LETTER

doi:10.1038/nature15365

Loss of *Karma* transposon methylation underlies the mantled somaclonal variant of oil palm

Meilina Ong-Abdullah[1], Jared M. Ordway[2], Nan Jiang[2], Siew-Eng Ooi[1], Sau-Yee Kok[1], Norashikin Sarpan[1], Nuraziyan Azimi[1], Ahmad Tarmizi Hashim[1], Zamzuri Ishak[1], Samsul Kamal Rosli[1], Fadila Ahmad Malike[1], Nor Azwani Abu Bakar[1], Marhalil Marjuni[1], Norziha Abdullah[1], Zulkifli Yaakub[1], Mohd Din Amiruddin[1], Rajanaidu Nookiah[1], Rajinder Singh[1], Eng-Ti Leslie Low[1], Kuang-Lim Chan[1], Norazah Azizi[1], Steven W. Smith[2], Blaire Bacher[2], Muhammad A. Budiman[2], Andrew Van Brunt[2], Corey Wischmeyer[2], Melissa Beil[2], Michael Hogan[2]†, Nathan Lakey[2], Chin-Ching Lim[3], Xaviar Arulandoo[3], Choo-Kien Wong[4], Chin-Nee Choo[4], Wei-Chee Wong[4], Yen-Yen Kwan[5], Sharifah Shahrul Rabiah Syed Alwee[5], Ravigadevi Sambanthamurthi[1] & Robert A. Martienssen[6]

Somaclonal variation arises in plants and animals when differentiated somatic cells are induced into a pluripotent state, but the resulting clones differ from each other and from their parents. In agriculture, somaclonal variation has hindered the micropropagation of elite hybrids and genetically modified crops, but the mechanism responsible remains unknown[1]. The oil palm fruit 'mantled' abnormality is a somaclonal variant arising from tissue culture that drastically reduces yield, and has largely halted efforts to clone elite hybrids for oil production[2–4]. Widely regarded as an epigenetic phenomenon[5], 'mantling' has defied explanation, but here we identify the *MANTLED* locus using epigenome-wide association studies of the African oil palm *Elaeis guineensis*. DNA hypomethylation of a LINE retrotransposon related to rice *Karma*, in the intron of the homeotic gene *DEFICIENS*, is common to all mantled clones and is associated with alternative splicing and premature termination. Dense methylation near the *Karma* splice site (termed the *Good Karma* epiallele) predicts normal fruit set, whereas hypomethylation (the *Bad Karma* epiallele) predicts homeotic transformation, parthenocarpy and marked loss of yield. Loss of *Karma* methylation and of small RNA in tissue culture contributes to the origin of mantled, while restoration in spontaneous revertants accounts for non-Mendelian inheritance. The ability to predict and cull mantling at the plantlet stage will facilitate the introduction of higher performing clones and optimize environmentally sensitive land resources.

The African oil palm (*E. guineensis*) is the most efficient oil-bearing crop, but demand for edible oils and biofuels, combined with sustainability concerns over dwindling rainforest reserves, has led to intense

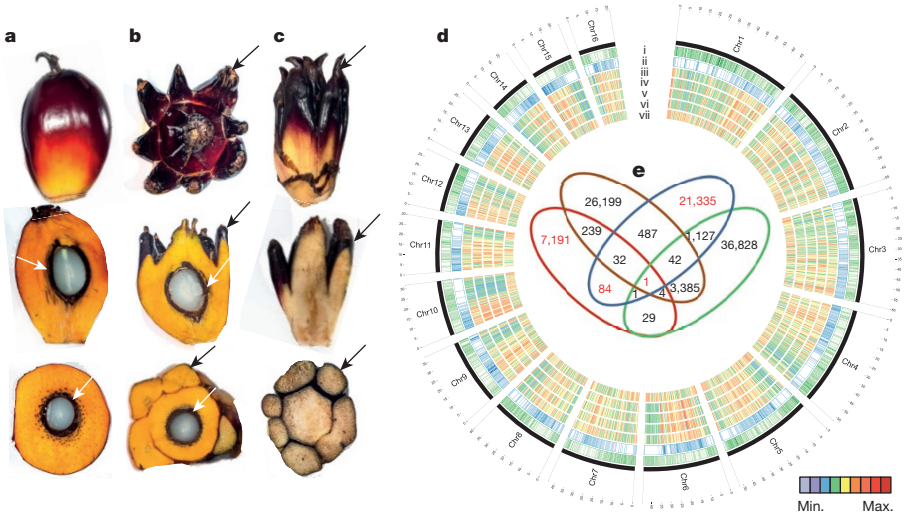


Figure 1 | Epigenome-wide association study (EWAS). **a**–**c**, Normal (**a**), fertile mantled (**b**) and parthenocarpic mantled (**c**) fruit shown as whole fruit (top), longitudinal sectioned (middle) and cross sectioned (bottom). Black arrows denote pseudocarpels; white arrows denote kernel. **d**, Circos plot of oil palm chromosomes. Track order: gene density (i); repeat density (ii); cytosine methylation density (whole-genome bisulfite sequencing) in ortet (iii); cytosine methylation densities (microarray) of ortet (iv), normal ramet (v) and mantled ramet (vi); differential cytosine methylation of normal minus mantled ramets (vii). Heatmaps represent average cytosine methylation densities in ~300-kb windows independent of sequence context. **e**, Venn diagram of microarray features differentially methylated between leaves from mantled and normal ramets ($P < 0.05$, two-sided Student *t*-test, Methods). Each set represents clonal lineages of given genotypes obtained from one source: source A (red, 15 mantled, 15 normal), source B (brown, 6 mantled, 14 normal), source C (blue, 14 mantled, 15 normal) and source D (green, 8 mantled, 10 normal). Red numbers indicate subsets including one of the four microarray features mapping to the *Karma* LINE element.

[1]Malaysian Palm Oil Board, 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia. [2]Orion Genomics, 4041 Forest Park Avenue, St Louis, Missouri 63108, USA. [3]United Plantations Berhad, Jendarata Estate, 36009 Teluk Intan, Perak, Malaysia. [4]Applied Agricultural Resources Sdn Bhd, No. 11, Jalan Teknologi 3/6, Taman Sains Selangor 1, 47810 Kota Damansara, Petaling Jaya, Selangor, Malaysia. [5]FELDA Global Ventures R&D Sdn Bhd, c/o FELDA Biotechnology Centre, PT 23417, Lengkuk Teknologi, 71760 Bandar Enstek, Negeri Sembilan, Malaysia. [6]Howard Hughes Medical Institute-Gordon and Betty Moore Foundation, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. †Present address: Thermo Fisher Scientific, 110 Miller Avenue, Ann Arbor, Michigan 48104, USA.

pressure to improve oil palm yield. Introduction of the *tenera* hybrid (*dura* × *pisifera*) increased oil yield by up to 30%, leveraging the *SHELL* gene that confers single gene heterosis[6,7]. Clones (ramets) of individual high-yielding *tenera* hybrid palms (ortets) provide a powerful shortcut to yield enhancement, with an additional 20–30% improvement[6]. Micropropagation through cell culture of immature apex leaf tissue (the 'heart of palm'), and plantlet regeneration on hormone-supplemented media (Methods), yields tens of thousands of genetically identical clonal palms. Unfortunately, shortly after the procedure was established, Tan Yap Pau of United Plantations, Malaysia, first noted a high frequency of homeotic floral phenotypes known as "mantling" among clonal ramets (T.Y.P., personal communication). Subsequently, Corley *et al.*[2] documented the occurrence of mantled palms after prolonged periods in culture. In mantled palms, staminodes of pistillate flowers and stamens of staminate flowers develop as pseudocarpels[8], often resulting in sterile parthenocarpic flowers with abortive fruit and very low oil yields (Fig. 1a–c and Extended Data Fig. 1). Pollination of mantled palms gave rise to variable numbers of mantled progeny, resembling rare naturally mantled variants known as *poissoni*, or *diwakkawakka* fruit forms[3,4]. The trait is non-Mendelian and sometimes reverts to normal[9] and so has long been considered epigenetic[5], with an overall decrease in DNA methylation found in mantled ramets[5,10]. The homeotic transformations observed in mantled palms resemble defects in B-function MADS-box genes, suggesting strong candidates for epigenetic modification[8]. However, decades of research into candidate retroelements[11,12] and candidate homeotic genes[8,12,13] have failed to identify epigenetic changes consistently found in somaclonal mantled palms.

We performed a genome-wide, unbiased, DNA methylation analysis (an epigenome-wide association study; EWAS) in search of loci epigenetically associated with the mantled phenotype, using a DNA microarray based on the *E. guineensis* (*pisifera*) reference genome[14] (Methods). DNA methylation density was measured in 1–2-kilobase (kb) intervals surrounding each feature by DNA-methylation-dependent comparative microarray hybridization[15] and statistical analyses (Methods). Genome-wide DNA methylation maps were constructed from parthenocarpic mantled ($n = 43$) or normal ($n = 54$) ramets, as well as ortets from which these ramets were derived ($n = 10$). These maps strongly resembled those constructed by whole-genome bisulfite sequencing in sample palms (Fig. 1d), demonstrating reproducibility.

At genome-wide resolution, the landscape of DNA methylation was remarkably consistent between ortets and ramets (Fig. 1d), with highest methylation within repetitive sequences[14]. However, thousands of loci were differentially methylated (Fig. 1d), most of which (~90%) were hypomethylated in mantled, consistent with previously reported reduced 5mC content[5,10]. Most hypomethylated loci (~75%) were transposons and repeats, while less frequent hypermethylated loci included genic sequences (Extended Data Fig. 2), resembling cell cultures of *Arabidopsis*[16]. Fifteen independent somaclonal lineages obtained from four independent sources were used to maximize genotypic diversity, and significant differentially methylated regions (DMRs) between normal and fully mantled samples were first identified within each source population (Methods). Results were then compared between populations (Fig. 1e). Although tens of thousands of DMRs were detected between mantled and normal clones in each population, 99.9% of these were exclusive to either one (94.4%) or two

