The oil palm VIRESCENS gene controls fruit colour and encodes a R2R3-MYB

Rajinder Singh1, Eng-Ti Leslie Low1, Leslie Cheng-Li Ooi1, Meilina Ong-Abdullah1, Rajanaidu Nookiah1, Ngoot-Chin Ting1, Marhalil Marjuni1, Pek-Lan Chan1, Maizura Ithnin1, Mohd Arif Abdul Manaf1, Jayanthi Nagappan1, Kuang-Lim Chan1, Rozana Rosli1, Mohd Amin Halim1, Norazah Azizi1, Muhammad A. Budiman2, Nathan Lakey2, Blaire Bacher2, Andrew Van Brunt2, Chunyan Wang2, Michael Hogan2, Dong He2, Jill D. MacDonald2, Steven W. Smith2, Jared M. Ordway2, Robert A. Martienssen3 & Ravigadevi Sambanthamurthi1

Oil palm, a plantation crop of major economic importance in Southeast Asia, is the predominant source of edible oil worldwide. We report the identification of the VIRESCENS (VIR) gene, which controls fruit exocarp colour and is an indicator of ripeness. VIR is a R2R3-MYB transcription factor with homology to Lilium LhMYB12 and similarity to Arabidopsis PRODUCTION OF ANTHOCYANIN PIGMENT1 (PAP1). We identify five independent mutant alleles of VIR in over 400 accessions from sub-Saharan Africa that account for the dominant-negative virescens phenotype. Each mutation results in premature termination of the carboxy-terminal domain of VIR, resembling McClintock’s C1-I allele in maize. The abundance of alleles likely reflects cultural practices, by which fruits were venerated for magical and medicinal properties. The identification of VIR will allow selection of the trait at the seed or early-nursery stage, 3-6 years before fruits are produced, greatly advancing introgression into elite breeding material.
Commerically grown oil palm (Elaeis guineensis) is an outbreeding diploid species (2n = 32) of West African origin1-3. We recently reported the genome sequences of E. guineensis and the South American oil palm, E. oleifera4, as well as the discovery of the oil palm SHELL gene, a homologue of SEEDSTICK (STK), responsible for oil palm fruit forms5. We next sought to identify the genetic basis of oil palm fruit colour.

Fruit colour is an important trait in terms of fruit harvesting and, therefore, oil yield. The majority of oil palms produce either nigrescens or virescens fruit type2. Nigrescens fruits are usually deep violet to black at the apex and yellow at the base when unripe, with minimal change in colour of the apex upon ripening (Fig. 1a,c). Virescens fruits are green when unripe, and change to orange when the bunch matures (Fig. 1b,c), reflecting degradation of chlorophyll and accumulation of carotenoids5. For nigrescens palms, harvesters rely on the presence of detached fruits on the ground to determine that bunches are ripe. However, as virescens fruits undergo a more profound colour change upon ripening, it is easier to identify ripe bunches, particularly in tall palms where they can be obscured by fronds, thus minimizing yield loss due to fallen fruits or harvesting of unripe bunches. Both nigrescens and virescens palms occur in natural groves. Although the virescens trait is dominant, the number of virescens palms found in natural populations is small, with frequencies ranging from below 1% in Nigeria and Angola2 to up to 50% in one location in Congo6. Virescens palms were used in ancient ceremonial rites7, explaining their occurrence among wild-type nigrescens palms, and ‘Ojuku’ trees matching the description of virescens palms were reportedly used in tribal sacrificial ceremonies in West Africa8.9.

Here, we identify the oil palm VIRESCENS gene and five independent, but remarkably similar mutant alleles of VIR. Phylogenetic analyses and transcriptome studies of virescens and nigrescens fruit suggest that VIR controls oil palm fruit exocarp pigmentation by coordinately regulating expression of genes involved in the anthocyanin biosynthetic pathway. The discovery of alleles responsible for the virescens phenotype, segregating within commercially relevant germplasm collections, has direct applications to the breeding and production practices of the predominant source of edible oil worldwide.

