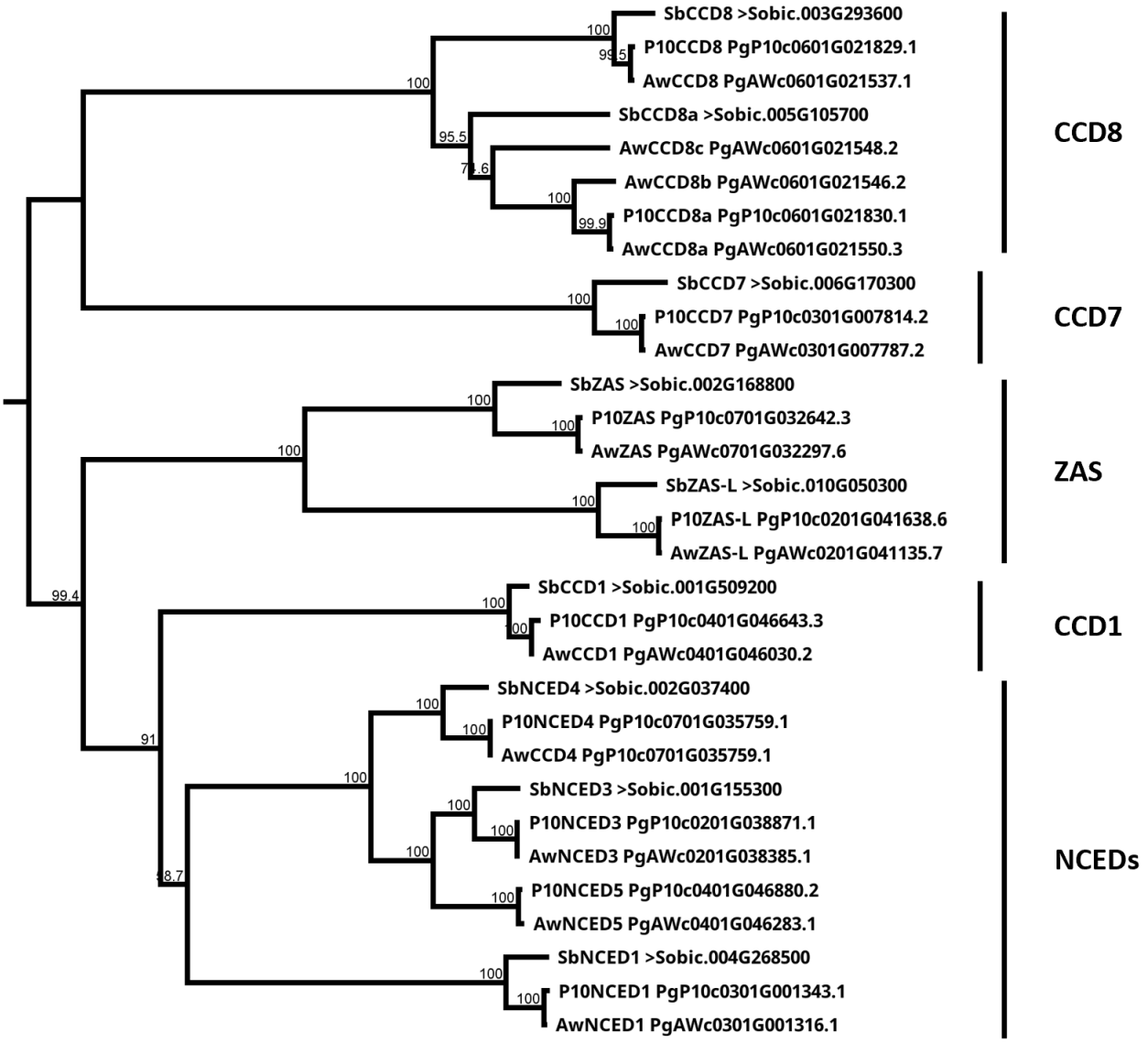
Supplementary Figure 5. **Phylogenetic tree for the MAKR3 gene family**.

SbMAKR3 is a candidate gene for Striga resistance in sorghum, previously detected through GWAS[8]. We named the closest homologs in pearl millet *P10MAKR3* and *AwMAKR3*, but found that they are CC-NB-LRR family genes and that the link to *AtMAKR3* was erroneous. Both pearl millet MAKR3s and the sorghum MAKR3 contain the CC, NB and LRR domains, while the passing resemblance to AtMAKR3 is spread out over the genes in small blocks. The automated annotation of *AwMAKR3* split the gene over two gene numbers, but a targeted manual annotation with Fgenesh+ (Softberry) shows the gene is present and intact in Aw.



Supplementary Figure 6. Differentially expressed genes under low phosphate and MP3 treatment.

a Differentially expressed genes in both P10 and Aw roots exposed to low phosphate were plentiful, while the application of MP3 (an artificial SL analog) affected much fewer genes. The number of genes impacted by both treatments was 233 for P10 and 333 for Aw. **b** Narrowing down the selection to genes strongly upregulated by phosphate starvation (over 2^5 or 32-fold) and upregulated by MP3 left only 30 genes for P10 and 42 for Aw. The resulting list for P10 still includes the predicted SL biosynthetic genes *D27*, *CCD8*, *MAX1-1400* and *CLAMT1b* (Supplementary Table 4).

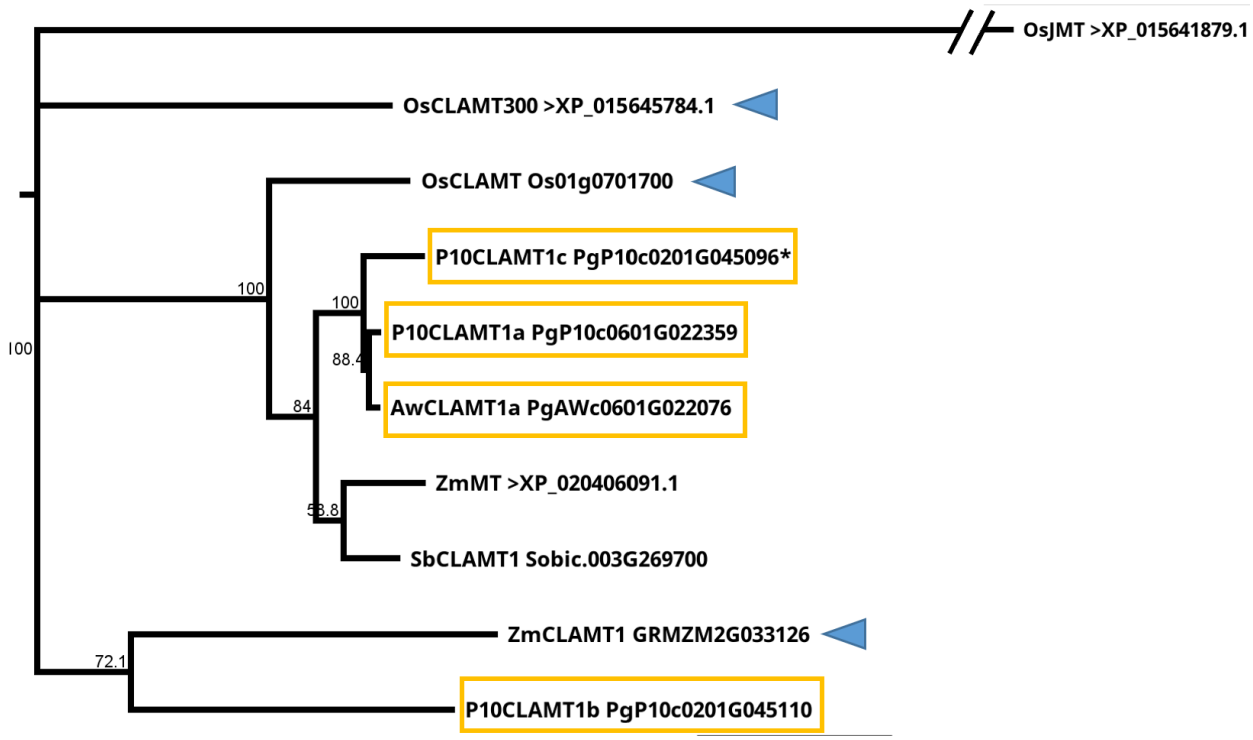


Supplementary Figure 7. **Pearl millet CCDs phylogenetic tree**.

Most members of the PgCCD family have a one-on-one homologous counterpart in sorghum. The only exception is *CCD8*, where P10 has one additional *CCD8*-like gene and Aw has three. Muscle 5.1 PPP alignment Jukes-Cantor, neighbor-joining Consensus tree, 1000 bootstrap.

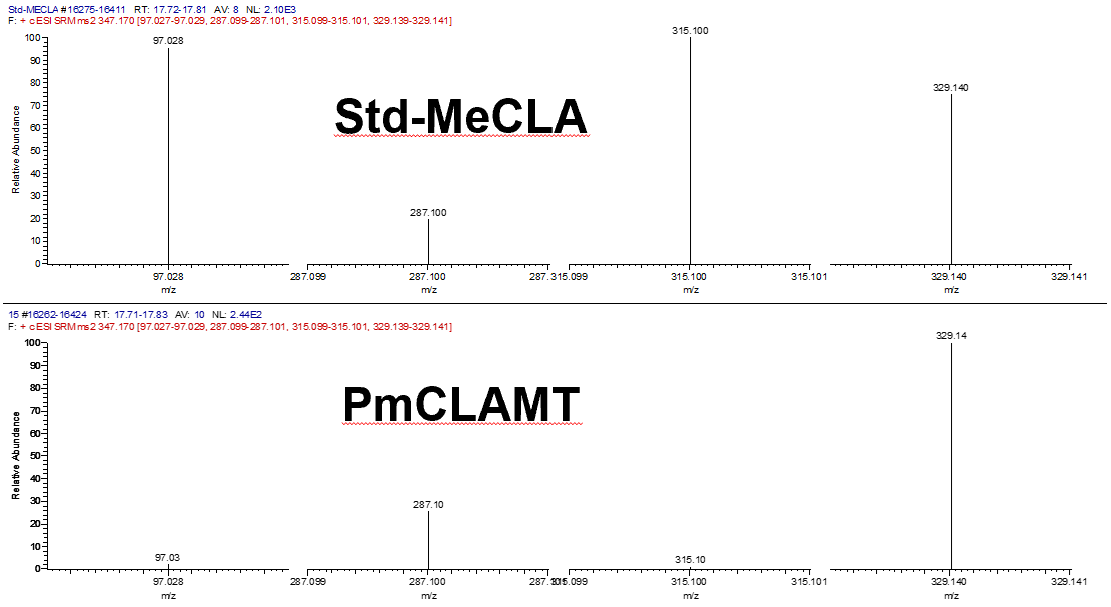
Supplementary Figure 8. **Pearl millet MAX1s phylogenetic tree**.