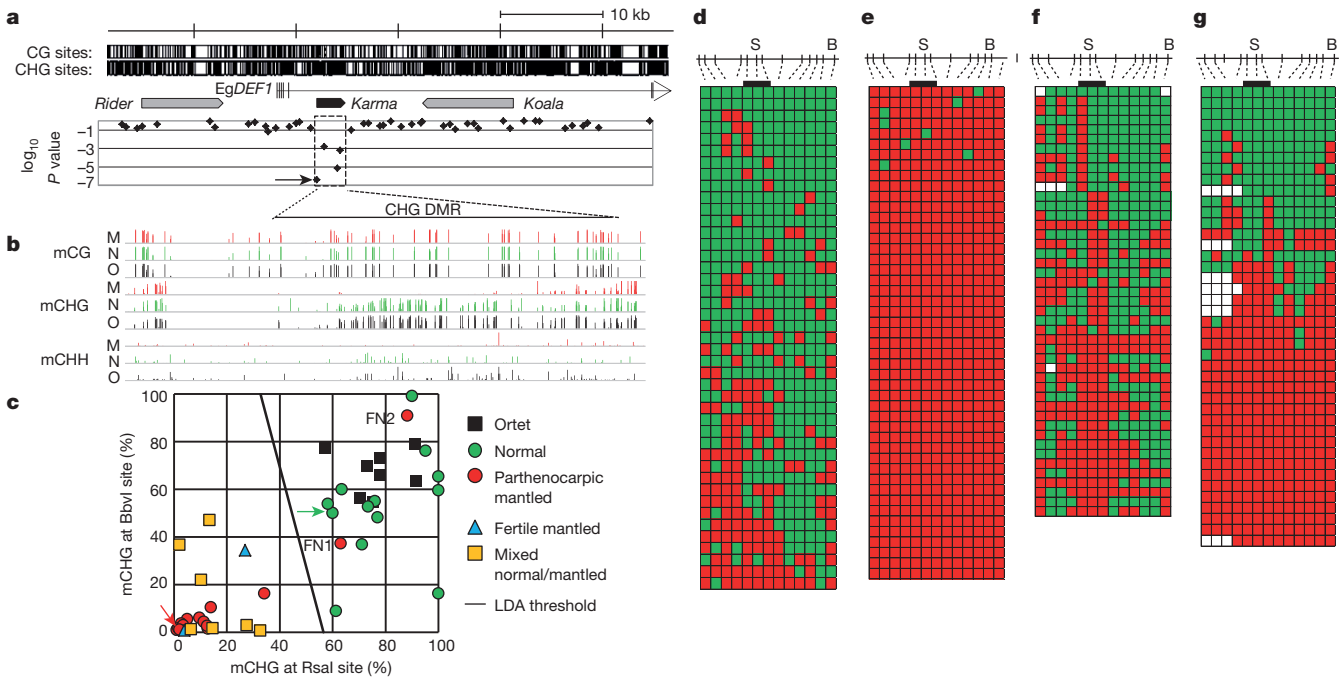


Figure 2 | Hypomethylation of *Karma* is associated with the mantled phenotype. **a**, Microarray feature data plotted below a map of the Eg*DEF1* gene (vertical ticks, exons; horizontal line, introns; arrow, direction of transcription) including locations of *Rider*, *Karma* (dashed box) and *Koala* retrotransposons. CG and CHG sites are shown at the top. $\log_{10}$ *P* values (54 normal versus 43 parthenocarpic mantled ramets) are plotted (two-sided Student's *t*-test). Arrow in *P* value plot denotes feature detected as hypomethylated in mantled ramets from all four sources (Fig. 1e). **b**, Genome-wide bisulfite sequencing of leaf samples from ortet (O, black, $n = 5$), normal ramets (N, green, $n = 5$) and parthenocarpic mantled ramets (M, red, $n = 5$). Mean methylation density per cytosine is plotted on a 0–100% scale for each cytosine context and sample type. CHG DMR, differentially CHG methylated region corresponding to *Karma*. **c**, CHG methylation monitored in 86 additional ortets, mantled and normal ramet leaf samples by restriction enzyme digestion and qPCR (Methods). Linear discriminant analysis was performed between normal ($n = 21$) and mantled ($n = 28$) samples with BbvI and RsaI restriction sites. FN1 and FN2, two false-negative mantled samples. Green and red arrows denote normal and mantled control samples, respectively. A similar analysis was performed on remaining normal ($n = 14$) and mantled ($n = 23$) samples with ScrFI restriction sites (Extended Data Fig. 4d). **d**–**g**, *Karma* bisulfite sequencing maps (antisense strand) of normal control (**d**), mantled control (**e**), FN1(**f**) and FN2 (**g**). Thirteen CHG sites are shown to scale above. 'S' denotes CHG at the *Karma* splice acceptor site (CAG/CTG); 'B' denotes the BbvI site. Bar, CHGs within the common microarray feature (Fig. 1e). Methylated and unmethylated CHG sites are indicated by green and red boxes, respectively. Open boxes denote low-quality base calls. Each row represents an individual Sanger DNA sequencing read.

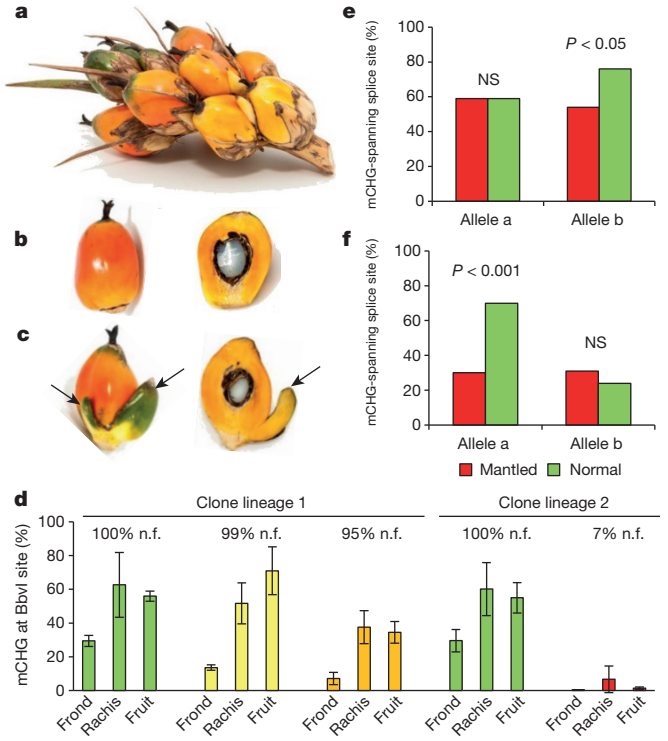


Figure 3 | *Karma* methylation in revertant palms. **a**–**c**, Spikelet from a revertant ramet (**a**) including normal (**b**) and fertile (**c**) mantled fruit with one or two pseudocarpels (arrows). **d**, Density of CHG methylation (percentage mCHG) at the BbvI site (Methods) in ramets yielding 100% normal fruit (n.f.) (green), revertant ramets yielding 99% (yellow) or 95% (orange) normal fruit and a mosaic ramet yielding 7% (red) normal fruit per bunch. Error bars denote s.d. (biological replicates of fronds ($n = 4$), rachis sections ($n = 8$) or fruit ($n = 2$)). **e**, **f**, Percentage mCHG for the three CHG sites found in the unique common microarray feature in normal (green) and subtly mantled (red) fruit from revertant ramets yielding 99% (**e**) or 95% (**f**) normal fruit per bunch (two-tailed Fisher's exact test; NS, not significant). Alleles were analysed separately based on a heterozygous single nucleotide polymorphism (SNP) within the bisulfite sequencing amplicon.

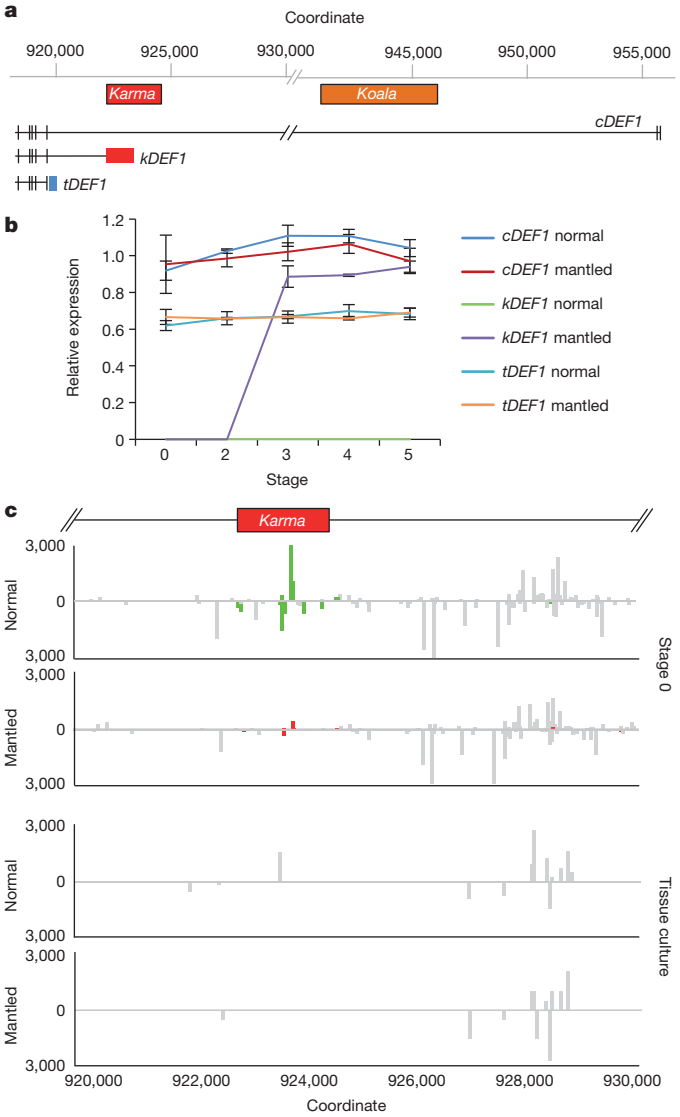


Figure 4 | Alternative splicing and loss of 24-nucleotide siRNA. **a**, Eg*DEF1* transcripts assembled from transcriptome sequencing (data not shown) and RT–PCR (Methods). Black boxes denote exons; blue box denotes intron 5 sequence included in the *tDEF1* transcript. Coordinates relative to the reference *pisifera* oil palm genome[14]. **b**, Quantitative RT–PCR of *cDEF1*, *tDEF1* and *kDEF1* transcripts in shoot apices (stage 0) and in early (stage 2) to late (stage 5) female inflorescences from normal and parthenocarpic mantled ramets. Error bars, s.d. between three replicate assays of three replicate tissue samples per phenotype, per stage. Expression relative to an endogenous reference gene is shown (Methods). **c**, 24-nucleotide siRNA accumulation in shoot apices (stage 0) from normal ($n = 5$) and parthenocarpic mantled ($n = 7$) ramets, and from second passage apical leaf tissue cultures re-cloned from normal ($n = 2$) or mantled ($n = 1$) ramets (Methods). Values expressed as fragments per kilobase per million mapped reads (FPKM). Bars above (sense) and below (antisense) the line indicate mapped normalized 24-nucleotide siRNAs that are not significantly different in abundance in normal and mantled (grey) or significantly differentially expressed in normal (green) relative to mantled (red) ($P < 0.05$, Student's *t*-test, two tailed, assuming equal variance).

(5.5%) of the four populations, indicating considerable genotypic variation in epigenetic response to tissue culture. A single microarray feature detected differential methylation between normal and mantled clones in all four populations (Fig. 1e). This feature lies within the ~35 kb intron 5 of Eg*DEF1* (Fig. 2a), the oil palm orthologue of the B-class MADS-box transcription factor genes, *Antirrhinum majus DEFICIENS* (*DEF*) and *Arabidopsis APETALA3* (*AP3*)[8,12,13].

Elaeis guineensis DEF1 spans ~40 kb on chromosome 12 (Fig. 2a). A *Ty1/copia Rider* retrotransposon lies upstream, while a *Ty3/gypsy* retrotransposon, *Koala*, is located within intron 5. Consistent with important earlier work[12], no DNA methylation difference within these retrotransposons was found in mantled clones across several populations (Fig. 2a and Extended Data Fig. 3). However, a third previously unreported repetitive element lies within intron 5, and has homology to rice *Karma* LINE elements. *Karma* is activated in rice embryogenic tissue culture, but only transposes in regenerated plants as trans-generational DNA hypomethylation of the element persists[17]. The 3.2-kb oil palm *Karma* element is flanked by a 13-base-pair (bp) target site duplication (TTCAAAATGATGA) and includes a defective reverse transcriptase open reading frame (ORF2) preceded by a splice acceptor ('^') and followed by a polyadenylation signal, resembling truncated *Karma* elements in rice[17,18] (Supplementary Fig. 1). The unique microarray feature, which consistently detected hypomethylation in mantled clones, serendipitously includes the predicted splice acceptor site (GAACAG^ATGC). All three additional microarray features mapping within the *Karma* element also detected significant hypomethylation in mantled clones (Fig. 2a and Extended Data Figs 3 and 4a–c).