**Results**

**Genetic mapping of the VIR locus.** Oil palm is an outbreeding species, and as such, a high degree of heterozygosity is expected. A population of 240 palms derived from the self-pollination of the tenera palm, T128 (0.151/128 × 0.151/128), from Malaysian Palm Oil Board’s (MPOB) Nigerian germplasm collection10–12 was used to generate a genetic linkage map3,4. In addition, a subset of 81 palms from six independent crosses (Supplementary Table 1) was used to confirm marker linkage (Methods). Markers were scored as co-dominant, segregating in a 1:2:1 ratio in most cases, while the virescens phenotype also showed the expected 3:1 segregation ratio in the mapping population (Supplementary Table 2, Methods). Three informative restriction-fragment length polymorphism (RFLP) markers were genotyped on the entire mapping population, and 197 SSR loci that were polymorphic in the mapping population were identified. Of 4,451 single-nucleotide polymorphisms (SNPs) screened, 711 were used in map construction. The locus for the virescens gene (VIR) was located on linkage group 1 (chromosome 1), with the RFLP marker MET16 being the most tightly linked (Supplementary Tables 3 and 4; Supplementary Fig. 1). Linkage of MET16 to the virescens trait was further tested in the 81 trees, resulting in 95% accuracy for distinguishing between nigrescens and virescens fruit traits (Supplementary Table 5).

Markers flanking the VIR candidate locus were mapped by sequence similarity to the E. guineensis (pisifera) reference genome assembly3 and localized to assembly scaffold 7 (p3-sc00007). A tiling path of bacterial artificial chromosome contigs corresponding to scaffold 7 was selected from a high-information content physical map of pisifera and sequenced. Additional SNP assays were designed from an improved assembly corresponding to scaffold 7 and genotyped (Methods). Markers mapping close to the VIR locus were identified (Supplementary Fig. 2) and markers SNP02708 and SNP02400 were positioned on each side of the VIR locus. The interval contained four potential candidate genes that impact fruit pigmentation in other species: a gene with homology to Lilium (lily) LhMYB12 and significant similarity to both Arabidopsis PRODUCTION OF ANTHOCYANIN PIGMENT 1 (PAP1) and AtMYB113, and three genes with significant similarity to Arabidopsis TRANSPARENT TESTA 12 (TT12), PURPLE ACID PHOSPHATASE 18 (PAP18) or the BHLH gene, ILR3.

**VIR mutations responsible for virescens fruit.** To extend beyond the E. guineensis reference genome sequence, we queried genome sequence assemblies of 12 independent T128 progeny palms (5 nigrescens and 7 virescens) derived from 20-fold raw sequence coverage (HISEQ 2000) per genome (Methods). Contigs from each assembly were mapped to the scaffolds that had been linked to genetic markers in the virescens genetic interval. In addition, the candidate genes above were each amplified by PCR, including exons and introns, and sequenced (Supplementary Note 1). The entire open-reading frame of the gene homologous to Lilium LhMYB12 and similar to Arabidopsis PAP1 and AtMYB113 was intact in all five nigrescens palms. However, all seven virescens palms were either heterozygous (n = 4) or homozygous (n = 3) for an A-to-T nonsense mutation in exon 3 of the identified
The exon 3 mutation results in a predicted truncation of the 21 carboxy-terminal amino acids within the transcriptional activation domain of the R2R3-MYB transcription factor (Fig. 2). Subsequently, the entirety of the gene was amplified and sequenced in 208 trees from the T128 cross (48 nigrescens and 160 virescens). In all, 158 trees were either heterozygous ($n = 99$) or homozygous ($n = 59$) for the nonsense mutation in exon 3, and 50 trees were homozygous wild type, for an overall concordance of this nonsense mutation (event 1) with fruit colour phenotype of 99% (Table 1). It is noted that a 1% discordance rate is well within the norms of phenotyping accuracy of breeding populations.

Although SNPs were identified in the other three candidate genes, the polymorphisms observed were not consistent with a functional mechanism affecting fruit colour phenotype of the 12 trees, and independent mutant alleles (see below) were not identified (Supplementary Note 1).

To further support the discovery of the VIR gene, we sequenced the entire gene in six independent breeding populations, as well as samples from germplasm collections (Table 1).