Pearl millet *MAX1* genes (boxed) are named after the closest sorghum homologs. The P10 and Aw versions for *PgMAX1-900* are identical at the protein level; this is also true for *PgMAX1-1400*. *PgMAX1-900* and *PgMAX1-1400* are part of the *MAX1-900/1400/1500* subfamily and thus likely use carlactone as a substrate. No clear prediction can be made for the exact enzymatic functions of PgMAX1-900 and PgMAX1-1400 based on this tree, since they don't group cleanly with the OsMAX1s, whose activity has been experimentally confirmed in more detail. Muscle 5.1 PPP alignment Jukes-Cantor, neighbor-joining Consensus tree, 1000 bootstrap; *Sobic.001G195100* as outgroup.



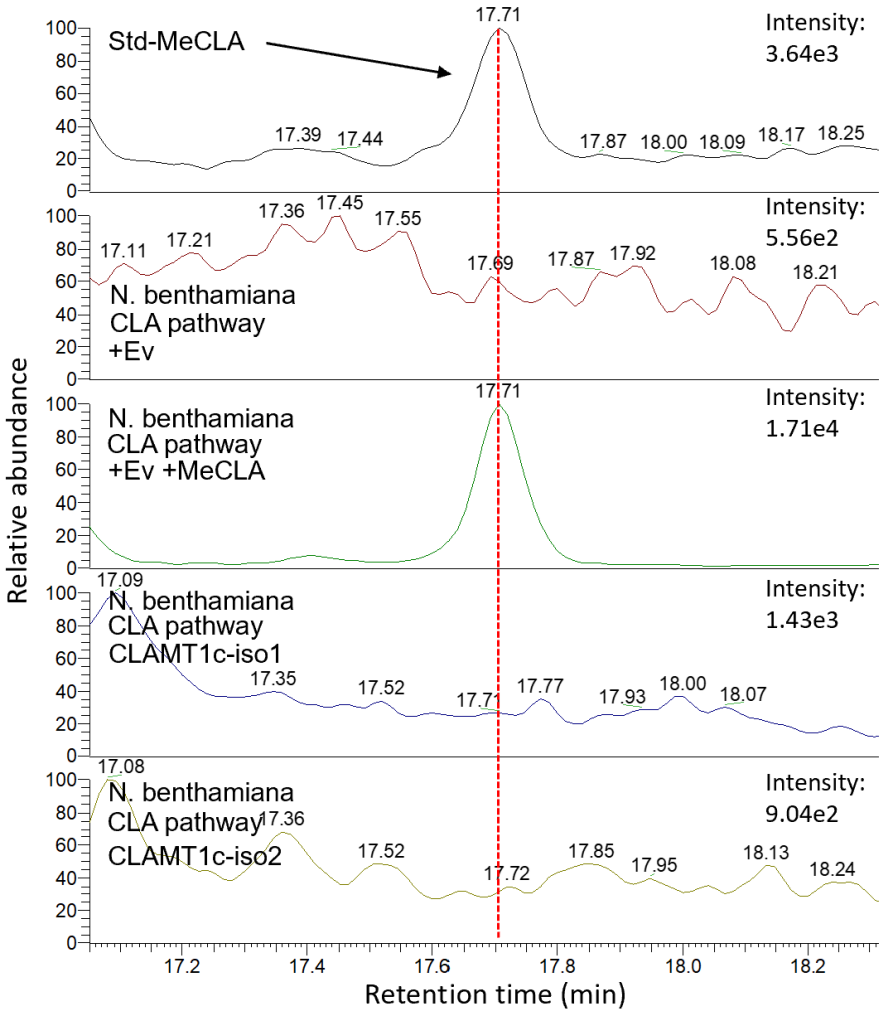
Supplementary Figure 9. **Phylogenetic tree of the *PgCLAMT* gene family**.

There are four *PgCLAMT*s (boxed); both P10 and Aw have a copy of *CLAMT1a* with a very similar sequence, while P10 has an additional two versions of CLAMT named *P10CLAMT1b* and *P10CLAMT1c*. The *CLAMT* genes with proven enzymatic activity from maize[12] and rice[13] are indicated with arrowheads. Muscle 5.1 PPP alignment Jukes-Cantor, neighbour-joining Consensus tree, 1000 bootstrap; *OsJMT* as outgroup.



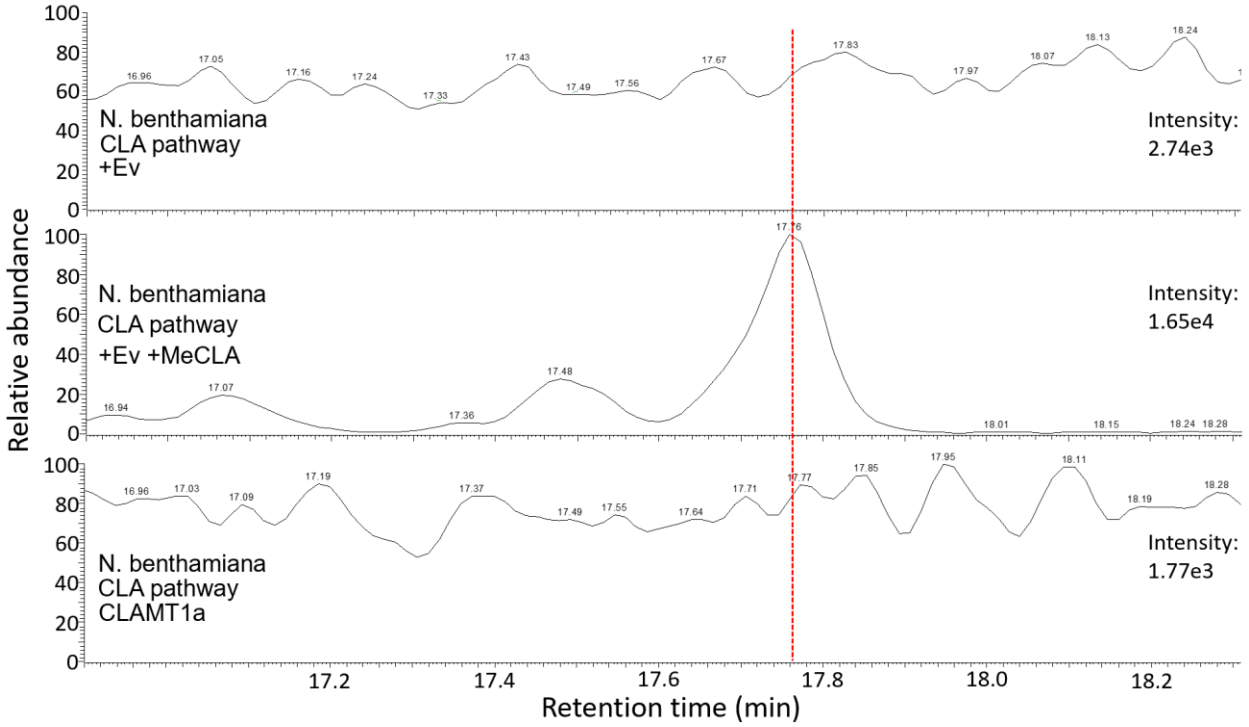
Supplementary Figure 10. Multiple Reaction Monitoring (MRM) comparison of MeCLA produced by P10CLAMT1b and an authentic MeCLA standard.

The MRM comparison of the MeCLA standard and the P10CLAMT1b produced MeCLA indicated in Fig1c.



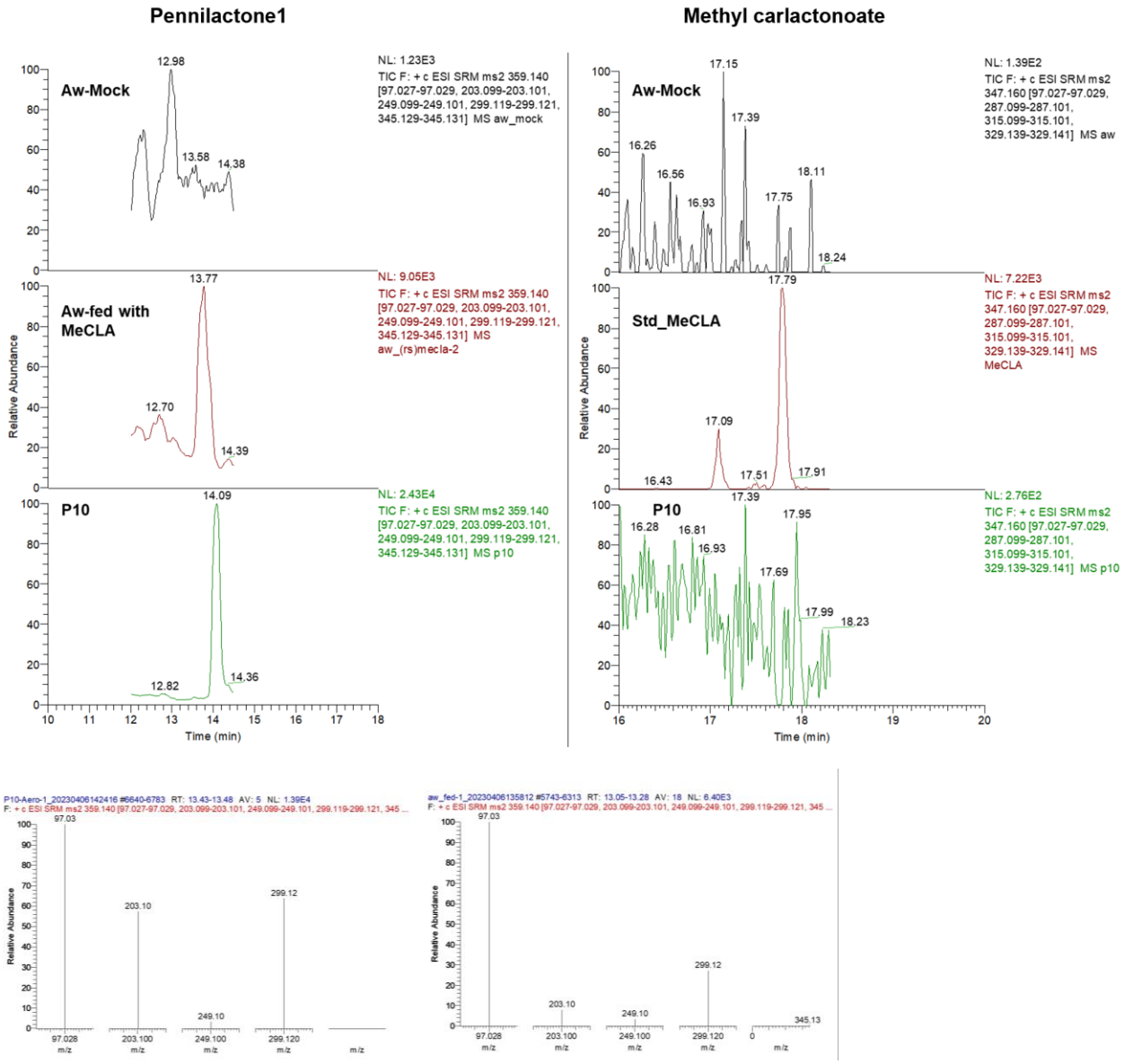
Supplementary Figure 11. Multiple Reaction Monitoring (MRM) analysis of the activity of CLAMT1c.