To verify *Karma* hypomethylation, sample trios comprising genetically identical ortet, parthenocarpic mantled and normal ramets from five independent clonal lineages were subjected to whole-genome bisulfite sequencing (Methods). CG methylation was unchanged across the Eg*DEF1* locus, but *Karma* CHG methylation (H = A, C

or T) was markedly reduced in mantled clones, revealing a DMR covering ~70 CHG sites. CHH methylation was much lower and only subtly reduced (Fig. 2b). To validate differential CHG methylation in unrelated clonal palms, quantitative PCR (qPCR) assays were used to quantify CHG methylation at BbvI and RsaI restriction sites within the DMR (Methods, Fig. 2c and Extended Data Fig. 4b) in a panel of 49 palms from 21 clonal lineages and 4 independent sources: 8 ortets and 13 normal clones, 19 parthenocarpic mantled clones, 2 fertile mantled clones and 7 partially revertant clones yielding bunches with both mantled and normal fruit. Linear discriminant analysis provided 93% sensitivity and 100% specificity for detection of mantling (Fig. 2c). Fronds from all seven of the revertant palms were scored as mantled, consistent with the observation that normal bunches arose late in development[2]. Similar results were obtained in 37 polymorphic palms using alternative restriction sites (Extended Data Fig. 4d). The two false-negative mantled palms (Fig. 2c) were further analysed by bisulfite sequencing of a region spanning the *Karma* splice acceptor site. While normal clones had dense CHG methylation, and mantled controls had lost all CHG methylation (Fig. 2d, e), the false-negative mantled samples lost CHG methylation near the splice acceptor site (Fig. 2f, g and Extended Data Fig. 4e), which was therefore sufficient to predict the mantled phenotype. Because of their strong predictive properties, we named the *mantled* hyper- and hypomethylated epialleles *Good Karma* and *Bad Karma*, respectively.

Two lineages of revertant palms had mixed bunches with both normal and mantled fruit[9], resembling epialleles in maize regulated by transposons[19]. The first lineage included two revertant ramets with 99% and 95% normal fruit per bunch, respectively, in which abnormal fruit had only one or two small pseudocarpels (Fig. 3a–c). A second lineage included a mosaic ramet with only 7% normal fruit. In all three ramets, CHG methylation at the BbvI site was low in fronds (Fig. 3d), consistent with other revertants (Fig. 2c). However, methylation was restored in fruit from the two revertant ramets, but not from the mantled mosaic ramet (Fig. 3d–f). As with similar epialleles in maize, *Linaria* and other plants[19–21], reversion of the abnormal phenotype accompanied by restoration of DNA methylation is strong evidence that *Karma* hypomethylation is the cause of the mantled phenotype. Differential methylation between individual mantled and normal fruit was not observed, however, probably reflecting non-cell autonomy of the B-class homeotic phenotype (Fig. 3a–d), also observed in *Antirrhinum* and *Arabidopsis*[22]. Bisulfite sequencing from normal and mantled fruit (Extended Data Fig. 5) revealed hyper- and hypomethylated reads at the splice acceptor site (Fig. 3e, f), suggesting that these fruit were indeed mosaic for hyper- and hypomethylated cells. In one mosaic palm, direct evidence for mosaicism was obtained from different samples of the same vegetative frond (Extended Data Fig. 6).

DNA methylation near splice acceptor sites affects alternative splicing, although the mechanism remains unclear[23]. To assess alternative splicing, Eg*DEF1* transcript models were built on the basis of transcriptome sequencing (data not shown), and were validated by reverse transcription PCR (RT–PCR; Methods). As previously reported[12], two forms of Eg*DEF1* transcripts were found in both normal and mantled inflorescences: the full-length Eg*DEF1* transcript (*cDEF1*), and a prematurely terminated transcript including exons 1–5 and 221 bp of intron 5 (*tDEF1*)[12] (Fig. 4a and Extended Data Fig. 4c). However, we identified a third alternative transcript in mantled female inflorescences (Fig. 4a). This novel transcript (*kDEF1*) was spliced from the donor site of intron 5 to the proximal *Karma* acceptor site, and is predicted to encode a truncated EgDEF1 peptide that terminates within the K domain of the MADS-box protein, and has a unique carboxy-terminal sequence (Extended Data Fig. 7a).

Expression of the three transcripts (Fig. 4b) was assessed in shoot apical meristem (stage 0) and four stages of immature female inflorescence development (Methods and Extended Data Fig. 7b–f). *kDEF1* expression was notably restricted to stages 3–5 of mantled (but not normal) female inflorescence, strongly suggestive of a role in the mantled phenotype. By contrast, *cDEF1* was detected at only slightly (although significantly) lower levels in mantled female inflorescences[12], while *tDEF1* was unchanged (Fig. 4b). Previously reported differences in the timing of *tDEF1* and *cDEF1* transcription within each stage[12] may be the consequences of *kDEF1* transcription (Methods).

In plants, 24-nucleotide small interfering RNAs (siRNAs) guide non-CG methylation, and we identified a cluster of antisense 24-nucleotide *Karma* siRNAs in shoot apical meristem (stage 0), which were reduced or absent in mantled (Fig. 4c) and in later stages of normal inflorescence (Extended Data Fig. 8 and Methods). In polyembryogenic tissue cultures derived from normal and abnormal clonal palms (Methods), small RNA (sRNA) underwent a switch from 24 to 21 nucleotides (Extended Data Fig. 9), resembling cell cultures and callus of *Arabidopsis*[16]. *Karma* methylation (Extended Data Fig. 10) and 24-nucleotide siRNA (Fig. 4c and Extended Data Fig. 8) were reduced in normal cultures between two and seven passages, and lost in abnormal cultures, suggesting a model for the origin of mantled: if meristems are the source of 24-nucleotide siRNA[24] (Fig. 4c), then leaf cells detached from the meristem would progressively lose 24-nucleotide siRNA and non-CG DNA methylation over time in culture. Antisense sRNA might influence exon trapping, as it does in humans[25,26], or else splicing might be associated with changes in chromatin, for example histone H3 Lys4 methylation[18]. *Good Karma* would only be restored during shoot generation if DNA methylation and small RNA were not entirely lost during tissue culture, or potentially if siRNAs were artificially applied during the tissue culture process. The 24-nucleotide siRNA could also facilitate non-Mendelian segregation[4,9], which resembles paramutation (the interconversion of heterozygous epialleles) in some respects[11,23,27].

Despite its importance, mantling has been the elusive target of molecular genetic investigation for the past three decades. We have demonstrated that the mantled trait is a consequence of epigenetic modification of the *Karma* transposable element within the B-class MADS-box Eg*DEF1* gene, which we have named *MANTLED*. B-class function in stamen identity is conserved in monocots, but similar to other monocot paleoAP3 genes, Eg*DEF1* overexpression fails to cause homeotic conversion in *Arabidopsis*[8], because of a diverged C-terminal exon[28]. Nonetheless, like *AP3* and *DEF*[29,30], the oil palm *MANTLED* gene is expressed in inner perianth and stamen primordia as the perianth initiates (stage 2), followed by stamen and stamenoid primordia (stage 3)[8]. The appearance of *kDEF1* transcripts during this transition (Fig. 4b) suggests that kDEF1 functions at this crucial window to induce the mantled phenotype.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 April; accepted 10 August 2015.
Published online 9 September 2015.

1. Stroud, H. *et al.* Plants regenerated from tissue culture contain stable epigenome changes in rice. *Elife* **2,** e00354 (2013).
2. Corley, R. H. V. in *CRC Handbook of Fruit Set and Development* (ed. Monselise, S. P.) 253–259 (CRC Press, 1986).
3. Zeven, A. C. The 'mantled' oil palm (*Elaeis guineensis* Jacq.). *J. W. Afr. Inst. Oil Palm Res.* **5,** 31–33 (1973).
4. Mgbeze, G. C. & Iserhienrhien, A. Somaclonal variation associated with oil palm (*Elaeis guineensis* Jacq.) clonal propagation: a review. *Afr. J. Biotechnol.* **13,** 989–997 (2014).
5. Jaligot, E., Rival, A., Beule, T., Dussert, S. & Verdeil, J. L. Somaclonal variation in oil palm (*Elaeis guineensis* Jacq.): the DNA methylation hypothesis. *Plant Cell Rep.* **19,** 684–690 (2000).
6. Corley, R. H. V. & Law, I. H. in *Plantation Management for the 21st Century* (ed. Pushparajah, E.) 279–289 (Incorp. Soc. Planters, 1997).
7. Singh, R. *et al.* The oil palm *SHELL* gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature* **500,** 340–344 (2013).
8. Adam, H. *et al.* Functional characterization of MADS box genes involved in the determination of oil palm flower structure. *J. Exp. Bot.* **58,** 1245–1259 (2007).
9. Rao, V. & Donough, C. R. Preliminary evidence of a genetic cause for the floral abnormalities in some oil palm ramets. *Elaies* **2,** 199–207 (1990).
10. Matthes, M., Singh, R., Cheah, S. C. & Karp, A. Variation in oil palm (*Elaeis guineensis* Jacq.) tissue culture-derived regenerants revealed by AFLPs with methylation-sensitive enzymes. *Theor. Appl. Genet.* **102,** 971–979 (2001).

