Improving cassava bacterial blight resistance by editing the epigenome

Kira M. Veley [1], Kiona Elliott[1,2], Greg Jensen[1], Zhenhui Zhong [3], Suhua Feng[3,4], Marisa Yoder[1], Kerrigan B. Gilbert [1], Jeffrey C. Berry [1], Zuh-Jyh Daniel Lin[1], Basudev Ghoshal[3,7], Javier Gallego-Bartolomé[3,8], Joanna Norton[4], Sharon Motomura-Wages[4], James C. Carrington [1], Steven E. Jacobsen [3,5,6] ✉ & Rebecca S. Bart [1] ✉

Pathogens rely on expression of host susceptibility (*S*) genes to promote infection and disease. As DNA methylation is an epigenetic modification that affects gene expression, blocking access to *S* genes through targeted methylation could increase disease resistance. *Xanthomonas phaseoli* pv. *manihotis*, the causal agent of cassava bacterial blight (CBB), uses transcription activator-like20 (TAL20) to induce expression of the *S* gene *MeSWEET10a*. In this work, we direct methylation to the TAL20 effector binding element within the *MeSWEET10a* promoter using a synthetic zinc-finger DNA binding domain fused to a component of the RNA-directed DNA methylation pathway. We demonstrate that this methylation prevents TAL20 binding, blocks transcriptional activation of *MeSWEET10a* in vivo and that these plants display decreased CBB symptoms while maintaining normal growth and development. This work therefore presents an epigenome editing approach useful for crop improvement.

Susceptibility (*S*) genes present a liability within a host genome, as their exploitation is a common mechanism used by diverse classes of pathogens[1]. However, mutating these loci may have negative consequences as these genes are often necessary for normal plant growth and reproductive development[2–5]. In many cases, pathogens actively upregulate expression of *S* genes during infection. Plant pathogenic bacteria within the *Xanthomonas* and *Ralstonia* genera express transcription activator-like (TAL) effectors that activate transcription of *S* genes by recognizing and binding specific, largely predictable, promoter sequences (effector binding elements, EBEs) within the host genome[4,6–9]. Previous research demonstrated that gene editing can be used to prevent the exploitation of *S* genes by pathogens and that this results in increased resistance[10–14]. Epigenomic modifications can also

influence gene expression. Because TAL binding can be inhibited by DNA methylation[15,16], we hypothesized that targeting methylation to a TAL effector binding site would block induction of the *S* gene and result in increased resistance as measured by decreased disease symptoms.

In eukaryotes, cytosine DNA methylation is an important mechanism of epigenetic gene regulation[17]. In plants, the RNA-directed DNA methylation (RdDM) pathway is responsible for establishing de novo 5-methylcytosine methylation. Previous targeted methylation efforts leveraged naturally occurring epialleles from Arabidopsis that affect flowering time[18,19]. These studies used CRISPR-Cas9-based tools combined with components of the methylation pathways, either the catalytic domain of the plant methyltransferase DOMAINS

[1]Donald Danforth Plant Science Center, Saint Louis, MO 63132, USA. [2]Division of Biological and Biomedical Sciences, Washington University in Saint Louis, Saint Louis, MO 63110, USA. [3]Department of Molecular, Cell and Developmental Biology, University of California at Los Angeles, Los Angeles, CA 90095, USA. [4]College of Tropical Agriculture & Human Resources, University of Hawaii at Manoa, Hilo, HI 96720, USA. [5]Eli & Edythe Broad Center of Regenerative Medicine & Stem Cell Research, University of California at Los Angeles, Los Angeles, CA 90095, USA. [6]Howard Hughes Medical Institute, University of California at Los Angeles, Los Angeles, CA 90095, USA. [7]Present address: Summerland Research and Development Centre, Agriculture and Agri-Food Canada, Summerland, BC V0H 1Z0, Canada. [8]Present address: Instituto de Biología Molecular y Celular de Plantas (IBMCP), CSIC-Universidad Politécnica de Valencia, 46011 Valencia, Spain. ✉e-mail: jacobsen@ucla.edu; rbart@danforthcenter.org

REARRANGED METHYLTRANSFERASE 2 (DRM2) involved in RdDM or the bacterial CG-methyltransferase MQ1[20,21]. In addition, a number of RdDM proteins are effective in targeting methylation when fused with artificial zinc-fingers (ZFs), including DRM2 and DEFECTIVE IN MERISTEM SILENCING 3 (DMS3), a protein that in Arabidopsis targets RNA polymerase V binding to chromatin to trigger RdDM[22,23]. While CRISPR-based tools are highly specific, ZF-based targeting systems are preferable for some applications, as they generally require smaller transgenes and are not constrained by the location of PAM sites.

Cassava bacterial blight (CBB) is caused by *Xanthomonas phaseoli* pv. *manihotis*[24] (*Xam*, aka. *Xanthomonas axonopodis* pv. *manihotis*) and occurs in all regions where cassava (*Manihot esculenta* Crantz) is grown[25,26]. Recently, a resistance gene was described that may provide some tolerance to certain strains of Xam[27], however, currently, there are no strong broad spectrum resistance genes available to breeders. Symptoms of an infection begin with dark green, angular, water-soaked lesions on leaves (Fig. 1A) after which *Xam* moves through the vasculature causing leaf wilting and eventual death of the plant. *Xam* secretes 20–30 effectors, including multiple TAL effectors, into cassava cells[28]. We previously demonstrated that one of these TAL effectors, TAL20, is a major virulence determinant for this pathogen. TAL20 induces ectopic *S* gene expression of a SWEET (Sugars Will Eventually be Exported Transporters) family sugar transporter, *MeSWEET10a*[6,29]. Expression of *MeSWEET10a* during infection is required for water-soaking phenotypes and full virulence[6]. In wild-type cassava, the promoter of *MeSWEET10a* is not methylated[30]. Thus, we hypothesized that targeted methylation at the TAL20 EBE would prevent ectopic expression of *MeSWEET10a* and provide resistance to CBB (Fig. 1B).

Here, we present epigenetic editing as a useful approach toward crop improvement. We show directed DNA methylation to the TAL20 EBE within the *MeSWEET10a* promoter using a ZF fused to DMS3. We demonstrate that cytosine methylation inhibits TAL20 binding to the EBE, preventing transcriptional activation of *MeSWEET10a* and decreasing bacterial blight symptoms in cassava plants.