The two P10CLAMT1c predicted isoforms did not convert CLA into MeCLA when tested in the same tobacco leaf transient expression assay as P10CLAMT1b (Figure 3).



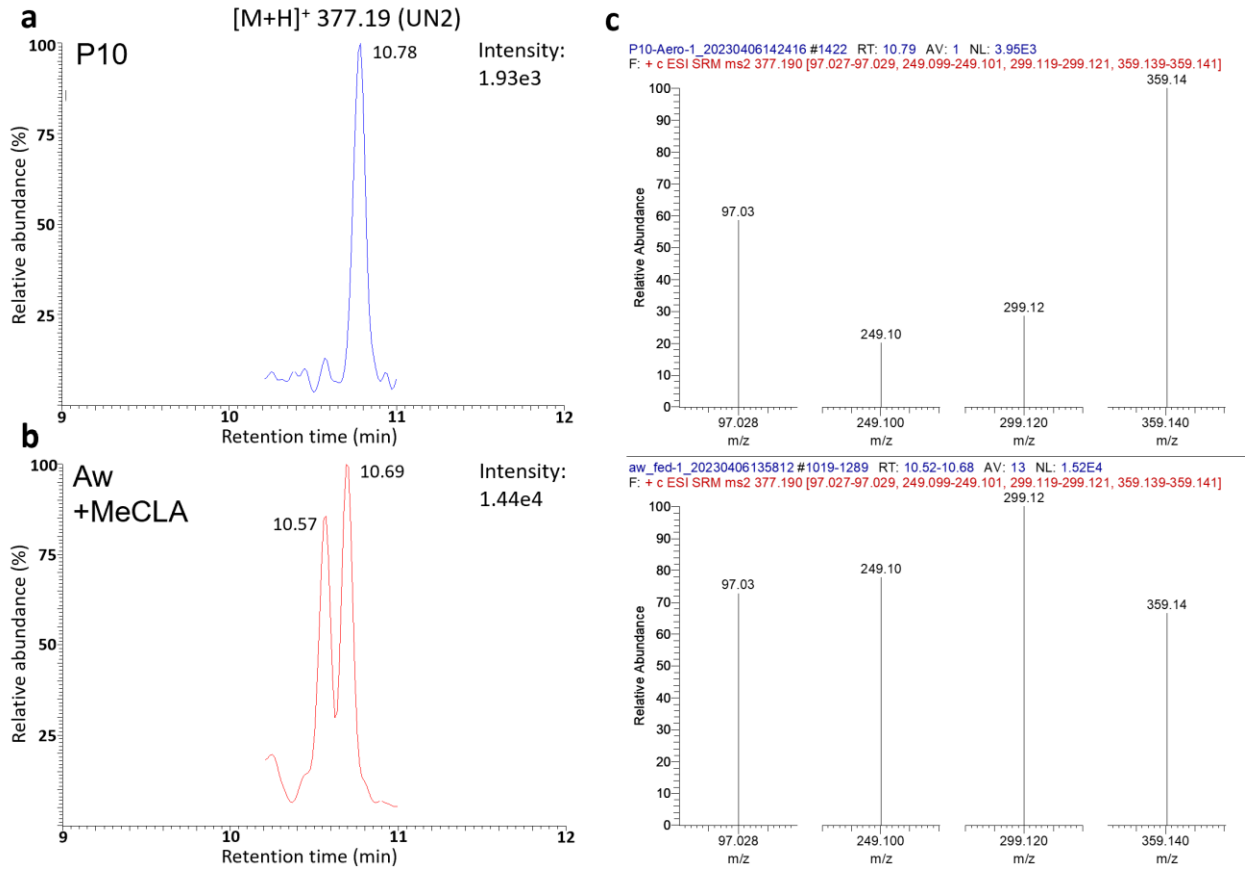
Supplementary Figure 12. Multiple Reaction Monitoring (MRM) analysis of the activity of CLAMT1a.

P10CLAMT1a did not convert CLA into MeCLA when tested in the same tobacco leaf transient expression assay as P10CLAMT1b (Figure 3).

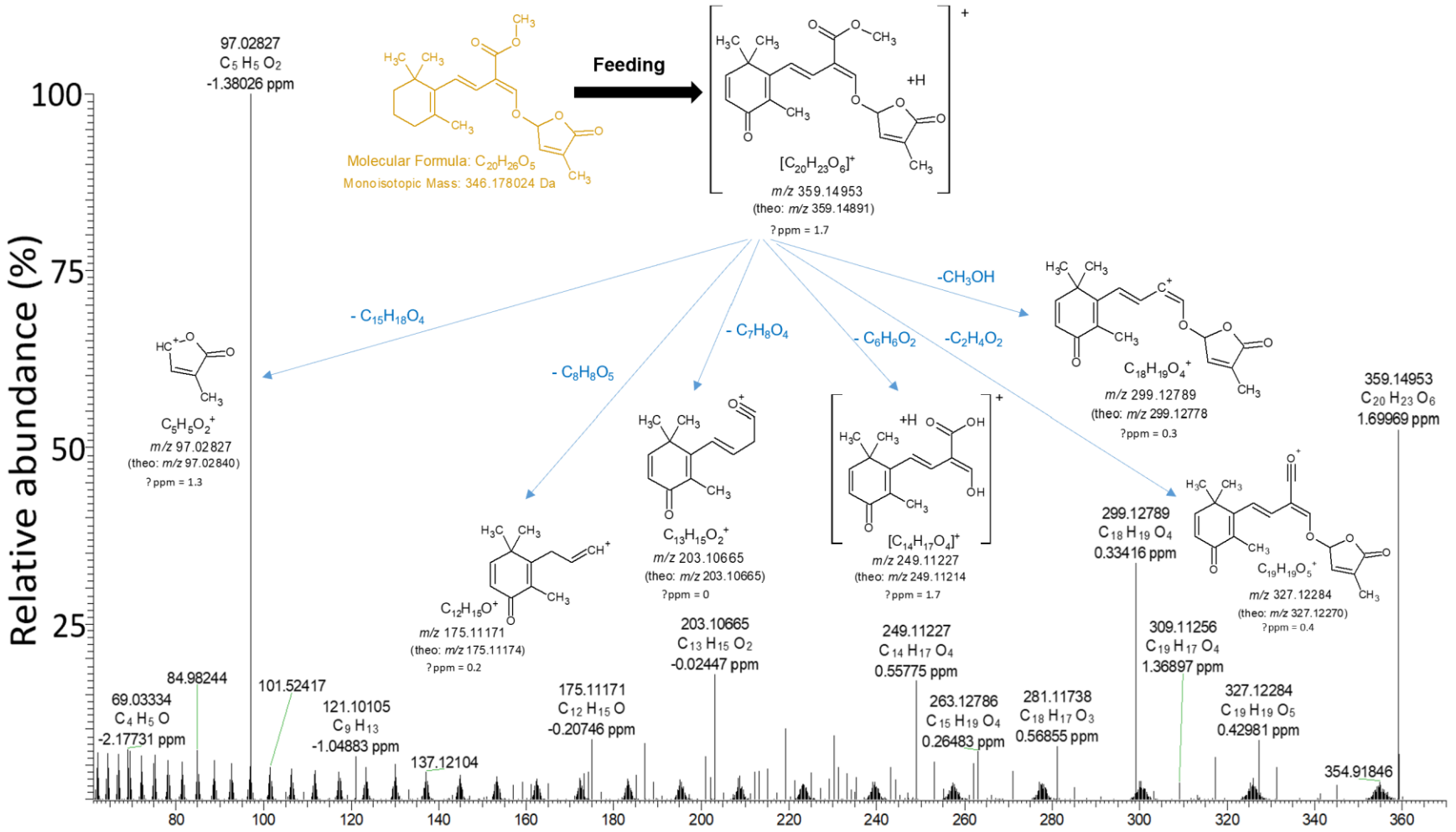


Supplementary Figure 13. Formation of Pennilactone 1 (PL1) by Aw upon feeding with methyl carlactonoate (MeCLA).

Neither PL1 nor MeCLA were detected in exudates of Aw plants (mock) without feeding. However, PL1 was released by Aw after feeding with MeCLA. In P10 exudates, PL1 was present, but not MeCLA. Multiple Reaction Monitoring (MRM) analysis of PL1 from P10 and the unknown compound of similar retention time produced by Aw when supplied with *rac*-MeCLA showed both of them as identical.

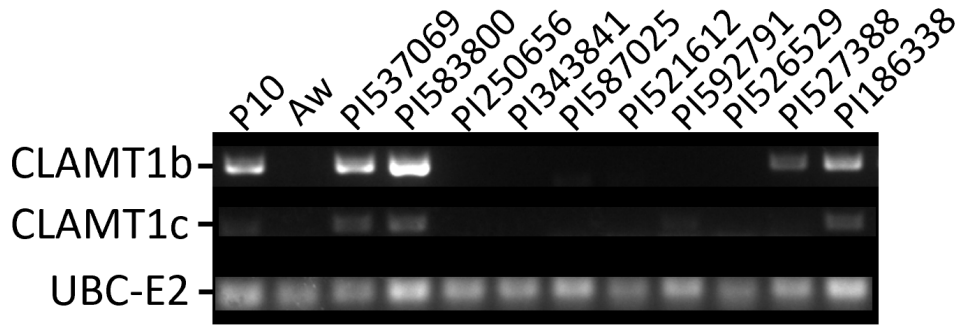


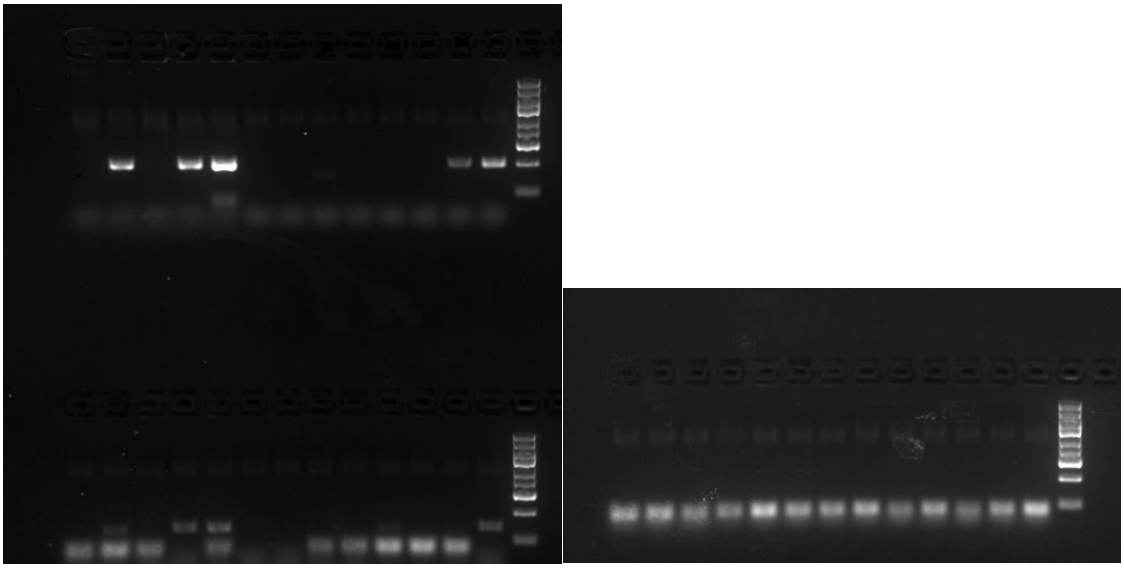
Supplementary Figure 14. Formation of PL2 by Aw upon feeding with methyl carlactonoate (MeCLA). **a** The new SL PL2 (*m/z* 377.19) produced by P10. **b** Formation of two compounds by Aw, with a retention time similar to that of PL2, upon feeding with *rac*-MeCLA. The peak eluting at 10.57 is an isomer of the PL2 eluting at 10.69 that corresponds to PL2 produced in P10. **c** Mass fragmentation of the Aw products (below) confirmed their being identical to PL2 (up).