11. Kubis, S. E., Castilho, A. M., Vershinin, A. V. & Heslop-Harrison, J. S. Retroelements, transposons and methylation status in the genome of oil palm (Elaeis guineensis) and the relationship to somaclonal variation. *Plant Mol. Biol.* **52,** 69–79 (2003).
12. Jaligot, E. *et al.* DNA methylation and expression of the *EgDEF1* gene and neighboring retrotransposons in *mantled* somaclonal variants of oil palm. *PLoS ONE* **9,** e91896 (2014).
13. Syed Alwee, S. *et al.* Characterization of oil palm MADS box genes in relation to the mantled flower abnormality. *Plant Cell Tissue Organ Cult.* **85,** 331–344 (2006).
14. Singh, R. *et al.* Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* **500,** 335–339 (2013).
15. Lippman, Z. *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430,** 471–476 (2004).
16. Tanurdzic, M. *et al.* Epigenomic consequences of immortalized plant cell suspension culture. *PLoS Biol.* **6,** e302 (2008).
17. Komatsu, M., Shimamoto, K. & Kyozuka, J. Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon Karma. *Plant Cell* **15,** 1934–1944 (2003).
18. Cui, X. *et al.* Control of transposon activity by a histone H3K4 demethylase in rice. *Proc. Natl Acad. Sci. USA* **110,** 1953–1958 (2013).
19. Martienssen, R., Barkan, A., Taylor, W. C. & Freeling, M. Somatically heritable switches in the DNA modification of Mu transposable elements monitored with a suppressible mutant in maize. *Genes Dev.* **4,** 331–343 (1990).
20. Cubas, P., Vincent, C. & Coen, E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature* **401,** 157–161 (1999).
21. Saze, H. & Kakutani, T. Heritable epigenetic mutation of a transposon-flanked *Arabidopsis* gene due to lack of the chromatin-remodeling factor DDM1. *EMBO J.* **26,** 3641–3652 (2007).
22. Perbal, M. C., Haughn, G., Saedler, H. & Schwarz-Sommer, Z. Non-cell-autonomous function of the *Antirrhinum* floral homeotic proteins *DEFICIENS* and *GLOBOSA* is exerted by their polar cell-to-cell trafficking. *Development* **122,** 3433–3441 (1996).
23. Regulski, M. *et al.* The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. *Genome Res.* **23,** 1651–1662 (2013).
24. Baubec, T., Finke, A., Mittelsten Scheid, O. & Pecinka, A. Meristem-specific expression of epigenetic regulators safeguards transposon silencing in *Arabidopsis*. *EMBO Rep.* **15,** 446–452 (2014).
25. Taniguchi-Ikeda, M. *et al.* Pathogenic exon-trapping by SVA retrotransposon and rescue in Fukuyama muscular dystrophy. *Nature* **478,** 127–131 (2011).
26. Alló, M. *et al.* Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nature Struct. Mol. Biol.* **16,** 717–724 (2009).
27. Arteaga-Vazquez, M. *et al.* RNA-mediated *trans*-communication can establish paramutation at the *b1* locus in maize. *Proc. Natl Acad. Sci. USA* **107,** 12986–12991 (2010).
28. Lamb, R. S. & Irish, V. F. Functional divergence within the *APETALA3/PISTILLATA* floral homeotic gene lineages. *Proc. Natl Acad. Sci. USA* **100,** 6558–6563 (2003).
29. Sommer, H. *et al.* Deficiens, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *EMBO J.* **9,** 605–613 (1990).
30. Jack, T., Brockman, L. L. & Meyerowitz, E. M. The homeotic gene *APETALA3* of *Arabidopsis* thaliana encodes a MADS box and is expressed in petals and stamens. *Cell* **68,** 683–697 (1992).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge the contributions of staff members of the Breeding and Tissue Culture Unit at MPOB for creating the valuable clonal lines, and for their extensive data collection and sampling efforts. We thank Genomics Unit at MPOB for conducting DNA fingerprinting to verify clonal lines. We thank The McDonnell Genome Institute at Washington University for genomic bisulfite sequencing and transcriptome sequencing support, and MOgene for microarray hybridizations. At Orion Genomics, we thank N. Sander, J. Reed, J. Brune, K. Soe, J. McDonald, C. Brown and B. Dove for technical support, and M.-F. Wu and M. Sachdeva for assistance with the manuscript and additional informatics support. We would also like to thank T. Dalmay for recommendations on sRNA library construction. We appreciate the constant support of the Director-General of MPOB, Datuk Dr. Yuen-May Choo, and the Ministry of Plantation Industries and Commodities, Malaysia.

Author Contributions M.O.-A. led the work on the *MANTLED* marker/gene. M.O.-A., R.Si., E.-T.L.L. and R.Sa. conceptualized the research programme. M.O.-A., R.Si., E.-T.L.L., R.N., N.L., S.W.S., J.M.O., R.Sa. and R.A.M. developed the overall strategy, designed experiments and coordinated the project. A.T.H., Z.I. and S.K.R. performed tissue culture on selected ortets and field-planted the ramets. Field data collection and fruit bunch census were conducted at various research stations by F.A.M., N.A.A.B., M.M., N.A., Z.Y. and M.D.A. M.O.-A., C.-C.L., X.A., C.-N.C., W.-C.W., S.S.R.S.A. and Y.-Y.K. identified samples for discovery and validation panels. M.O.-A., C.-C.L. and X.A. identified materials used in the mosaic experiments. M.O.-A., S.-E.O., S.-Y.K., N.S. and N.A. conducted laboratory experiments, histological staging of inflorescences and data analyses. N.J. and S.W.S performed microarray analyses. B.B. and M.A.B. prepared fractions for microarray hybridizations. B.B. designed and analysed qPCR experiments. B.B. and M.B. performed qPCR assays. A.V.B. designed and analysed clone-based bisulfite sequencing experiments, and A.V.B. and M.B. performed bisulfite sequencing assays. M.B. designed and performed qRT–PCR experiments. C.W. and J.M.O. analysed transcriptome data. K.-L.C., N.A., S.W.S., M.H., C.W. and A.V.B. provided bioinformatics support. M.O.-A., R.Si., E.-T.L.L., R.N., N.L., S.W.S., J.M.O., R.Sa. and R.A.M. prepared and revised the manuscript.

Author Information Microarray data have been deposited in the NCBI Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE68410. Small RNA sequence data from the region of interest have been deposited in the NCBI Sequence Read Archive (SRA) database under the accession numbers SAMN03569290–SAMN03569351. Whole-genome bisulfite sequence data have been deposited in the NCBI SRA under accession numbers SAMN03569063–SAMN03569077. The cDNA sequence of the *kDEF1* transcript has been deposited in GenBank under accession number KR347486. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.A.M. (martiens@cshl.edu) or R.Sa. (raviga@mpob.gov.my).

METHODS

Samples and tissue culture. Leaf tissue used in the EWAS discovery and validation panels was sampled from oil palm clones (known as ramets) derived from *tenera* mother palms (known as ortets) from various genetic backgrounds. The female parents of the ortets were predominantly of a Deli *dura* background while the *pisifera* male parents were derived from La Me, Yangambi, AVROS and Binga genotypes. These parental lines make up the major genetic backgrounds of the oil palm populations in Malaysia. These palms were collected from the Malaysian Palm Oil Board (MPOB), United Plantations Berhad, FELDA Global Ventures R&D Sdn Bhd and Applied Agricultural Resources Sdn Bhd (Extended Data Fig. 3). The ramets were derived from explants excised from non-chlorophyllous leaves of their respective ortets cultured on hormone-supplemented media[31–34]. The oil palm tissue culture process involves callus initiation and polyembryoid generation followed by shoot and root initiation[35]. In large scale production, the tissue culture process takes 48–52 months to complete. Once established, ramets undergo acclimatization in the nursery for 3–4 months, are moved to the field nursery for another 8–9 months, and finally are field planted. At every stage of the tissue culture process, off-types are culled. Flower and fruit bunch census are taken at the onset of flowering (2–3 years after field planting) and in subsequent years. Normal and mantled ramets were identified based on the census data collected.

Recloned tissue culture materials were generated using the standard protocol described above. Normal and abnormal ramets from identical genetic backgrounds were subjected to the tissue culture process and sub-sampling was carried out at the polyembryogenic stage, namely at subculture passage two (SC2) and seven (SC7), representing short and prolonged exposure in culture on hormone-supplemented media, respectively.

Female inflorescence samples used in staging the developmental phase were obtained from a total of 31 clonal palms with ages ranging from 3 to 10 years after field-planting. As a means to categorize the developmental phases, these inflorescence samples were histologically analysed and classified as stage 0, shoot apical meristem; stage 2, initiation of perianth organs; stage 3, development of perianth organs and initiation of reproductive organs; stage 4, development of reproductive organs; stage 5, fully formed reproductive organs according to ref. 36.

***E. guineensis* genome microarray design.** The microarray design was based on the *E. guineensis* genome[14] and contains more than 1 million 60-base probes. Probes were selected from non-overlapping 1.5-kb windows across all scaffolds of the P1 *pisifera* genome build, choosing the 60-mer with the lowest composite 15-mer frequency count within the genome. When compared to the publically available EG5 genome assembly[14], 860,861 probes from the microarray matched at 100% stringency (81,194 probes falling on exons and the remaining covering intronic and intergenic regions of the genome). The microarrays were manufactured by Roche NimbleGen using the HX1 platform.

DNA methylation-dependent fractionation. Genomic DNA (60 μg) was mechanically hydrosheared to 1–4-kb fragments. Sheared DNA was divided into four equal portions, two of which were digested with 10 U μg^{-1} McrBC (New England Biolabs) under manufacturer's recommended conditions and the other two were mock-treated without adding enzyme. After digestion, DNAs were treated with proteinase K (50 mg ml^{-1}) for 1 h at 50 °C and then precipitated with ethanol under standard conditions. Resuspended DNAs were resolved by agarose gel electrophoresis, and DNA in the 1–4-kb size range was excised from gels and extracted. McrBC requires that two methylated half sites (RmC, where R = A or G) lie within 40–3,000 bp of each other, and cutting occurs in the proximity of one of the half sites[37]. Because 1–4-kb fragments are treated with McrBC, and undigested fragments are isolated, hybridization of a microarray probe complimentary to sequence distant from the methylation site results in a 'wingspan' effect in which probes are able to detect DNA methylation from a distance up to ~1.5 kb[38]. For each fraction, 200 ng was used for cyanogen dye labelling (Cy3 or Cy5). For each sample, four microarrays were hybridized in a duplicated dye swap design to differentially labelled untreated and DNA methylation depleted fractions. Sample size was chosen to allow several clonal lineages including both normal and parthenocarpic samples from each of four independent sources, but no statistical methods were used to determine sample size.

Microarray data processing, normalization and statistical analysis. Among the ~860,000 microarray probes matching the oil palm genome with 100% sequence stringency, a subset of ~460,000 uniquely mapped probes was selected to reduce noise from non-specific hybridizations. For data processing, corrections on spatial non-uniformity of fluorescent signal intensities, done separately for the Cy3 and Cy5 dye channels, were made. Data were then normalized by background subtraction using negative control probes, followed by scaling. After normalization, median log signal of the control probes for each dye of each array was set to zero. MAD (median absolute deviation) log signal of all the probes on an array was used as a constant for a given treatment (untreated or DNA methylation depleted). DNA methylation was measured as the sample average (over the two pairs of dye-swapped technical replicates) of normalized $\log_2$ ratios of untreated over methylation-depleted DNA. Statistical analysis was first conducted within samples derived from each source independently. Within each group, a two-sided *t*-test was performed between normal and mantled phenotypes. Using a cut-off of $P = 0.05$, one probe that was significantly differentially methylated in all the four groups was identified. To confirm this finding with a different statistical approach, quantile normalization of methylation measurements (the sample average of normalized log ratios of untreated over methylation depleted DNA) was performed on all samples together, and then a *t*-test on all normal versus all mantled samples was conducted. This process identified the same probe that was found in the initial analysis, as well as an additional three immediately neighbouring probes. The experiments were not randomized, and investigators were not blinded to allocation during experiments and outcome assessment.