Results

Establishing methylation at the *MeSWEET10a* promoter

We first tested the effect of methylation on TAL20 binding to the EBE in vitro using an electrophoretic mobility shift assay (EMSA) (Supplementary Fig. 1). As shown previously[6], purified 6xHis-TAL20$_{Xam668}$

binds the unmethylated EBE oligo (lane 2) and, as expected, providing an excess of unlabeled, unmethylated oligo outcompeted for binding (lane 5). Conversely, 6xHis-TAL20$_{Xam668}$ does not efficiently bind the methylated oligo (lane 3), nor did the methylated oligo compete for TAL20 binding, even when supplied in excess (lanes 6 and 7). Therefore, the EBE of *MeSWEET10a* was deemed a good candidate for engineering epigenetic disease resistance. Of the RdDM factors tested in *Arabidopsis thaliana*, DMS3 was shown to be the most effective at delivering methylation to a given locus[23]. Transgenic cassava plants were generated expressing an artificial ZF targeting the EBE of *MeSWEET10a* fused to Arabidopsis DMS3 (Fig. 2A, Supplementary Figs. 2–7). The targeted region of the *MeSWEET10a* promoter was assessed for methylation using amplicon-based bisulfite sequencing
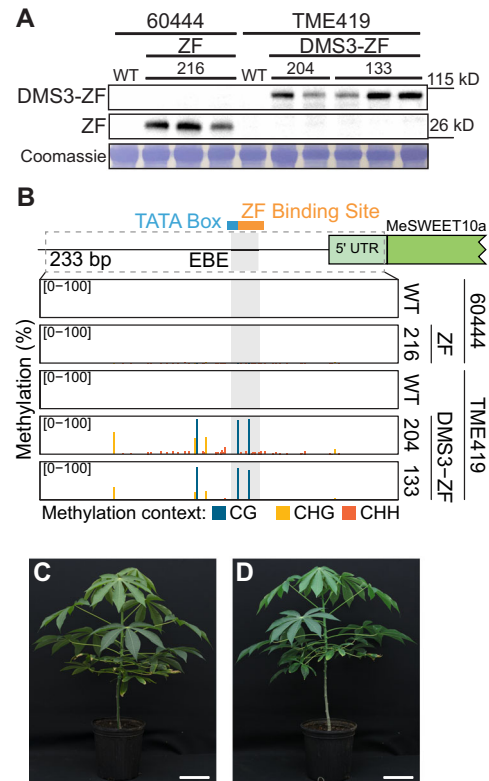
Fig. 2 | DMS3-ZF expression results in CG methylation at the EBE in vivo.
A Expression of transgenes in individual plants from two independent DMS3-expressing transgenic lines (133 and 204) as well as a ZF-only negative control line (216). Cassava variety information for each sample is given above the lanes. First two rows: representative western blots (anti-FLAG) showing expression of the ZF (ZF-3xFLAG) protein with (top) and without (middle) DMS3. Relevant size standards are shown to the right (kD). Bottom: Coomassie Brilliant Blue stained Rubisco, loading control. Uncropped images of the western and Coomassie are shown in Supplementary Fig. 5A. **B** Representative PCR-based bisulfite sequencing (ampBS-seq) results from samples shown in (**A**). Details results for these samples are shown in Supplementary Fig. 5A. Top: Graphical depiction of *MeSWEET10a* promoter region assessed for methylation. The EBE (gray), a presumed TATA box (blue) and the ZF binding site (orange) are indicated. The predicted 5′ UTR and *MeSWEET10a* transcriptional start site are shown in green. The area within the dotted lined box (233 bp) was subjected to ampBS-seq. Bottom: CG, CHG, and CHH DNA methylation levels (percent, y-axis) of the *MeSWEET10a* promoter (EBE, gray) measured by ampBS-seq with and without DMS3-ZF. Methylation was called for cytosines with >500 reads mapped. The results in (**A** and **B**) are representative examples of the six independent experiments presented in Supplementary Figs. 2–7. **C** Representative wild-type (TME419) and (**D**) DMS3-ZF-expressing (line #133) plants. Scale bar = 14 cm. Additional data quantifying growth in the field are shown in Supplementary Fig. 8. Source data are provided as a Source Data file.
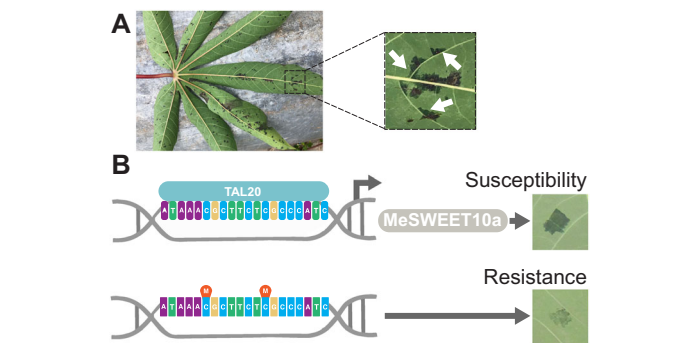
Fig. 1 | Overview of strategy to improve disease resistance using epigenetic modification. A Left: Example of CBB-infected cassava in a field (Uganda, 2019). Right: selected region from leaf on the left. Water-soaked regions are indicated with white arrows. **B** Graphical representation of epigenetic CBB disease resistance strategy. Top: In WT plants, TAL20 from *Xam* binds a specific sequence (EBE) present in the promoter of the *S* gene *MeSWEET10a*. Upon binding, TAL20 induces the ectopic expression of MeSWEET10a, a sugar transporter, which is required for establishment of disease. A typical 'water-soaked' lesion is shown to the right, an early indicator of CBB. Bottom: DNA methylation prevents TAL20 from binding the EBE. *MeSWEET10a* expression is not induced, and disease symptoms are reduced. Source data are provided as a Source Data file.
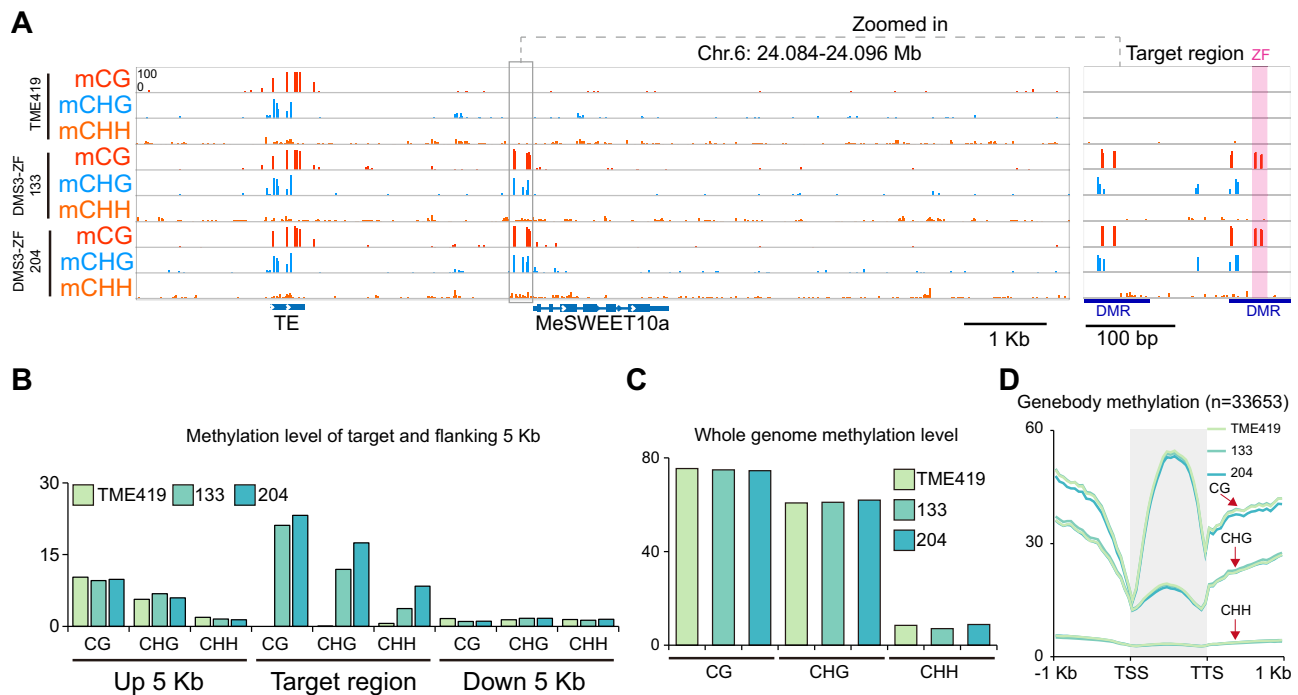
Fig. 3 | Targeted methylation at MeSWEET10a shows high specificity.
A Distribution of CG, CHG, and CHH methylation over the targeted and flanking 5 kb regions of the *MeSWEET10a* promoter in wild-type TME419 and two independent DMS3-ZF lines (transgenic lines 133 and 204). Gray box indicates the targeted region within the *MeSWEET10a* promoter. ZF binding site is shaded within the pink box. Two detected hypermethylated differentially methylated regions (DMR) are also indicated. **B** Summary of CG, CHG, and CHH methylation level of targeted and flanking 5 kb regions of the *MeSWEET10a* promoter in wild-type TME419 and the two independent DMS3-ZF lines. **C** Whole genome CG, CHG, and CHH methylation level of wild-type TME419 and the two representative DMS3-ZF lines. **D** Metaplot of CG, CHG, and CHH methylation over the gene of protein coding genes (*n* = 33,653) and flanking 1 kb regions in wild-type TME419 and the two independent DMS3-ZF lines. Transcriptional start sites (TSS) and transcriptional termination sites (TTS) are indicated. Source data are provided as a Source Data file.