Supplementary Figure 15. Proposed structure of Pennilactone 1 (PL1) based on mass fragmentation.

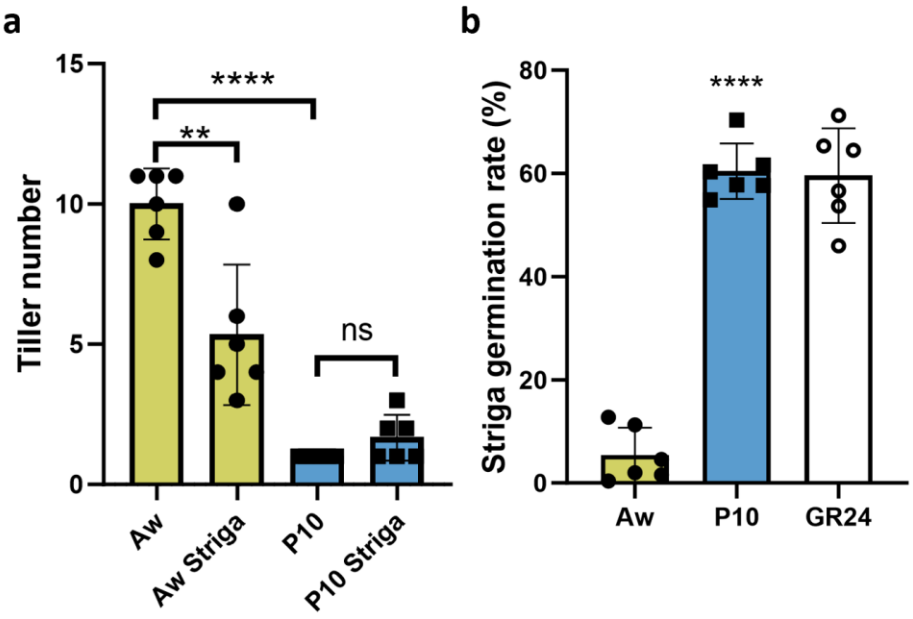
Proposed structure of Pennilactone 1 (PL1) ($[M+H]^+$ 359.14953 in positive mode) was calculated from MS/MS fragmentation, identified from Aw fed with MeCLA as well as P10 root exudate.





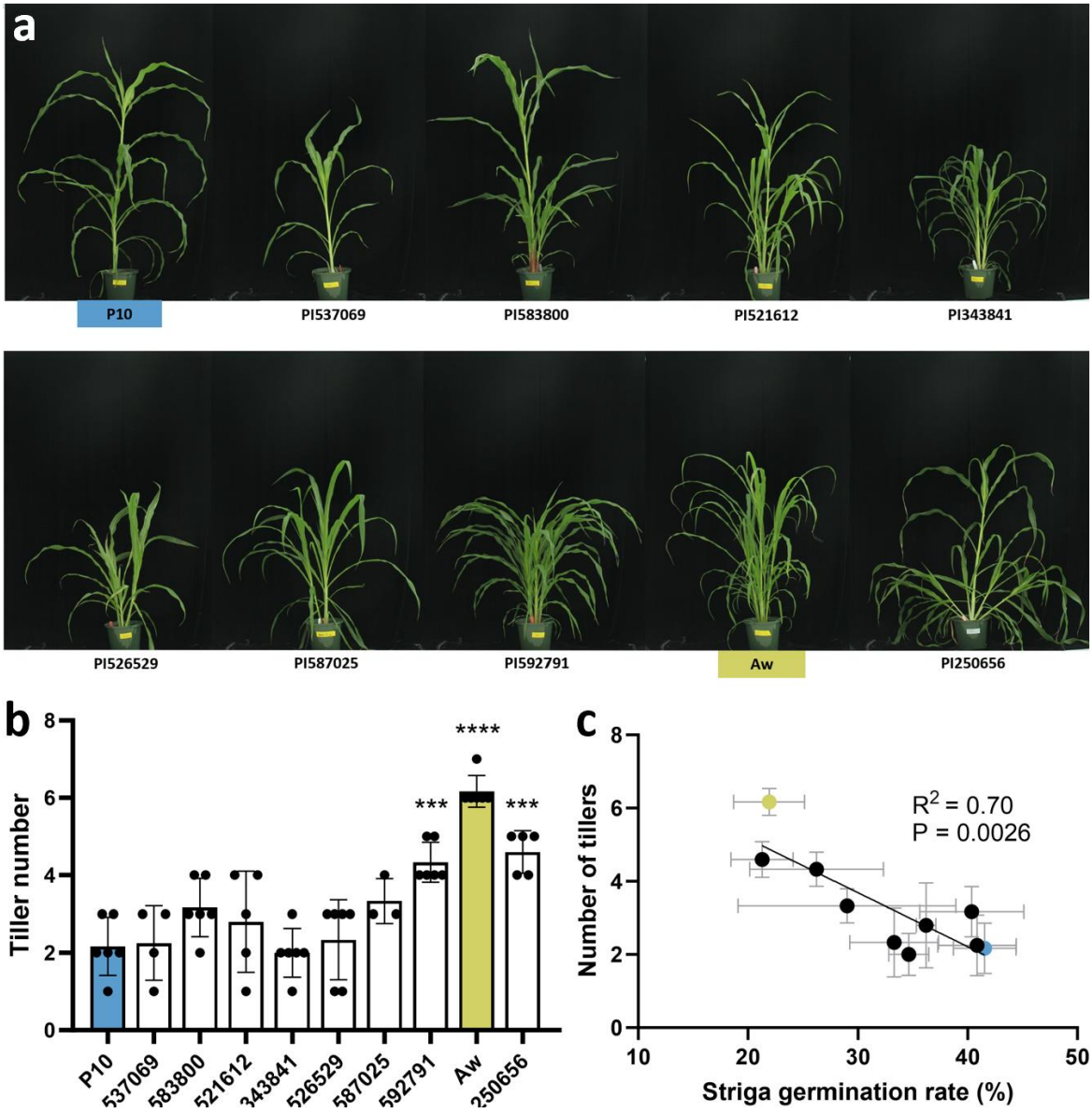
Supplementary Figure 16. Genotyping 10 accessions for the presence of the *CLAMT1b/c* fragment.

PCR-Genotyping of the 10 lines as received, for the presence of the *CLAMT1b* and *CLAMT1c* genes, showed the presence of both genes in P10, PI537069 and PI583800, as expected form their genome sequence. The other eight lines did not show the presence of *CLAMT1b* and *CLAMT1c*, in accordance with their published genomic sequences, except for PI527388 and PI186338. Therefore, the PI527388 and PI186338 seeds we received cannot be considered pure lines and were thus excluded from further analysis. In accordance with the guidelines for gel electrophoresis pictures, the unaltered images are provided with the figure.



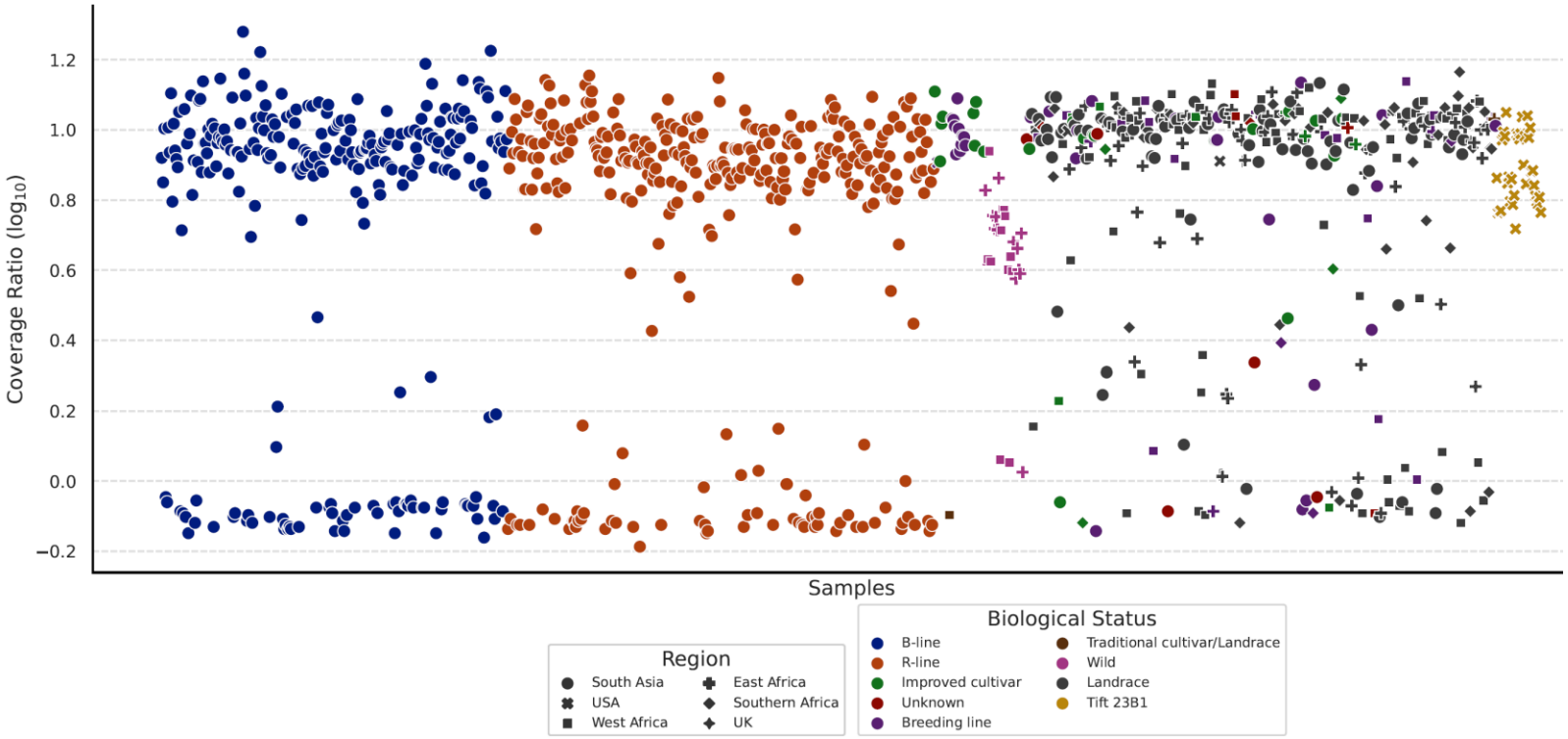
Supplementary Figure 17. Additional phenotyping of Aw and P10 under greenhouse conditions.

a The tiller number of Aw decreased from an average of 10 under normal growth conditions to 5 when growing in Striga infested soil. P10 did not produce any additional tillers under normal conditions, while it sometimes developed additional tillers when grown in Striga infested soil, but this increase was not significant. **b** The Striga seed germination rate induced by P10 root exudate was not only significantly higher than that of Aw, as shown in main Figure 1, but it also reached the rate of the positive control (GR24 treatment). Error bars represent the mean ± s.d. Significant differences were tested using a two-tailed *t*-test (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$). Source data are provided as a Source Data file.