Code availability. Computer code for microarray data processing and normalization is available for download at http://www.oriongenomics.com/files/methylscope_processing.tar.

qPCR DNA methylation assays. Primer pairs were designed to amplify two *Karma* element regions as diagrammed in Extended Data Fig. 4a, b. A 633-bp amplicon included methylation-sensitive restriction sites containing CHG positions 188 bp (BbvI) and 375 bp (ScrFI) downstream of the *Karma* splice site CHG. A 632-bp amplicon that amplified a region near the centre of the *Karma* element included a RsaI methylation-sensitive restriction site. Primer pairs were confirmed to amplify a single band of the correct size by agarose gel electrophoresis. For qPCR DNA methylation assays, 100 ng of genomic DNA was digested with 10 U μg^{-1} of the indicated restriction enzyme under standard conditions. An equal amount of genomic DNA was mock-treated in a reaction lacking enzyme. Digestion reactions were incubated at 37 °C for 16 h. qPCR was carried out using 10 ng each of the mock-treated and enzyme-digested samples in 1 × Roche SYBR Green Master Mix on a Roche LC480 instrument. qPCR amplifications were performed in duplicate. For each duplicate mock/digested amplification pair, the ΔC_t value was calculated as the digested C_t minus the mock C_t and duplicated ΔC_t values were averaged. DNA methylation density was calculated as percentage dense methylation = $2^{(-\Delta Ct(\text{digested} - \text{mock}))}$. Samples were genotyped by restriction digestion of PCR amplicons with either BbvI or RsaI to confirm that all samples used for DNA methylation validation included intact restriction sites on both alleles. Enzyme digestions were quality controlled by performing qPCR assays monitoring three independent invariantly unmethylated endogenous genomic loci and one invariantly methylated endogenous genomic locus. All quality control passed digestions reported < 5% methylation of the unmethylated controls and > 95% methylation of the methylated control. Primer sequences are available on request.

Whole-genome bisulfite sequencing. Genomic DNA (1 μg) from each of 15 mature leaf samples (5 lineage trios of ortet, normal ramet and mantled ramet) was used to construct TruSeq fragment libraries (Illumina), and up to 500 ng of adapted library molecules were bisulfite converted using the EZ DNA Methylation-Lightning Kit (Zymo Research). Each library was sequenced in one lane of a HiSeq 2000 flow cell to generate paired 100-bp reads. Reads were mapped to an *in silico* bisulfite converted reference *E. guineensis* (*pisifera*) genome. For each cytosine context (CG, CHG or CHH), the number of mapped reads corresponding to unconverted cytosines relative to the total number of reads including the particular base was used to calculate the percent methylation at each cytosine position.

Clone based bisulfite sequencing. Because all cytosines are potential sites for DNA methylation in plants, and because whole-genome bisulfite sequencing demonstrated that CG methylation is maintained at high levels in both normal and mantled ramets, bisulfite sequencing amplicon primers were designed to include CG dinucleotides, but exclude CHG and CHH trinucleotides. Within primer sequences, CG dinucleotides were assumed to be methylated. The amplicon (amplified from the antisense strand) contained 13 CHG sites, including the CHG site at the *Karma* splice acceptor site. Then 2 μg of each sample was bisulfite-converted as described for whole-genome bisulfite sequencing. In total, 30 ng converted DNA was used for PCR amplification in 1× HiFi Hotstart Uracil+ Ready Mix (Kappa). Amplicons were cloned using the TOPO TA Cloning Kit (Invitrogen) following A-tailing by Klenow treatment. For each sample, 48 white colonies were individually picked, propagated and plasmid DNA extracted. Plasmid inserts were PCR-amplified and Sanger-sequenced (ABI 3730) using vector-specific primers. Sequencing was performed on 48 clones per sample, and reads from plasmids not including the amplicon insert are not shown. Sequences were base called in CONSED and methylation densities at each CHG site were calculated. Where possible, heterozygous non-cytosine SNPs were scored so that each allele could be analysed independently. In cases where a polymorphism changed a CHG site to either a CG or CHH site, the non-CHG variant was not included in calculations of CHG methylation. Because CHH methylation was

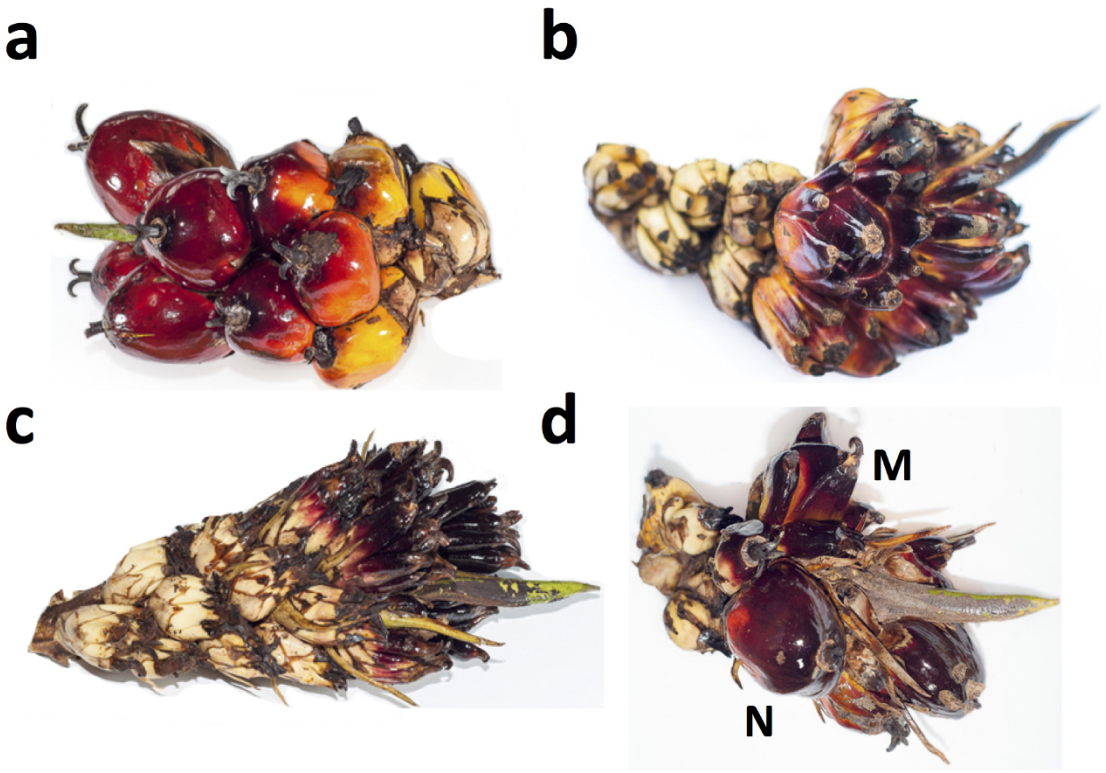
determined to be consistently very low in both normal and mantled ramets, conversion of CHH sites within the amplicon was used to control for bisulfite conversion rates. All samples analysed displayed <4% methylation of CHH sites, demonstrating that bisulfite conversion was >96% complete in all samples. Primer sequences are available on request.

mRNA and siRNA sequencing. Transcriptome sequencing was performed on shoot apex, early inflorescence (<2 cm) and late stage inflorescence (three normal and three mantled female inflorescence biological replicates each). About 2–3 μg total RNA was used to construct individually barcoded Illumina TruSeq stranded libraries. Libraries were pooled in sets of four, and each pool was sequenced in one lane of a HiSeq 2000 flow cell to generate paired 100-bp reads. sRNA fractions of female shoot apex tissue at stage 0 and female inflorescence tissue at stages 2, 3, 4 and 5 (7 mantled and 5 normal biological replicates at stage 0, 6 mantled and 8 normal biological replicates each at stages 2 and 3, 7 mantled and 5 normal biological replicates at stage 4, and 5 mantled and 4 normal biological replicates at stage 5), as well as second-passage tissue cultures recloned from mantled ($n = 1$) or normal ($n = 2$) ramets, were used to construct Illumina TruSeq sRNA libraries and sequenced following the same strategy as mRNA sequencing. mRNA sequencing data was used to construct gene models for all observed Eg*DEF1* alternative transcripts. siRNA reads mapping to the genomic scaffold including Eg*DEF1* were identified and normalized as fragments per 1,000 reads mapped to the genome (FPKM). FPKM values for each 24-mer were compared between biological replicates of normal and mantled samples by a two-tailed Student's *t*-test, assuming equal variance.

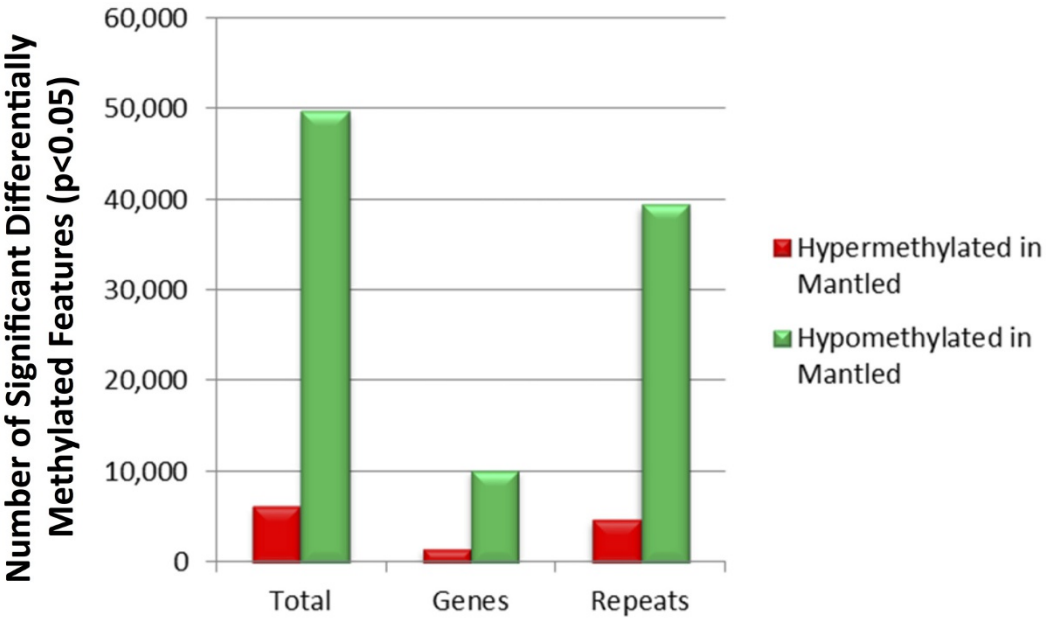
qRT–PCR assays. To specifically quantify *cDEF1* expression, a forward primer spanning the junction of exons 1 and 2 was used with a reverse primer within exon 7 (Extended Data Fig. 7b–d). The same forward primer was used with a reverse primer including intron 5 sequence to specifically quantify *tDEF1* expression. Finally, a forward primer spanning the junction of exons 4 and 5 was used with a reverse primer within *Karma*, downstream of the exon 5/*Karma* splice junction, to specifically quantify *kDEF1* expression. Assays were optimized using normal and mantled late stage inflorescence total RNA, and cDNAs were Sanger sequenced to confirm the identity of the amplicons. Standard curves generated from serially diluted cDNAs were generated for each primer pair, as well as for two internal oil palm reference gene assays[39] (Extended Data Fig. 7e, f). Gene expression was quantified in developing inflorescence stages 0, 2, 3, 4 and 5. All first-strand cDNA reverse transcription reactions were performed from 1 μg total RNA using a cocktail of reverse primers specific to Eg*DEF1* exons 6 and 7, as well as 3′ regions of *Karma*. For each stage, three technical replicates were performed for each of the three biological replicates per phenotype, per stage. qRT–PCR reactions were performed using 1 μl first-strand cDNA in 1× Roche SYBR Master Mix on a Roche LC480 instrument. Cycle thresholds above 33 cycles were not included in calculations, and detectable expression was calculated only for samples in which expression was detected in at least 2 of 3 technical replicates. Expression levels were quantified by extrapolation from the standard curve for each assay, and expression levels relative to the reference gene were calculated. Primer sequences are provided in Extended Data Fig. 7d.