**Table 1 | Summary of VIR genotypes.**

| Phenotype | Genotype | Genotype/phenotype concordance $^*$ |
|-----------|----------|-------------------------------------|
| Mapping population T128 | | 99.0% |
| Breeding populations | | |
| DT35 | 11 11 22 | 100.0% |
| DT38 | 8 12 20 | 100.0% |
| DP454 | 9 12 21 | 100.0% |
| TT108 | 6 6 12 | 100.0% |
| AVROS | 43 — 43 | 100.0% |
| MPO8 PK75 | 10 11 21 | 100.0% |
| Total | 87 52 139 | 100.0% |
| Germplasm collections | | |
| Angola | 261 48 309 | 99.7% |
| Madagascar | 27 — 27 | 100.0% |
| Tanzania | 47 12 59 | 96.6% |
| Ghana | 8 15 23 | 100.0% |
| Congo | 3 7 10 | 90.0% |
| Cameroon | 5 3 8 | 100.0% |
| Nigeria | 2 2 4 | 100.0% |
| Total | 353 87 440 | 99.2% |

$^*$Genotype/phenotype concordance calculated as ((number of virescens-phenotyped trees genotyped as either heterozygous or homozygous for events 1, 2, 3, 4 or 5) + (number of nigrescens phenotyped trees genotyped as wild type)) divided by the total number of trees sequenced.

$^*$Nigrescens fruit exocarp colour phenotype.

$^*$Virescens fruit exocarp colour phenotype.

$^*$Wild-type (espresss) genotype.
The breeding populations included 139 trees, where the fruit colour phenotype was known (DT35, DT38, DP454, TT108, MP0B PK575 and a collection of palms from the AVROS background). In addition, 440 trees from Angola, Madagascar, Tanzania, Ghana, Congo, Cameroon and Nigeria were analysed. In the breeding populations, all 52 virens trees, but none of the 87 nigrescens trees were found to be either heterozygous or homozygous for the event 1 nonsense mutation in exon 3 (Table 1). However, among the germplasm collections, the event 1 mutation was detected in only 5 of 87 virens trees, all of which were from either the Ghana or Nigeria collections. Instead, four independent, but closely related mutations were identified in the other germplasm collections from sub-Saharan Africa. First, a G-to-T nonsense mutation (event 2) was detected in exon 3, 30 base pairs (bp) 5' to event 1 (Fig. 2; Supplementary Figs 3 and 4). This mutation results in a predicted truncation of the 31 carboxy-terminal amino acids within the transcriptional activation domain. Event 2 was heterozygous or homozygous in 68 trees from the Angola (n = 45), Tanzania (n = 14), Ghana (n = 4) or Congo (n = 5) collections (Table 1; Fig. 3). Next, a G-to-A nonsense mutation (event 3) was detected in exon 3, 113 bp 5’ to event 1 (Fig. 2; Supplementary Figs 3 and 4). This mutation results in a predicted truncation of the 59 carboxy-terminal amino acids. The event 3 mutation was heterozygous in 10 trees from Angola (n = 1), Ghana (n = 8) or Congo (n = 1) (Table 1; Fig. 3). A fourth mutation (event 4) is a 2-bp deletion beginning 11 bp 3’ to event 3, resulting in translation frameshift at the 55th carboxy-terminal amino acid (Fig. 2; Supplementary Figs 3 and 4), and was heterozygous in three trees from Angola and Congo (Table 1; Fig. 3). Finally, a heterozygous rearrangement (event 5) resulting in a translational frameshift and premature truncation was detected in three of three virens trees from Cameroon (Table 1; Figs 2 and 3; Supplementary Figs 3 and 4). The mutation is a 195-bp deletion with a 21-bp duplication, which results in the truncation of 75 carboxy-terminal amino acids and a single amino-acid conversion before reading a new stop codon. Considering all five single-gene mutations, the concordance between genotype and fruit colour is 99.2% (Table 1). The identification of five independent genetic mutations, each resulting in remarkably similar premature truncation, provides strong evidence for the identification of the VIR gene. C-terminal truncations of related genes in the R2R3-MYB family, most notably the maize C1 gene, have similarly dominant-negative allelic forms. Furthermore, sequence similarity searches (BLAST) of the genome of the South American oil palm, E. oleifera, which does not produce the deep-violet coloured fruits similar to wild-type E. guineensis, do not identify an intact VIR gene.