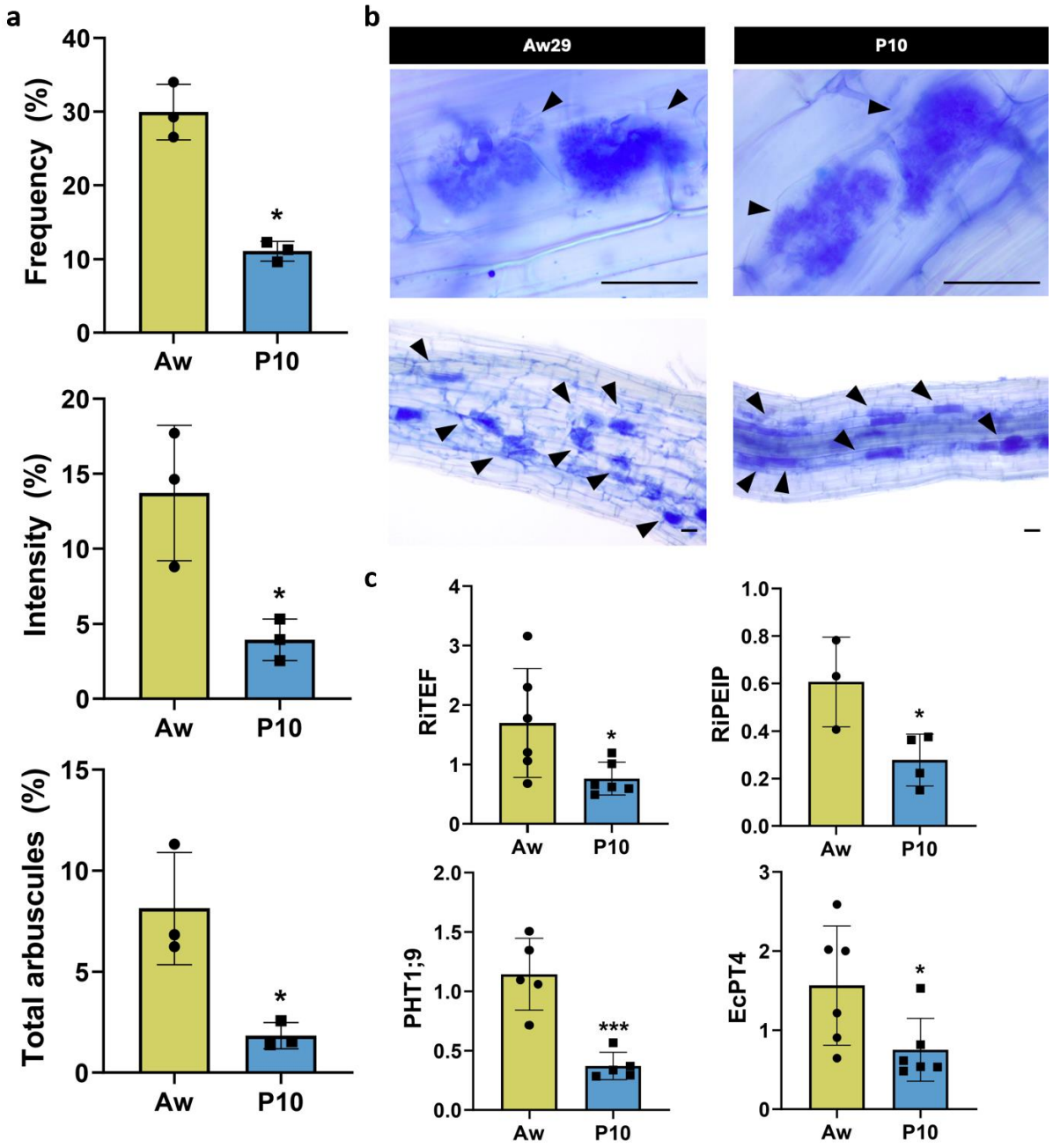
Supplementary Figure 18. Phenotypes of pearl millet panel accessions.

a The pearl millet panel lines, presented in the order of decreasing induction of Striga seed germination by root exudates (Figure 4). The first three lines (P10, PI537069 and PI583800) have the *CLAMT1b* gene and produced pennilactone 1--4. **b** Tiller number was only significantly higher for the three lines with the lowest induction of Striga seed germination (PI592791, Aw and PI250656), compared with P10. **c** The correlation between Striga germination rate and number of tillers is negative. With an R^2 of 0.70 for a linear regression line with a *P*-value of 0.0026 the correlation is significant, but not strongly predictive. Therefore, while the PL-branch SLs produced by pearl millet lines are likely affecting tillering, there are additional factors in tiller number determination obscuring a more close correlation. Error bars represent the mean ± s.d. Significant differences were tested using a two-tailed *t*-test (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$). Source data are provided as a Source Data file.



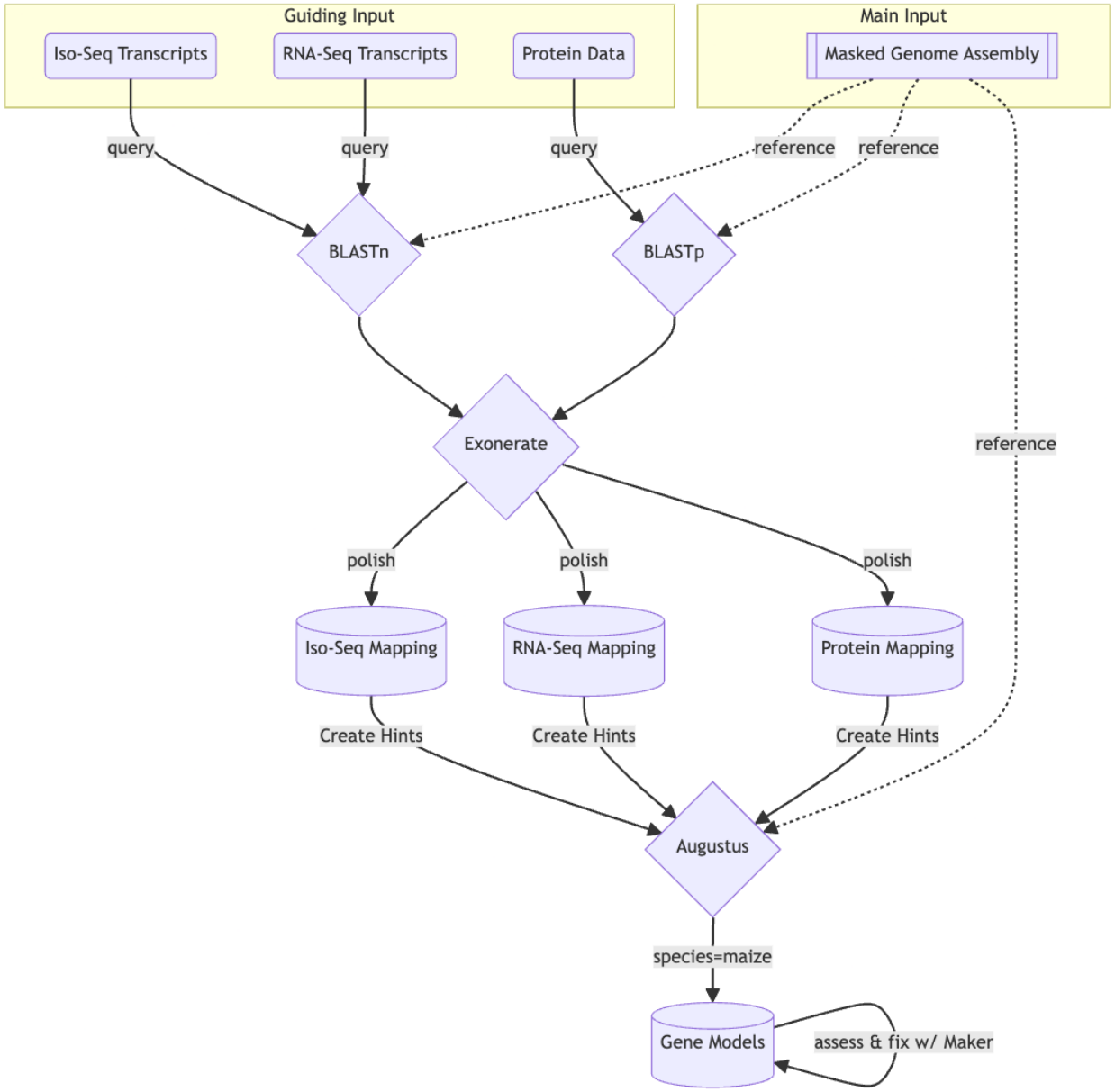
Supplementary Figure 19. **Distribution of the CLAMT region in a wide selection of pearl millet accessions.**

Using resequencing data from a wide selection of pearl millet accessions[13], the ratio of the coverage of the P10 chromosome 2 and the coverage of the 0.7Mbp CLAMT region was plotted for each accession. This results in a value of about 1 when the CLAMT region is present, as the coverage is approximately the same as the average coverage of chromosome 2, while the value is much higher when the CLAMT region is absent. These values are plotted as log_{10} of the ratio, where outcomes at or below 0.2 are deemed positive for the CLAMT region, 0.2-0.6 as inconclusive and 0.6 and above as negative, totaling 175, 38 and 822 accessions respectively. The graph shows a clear separation between the two populations and a wide distribution of both accessions with the CLAMT region and without, regardless of provenance. The method of resequencing, RAD or WGS, does not seem to affect the ratio in a significant way. As a control, we included shotgun sequencing data from TIFT-23B1, which does not contain the CLAMT region, and indeed it shows only ratio values above 0.6.