We found only a subtle decrease in expression of *cDEF1* in female mantled relative to normal inflorescence at stage 3 (Fig. 4b). A more significant decrease was previously reported[8] but was not detected consistently[40]. More recently, an increase in the ratio of *tDEF1* to *tDEF1* + *cDEF1* in samples from early and late time points at each stage of inflorescence development was reported[12]. However, when absolute values of (tDEF1/tDEF1 + cDEF1) from early and late samples from any given stage are pooled (see figure 6b and Supplementary figure 8b in ref. 12) then only modest increases in expression are observed in mantled samples in agreement with our results (Fig. 4b).

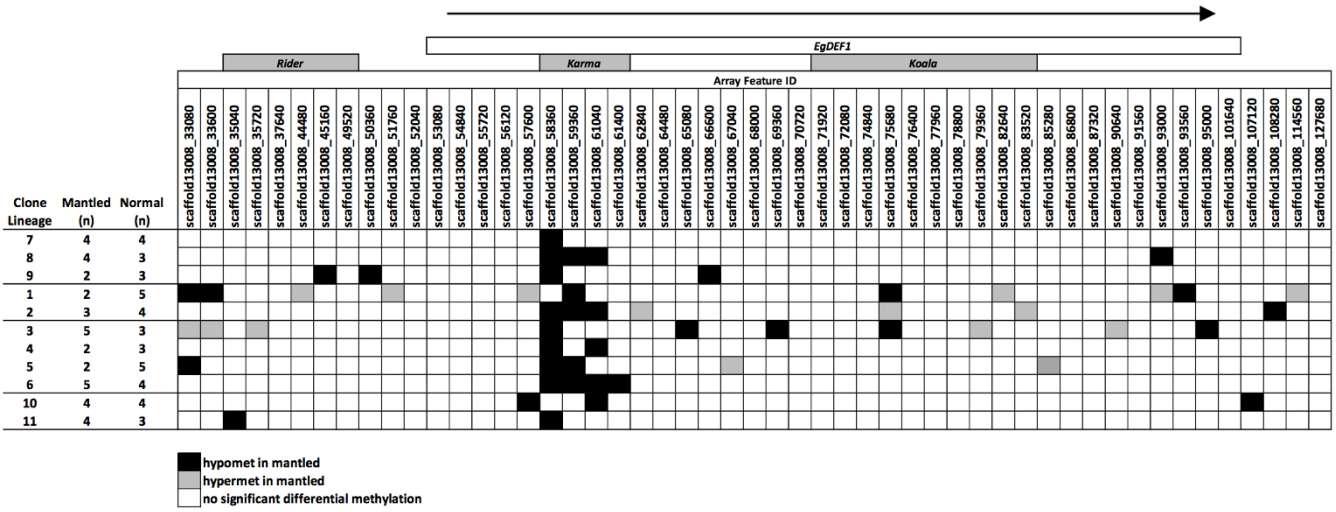
31. Jones, L. H. Propagation of clonal oil palms by tissue culture. *Planter* **50,** 374–381 (1974).
32. Rabechault, H. & Martin, J.-P. Multiplication vegetative du palmier a huile (*Elaeis guineensis* Jacq.) a l'aide de cultures de tissues foliares. *C. R. Acad. Sci. Paris. Ser. D* **283,** 1735–1737 (1976).
33. Durand-Gasselin, T., Le Guen, V., Konan, K. & Duval, Y. Oil palm (*Elaeis guineensis* Jacq.) plantations in Cote-d'Ivoire obtained through *in vitro* culture. First results. *Oléagineux* **45,** 10–11 (1990).
34. Kushairi, A. *et al.* Current status of oil palm tissue culture in Malaysia. In *Proceedings of the Clonal and Quality Replanting Material Workshop. Towards Increasing the Annual National Productivity by One Tonne FFB/ha/year* 3–14 (Malaysian Palm Oil Board, 2006).
35. Rohani, O. *et al.* in *Advances in Oil Palm Research* (eds Yusof, B., Jalani, B. S. & Chan, K. W.) 238–283 (Malaysian Palm Oil Board, 2000).
36. Adam, H. *et al.* Determination of flower structure in *Elaeis guineensis*: do palms use the same homeotic genes as other species? *Ann. Bot.* **100,** 1–12 (2007).
37. Stewart, F. J. & Raleigh, E. A. Dependence of McrBC cleavage on distance between recognition elements. *Biol. Chem.* **379,** 611–616 (1998).
38. Ordway, J. M. *et al.* Comprehensive DNA methylation profiling in a human cancer genome identifies novel epigenetic targets. *Carcinogenesis* **27,** 2409–2423 (2006).
39. Chan, P. L. *et al.* Evaluation of reference genes for quantitative real-time PCR in oil palm elite planting materials propagated by tissue culture. *PLoS ONE* **9,** e99774 (2014).
40. Jaligot, E. *et al.* Epigenetic imbalance and the floral developmental abnormality of the *in vitro*-regenerated oil palm *Elaeis guineensis*. *Ann. Bot.* **108,** 1453–1462 (2011).



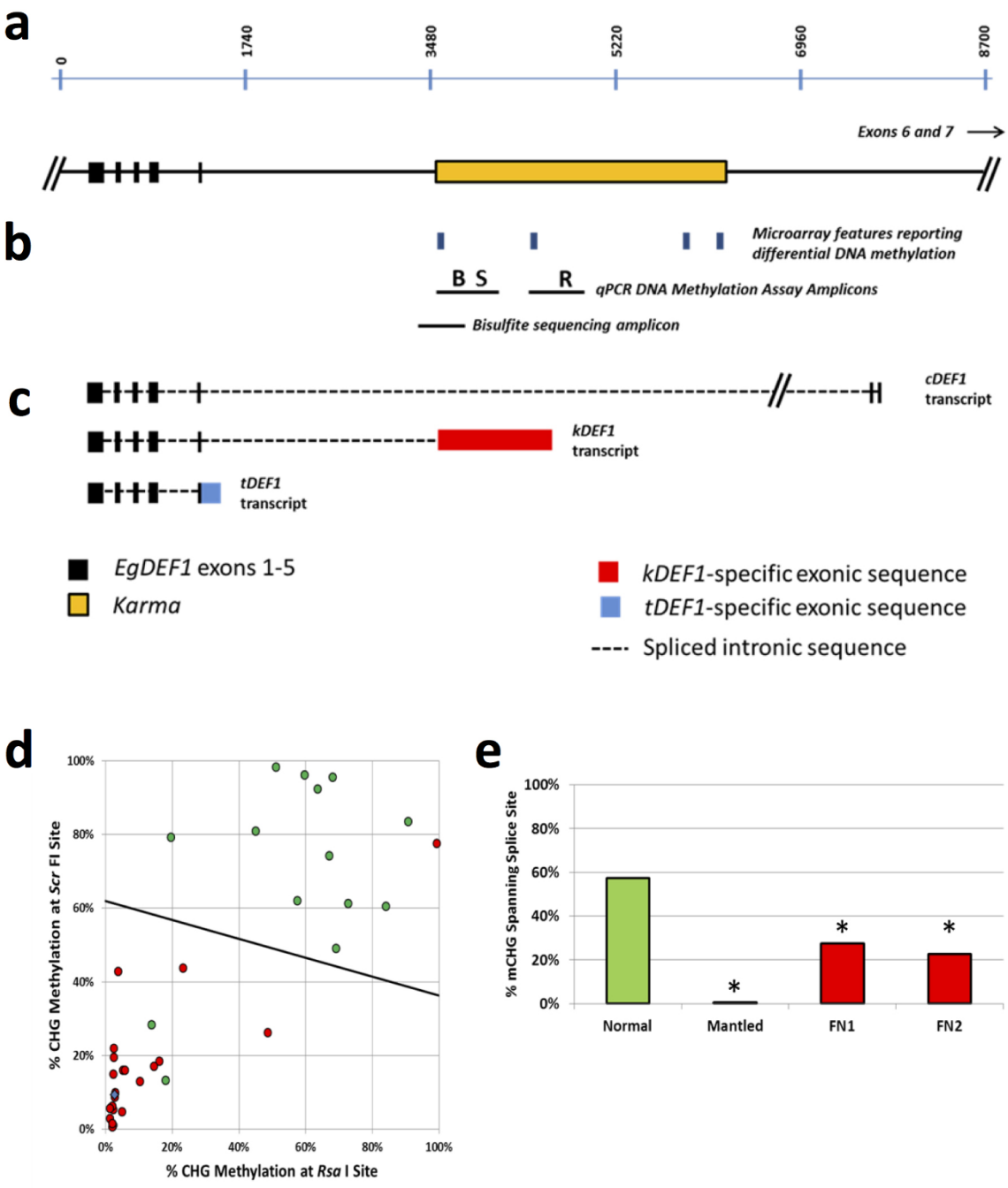
Extended Data Figure 1 | Spikelets from clonal palms of different fruit form phenotypes. **a**, Spikelets from a normal ramet. **b**, Spikelet from a fertile mantled ramet. **c**, Spikelet from a parthenocarpic mantled ramet. **d**, Spikelet from a revertant ramet displaying both normal (N) and mantled (M) fruits in the same spikelet.



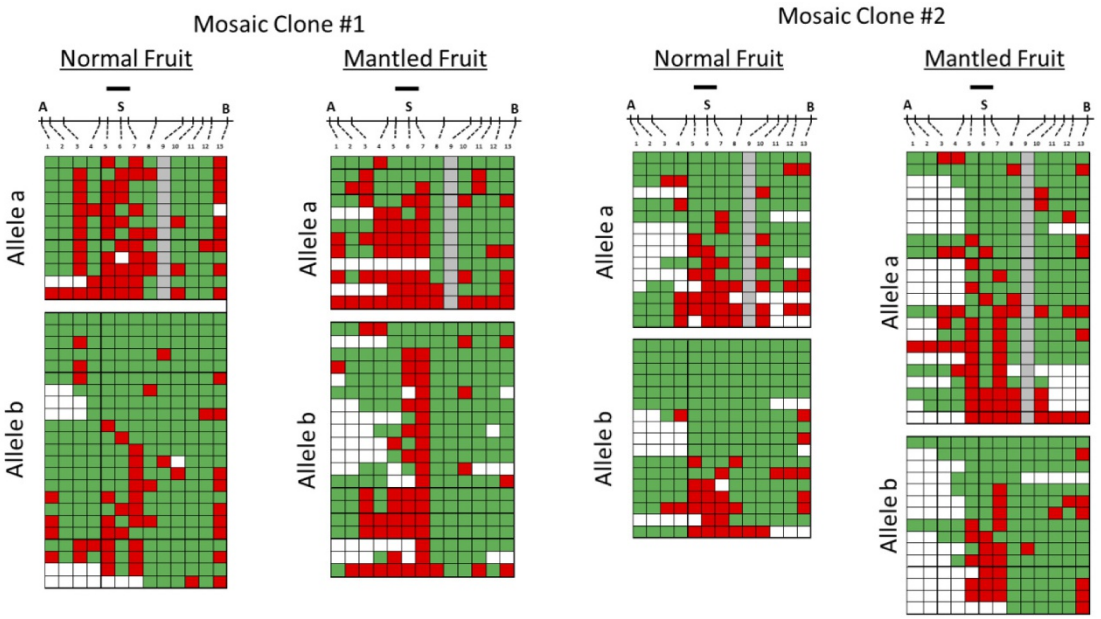
Extended Data Figure 2 | Annotation of genome-wide differentially methylated loci. Sequences of microarray features reporting significant differential DNA methylation between fully normal and fully mantled leaf DNA samples of one or more clonal lineages ($P < 0.05$, two-sided Student's *t*-test, Methods) were mapped to the reference *E. guineensis pisifera* genome[14]. Numbers of biological replicates per clonal lineage are provided in Extended Data Fig. 3. Features were assigned to gene and repeat classes according to annotations of genomic elements mapped within 3 kb of the microarray feature sequence, as this is the distance at which McrBC is capable of monitoring DNA methylation density. The repeat class includes all repetitive sequences, including transposons and *pisifera*-specific repetitive sequences[14]. Features mapping within 3 kb of both a gene and a repeat were assigned to both classes. The number of features reporting hypermethylation (red) and hypomethylation (green) are plotted.