(ampBS-Seq) (Fig. 2B, Supplementary Figs. 2–7). Plants expressing the ZF alone showed no methylation at or around the EBE of *MeSWEET10a*. In contrast, all plants expressing DMS3-ZF contained high levels of ectopic CG methylation within and near the ZF target, including the only two CG sites within the EBE (Fig. 2B, Supplementary Figs. 2–7). The specific analysis that the plants were used for in each Supplementary Fig. 2–7 are indicated in the title of the figure. Plant phenotypes were evaluated under greenhouse and field conditions using our standard methods (see methods for details). These analyses indicated that methylation did not grossly affect root and shoot growth and development (Fig. 2C, D, Supplementary Fig. 8). Whole Genome Bisulfite Sequencing (WGBS) was used to assess specificity of the DMS3-ZF. High levels of de novo DNA methylation were observed within the target region, similar to the ampBS-seq result (Fig. 2B, Fig. 3A), while DNA methylation within an extended region of 5 kb on either side was comparable between the DMS3-ZF-expressing lines and the ZF alone control (Fig. 3A, B). In addition, CG, CHG, and CHH methylation across all chromosomes, as well as within genes, was similar between the DMS3-ZF-expressing lines and ZF alone control (Fig. 3C, D, Supplementary Fig. 9, and Supplementary Table 1). We investigated potential off-targets by predicting ZF binding sites with up to 4 mismatches within the target sequence (5′- CGCTTCT CGCCCATCCAT – 3′). This yielded 1542 potential off-target regions genome-wide (Supplementary Fig. 10A). We compared methylation levels of these potential off-targets, plus 100 bp flanking sequences, in WT and the two DMS3-ZF-expressing lines 133 and 204. No significant differences were detected (Supplementary Fig. 10B). As a complementary approach to identify potential non-specific activity, differentially methylated regions (DMRs) (CG variation > 0.4, CHG variation > 0.2, and CHH variation > 0.1, respectively) were obtained by comparing the two DMS3-ZF lines with the WT control. This method
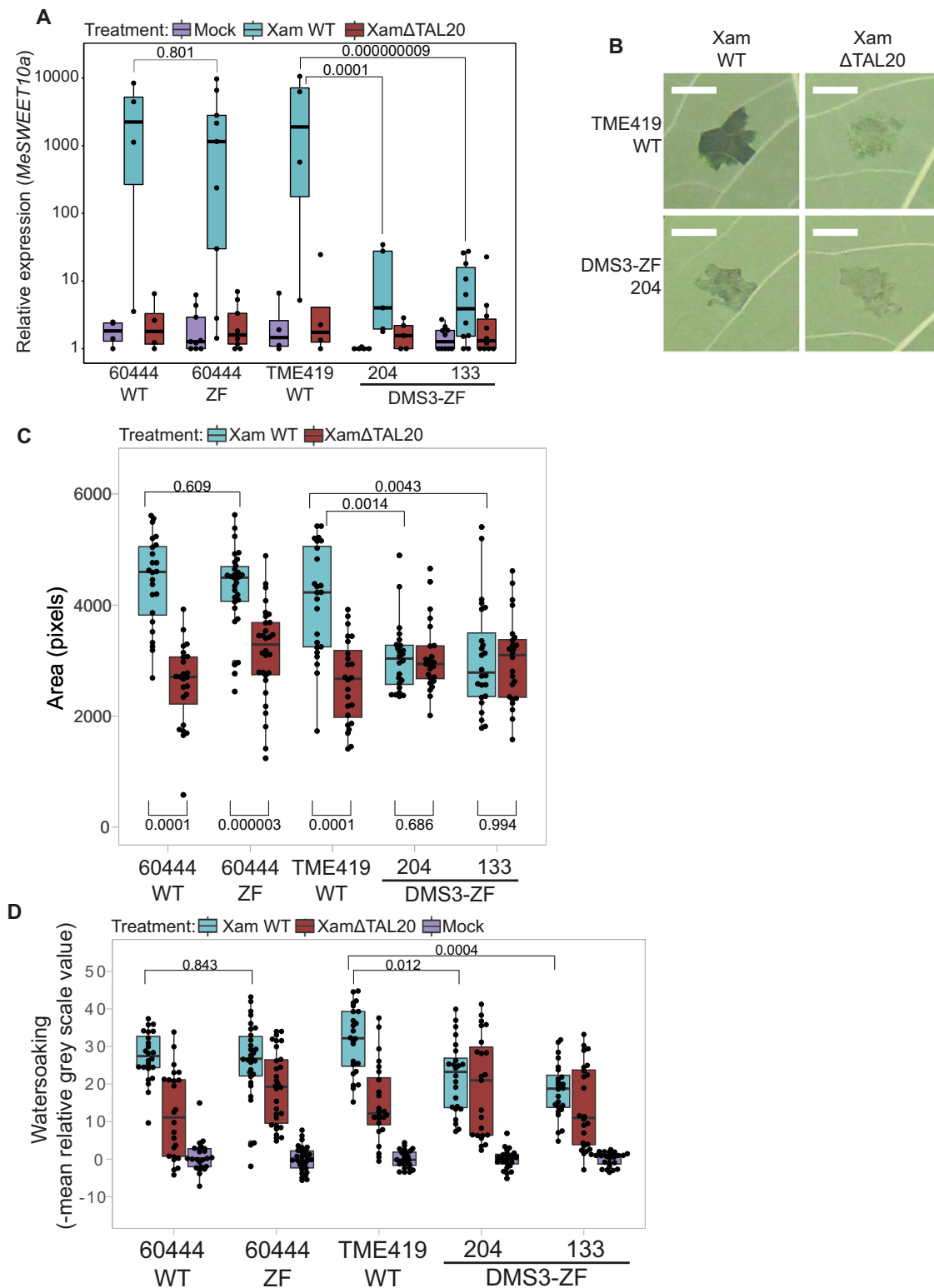
successfully detected two hypermethylated DMRs in the targeted MeSWEET10a region (Fig. 3A). However, most of the other detected DMRs were hypoDMRs (i. e., regions showing decreased methylation in the DMS3-ZF lines) rather than hyperDMRs (increased methylation in the DMS3-ZF lines). An overall decrease in genome methylation has been attributed to the plant tissue culture process in cassava[31] and in many other plant species including maize[32], rice[33] and oil palm[34]. Cross referencing the observed DMRs with the 1542 potential off-targets revealed that very few of DMRs (particularly hyperDMRs) fell into off-target regions, and the chance was indistinguishable from the random control regions (Supplementary Fig. 10C, Supplementary Table 2). Together, these results demonstrate a reasonable level of specificity for targeted DMS3-ZF-mediated methylation in cassava.