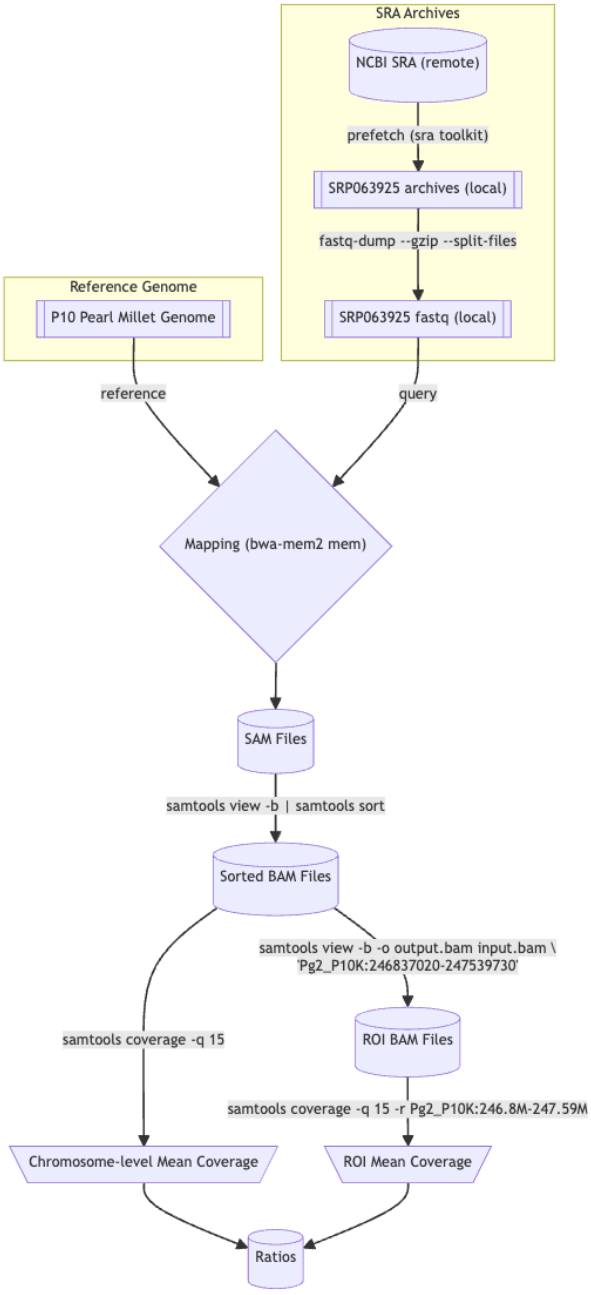


Supplementary Figure 20. Mycorrhizal colonization of P10 and Aw.

a Morphological evaluation shows that P10 has a significant reduction in both frequency and intensity of colonization, compared to Aw. F%: frequency of mycorrhizal colonization; M%: intensity of mycorrhizal colonization; A%: total number of arbuscules. **b** Arbuscule formation at 40 day-post inoculation. Arrows indicate arbuscule-containing cells (Scale bars, 50 μm). **c** The colonization by the AM fungus *Rhizophagus irregularis* was quantified by measuring the expression of fungal genes (*RiEF* and *RiPEIP1*) and plant AM marker genes (*SiPHT1;9* and *EcPT4*), presented here as normalised expression values. Data are mean ± SE (n>=4). Significant values (by One-way ANOVA) are shown as follows: **P* < 0.05; ***P* < 0.01, *** *P* < 0.001. Source data are provided as a Source Data file.



Supplementary Figure 21. Workflow for Aw and P10 genome annotation.



Supplementary Figure 22. Workflow for analysis of resequencing data.

Supplementary Table 1. **Extended genome assembly and annotation statistics for Aw and P10**.

	Aw				P10			
	Length	***GC %***	***GC Skew %***	***Num of gaps*** [1]	***Length***	***GC %***	***GC Skew %***	***Num of gaps*** [1]
Chr1	314,009,907	48.86	0.16	0	307,615,467	48.85	0.11	0
Chr2	276,062,833	49.1	-0.25	2	281,688,668	49.05	-0.11	0
Chr3	325,849,051	49.21	0.09	0	334,621,887	49.22	-0.01	1
Chr4	243,773,284	48.54	0.17	2	251,295,635	48.56	0.12	1
Chr5	174,422,329	49	-0.15	1	177,041,146	49.11	0.04	0
Chr6	286,409,356	49.3	0.06	0	282,264,073	49.47	0.07	1
Chr7	283,920,974	48.96	0.02	0	281,846,746	48.91	-0.03	0
ChrUn	10,392,345	53.28	1.31	152	9,861,679	47.74	1.18	235
	Genes/mRNA	***CDS / Exons***	***5' UTRs***	***3' UTRs***	***Genes/mRNA***	***CDS / Exons***	***5' UTRs***	***3' UTRs***
HC	38,920/62,477	284,910/312,473	39,586	38,211	40,689/66,411	295,072/320,789	38,568	37,083
LC	22,615/25,091	44,193/44,253	184	75	23,248/26998	49,413/49,470	175	83
BUSCO [2]	C:93.0%[S:58.7%,D:34.3%],F:2.5%,M:4.5%,n:4896				C:93.5%[S:59.5%,D:34.0%],F:2.4%,M:4.1%,n:4896			

HC is High Confidence genes and LC is low confidence genes.

[1] Each gap is represented with 100 Ns.

[2] Complete; S: Single BUSCOs; D: Duplicate BUSCOs; F: Fragmented; M: Missing; n: 4896 reference genes from Poales

Supplementary Table 2. Repeat content of the Aw and P10 genomes.

		Aw		P10	
		bp	%	bp	%
Class 1	LTR_Copia	374306143	19.55	337394745	17.52
	LTR_Gypsy	835647004	43.64	870042961	45.17
	LTR_unknown	234813472	12.26	196586386	10.21
Class 2	DNA_HAT	11691551	0.72	11672174	0.73
	DNA_CACTA	46761294	2.45	48386023	2.52
	DNA_Harbinger	15468933	0.98	16215262	0.99
	DNA_Mutator	26438164	2.38	25044840	2.38
	DNA_Mariner	15194248	0.91	16458865	0.99
	Total	1587408978	82.90%	1550644898	80.50%

The amount of TE related sequences is highly similar in the two accessions: 1.59 and 1.55 Gbp corresponding to 82.90% and 80.50% of the total genome size for Aw and P10, respectively. These values are greater than those reported by Varshney et al [1] , i.e. 1.22 Gbp and 77.2 %, but they entirely meet the expectation of Varshney et al made for a true minimum TE fraction of 80% of the entire genome[1]. This testifies the greater efficiency of long read based sequencing techniques in capturing the repetitive fraction of genomes. The most abundant class of TE in Pearl millet is that of Long Terminal Repeat (LTR) retrotransposons totaling to 75.45% and 72.9% in Aw and P10 genomes, respectively with the Ty3-gypsy superfamily (43.64% and 45.17%) largely more abundant that the Ty1-copia one (12.26% and 17.21%). DNA TEs represent 7.44% and 7.66% of the genome sequence in Aw and P10, respectively with CACTA and Mutator families being the more abundant (always representing more than 2% of the genome sequence).

Supplementary Table 3. Comparison to published pearl millet genome assemblies.

Cultivar	Bioproject	Accession	# Gaps	BUSCO Complete (%)	Complete (%) [Singletons, Doubletons]	Fragmented (%)	Missing (%)	Notes [&]	Reference
Aw	PRJEB71762		5	98.5	[S:95.1,D:3.4]	0.20	1.30		This paper
P10	PRJEB71762		3	98.6	[S:95.4,D:3.2]	0.20	1.20		This paper
Tift-2023			296	98.3	[S:95.2,D:3.1]	0.20	1.50		4
ICMR 06777			26	98.4	[S:95.3,D:3.1]	0.20	1.40		4
843 B			143	98.4	[S:95.4,D:3.0]	0.20	1.40		4
PI537069	PRJNA749489	GCA_020739565.1	91	96.2	[S:91.9,D:4.3]	0.20	3.60		2
Tifleaf 3	PRJNA749489	GCA_020739585.1	393	94.9	[S:87.6,D:7.3]	0.30	4.80		2
PI526529	PRJNA749489	GCA_020739535.1	329	94.1	[S:88.7,D:5.4]	0.30	5.60		2
PI583800	PRJNA749489	GCA_020739575.1	1202	95.3	[S:91.0,D:4.3]	0.40	4.30		2
PI521612	PRJNA749489	GCA_020739525.1	832	96.6	[S:89.2,D:7.4]	0.50	2.90		2
PI587025	PRJNA749489	GCA_021560375.1	1617	97	[S:92.1,D:4.9]	0.50	2.50		2
PI186338	PRJNA749489	GCA_027789755.1	872	96.8	[S:91.6,D:5.2]	0.40	2.80		2
PI527388 [R]	PRJNA749489	GCA_027789915.1	2186	95.7	[S:91.9,D:3.8]	0.40	3.90	Contam. warning	2
PI343841	PRJNA749489	GCA_027745475.1	1809	96.8	[S:91.9,D:4.9]	0.30	2.90		2
PI250656 [R]	PRJNA749489	GCA_027745265.1	2186	95.7	[S:91.9,D:3.8]	0.40	3.90	Contam. warning	2
Tift 23D2B1-P1-P5	PRJEB57746	GCA_947561735.1	120	96.1	[S:93.1,D:3.0]	0.30	3.60	NCBI Ref	3
Tift-2018	PRJNA294988	GCA_002174835.2	201600	NA	NA	NA	NA		1

BUSCO settings: BUSCO version 5.1.2; BUSCO DB: poales_odb10; # BUSCOs: 4896; Mode: Genome.

ℝ: In a private email communication, NCBI staff have confirmed that despite the similarity in lengths, the sequences are different.

&: Notes taken from source repository e.g. NCBI Genome

Supplementary Table 4. List of genes from P10 and Aw, which were both highly induced (>2^5) by low phosphate conditions and then induced significantly more by the addition of MP3 treatment.