Phylogeny, expression and function of VIR. The R2R3-MYB family includes >100 genes in Arabidopsis14,15 and >80 genes in maize16. The family includes two sets of imperfect repeats (R2 and R3), each including three alpha-helices forming a helix-turn-helix motif17. The R2R3 proteins are members of regulatory networks controlling development, metabolism and responses to biotic and abiotic stresses18. Phylogenetic analysis of the R2R3-MYB domain of VIR relative to MYB family members from various plant species indicates that VIR is most closely related to monocot Lilium LhMYB12 (Fig. 4; Supplementary Figs 5 and 6). Although oil palm and Lilium are monocots, VIR and LhMYB12 cluster together within a distinct subgroup that is more similar to dicot cacao TcMYB113 and Arabidopsis PAP1, PAP2 and AtMYB113 than to monocot maize and rice C1. This classification is consistent with previous phylogenetic comparisons of LhMYB12, which place this MYB family protein in a subgroup with dicot MYB proteins including Arabidopsis PAP1 and PAP2, apple MYB10 and petunia AN2 and separate from a subgroup including monocot maize C1 (ref. 19). LhMYB12, PAP1 and AtMYB113 control accumulation of anthocyanins by regulation of expression of biosynthetic genes20-22. Expression levels of LhMYB12 are positively correlated with tepal anthocyanin pigmentation in Asiatic hybrid lilies19. Cacao TcMYB113 was recently identified as a likely candidate for regulation of green/red pod colour23. Overexpression of Arabidopsis PAP1 results in intense purple pigmentation in many vegetative organs throughout development, and ectopic expression in tobacco results in purple-pigmented plants24. Overexpression of AtMYB113 in Arabidopsis results in elevated pigment production, and downregulation of AtMYB113, AtMYB114, PAP1 and PAP2 results in anthocyanin deficiency22. Furthermore, overexpression of Arabidopsis Myb114 lacking the transactivation domain results in dominant anthocyanin deficiency22. The phylogenetic placement of VIR and LhMYB12 within a clade including mostly dicot MYB family proteins suggests that these MYB proteins represent a class of pigment-related regulators for which there are no extant orthologues in model monocots such as corn and rice. Although VIR is a member of a different clade of MYB proteins than maize C1, all five VIR mutations are intriguingly similar to McClintock’s maize C1-I allele in which a frameshift mutation in the carboxy-terminal region of C1 generates a dominant-negative protein resulting in reduced pigmentation24. Further, the last 20 amino acids of the oil palm protein are conserved in Lilium, but deleted from all 5 dominant alleles of VIR. These 20 amino acids share similarity with the C-terminal domain of C1 (Supplementary Fig. 6). These findings suggest that similar C-terminal truncation mechanisms result in anthocyanin deficiencies in oil palm.

In order to examine anthocyanin deficiency in virens fruits, we performed a combination of metabolic and gene expression analyses. Spectrophotometric and chromatographic (high-performance liquid chromatography (HPLC)) analyses of acidified
methanol extracts of exocarp confirmed the presence of anthocyanins in nigrescens, but absence in virescens fruit (Fig. 5). Gene expression in nigrescens and virescens whole fruits at 8 weeks after anthesis (WAA) was analysed by transcriptome sequencing (Fig. 6; Supplementary Table 6 and Methods). The oil palm fruit typically exhibits biphasic growth with an initial growth spurt between 4 and 9 WAA. Further, significant biochemical changes are observed starting at 8 WAA and up to 10 WAA during the transition phase between a metabolic sink and a storage sink. Therefore, 8 WAA was chosen to examine expression of anthocyanin biosynthetic genes, avoiding later stages when expression of other mesocarp genes occurs that share the phenylpropanoid pathway, such as those involved in polyphenol biosynthesis. Transcriptome reads were annotated based on sequence comparisons with the rice proteome where possible. Transcriptome reads with substantial sequence similarity to biosynthetic genes in the anthocyanin phenylpropanoid pathway were identified (Supplementary Table 7). Gene annotations are based solely on cross-species sequence comparisons and represent putative orthologues of anthocyanin pathway genes. Arabidopsis flavonoid enzymes can be divided into ‘early’ and ‘late’ groups that regulate distinct temporal stages of the pathway. Late genes initiate at the dihydroflavonol reductase step, with downstream genes being regulated by Myb/bHLH/WD-repeat proteins. However, in maize there is no early/late split. At 8 WAA, nigrescens fruits display higher expression of VIR, as well as anthocyanin pathway genes starting at the trans-cinnamate 4-monoxygenase (C4H) step and