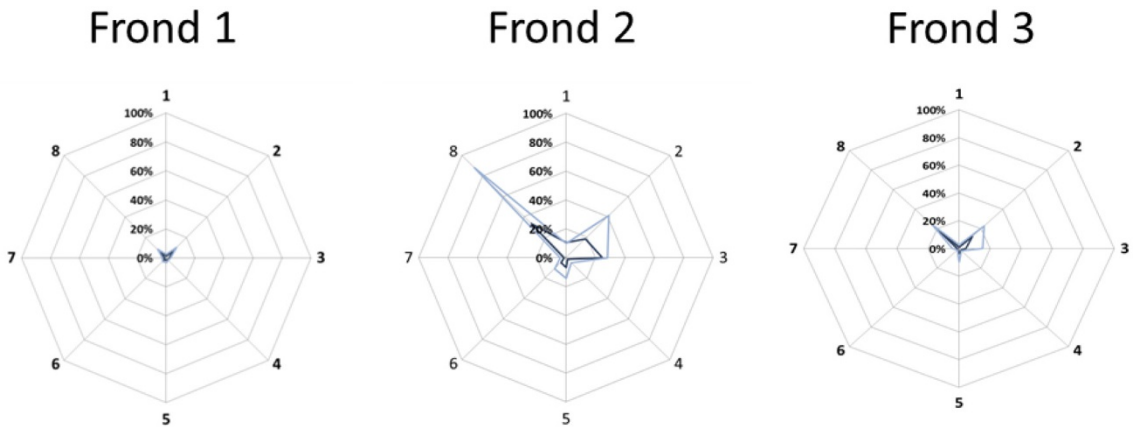
Extended Data Figure 3 | Summary of DNA methylation changes predicted by EWAS within clonal lineages. Rows indicate independent clonal lineages from four oil palm industry sources (source A–D, as indicated in Fig. 1e). Clone lineages map to sources as follows: 7–9, source A; 1 and 2, source B; 3–6, source C; 10 and 11, source D. The numbers of fully normal and fully mantled palms per lineage represented are indicated to the left. Columns represent each microarray feature mapping to the Eg*DEF1* (open box at top) and upstream region. The relative positions of *Rider*, *Karma* and *Koala* elements are indicated. The arrow indicates the direction of Eg*DEF1* transcription. Features reporting significant hypomethylation or hypermethylation in mantled relative to normal clones are indicated as black and grey boxes, respectively. White boxes indicate features reporting no significant DNA methylation difference. Only clonal lineages including more than 1 ramet per phenotype are shown to determine statistical significance within each clonal lineage ($n = 41$ normal; $n = 37$ parthenocarpic mantled). Ramets from four additional clonal lineages were included in the source-by-source analysis shown in Fig. 1e.



Extended Data Figure 4 | DNA methylation assays and supporting DNA methylation data. **a**, Diagram of the Eg*DEF1* gene including the *Karma* element within intron 5 (orange box). Black boxes represent exons and the horizontal line represents introns. Scale bar is in base pair units. **b**, Blue tick marks represent the relative positions of the four microarray features reporting significant hypomethylation of mantled clones in all source lineages. The left-most feature includes the *Karma* splice acceptor site. Horizontal lines labelled B (BbvI) and R (RsaI) indicate the relative positions of amplicons used for qPCR-based CHG methylation assays. The BbvI amplicon also includes a ScrFI site (S) used in **d**. The relative position of the bisulfite sequencing amplicon used to determine *Karma* splice site CHG methylation is shown below the qPCR amplicons. **c**, Diagrams of the three alternatively spliced Eg*DEF1* transcripts. Black boxes represent exons included in each transcript. The dotted lines represent intronic sequences spliced out of the mature mRNA transcripts. The red box represents *Karma* element sequence spliced to Eg*DEF1* exon 5 in the *kDEF1* transcript. The blue box represents Eg*DEF1* intron 5 sequence included in the *tDEF1* transcript that does not use the exon 5 splice donor site. **d**, In addition to adult leaf samples analysed by BbvI and RsaI qPCR assays (Fig. 2c), 37 samples were found to have a SNP in the BbvI site and were therefore analysed by ScrFI and RsaI qPCR assays (Methods and Extended Data Fig. 4b). Linear discriminant analysis was performed between normal ($n = 14$) and mantled ($n = 22$ parthenocarpic mantled; $n = 1$ fertile mantled) samples. Combining these results with those shown in Fig. 2c, sensitivity and specificity for detection of mantling are each 94%. **e**, Bisulfite sequencing of controls, FN1 and FN2 (Fig. 2c). mCHG density was calculated for the three CHG sites covered by the unique common microarray feature (Figs 1e and 2d–g). FN1, FN2 and the mantled control were significantly hypomethylated relative to the normal control (*$P < 0.0001$, two-tailed Fisher's exact test).



Extended Data Figure 5 | Clone based bisulfite sequencing maps of normal and mantled phenotype fruits from epigenetic mosaics. The heatmap format is as described in Fig. 2d–g. Grey boxes indicate a site in which a SNP on allele a results in a CHG to CHH site conversion. Mosaic clone 1 represents a revertant clone yielding 95% normal fruit. Mosaic clone 2 represents a revertant clone yielding 99% normal fruit. Alleles were analysed independently based on a SNP not affecting a potentially methylated base. Statistical analyses of methylation at the three CHG sites spanning the *Karma* splice site are shown in Fig. 3e, f.



Extended Data Figure 6 | CHG methylation in rachis sectors of an oil palm yielding 7% normal fruit (clone lineage 2 in Fig. 3d). Rachis of three successive fronds was dissected into 8 equal sectors. DNA methylation in each sector per frond was measured by BbvI and ScrFI assays, as described in Methods. Average DNA methylation density measurements of three technical replicates per frond, per sector, per assay are plotted on a radial graph representing the 8 rachis sections around the palm trunk (ScrFI assay, light blue; BbvI assay dark blue). Sector numbering was ratcheted for frond 2 versus 1, and frond 3 versus 2 based on the R^2 best fit of CHG methylation density around the circumference of the palm to correct for out-of-register numbering of rachis sectors between successive fronds (data not shown). Consistent with the fact that this oil palm yields only 7% normal phenotype fruit, most DNA methylation measurements are consistent with the mantled phenotype. However, sectors 8 and 2 display gains of CHG methylation in rachis sectors of all three fronds, and reach or approach normal levels in sectors 8 and 2 of frond 2, thus demonstrating mosaicism directly.

a

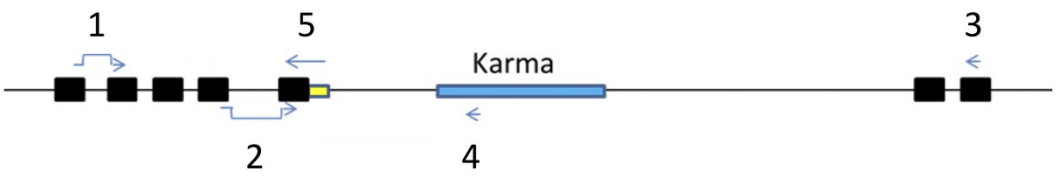
1 : MGRGKIEIKKIENPTNRQVTYSKRRTGIMKKAKELTVLCDAEVSLIMFSSTGKFSEYCSPLSDT : 64 kDEF1
1 : MGRGKIEIKKIENPTNRQVTYSKRRTGIMKKAKELTVLCDAEVSLIMFSSTGKFSEYCSPLSDT : 64 cDEF1

65 : KTIFDRYQQVSGINLWSAQYEKMQNTLNHLREINQNLRREIRQRMGEDLDSLGIHELRGLEQNL : 128
65 : KTIFDRYQQVSGINLWSAQYEKMQNTLNHLREINQNLRREIRQRMGEDLDSLGIHELRGLEQNL : 128

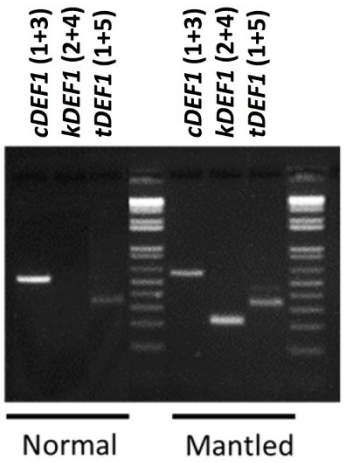
129 : DEALKVVRHRKYHVITTQTDTYKKKMHLKSALDHLLK : 156 +12
129 : DEALKVVRHRKYHVITTQTDTYKKKLKNSNEAHKNLLHELEMKDEHPVYGFVDDDPSNYAGALA : 192

193 : LANGASHMYAFRVQPSQPNLHRMGFGSHDLRLA* : 226

b



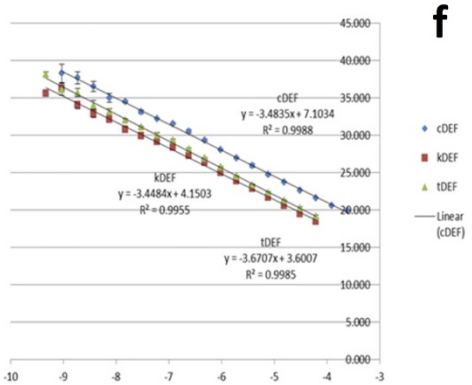
c



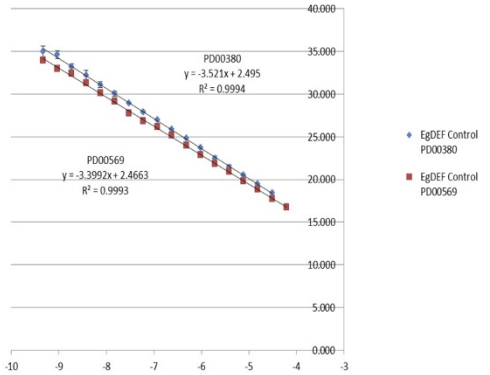
d

Primer	Sequence
1	CTTTCCGACACCAAGACCA
2	GTTCGTCACAGAAAATACCATGT
3	CAAGTAGCGGATAGAGAGGCTTAC
4	TCTTCTGATCGCCTTGCTAGA
5	ACCGGATCAAAGACCGTAAAG