Cytosine methylation prevents *MeSWEET10a* transcriptional activation by TAL20 in vivo
We next tested the hypothesis that CG methylation at the TAL20 EBE would prevent activation of *MeSWEET10a* in vivo. Compared to controls, plants from independent DMS3-ZF-expressing lines showed a significant lack of *MeSWEET10a* induction in response to *Xam* infection (Fig. 4A). The suppression of *MeSWEET10a* induction occurred in independently derived transgenic lines and across independent experiments. This suppression did not correlate with increased expression of DMS3-ZF (Supplementary Fig. 11). These data indicate that methylation prevents TAL20-mediated activation of *MeSWEET10* expression.

Epigenomic editing at *MeSWEET10a* decreases cassava bacterial blight symptoms
To quantify the effect of methylation at the *MeSWEET10a* EBE on CBB disease in vivo, we measured water-soaked lesions (Fig. 1A, Fig. 4B,

In addition to a decrease in lesion size, the intensity of water-soaking at and around the infiltration site was also reduced in EBE methylated plants. Gray-scale values from images of lesions were compared to mock-inoculated sites on the same leaf, where more water-soaking equates to a darker appearance. Water-soaking intensity was reduced in plants with methylation at the EBE (Fig. 4D). These results demonstrate that methylation of the EBE site within the promoter of the *S* gene *MeSWEET10a* results in a reduction in CBB-related symptoms.

Supplementary Fig. 12), which are the most obvious early disease symptom and requires induction of *MeSWEET10a*[6,35]. Quantitative analysis of images from *Xam*-infiltrated leaves revealed that plants with methylation at the EBE had a reduced infected area compared to wild-type (Fig. 4C). Further, the area of infection in DMS3-ZF leaves infiltrated with wild-type *Xam* was comparable to leaves infected with *Xam* lacking TAL20 (XamΔTAL20). No significant difference in bacterial population was observed in the DMS3-ZF lines (Supplementary Fig. 13).

Fig. 4 | Effect of methylation on CBB disease phenotypes in cassava.
A Normalized *MeSWEET10a* expression (*y*-axis, log10 scale) in WT and transgenic lines as determined by RT-qPCR. The cassava genes GTPb (Manes.09G086600) and PP2A4 (Manes.09G039900) were used as internal controls. Boxes are colored according to Xanthomonas treatment. Biological replicates (black dots) included in each background (*x*-axis) are as follows: $n = 4, 9, 4, 5, 10$ examined over four independent experiments. The *n* included in each treatment group for each background are consistent. Horizontal black line within boxes indicates the value of the median while the box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range ($1.5 \times$ IQR) from the 25th and 75th percentiles. Two-sided Welch's *t*-test *p* values are noted above brackets within plot. **B** Representative images of water-soaking phenotype of leaves from TME419 WT and DMS3-ZF-expressing plants. Images were taken 4 days post-infection with either *Xam668* (XamWT) or a *Xam668* TAL20 deletion mutant (Xam∆TAL20). Scale bar = 0.5 cm. **C** Observed area (pixels, *y*-axis) of water-soaking from images of Xam-infiltrated leaves (backgrounds, *x*-axis) 4 days post-infiltration. Images from (**B**) and Supplementary Fig. 11 are included in dataset. Calculated *p* values (Kolmogorov–Smirnov test) are shown above brackets within plot. **D** Intensity of water-soaking phenotype (*y*-axis) of region measured in (**C**). The negative mean gray-scale value for the water-soaked region relative to the average of the mock-treated samples within the same leaf is reported (see methods for details). Calculated *p* values (two-sided Kolmogorov–Smirnov test) are shown above brackets within plot. Both box plots: Biological replicates (black dots) included in each background (*x*-axis) are as follows: $n = 24, 30, 24, 24, 24$ examined over three independent experiments. Horizontal black line within boxes indicates the value of the median while the box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range ($1.5 \times$ IQR) from the 25th and 75th percentiles. Source data are provided as a Source Data file.