P10 lowP/MP3 up	description from similar proteins	Aw lowP/MP3 up
PgP10c0101G010112	Tropinone reductase	PgAWc0101G010023
	Obtusifoliol demethylase	PgAWc0101G011320
	Chalcone synthase	PgAWc0101G011727
PgP10c0101G011332	D27	
PgP10c0101G012222	probable carboxylesterase	
PgP10c0101G012240	Abscisic acid 8'-hydoxylase 3	
PgP10c0101G017802	benzyl alcohol O-benzoyltransferase	PgAWc0101G017579
	hypothetical protein	PgAWc0201G036589
	F-box/FBD/LRR-repeat protein At1g13570-like	PgAWc0201G036590
	Ergosterol biosynthetic protein 28	PgAWc0201G038150
	cytochrome P450 89A2	PgAWc0201G038988
PgP10c0201G040400	ent-copalyl diphosphate synthase 1, chloroplastic	PgAWc0201G039931
	protein CYCLOPS-like	PgAWc0201G040446
PgP10c0201G042842	putative 2-oxoglutarate-dependent dioxygenase	PgAWc0201G042298
PgP10c0201G042843	ent-isokaurene C2-hydroxylase	PgAWc0201G042299
	unknown	PgAWc0201G044004
	unknown	PgAWc0201G044005
PgP10c0201G045110	CLAMT1b	
	bidirectional sugar transporter SWEET15	PgAWc0301G003590
	UDP-glycosyltransferase 85A2	PgAWc0301G005124
PgP10c0301G006912	ABC transporter B family member 15	PgAWc0301G006836
	hypothetical protein	PgAWc0401G047069
	S-norcoclaurine synthase 1	PgAWc0401G047821
PgP10c0401G048472	2-oxoglutarate-dependent dioxygenase - S-norcoclaurine synthase 1	
PgP10c0401G048625	2-oxoglutarate-dependent dioxygenase - S-norcoclaurine synthase 1	
PgP10c0401G048754	indole-2-monooxygenase	PgAWc0401G048099
PgP10c0401G049555	aquaporin NIP3-3	
PgP10c0401G051155	CYP706	PgAWc0401G050240
PgP10c0401G051273	uncharacterized protein	
	2-alkenal reductase like	PgAWc0401G052246
PgP10c0401G053327	ent-copalyl diphosphate synthase 2	PgAWc0401G052458
PgP10c0501G055622	NADP-dependent alkenal double bond reductase P2-like	
PgP10c0501G055754	cytochrome P450 93A3	
PgP10c0501G056021	cytochrome P450 93A3	
	squamosa promoter-binding-like	PgAWc0501G055634
PgP10c0501G058210	serine carboxypeptidase-like 34	PgAWc0501G057204
	cytochrome P450 76M5	PgAWc0501G058053
	serine/threonine receptor-like kinase NFP	PgAWc0501G059262
PgP10c0601G019773	Tropinone reductase-like protein	
PgP10c0601G021829	CCD8	PgAWc0601G021537
PgP10c0601G022089	ABC transporter G family member 38-like	PgAWc0601G021810
PgP10c0601G022322	germin-like protein 1-2	PgAWc0601G022036
PgP10c0601G022366	MAX1-1400	PgAWc0601G022078
PgP10c0601G023133	putative fatty-acid--CoA ligase	PgAWc0601G022896
PgP10c0601G027441	hypothetical protein	
	short-chain dehydrogenase TIC 32, chloroplastic	PgAWc0701G027593
	ABC transporter B family member 4	PgAWc0701G027594
	cytochrome P450 709B2	PgAWc0701G028056
PgP10c0701G029515	cytochrome P450 716B1-like	PgAWc0701G029274
PgP10c0701G029516	cytochrome P450 716B1	PgAWc0701G029275
PgP10c0701G031847	putative amidase At4g34880-like	PgAWc0701G031529
	unknown	PgAWc0701G032050
	ZAS	PgAWc0701G032297
	epoxide hydrolase A-like	PgAWc0701G034036
	putative 1-deoxy-D-xylulose-5-phosphate synthase 2, chloroplastic	PgAWc0701G034758

The 18 genes in common between P10 and Aw are given a grey background. Homologs of enzymes reported to be part of the strigolactone biosynthesis pathway in maize are highlighted in yellow. In the case of *CCD8*, *CLAMT* and *MAX1* these are not the only homologous options in pearl millet (see Supplementary Figures 7, 8 and 9), but we can conclude from this co-expression that these are the most likely candidates to be active genes in the strigolactone biosynthetic pathway. The candidate genes for the production of PL1 and PL2 from methyl carlactonoate can be upregulated in both P10 and Aw, however for the genes involved in producing PL3 and PL4 this is not expected, since these were not detected after feeding Aw with methyl carlactonoate.

Supplementary Table 5. Presence (green) or absence (red) of genes in the *CLAMT* gap and flanking areas of chromosome 2 for P10, Aw, and the 10 sequenced pearl millet lines described by Yan et al. [2].

		flavonoid methyltra	WVD2-like	CLAMT1c	acyl trans-ferase	CYP51	CLAMT1b	Glucosi-dase	SEC14
P10									
Aw									
PI592791	Tifleaf 3	98	95					96	97
PI537069	Baoudarache	99	99	99.94	99.77	99.90	100	100	99
PI583800	ICMV 88908	99	99	99.26	99.77	99.93	100	96	99
PI343841	-	99	98					97	99
PI527388	11	98	96					95	97
PI526529	AMM 1227	98	96					96	97
PI521612	Ngululu	100	99					96	97
PI587025	Dokhn	98	96					97	97
PI250656	K551	98	96					95	97
PI186338	-	98	96					95	97

Numbers in each cell indicate the percentage of the protein sequence that is identical to P10, showing that the CLAMT1b proteins produced by P1537069 and P1583800 are identical to the P10 CLAMT1b.

Supplementary Table 6. Accurate Mass $[M+H]^+$ of natural strigolactones in *Orbitrap ID-X Tribrid* Mass Spectrometer .

Strigolactones	Molecular formula	$[M+H]^+$
Orobanchol	$C_{19}H_{22}O_6$	347.14891
Oro-Acetate	$C_{21}H_{24}O_7$	389.15948
UN1	$C_{20}H_{23}O_6$	359.14893
UN2	$C_{20}H_{25}O_7$	377.15939
UN3	$C_{20}H_{23}O_7$	375.14386
UN4	$C_{19}H_{32}O_{12}$	452.19260

Supplementary Table 7. Gene sequences used in this study.

Gene	Coding sequence
CLAMT1a	ATGGCGTCCTCGCTGCTCCACTGCTCCGACAAGCTCCCGTTCATGGACGTGGAGACAATCCTCCACATGAAAGAGG GGCTTGGCGAGACCAGCTACGCGCAGAACTCCTCTCTTCAGAAGCGGGGCATGGACACGCTGAAGAGCCTCATCAC CAACTCGGCGACGGACGTGTACATCTCGCAGATGCCGGAGAGGTTCACGGTGGCCGACCTGGGCTGCTCGTCGGGC CCGAACGCACTGTGCCTCGTCGAGGACATCGTCGGGAGCATCGGCCGGGTGTGCGGCCGGTCGTCGCAGCCGCCGC CCGAGTTCTCGGTGCTCCTCAACGACCTCCCGACCAACGACTTCAACACCATCTTCTTCAGCCTGCCGGAGTTCAC CGACCGGCTCAAGGCCGCCGCCGAGACCGACGAGTGGGGCCGGCCGATGGTGTTCCTGTCCGGCGTGCCGGGGTCT TTCTACGGGAGGCTCTTCCCCAGGAAGAGCGTGCACTTCATCTGCTCCTGCTCCAGCCTGCACTGGCTCTCCCAGG TCCCGCCGGGGCTCTTCGACGAGGCCACGGGCACGCCCATCAACAAGGGGAAGATGTACATCTCGAGCTCCAGCCC GCTCGCCGTGCCGACGGCCTACCTGAGGCAGTTCCAGAGGGACTTCGGCCTGTTCCTCAGATCGCGCGCCGCCGAG GTCGTCGCCGGCGGCCGGATGGTACTGGCCATGCTCGGCAGGCAGACCGAGGGCTACATCGACAGGCGAACCACCT TCCTCTGGGAGCTCCTCTCCGAGTCGTTCGCCTCGCTCGTGGCACAGGGGTTGGTGGCCCAGGAGAAGGTGGACGC GTACAACGTGCCGTTCTACGCGCCGTCGATCCAGGAGGTGGAGGAGGAGGTGCGGCGAGAGGGGTCGTTCCGGCTC GACTACGTGCAGACGTACGAGATCAACCTGAGCAGCAGCGGTGACGCCAAGGAGGACGGCCGGACGGTGTCCATGG CGATCAGGGCCATCCAGGAGTCCATGCTGAACCACCACTTCGGCCCAGACATTGTCGACGCGCTCTTTCACAGGTA CACGGAACTCGTCACCGAGTCCATGGAGAGGGAGGAGGTCGAAAGCGTTCAGATTGGGGTCGTCCTCACAAGGTTA TGA
CLAMT1b	ATGGCTTCTTCCTCCAATTGCTCCAACAAACACCCTCCCATGATGAATGTGGAGGCCGTCCTCCACATGAAGGCAG GCCTCGGTGAGAACAGTTACGCGATGAACTCTGTGATTCAGAAAAAAGGCATGGACACCGTGAGGAGCCTTGTCAC CGATTCAGCGACACAACTGTACTTGTCACTCAAACCAGAGAGGTTCACCATGGCCGACCTTGGGTGCTCCTCGGGG ACGAACGCGCTGGGCATGGTGGAGGACGTCGTCAGGAGCGTCGGCAAGGCCTGCTGTGGCTGCGGGTCACCGCCGC CAGAGTTCTCGGTGCTCCTGAACGATCTCCCTACCAACGACTTCAACACCGTCTTCTCCCGGGTACCAGAGTTCAC AGCCAAGCTGAAAGCCGACGCCCCGTCGCCGATGATCTTCGTCTCTGGCGTCCCGGGATCCTTCTACGGAAGGCTG TTCCCTAGCAGGAGCGTGCACTTGGCTTGCTCCTTCGCTAGCCTGCACTGGCTCTCTCAGGTTCCTCGACTTTCTG ATGAGACGAACACGCCCATGAACAAGGGGAAGATGTTCATATCAAGCACGAGCCCTCCTGCCGTGCCGGCGGCCTA CCTGAGGCAGTTCCAGAGAGACTTCACCGTTTTCCTCGAGTCACGAGCCGCGGAGGTTGTCCCCGGTGGCCGCATA GTTGTGTCGATGCTAGGGCGTCAAACCGAGGGGTACACGGACATGAAGACCACCCTCTTGTGGGATCTTCTTTCCG AGTCGCTTGCCGCGATGGTTTCACAGGGCTTGGCCGAGCAGGAGAAGGTCGATGCCTACGACGTTCCCTTCTACGC GCCAAGCTTACGGGAGATCAAGGAAGTGGTGAGCAAGGAAGGTTCCTTCAGCCTCAACTGCGTGAGGACGTACGAG GCAACCCATGGCGGGAGCGACGCCAAAAGAGACGGCAAGATGCTAGCTATGACGGCCAGGGCCGTCCATGAGTCGA TGATGAGCCACCACTTCGGGCCAGGCATCGTGGAGCCGCTGTTCCACAAGTACGGCGAGTTGGTTGCCCAGTTCAT GGAGACGGGAGAGGTCAAGAGTGTCCAGATAGGGGTGGTCCTGACAAGGTTGTTGTGA
CLAMT1c -iso1	ATGGCGTCCTCGCTGCTCCACTGCTCCGACAAGCTCCCGTTCATGGACGTGGAGACCATCCTCCACATGAAAGAGG GGCTCGGCGAGACCAGCTACGCGCAGAACTCCTCGCTTCAGAAGCGGGGCATGGACACGCTGAAGAGCCTTATCAC CAACTCGGCGACGGACGTGTACATCGCGCAGATGCCGGAGAGGTTCACGGTGGCCGACCTGGGCTGCTCGTCGGGC CCGAACGCGCTGTGCCTCGTCGAGGACATCGTCGGGAGCATCGGCCGGGTGTGCGGCCGGTCGTCGCAGCCGCCGC CCGAGTTCTCGGTGCTCCTCAACGACCTCCCGACCAACGACTTCAATACCATCTTTTTCAGCCTGCCGGAGTTCAC CGACCGGCTCAAGGCCGCCGCCGAGACCGATGAGTGGGGCCGGCCGATGGTGTTCCTGTCCGGCGTGCCGGGGTCC TTCTACGGGAGGCTCTTCCCAAGGAAGAGCGTGCACTTCATCTGCTCCTACTCCAGCTTGCACTGGCTCTCCCAGG TCCCGCCGGGGCTCTTCGACGAGGCCACGGGCACGCCCATCAACAAGGGGAAGATGTACATCTCGAGCTCCAGCCC GCTCGCCGTGCCGACAGCCTACCTGAGGCAGTTCCAGAGGGACTTCGGCCTGTTCCTCAGATCACGCGCCGCCGAG GTCGTCGCCGGCGGCCGGATGGTCCTCGCCATGCTCGGCAGGCAGACCGAGGGCTACATCGACAGGCGAACCACCT TCCTCTGGGAGCTCCTCTCCGAGTCGTTCGCCTCGCTCGTGGCACAGAACGCCGAAGAACCTGAACTTGATTACTG A
CLAMT1c -iso2	ATGGCGTCCTCGCTGCTCCACTGCTCCGACAAGCTCCCGTTCATGGACGTGGAGACCATCCTCCACATGAAAGAGG GGCTCGGCGAGACCAGCTACGCGCAGAACTCCTCGCTTCAGAAGCGGGGCATGGACACGCTGAAGAGCCTTATCAC CAACTCGGCGACGGACGTGTACATCGCGCAGATGCCGGAGAGGTTCACGGTGGCCGACCTGGGCTGCTCGTCGGGC CCGAACGCGCTGTGCCTCGTCGAGGACATCGTCGGGAGCATCGGCCGGGTGTGCGGCCGGTCGTCGCAGCCGCCGC CCGAGTTCTCGGTGCTCCTCAACGACCTCCCGACCAACGACTTCAATACCATCTTTTTCAGCCTGCCGGAGTTCAC CGACCGGCTCAAGGCCGCCGCCGAGACCGATGAGTGGGGCCGGCCGATGGTGTTCCTGTCCGGCGTGCCGGGGTCC TTCTACGGGAGGCTCTTCCCAAGGAAGAGCGTGCACTTCATCTGCTCCTACTCCAGCTTGCACTGGCTCTCCCAGG TCCCGCCGGGGCTCTTCGACGAGGCCACGGGCACGCCCATCAACAAGGGGAAGATGTACATCTCGAGCTCCAGCCC GCTCGCCGTGCCGACAGCCTACCTGAGGCAGTTCCAGAGGGACTTCGGCCTGTTCCTCAGATCACGCGCCGCCGAG GTCGTCGCCGGCGGCCGGATGGTCCTCGCCATGCTCGGCAGGCAGACCGAGGGCTACATCGACAGGCGAACCACCT TCCTCTGGGAGCTCCTCTCCGAGTCGTTCGCCTCGCTCGTGGCACAGGTGGGTGATCGAGCTTGCCATTCGAGAAG GTGGACGCGTACAACGTGCCGTTTCTACGCGCCGTCGACCCAGCAGGTGGAGGAGGAGGTGCGGCGAGAGGGGTCG TTCCGGCTCGACTACGTGCAGACGTACGAGATCAACCTGAGCAGCAGCGGTGACGCCAAGGAGGACGGCCGGACGG TGTCCATGGCGATCAGGGCCATCCAGGAGTCCATGCTCAGCCACCACTTCGGCCCGGACATCGTCGATGCGCTCTT CCACAGTAACGAGTACTGGTTCTGGATGTAA