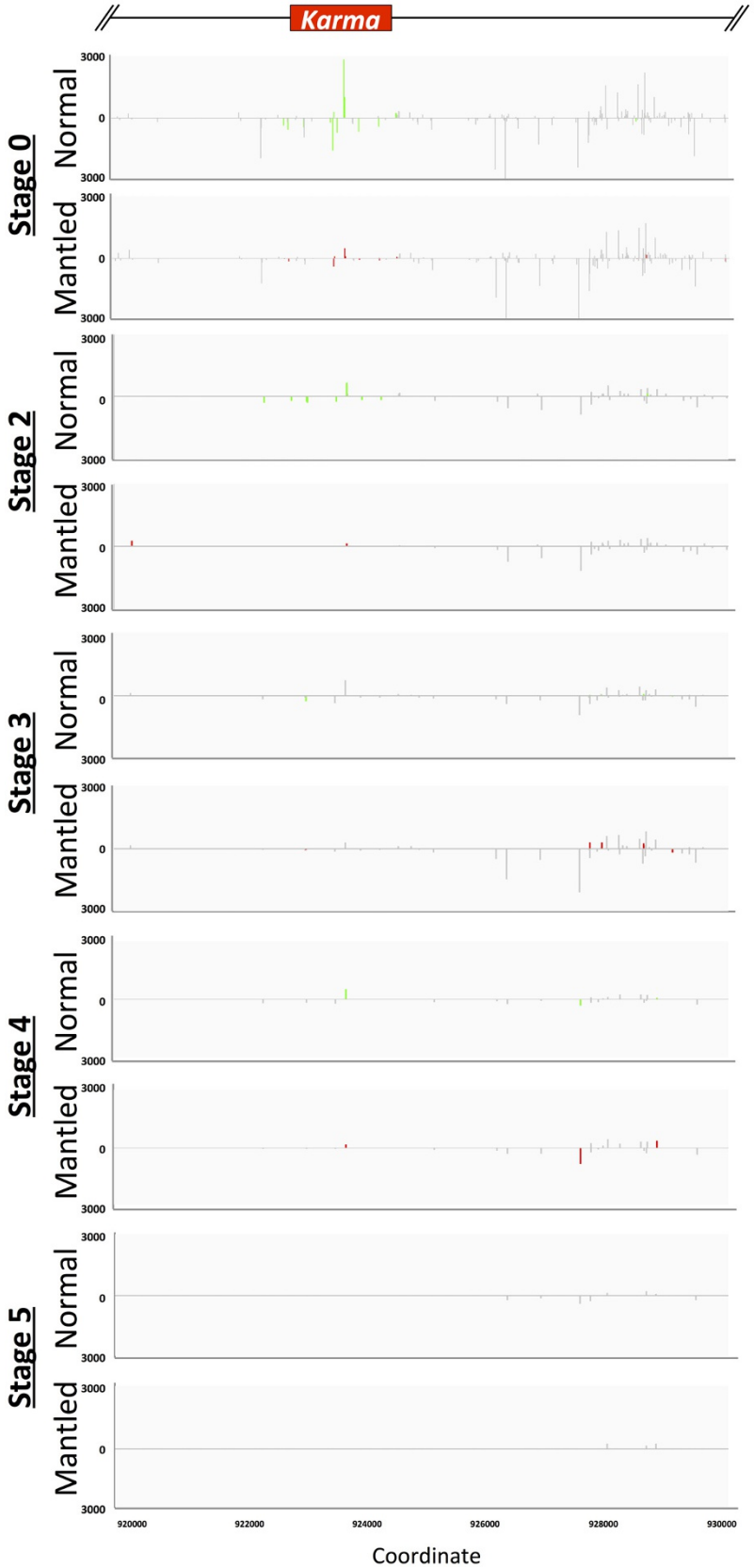
e



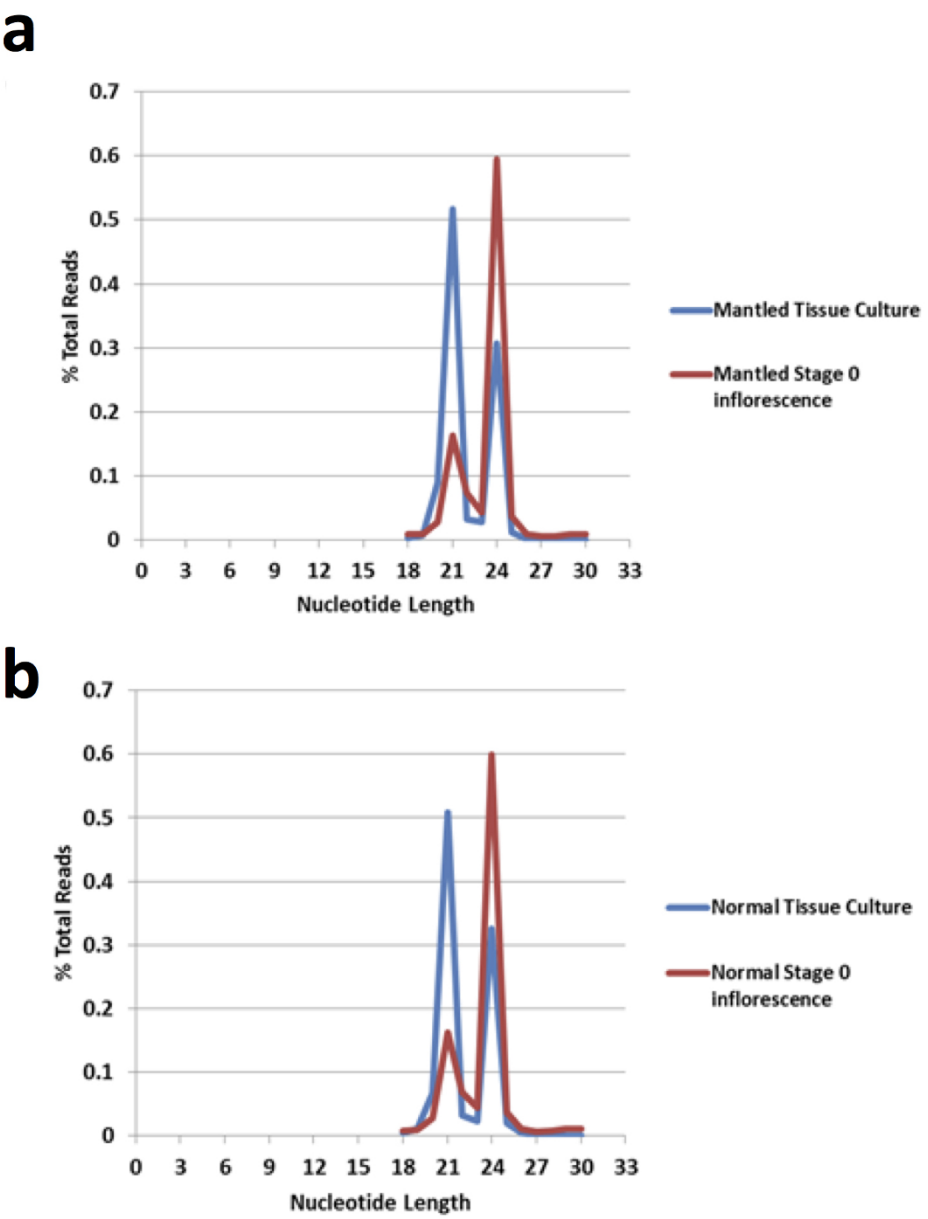
f



Extended Data Figure 7 | Protein sequences and summary of qRT–PCR assay designs. **a**, Residues highlighted in red are encoded by *Karma* sequence splice to exon 5 of Eg*DEF1*. The alternate splicing event disrupts the transcription activation domain of EgDEF1. Twelve variant amino acids are coded by *Karma* sequencing, followed by a stop codon. **b**, Diagram of Eg*DEF1* locus including positions of qRT–PCR primers. *cDEF1* transcripts were detected using primer a (spanning the splice junction of exons 1 and 2) and primer c (internal to exon 7). *kDEF1* transcripts were detected using primer b (spanning the splice junction of exons 4 and 5) and primer d (internal to *Karma* ORF2). *tDEF1* transcripts were detected using primer a and primer e (spanning the 3′ end of exon 5 and including *tDEF1*-specific intron 5 sequence). **c**, All assays were confirmed to give a single band of the correct size by agarose gel electrophoresis. Amplicons were Sanger sequence verified. Note that no band is amplified using the *kDEF1* primer pair in samples from normal inflorescence, consistent with lack of expression of *kDEF1* in normal inflorescence. **d**, Sequences of primers diagrammed in **b**. **e**, **f**, Standard curves for qRT–PCR assays. PCR amplicons including each qRT–PCR amplified sequence were serially diluted and quantified in triplicate by qPCR using the indicated primer pairs. Dilutions (*x* axis) were plotted against the measured cycle threshold (*y* axis). **e**, Standard curves for *cDEF1* (blue), *kDEF1* (red) and *tDEF1* (green). Line equations were used to calculate the efficiency of each primer pair. The efficiency of each primer pair was used in calculations for quantification of expression of each associated transcript. **f**, Standard curves for two endogenous oil palm control genes. The efficiency of each primer pair was used in calculations for quantification of expression of each associated transcript. Expression of each alternative transcript was calculated relative to the control PD00569 control. Control qRT–PCR primers are described previously[39].



Extended Data Figure 8 | Antisense 24-nucleotide siRNA analysis of inflorescence development. siRNA expression at inflorescence stages 0 (shoot apical meristem), 2, 3, 4 and 5, was analysed by Illumina siRNA sequencing (Methods). FPKM normalized expression values for each measured 24-nucleotide siRNA are plotted in scale with the genomic elements diagrammed at the top of the figure. Grey bars indicate detected 24-nucleotide siRNAs that are not significantly differentially expressed between normal relative to mantled tissues ($P > 0.05$, Student's *t*-test, two-tailed assuming equal variance). Differentially expressed 24-nucleotide siRNAs are plotted as green or red bars for normal or mantled tissues, respectively. Bars above and below the zero line represent sense and antisense siRNAs, respectively, and are plotted on the same scale in both directions.



Extended Data Figure 9 | Relative abundance of 21- and 24-nucleotide sRNA in normal and mantled reclones and stage 0 inflorescence. **a**, Distribution of sRNA lengths derived from mantled reclone (blue) and stage 0 inflorescence (red). **b**, Distribution of sRNA lengths derived from normal reclone (blue) and stage 0 inflorescence (red). Read lengths of sRNA sequencing reads are plotted as the percentage of total reads for each incremental sRNA nucleotide length.

Lineage	Ramet Phenotype	Recloned Culture	BbvI	RsaI
1	Normal	SC2	33%	95%
1	Normal	SC7	18%	87%
2	Normal	SC2	76%	68%
2	Normal	SC7	34%	26%
1	Mantled	SC2	24%	22%
1	Mantled	SC7	4%	5%
2	Mantled	SC2	2%	3%
2	Mantled	SC7	1%	2%

Extended Data Figure 10 | CHG methylation in recloned tissue cultures. Tissue cultures were reconstituted from normal and mantled ramets from two clonal lineages ('clones of clones'). Methylation at three CHG sites across the *Karma* DMR was quantified by qPCR assays at two (SC2) and seven (SC7) passages in tissue culture. Cultures derived from normal ramets displayed higher CHG methylation than those derived from mantled ramets. In both normal and mantled reclones, CHG methylation generally decreased with time in culture. At SC2, the time point at which 24-nucleotide siRNAs were measured (Fig. 4c), the culture from normal ramet lineage 1 had lost methylation at the BbvI (the site nearest the *Karma* splice acceptor site).