Discussion

In summary, we show that DMS3 paired with an artificial ZF caused targeted, de novo establishment of methylation at the EBE of *MeSWEET10a*, blocking TAL20-induced ectopic expression of *MeSWEET10a* and resulted in decreased CBB disease symptoms. Cassava is typically clonally propagated, and we have monitored methylation over time throughout the course of this study (3–4 years maintenance in tissue culture and in the greenhouse). We conclude that the methylation is stable in the presence of the transgene through clonal propagation. Because the components of the technology presented here do not derive from cassava (Arabidopsis DMS3 paired with a synthetic zinc finger), it is reasonable to hypothesize that this approach may be similarly successful in other crop species. Gene editing represents an existing, powerful strategy that can achieve similar outcomes to the work presented here, and we expect this will continue to be the favored approach for many applications[5,10–14,36–40]. At the same time, as targeted methylation technologies continue to mature, there will likely be cases where it may be preferred. For example, DNA methylation may be applied as a gene expression dial, opposed to an on/off switch. Similarly, as DNA methylation does not change the DNA sequence, it may be possible to use DNA methylation changes to achieve expression patterns in a tissue- or environment-specific manner. This might be achieved in transgenic plants by driving the methylation reagents using a tissue- or environment-specific promoter. In addition, a key near-term future goal will be to study the heritability of DNA methylation in the absence of the targeting transgene[41,42]. It seems likely that the size of the region with de novo methylation, the type of methylation (CG, CHG, CHH), and the specific location in the genome may all affect heritability. These factors will each need to be investigated, and while there is value in pursuing this work in the crop of interest, we expect that much can be learned using model systems. Ultimately, the practicality of this strategy will be determined by the heritability and stability of epigenetic edits, and this is likely to be application- and target-specific. Nonetheless, the transgenic epigenetic approach shown here may be useful in a variety of other crop improvement programs.

Methods
Construct design and cloning
The genomic coding sequence of Arabidopsis *DMS3* (AT3G49250), including promoter, was cloned into the pEG302 backbone[23,43]. The ZF targeting the EBE of *MeSWEET10a* was designed to target the sequence 5′-CGCTTCTCGCCCATCCAT-3′[22]. CodeIT was used to generate the amino acid in silico (linker type normal). The resulting amino acid sequence was plant codon optimized, synthesized by IDT and cloned at the 3′ end of *DMS3* (AT3G49250) in pEG302. A hygromycin resistance cassette was cloned into an available PmeI site for selection in cassava.

Production of plants and growth conditions
The transgenic cassava line expressing the ZF with no methylating factor was generated for a different project, and thus a different background was used: 60444 (ZF∆DMS3) vs. TME419 (DMS3-ZF). The promoter region of interest is identical in both genetic backgrounds, as verified by Sanger sequencing (Supplementary Fig. 14). Transgenic cassava lines of cultivar TME419 and 60444 were generated and maintained in vitro as described previously[44] with some modifications, as follows: transgenic FEC cells were selected for resistance using hygromycin (20 mg/mL) on spread plates (also containing 125 mg/L cefotaxime). On stage 1, 2, and 3 plates, 20 mg/mL hygromycin was used as well. Growth chamber settings used for tissue culture: $28\,°C \pm 1\,°C$, $75\,\mu mol\,m^{-2}\,s^{-1}$ light; 16 h light/8 h dark. In vitro plantlets were propagated and established in soil on a misting bench with domes. Once established, plants were removed from the misting bench and acclimated to greenhouse conditions (28 °C; humidity >50%; 16 h light/8 h dark; 1000 W light fixtures supplementing natural light levels below 400 W/m²). 1000 W light fixtures supplementing natural light levels below 400 W/m²). For field-grown plants, individuals were first established in the greenhouse (natural light and temperature, 17–28 °C, humidity >50%) for 29 weeks then transitioned outside and planted 3 feet apart.

PCR bisulfite sequencing library preparation and analysis
PCR-based BS-seq (ampBS-seq) of the *MeSWEET10a* promoter was performed[21]. Leaf tissue and DNA extracted using the GenElute™ Plant Genomic DNA Miniprep Kit (Sigma) and sodium bisulfite treated using the Epitect Bisulfite Conversion kit (QIAGEN). The region of interest within the *MeSWEET10a* promoter was amplified using the primers in Supplementary Table 3, giving a 233 bp product. Libraries were generated from purified PCR products amplified from the bisulfite-treated DNA, using an Ovation Ultralow V2 kit (NuGEN) or a Kapa DNA HyperPrep Kit (Roche) in combination with IDT UD barcodes (Illumina). iSeq 100 (Illumina) was used to sequence the libraries. Raw sequencing reads were aligned to the region of interest using the epigenomic analysis plugin for CLC Genomics Workbench 20.0 (QIAGEN). Reads shorter than 125 bp or longer than 152 bp were discarded. Reads were mapped to the following sequence: ATATTGTTTTTTT TTTTTTTAAAAAAAATAATAAAAGAAATAAGGTTACTGTTACATTGAC ATATTTTATTCACTTTAATCATGCATGCAACTTGACTTCATTCCGTTC CCTGGATTCCTCCCCTATATAAACGCTTCTCGCCCATCCATCATTGCA CAACATAGCTAGAGTTTCCTCTTGAGAAAGAGAGTTTCCTCTGCATAA GGGAAAGAGAGTTTTTATTATAGTYGGA. For mapping, the following parameters were used: non-directional, no masking, match score = 1, mismatch cost = 2, affine gap cost, insertion open cost = 6, insertion extend cost = 1, deletion open cost = 6, deletion extend cost = 1, length fraction = 0.9, similarity fraction = 0.9. Methylation was called for cytosines with >500 reads mapped and the resulting methylation level

was plotted using the provided script (MeSWEET10a_07272021-for_publication.0920.R).

Whole genome bisulfite sequencing library preparation and analysis

Genomic DNA from samples were end-repaired and ligated with TruSeq LT DNA single adapters (Illumina) using Kapa DNA HyperPrep kit (Roche). Adapter-ligated DNA was converted with EpiTect Bisulfite Conversion Kit (Qiagen). Converted DNA was PCR-amplified by MyTaq polymerase (Bioline) for 12 cycles. Library quality and size were determined by D1000 ScreenTape (Agilent) and libraries were subsequently purified by AMPure XP beads (Beckman Coulter). Library concentrations were measured with Qubit dsDNA Broad-Range Assay kit (ThermoFisher). Libraries were sequenced on NovaSeq 6000 sequencer (Illumina). WGBS reads were mapped to the TME7 phase0 genome with BSMAP (v2.90) allowing one mismatch and one best hit (-v 1 -w 1)[45,46]. Duplicated reads were removed with SAMtools (v1.3.1)[47]. Reads with three or more consecutive methylated CHH sites were considered as unconverted reads and removed in the following analysis. Conversion rate was estimated by calculating methylation level of the chloroplast genome. DNA methylation level at each cytosine was calculated by the number of methylated C vs. total C and T account. Methylation levels of each chromosome were determined by calculating CG, CHG, and CHH level of 100 kb bins. Differentially Methylated Regions (DMRs) were called using BSMAP (v2.90) with $p < 0.01$, where the differences in CG, CHG, and CHH methylation were at least 0.4, 0.2, and 0.1, respectively. Methylation track files were visualized with Integrative Genomics Viewer (IGV, v3.0).