Figure 4 | Phylogenetic analysis of VIR and various R2R3-MYB family members. Node numbers represent percentage of bootstrap replicates containing each clade. Placement of the Elaesis guineensis VIR gene is designated by the arrow. Scale represents number of amino-acid changes per position within the alignment.
Anthocyanins are known to absorb strongly around this wavelength. This peak was not observed in conditions. The was brilliant red in acidic conditions and turned green under alkaline conditions. The extract, however, was light orange and did not absorb at about 520 nm. This peak was not observed in virescens. Anthocyanins are known to absorb strongly around this wavelength. (Fig. 6; Supplementary Fig. 7; Supplementary Table 7). These results suggest that the truncating mutations result in coordinated dominant inhibition of MYB-regulated target gene expression at all steps of the anthocyanin pathway. Transcripts from four independent genes, two with homology to C4H and two with homology to C2, were coordinately regulated. *Annotation of transcripts by comparison with the rice proteome did not reveal transcripts annotated as dihydroflavonol reductase (DFR). However, a differentially expressed transcript has closest homology to the grape DFR gene and is highly homologous to Brachypodium and Medicago DFR.

**Discussion**

Our findings establish that the oil palm VIR gene controls fruit colour and that any one of five independent, but closely related, dominant mutations in the gene can cause the virescens fruit colour phenotype. Further demonstration of the effect of the VIR truncation mutations by transgenic approaches in model organisms, as well as the possible contributions of additional unidentified genetic variants to fruit colour phenotype are areas for future research. However, the discovery of the genetic basis of the role of VIR in the virescens phenotype paves the way for development of genetic testing for fruit colour well before planting and for the introgression of the desirable trait into elite breeding materials. For example, the identification of the VIR gene allows differentiation of the homozygous and heterozygous forms of virescens palms, as early as the seedling stage, and together with the recent identification of SHELL, allows breeders to develop paternal (pisifera) lines that are homozygous for virescens for use in breeding programmes or for commercial seed production. All five alleles of VIR from equatorial Africa have mutations resulting in premature C-terminal truncations of the VIR protein, and their prevalence is unprecedented. This likely reflects dominant-negative inheritance (which makes novel alleles conspicuous) and cultural practices that retain the alleles for ritual purposes. The utility of these alleles will have important impacts on fruit harvesting practices, to improve oil yields and lead to improved land utilization.

**Methods**

**Plant materials and germplasm collection.** The mapping family used was derived from the self-pollination of a high-iodine value virescens tenera palm T128 (accession number MPOB 371), which has been described in detail. An additional 108 palms derived from six families of different genetic backgrounds...
Genetic mapping. A total of 240 palms of the mapping family were available for DNA extraction at the start of this study. Of these, 32 palms could not be phenotypically confirmed, as the palms had been proven to be ’virescens’ and nigrescens exocarp before the fruit excocarp colour could be determined or re-confirmed. Of the 208 palms that were successfully phenotyped, 160 were identified as *virescens* palms and 48 as *nigrescens* palms. However, all 240 available palms were genotyped by 4,451 SNP markers using the Illumina Select assay (Illumina), 3 RFLP and 197 SSR markers. The genotype data were formatted for mapping according to an F2 population. Markers showing segregation profile of 1:2:1 were used in the map construction. Two sets of genotype data were created, in which one was the converse of the other to account for phase differences in the T128 ’selfed’ F2 population. Genotyping was then conducted using JoinMap 4.0. Markers that exhibited severe distortion (P < 0.0001) and markers having >10% missing data were excluded. Both sets of genotype data were grouped at a recombination frequency of <0.2. Markers exhibiting nearest neighbour stress value >2 (CM) were identified and excluded from the analysis. Markers contributing to insufficient linkages were also removed. The T128 co-dominant map constructed comprised 16 groups, and VIR was placed on linkage group 1.