Supplementary Table 8. Primers used in this study.

Purpose	Primer name	Sequence (5'-3')	Reference
Cloning	CLAMT1aF	ATGGCGTCCTCGCTGCTCCAC	
	CLAMT1aR	TCATAACCTTGTGAGGACGAC	
	CLAMT1a-bamH1F	ATTGGATCCATGGCGTCCTCGCTGCTCCAC	
	CLAMT1a-not1R	ATTGCGGCCGCTCATAACCTTGTGAGGACGAC	
	CLAMT1bF	ATGGCTTCTTCCTCCAATTGCTCC	
	CLAMT1bR	TCACAACAACCTTGTCAGGACCAC	
	CLAMT1b-Xba1F	AATTCTAGAATGGCTTCTTCCTCCAATTGCTCC	
	CLAMT1b-not1R	AATGCGGCCGCTCACAACAACCTTGTCAGGACCAC	
	CLAMT1cF	ATGGCGTCCTCGCTGCTCCAC	
	CLAMT1ciso1R	CACAGGAAACAGCTATGACC	
	CLAMT1ciso2R	TTACATCCAGAACCAGTACTCG	
	CLAMT1c-bamH1F	ATTGGATCCATGGCGTCCTCGCTGCTC	
	CLAMT1ciso1-Not1R	AATGCGGCCGCCACAGGAAACAGCTATGACC	
	CLAMT1ciso2-Not1R	AATGCGGCCGCTTACATCCAGAACCAGTACTCG	
	AtCLAMTF	ATGGATAAGAAGGATATGGAG	
	AtCLAMTR	TCAGAGCTTCTTTCTTAGGAC	
	AtCLAMT-BamH1F	ATTGGATCCATGGATAAGAAGGATATGGAG	
	AtCLAMT-Not1R	ATTGCGGCCGCTCAGAGCTTCTTTCTTAGGAC	
	AtMAX1-BamH1F	AATGGATCCATGAAGACGCAACATCAATGGT	
	AtMAX1-Not1R	AATGCGGCCGCTCAGAATCTTTTGATGGTTCTGAGCT	
Genotyping	CLAMT1bF	CACAGCCAAGCTGAAAGCCG	
	CLAMT1bR	CGTGCTTGATATGAACATCTTCCC	
	CLAMT1cF	CGACTTCAATACCATCTTTTTCAGC	
	CLAMT1cR	CCAGTGCAAGCTGGAGTAGG	
	UBC-E2F	ACCGCCTGACAATCCCTATG	[14]
	UBC-E2R	GGGATAGTCTGGCGGAAAATG	[14]
RT-PCR	*EcPT4 F*	ACGCCTACGACCTATTCTGCATCACC	[15]
	EcPT4 R	GCCGAACACGAG CTGGCCTATCA	[15]
	RiEFαf	GCTATTTTGATCATTGCCGCC	[16]
	RiEFαr	TCATTAAAACGTTCTTCCGACC	[16]
	RiPEIP1	AAGAAAGTAAACGTGTGGCT	[16]
	RiPEIP1	TAACACTCATCTCGGGACTG	[16]
	SiPHT1;9 F	CAAGGAGATAAACGCCCTGAC	[17]
	SiPHT1;9 R	ACCGATGAGCTGGATCAGGTA	[17]
	TUA F	GAGCGTCTGTCTGTTGACTATG	[18]
	TUA R	GTGGACAGGACACTGTTGTATG	[18]

Supplementary references

1. Varshney, R. K., Shi, C., Thudi, M., Mariac, C., et al. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nature Biotechnology*, **35**, 969-976 (2017).

2. Yan, H., Sun, M., Zhang, Z. et al. Pangenomic analysis identifies structural variation associated with heat tolerance in pearl millet. *Nature Genetics*, **55**, 507-518 (2023).

3. Salson, M. et al. An improved assembly of the pearl millet reference genome using Oxford Nanopore long reads and optical mapping. *G3 Genes|Genomes|Genetics* **13**, jkad051 (2023).

4. Ramu, P., Srivastava, R.K., Sanyal, A., et al. Improved pearl millet genomes representing the global heterotic pool offer a framework for molecular breeding applications. *Communications Biology*, **6**, 1-11 (2023).

5. Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol Biol*. **1962**, 227-245 (2019).

6. Swain MT, Tsai IJ, Assefa SA, Newbold C, Berriman M, Otto TD. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. *Nat Protoc.* **7**, 1260-84 (2012).

7. Tange, O. GNU Parallel 20230822 ('Chandrayaan'). Zenodo https://doi.org/10.5281/zenodo.8278274 (2023).

8. Mallu TS, Irafasha G, Mutinda S, Owuor E, Githiri SM, Odeny DA, Runo S. Mechanisms of pre-attachment Striga resistance in sorghum through genome-wide association studies. *Mol Genet Genomics*. **297**, 751-762 (2022).

9. Kavuluko J, Kibe M, Sugut I, Kibet W, Masanga J, Mutinda S, Wamalwa M, Magomere T, Odeny D, Runo S. GWAS provides biological insights into mechanisms of the parasitic plant (Striga) resistance in sorghum. *BMC Plant Biol.* **21**, 392 (2021).

10. Badu-Apraku, B., Adewale, S., Paterne, A. A., Gedil, M., Toyinbo, J., & Asiedu, R. Identification of QTLs for grain yield and other traits in tropical maize under Striga infestation. *PLoS One*, **15**, e0239205 (2020).

11. Hu L, Wang J, Yang C, Islam F, Bouwmeester HJ, Muños S, Zhou W. The effect of virulence and resistance mechanisms on the interactions between parasitic plants and their hosts. *Int J Mol Sci.* **21**, 9013 (2020).

12. Li C, Dong L, Durairaj J, Guan JC, Yoshimura M, Quinodoz P, Horber R, Gaus K, Li J, Setotaw YB, Qi J, De Groote H, Wang Y, Thiombiano B, Floková K, Walmsley A, Charnikhova TV, Chojnacka A, Correia de Lemos S, Ding Y, Skibbe D, Hermann K, Screpanti C, De Mesmaeker A, Schmelz EA, Menkir A, Medema M, Van Dijk ADJ, Wu J, Koch KE, Bouwmeester HJ. Maize resistance to witchweed through changes in strigolactone biosynthesis. *Science* **379**, 94-99 (2023).

13. Haider, I., Yunmeng, Z., White, F., Li, C., Incitti, R., Alam, I., Gojobori, T., Ruyter-Spira, C., Al-Babili, S., & Bouwmeester, H. J. Transcriptome analysis of the phosphate starvation response sheds light on strigolactone biosynthesis in rice. *Plant Journal*, **114**, 355-370 (2023).

14. Shivhare R, Lata C. Selection of suitable reference genes for assessing gene expression in pearl millet under different abiotic stresses and their combinations. *Sci Rep.* **6**, 23036 (2016).

15. Pudake RN, Mehta CM, Mohanta TK, Sharma S, Varma A, Sharma AK. Expression of four phosphate transporter genes from Finger millet (*Eleusine coracana* L.) in response to mycorrhizal colonization and Pi stress. 3 *Biotech*. **7**, 17 (2017).

16. Fiorilli, V., Belmondo, S., Khouja, H.R. et al. RiPEIP1, a gene from the arbuscular mycorrhizal fungus Rhizophagus irregularis, is preferentially expressed in planta and may be involved in root colonization. *Mycorrhiza* **26**, 609–621 (2016).

17. Ceasar, S. A., Hodge, A., Baker, A., & Baldwin, S. A. Phosphate concentration and arbuscular mycorrhizal colonisation influence the growth, yield and expression of twelve PHT1 family phosphate transporters in foxtail millet (*Setaria italica*). *PLoS One*, **9**, e108459 (2014).

18. Saha P, Blumwald E. Assessing reference genes for accurate transcript normalization using quantitative real-time PCR in pearl millet [*Pennisetum glaucum* (L.) R. Br]. *PLoS One*. **9**, e106308 (2014).