Off-target analysis

ZF off-target sites were predicted by Cas-OFFinder by searching cassava genome with target site (5′-CGCTTCTCGCCCATCCAT-3′) allowing 4 mismatches and no PAM site[48]. Methylation levels of off-target sites and 100 bp flanking sequences (50 bp upstream and 50 bp downstream) were compared between samples. Sequence consensus analysis of potential off-target sites was performed with WebLogo (v3.6.0)[49].

SDS-PAGE gel and western blotting

All individuals used for qPCR and disease assay experiments were checked for construct expression (Supplementary Figs. 3 (qPCR1), 4 (qPCR2), 5 (Water-soaking), 6 (qPCR3), 7 (Water-soaking, qPCR4), 8 (Water-soaking)). Tissue (100 mg) was collected from each individual plant to be used in an experiment in a 1.5 mL microcentrifuge tube and frozen in liquid nitrogen along with four 3 mm glass beads. A tissue lyser (TissueLyser II, QIAGEN) was used to homogenize the sample (30/s frequency, 2 min, performed twice with sample rotation between runs). 400 μL of 1× Laemmli Sample buffer (0.0625 M Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 100 mM DTT) was added to each sample, then samples were boiled 90–100 °C for 4 min, centrifuged for 2 min and then either run immediately on a gel or frozen for storage at −80 °C. For SDS-PAGE, 10 μL of each sample was run on a 12% TEO-Tricine Precast Gel (RunBlue™, AbCam, ab119203) then transferred to PVDF membrane (ThermoFisher Scientific). For western blotting, monoclonal ANTI-FLAG® M2-Peroxidase (HRP) antibody (A8592, Sigma) was used (1:2000 dilution in 5% milk in TBST). HRP activity was assessed using SuperSignal™ West Femto substrate (ThermoFisher Scientific). Chemiluminescence was detected using the ChemiDoc XRS + System equipped with Image Lab 6.1 software (Bio-Rad). The molecular weight of protein bands was estimated by comparison with the BenchMark™ Pre-stained Protein Ladder (Thermo Fisher Scientific). The abundance of the 3×FLAG-ZF proteins were quantified from the band intensities also using Image Lab Software. Coomassie staining was performed on membranes using Coomassie brilliant blue (0.025% (w/v) Coomassie brilliant blue R-250 (ThermoFisher Scientific) in 40% methanol/7%

acetic acid (v/v). Staining was performed for 3 min followed by destain (solution: 50% methanol/7% acetic acid (v/v)) for 30 min. Then membranes were rinsed with water and allowed to dry overnight before imaging.

Electrophoretic mobility shift assay

For protein purification, $TAL20_{Xam668}$ was cloned into pDEST17 (Thermo Fisher Scientific) using Gateway™ technology, resulting in the addition of a 6xHis-tag at the N-terminus of the protein upon expression. 6xHis-TAL20$_{Xam668}$ was expressed and purified from *E. coli* BL21 cells using HisPur™ Ni-NTA Spin Columns (Thermo Fisher Scientific). Protein size and purity was assessed by SDS-PAGE (4–12% gradient gel) followed by western blot (Supplementary Fig. 1A) using an anti-His antibody (His-Tag (27E8) Mouse mAb (HRP Conjugate), Catalog #9991, Cell Signaling Technology). For the EMSA, the sequence previously shown to be able to be bound by TAL20$_{Xam668}$[6], with and without methylation, was synthesized (Integrated DNA Technologies) and biotin-labeled (Pierce™ Biotin 3′ End DNA Labeling Kit, Thermo Fisher Scientific). The sequence used is as follows, with methylated cytosines indicated by underlining: sense 5′-CCCTATA TAAAC̲GCTTCTCGC̲CCATCCATCATT-3′, anti-sense: 3′-GGGATATAT TGC̲GAAGAGC̲GGGTAGGTAGTAA-5′. EMSA was performed with the Light Shift® Chemiluminescent EMSA Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. Biotin-labeled oligonucleotides were mixed with 0, 20, 200, 300, and 1000 fmol of purified TAL20$_{Xam668}$ effector protein as indicated in the chart in Supplementary Fig. 1B. All binding reactions were incubated at room temperature for 20 min. For competition assays, unlabeled oligonucleotides were pre-incubated with other reaction components for an additional 20 min. Gel electrophoresis was performed on a 6% native polyacrylamide gel.

Bacterial inoculations

Xanthomonas strains were grown on plates containing necessary antibiotics for 2–3 days at 30 °C. The *Xanthomonas* strains used and their antibiotic resistances are as follows: *Xam668* (rifampicin 50 μg/mL) and *Xam668ΔTAL20* (suicide vector knockout[6], tetracycline 5 μg/mL, rifampicin 50 μg/mL). Bacteria were scraped from the plates into 10 mM MgCl$_2$ to either OD$_{600}$ = 1.0 (RT-qPCR) or OD$_{600}$ = 0.01 (water-soaking) concentrations according to established lab protocols. Leaves (2–3 weeks after plants were transferred to a greenhouse) were inoculated with a 1.0 mL needleless syringe using one bacterial strain per lobe, three injection points on each of 2–3 leaves. After inoculation, plants were kept under fluorescent light (12 h day/night light cycle) at 27 °C for either 48 h (RT-qPCR and ampBS-seq) or 4 days (water-soaking).

RNA extraction and RT−qPCR analysis

For RNA, 48 h post-infiltration, samples were taken using an 8 mm cork borer at each infiltration site, frozen in liquid nitrogen, ground to a fine powder (3 mm glass beads added to tissue, ground using tissue lyser (TissueLyser II, QIAGEN), 30/s frequency, 2 min, performed twice with sample rotation between runs) and total RNA was extracted (Spectrum™ Plant Total RNA Kit, Sigma). Technical replicates were pooled, thus each RNA sample contained three injection sites from one leaf. 1.0 μg DNase-treated RNA was reverse-transcribed to cDNA using SuperScript™ III Reverse Transcriptase (Thermo Fisher Scientific) according to the manufacturer's instructions.

Quantitative reverse transcription PCR (RT-qPCR) was done with SYBR Select Master Mix (Thermo Fisher Scientific) on a CFX384 Touch Real-Time PCR Detection System (Bio-Rad). Primers specific for cassava GTPb (Manes.09G086600) and PP2A4 (Manes.09G039900) genes were used as the internal controls to calculate relative expression levels (Supplementary Table 1)[50,51]. Normalized relative expression was then calculated[52].

Water-soaking area and intensity quantification, image analysis

Cassava leaves were infiltrated with either 10 mM $MgCl_2$ alone (mock) or containing *Xanthomonas* (*Xam668* strains with and without TAL20) as described above and imaged using a Canon EOS Rebel T5i camera with a 15–85 mm lens at 0- and 4-days post-inoculation. Images were gray balanced using a X-Rite Passport by estimating the saturation of the six gray chips and lowering the brightness accordingly. Gray corrected images were uploaded to FIJI[53] and duplicated. In order to define the infected area, visible water-soaking spots were manually outlined on the duplicate image using the pencil tool (color: #ff00b6 and size 2). The outlined images were converted from RGB to LAB and split to obtain the A color channel. The A channel images were thresholded, converted to a mask and the mask for each spot was added to the ROI manager using the analyze particle tool. The ROI masks were applied to the original RGB gray corrected images. In order to define the background color of a given leaf, mock-infiltrated spots (no water-soaking) were added to the ROI manager using an arbitrarily sized rectangle selection tool consistently set to a W = 51 and H = 61. Area and gray scale mean data were obtained for each infiltrated spot using the FIJI measure tool. Gray balancing using the X-Rite Passport is a coarse adjustment and finer standardization is performed post image processing using statistical methods. Gray value standardization was achieved by estimating the grand mean of all gray values in each image and centering those values to the grand mean of all images by using model residuals. Further, the residuals within each image were re-centered with respect to the mock treatment by subtracting the mean gray value of mock from all other features.

Bacterial growth assay

Cassava leaves were infiltrated with either 10 mM $MgCl_2$ alone (mock) or containing *Xanthomonas* (*Xam668* strains with and without TAL20) as described above. Leaf punch samples were taken at the site of infiltration using a 5 mm cork borer at 0- and 4-days post-inoculation. For day-0 samples, infiltrated spots were allowed to dry down prior to processing. Individual leaf punches were transferred to 2 mL Eppendorf Safelock tubes with 200 µL of 10 mM $MgCl_2$ and three disposable 3 mm Propper solid glass beads. Samples were ground with a Qiagen Tissuelyzer at 28 hZ for 3 min. 200 µL of the ground sample were transferred to the first column of a labelled 96-well plate. Serial dilutions were performed by transferring 20 µL of the non-dilute sample (10^1) to the next well containing 180 µL of 10 mM $MgCl_2$. Samples were serially diluted to 10^4 for day-0 and 10^6 for day-4. Ten µL of each serial dilution was spread onto labelled quadrants of a NYG plate with cycloheximide and the appropriate antibiotics for the infiltrated bacterial strain. Plates were incubated at 30 °C for 2–3 days, and the number of colonies were counted. Colony forming Units (CFU) reported in this manuscript were transformed by sample area (CFU/cm^2 where cm^2 = 0.52).

Storage root production and measurements

Storage root production and analysis was conducted with some modification as described[54]. Briefly, plants were maintained in small, 2 inch pots for 11 weeks in order to induce storage root production. The same greenhouse conditions were used as described above. After 11 weeks, roots were removed from pots, washed to remove soil, and storage roots isolated from the root mass and weighed for fresh weight measurements. For starch content estimation (~80% of storage root weight), the storage roots were peeled, cut into ~2 cm cubes, placed into a 50 mL conical tube, and frozen in dry ice and stored at −80 °C prior to lyophilization. Samples were lyophilized for 2 days, then weighed to obtain dry weight.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All amplicon-based bisulfite sequencing data, western blots, R code for generating plots, vector sequences, and images used for imaged-based analysis of water-soaking disease symptoms are available in a FigShare repository [https://10.0.23.196/m9.figshare.16887934]. Whole genome bisulfite sequencing (WGBS) data is also available at GEO [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE187022]: GSM5667182 (WGBS data of wild type TME419), GSM5667183 (WGBS data of DMS3-ZF line 133), GSM5667184 (WGBS data of DMS3-ZF line 204). Source data are provided with this paper.

References

1. van Schie, C. C. N. & Takken, F. L. W. Susceptibility genes 101: how to be a good host. *Annu. Rev. Phytopathol.* **52**, 551–581 (2014).
2. Chu, Z. et al. Promoter mutations of an essential gene for pollen development result in disease resistance in rice. *Genes Dev.* **20**, 1250–1255 (2006).
3. Yang, B., Sugio, A. & White, F. F. Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *PNAS* **103**, 10503–10508 (2006).
4. Antony, G. et al. Rice xa13 Recessive Resistance to Bacterial Blight Is Defeated by Induction of the Disease Susceptibility Gene Os-11N3. *Plant Cell* **22**, 3864–3876 (2010).
5. Oliva, R. et al. Broad-spectrum resistance to bacterial blight in rice using genome editing. *Nat. Biotechnol.* **37**, 1344–1350 (2019).
6. Cohn, M. et al. Xanthomonas axonopodis virulence is promoted by a transcription activator-like effector-mediated induction of a SWEET sugar transporter in cassava. *Mol. Plant Microbe Interact.* **27**, 1186–1198 (2014).
7. Hu, Y. et al. Lateral organ boundaries 1 is a disease susceptibility gene for citrus bacterial canker disease. *PNAS* **111**, E521–E529 (2014).
8. Cox, K. L. et al. TAL effector driven induction of a SWEET gene confers susceptibility to bacterial blight of cotton. *Nat. Commun.* **8**, 1–14 (2017).
9. Phillips, A. Z. et al. Genomics-enabled analysis of the emergent disease cotton bacterial blight. *PLOS Genet.* **13**, e1007003 (2017).
10. Nekrasov, V. et al. Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Sci. Rep.* **7**, 482 (2017).
11. Peng, A. et al. Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene CsLOB1 promoter in citrus. *Plant Biotechnol. J.* **15**, 1509–1519 (2017).
12. Zhang, Y. et al. Simultaneous modification of three homoeologs of TaEDR1 by genome editing enhances powdery mildew resistance in wheat. *Plant J.* **91**, 714–724 (2017).
13. Gomez, M. A. et al. Simultaneous CRISPR/Cas9-mediated editing of cassava eIF4E isoforms nCBP-1 and nCBP-2 reduces cassava brown streak disease symptom severity and incidence. *Plant Biotechnol. J.* **17**, 421–434 (2019).
14. Xu, Z. et al. Engineering Broad-Spectrum Bacterial Blight Resistance by Simultaneously Disrupting Variable TALE-Binding Elements of Multiple Susceptibility Genes in Rice. *Mol. Plant* **12**, 1434–1446 (2019).
15. Bultmann, S. et al. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res.* **40**, 5368–5377 (2012).
16. Valton, J. et al. Overcoming Transcription Activator-like Effector (TALE) DNA Binding Domain Sensitivity to Cytosine Methylation*♦. *J. Biol. Chem.* **287**, 38427–38432 (2012).
17. O'Malley, R. C. et al. Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. *Cell* **165**, 1280–1292 (2016).
18. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet* **11**, 204–220 (2010).

19. Zhang, H., Lang, Z. & Zhu, J.-K. Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **19**, 489–506 (2018).

20. Papikian, A., Liu, W., Gallego-Bartolomé, J. & Jacobsen, S. E. Site-specific manipulation of Arabidopsis loci using CRISPR-Cas9 SunTag systems. *Nat. Commun.* **10**, 1–11 (2019).

21. Ghoshal, B., Picard, C. L., Vong, B., Feng, S. & Jacobsen, S. E. CRISPR-based targeting of DNA methylation in Arabidopsis thaliana by a bacterial CG-specific DNA methyltransferase. *Proc. Natl. Acad. Sci. USA.* **118**, e2125016118 (2021).

22. Johnson, L. M. et al. SRA- and SET-domain-containing proteins link RNA polymerase V occupancy to DNA methylation. *Nature* **507**, 124–128 (2014).

23. Gallego-Bartolomé, J. et al. Co-targeting RNA Polymerases IV and V Promotes Efficient De Novo DNA Methylation in Arabidopsis. *Cell* **176**, e19 (2019).

24. Constantin, E. C. et al. Genetic characterization of strains named as Xanthomonas axonopodis pv. dieffenbachiae leads to a taxonomic revision of the X. axonopodis species complex. *Plant Pathol.* **65**, 792–806 (2016).

25. López, C. E. & Bernal, A. J. Cassava Bacterial Blight: Using Genomics for the Elucidation and Management of an Old Problem. *Tropical Plant Biol.* **5**, 117–126 (2012).

26. Lin, Z. J. D., Taylor, N. J. & Bart, R. Engineering Disease-Resistant Cassava. *Cold Spring Harb. Perspect. Biol.* **11**, a034595 (2019).

27. Bickhart, D. M. et al. Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. *Nat. Biotechnol.* 1–9 https://doi.org/10.1038/s41587-021-01130-z (2022).

28. Bart, R. et al. High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *PNAS* **109**, E1972–E1979 (2012).

29. Cohn, M., Morbitzer, R., Lahaye, T. & Staskawicz, B. J. Comparison of gene activation by two TAL effectors from Xanthomonas axonopodis pv. manihotis reveals candidate host susceptibility genes in cassava. *Mol. Plant Pathol.* **17**, 875–889 (2016).

30. Wang, H. et al. CG gene body DNA methylation changes and evolution of duplicated genes in cassava. *PNAS* **112**, 13729–13734 (2015).

31. Lim, Y.-W. et al. Mutations in DNA polymerase δ subunit 1 co-segregate with CMD2-type resistance to Cassava Mosaic Geminiviruses. *Nat. Commun.* **13**, 1–11 (2022).

32. Stelpflug, S. C., Eichten, S. R., Hermanson, P. J., Springer, N. M. & Kaeppler, S. M. Consistent and heritable alterations of DNA methylation are induced by tissue culture in maize. *Genetics* **198**, 209–218 (2014).

33. Stroud, H. et al. Plants regenerated from tissue culture contain stable epigenome changes in rice. *eLife* https://elifesciences.org/articles/00354. https://doi.org/10.7554/eLife.00354 (2013).

34. Ong-Abdullah, M. et al. Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature* **525**, 533–537 (2015).

35. Aung, K., Jiang, Y. & He, S. Y. The role of water in plant-microbe interactions. *Plant J.* **93**, 771–780 (2018).

36. Li, T., Liu, B., Spalding, M. H., Weeks, D. P. & Yang, B. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat. Biotechnol.* **30**, 390–392 (2012).

37. Wang, F. et al. Enhanced Rice Blast Resistance by CRISPR/Cas9-Targeted Mutagenesis of the ERF Transcription Factor Gene OsERF922. *PLoS One* **11**, e0154027 (2016).

38. Blanvillain-Baufumé, S. et al. Targeted promoter editing for rice resistance to Xanthomonas oryzae pv. oryzae reveals differential activities for SWEET14-inducing TAL effectors. *Plant Biotechnol. J.* **15**, 306–317 (2017).

39. Chandrasekaran, J. et al. Development of broad virus resistance in non-transgenic cucumber using CRISPR/Cas9 technology. *Mol. Plant Pathol.* **17**, 1140–1153 (2016).

40. Wang, Y. et al. Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nat. Biotechnol.* **32**, 947–951 (2014).

41. Schmitz, R. J. et al. Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. *Science* **334**, 369–373 (2011).

42. Zheng, X. et al. Transgenerational Variations in DNA Methylation Induced by Drought Stress in Two Rice Varieties with Distinguished Difference to Drought Resistance. *PLoS ONE* **8**, e80253 (2013).

43. Law, J. A. et al. A Protein Complex Required for Polymerase V Transcripts and RNA- Directed DNA Methylation in Arabidopsis. *Curr. Biol.* **20**, 951–956 (2010).

44. Taylor, N. et al. A High-throughput Platform for the Production and Analysis of Transgenic Cassava (Manihot esculenta) Plants. *Tropical Plant Biol.* **5**, 127–139 (2012).

45. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinforma.* **10**, 232 (2009).

46. Mansfeld, B. N. et al. Large structural variations in the haplotype-resolved African cassava genome. *Plant J.* https://doi.org/10.1111/tpj.15543 (2021).

47. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

48. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).

49. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004).

50. Moreno, I., Gruissem, W. & Vanderschuren, H. Reference genes for reliable potyvirus quantitation in cassava and analysis of Cassava brown streak virus load in host varieties. *J. Virol. Methods* **177**, 49–54 (2011).

51. Hu, M., Hu, W., Xia, Z., Zhou, X. & Wang, W. Validation of Reference Genes for Relative Quantitative Gene Expression Studies in Cassava (Manihot esculenta Crantz) by Using Quantitative Real-Time PCR. *Front Plant Sci.* **7**, 680 (2016).

52. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**, R19 (2007).

53. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

54. Beyene, G., Chauhan, R. D., Gehan, J., Siritunga, D. & Taylor, N. Cassava shrunken-2 homolog MeAPL3 determines storage root starch and dry matter content and modulates storage root post-harvest physiological deterioration. *Plant Mol. Biol.* https://doi.org/10.1007/s11103-020-00995-z (2020).

Acknowledgements

Author contributions

Conceptualization: S.M.W., J.C.C., S.E.J., R.S.B.; Methodology: K.M.V., K.E., G.J., Z.Z., S.F., M.Y., K.B.G., J.C.B., Z.J.D.L., B.G., J.G.B., J.C.C., S.E.J., R.S.B.; Investigation: K.M.V., K.E., G.J., Z.Z., S.F., M.Y., K.B.G., J.C.B., Z.J.D.L., B.G., J.G.B., J.N.; Visualization: K.M.V., K.E., Z.Z., M.Y., K.B.G., J.C.B.; Funding acquisition: S.M.W., J.C.C., S.E.J., R.S.B.; Project

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-35675-7.

Correspondence and requests for materials should be addressed to Steven E. Jacobsen or Rebecca S. Bart.

Peer review information *Nature Communications* thanks Alvaro Perez-Quintero and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.