**Fruit colour phenotyping.** The fruit excocarp colour was determined on ripe bunches having at least one loose fruit per bunch (irrespective of plant height). The bunch was harvested from the tree and a minimum of five fruitlets was stripped from the bunch. Visual observation was made of the excocarp, and fruits were classified as *nigrescens* (reddish to deep violet) or *virescens* (orange) as seen from the apex. In this study, at least two independent attempts were made to determine fruit colour of the mapping family as well as the breeding populations. With respect to the germplasm collection, fruit colour observations were made only once.

**Genome and transcriptome sequencing.** Twelve independent T128 progeny palms (five *nigrescens* and seven *virescens*) were sequenced to ~20X raw sequence coverage by HiSeq 2000 (Illumina). For transcriptome sequencing, RNA was extracted from 10 to 20 fruits from 2 trees (1 *nigrescens* and 1 *virescens* per tree) and sequenced at 8X AWA. Three replicate RNA extractions were performed for each fruit pool. TrueSeq (Illumina) libraries were constructed and sequenced by HiSeq 2000, generating 1/8 lane of reads per phenotype replicate.

**VIR Sanger sequencing.** The entirety of the VIR gene was amplified by PCR from oil palm genomic DNA using a forward primer sequence, 5’-CGGATCCTCGGTTGACACCACAACCA-3’, and reverse primer sequence, 5’-CTCCATCCCTGAGGCAAGAG GT-3’, generating a single ~2.9 kb amplicon. Forward and reverse primers included M13 Forward or M13 Reverse sequence tags, respectively. Amplicons were treated with exonuclease 1 (New England Biolabs) and shrimp alkaline phosphatase (Affymetrix) under standard conditions. Amplicons were sequenced using a combination of M13 primers and internal primers (internal primer sequences available upon request). Sequencing was performed on an ABI 3730 capillary DNA sequencer using big dye terminator VS 3.1 chemistry (Life Technologies). Local assemblies of each amplicon were constructed with PHRAP and reviewed in CONSED. Consensus sequence for each palm was aligned to the reference *pisifera* genome sequence. Data were analysed to determine the integrity of the coding sequence and resulting putative translated polypeptide for each palm. A large percentage of the palms analysed were part of the 110,000 diverse germplasm collection available at MPOB.

**Phylogenetic analysis.** A collection of R2R3 MYBs from previously studied plant species were selected based on their similarity to the VIR protein. These sequences were aligned using the ClustalX program, and the highly conserved R2R3 domains were then processed using the Neighbor joining method with 1,000 bootstrap replicates.

**Pigment extraction.** Acidified methanol (1% HCl, v/v) was added to ground excocarp slices of *E. guineensis* (15WVAA *nigrescens* and *virescens* fruits) and stirred to ensure efficient extraction of pigments. The extracts were centrifuged at 3,000 g in an Eppendorf 5810R centrifuge to remove debris. The supernatants were removed and filtered before further analysis. Spectrophotometric and chromatographic analyses were carried out to determine the presence, if any, of anthocyanins. Equations for the oil palm *nigrescens* and *nigrescens* exocarp materials were used for extraction under identical conditions.
22. Gonzales, A. et al. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in Arabidopsis seedlings. *Plant J.* 53, 814–827 (2008).
23. Motamayor, J. C. et al. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biol.* 14, r53 (2013).
24. Goff, S. A., Cone, K. C. & Fromm, M. E. Identification of functional domains in the maize transcriptional activator C1: comparison of wild-type and dominant inhibitor phenotypes. *Genes Dev.* 5, 298–309 (1991).
25. Kok, S. Y. et al. Biochemical characterization during seed development of oil palm (Elaeis guineensis). *J. Plant Res.* 126, 539–547 (2013).
26. Pelletier, M. K., Murrell, J. R. & Shirley, B. W. Characterization of flavonol synthase and leucoanthocyanidin dioxygenase genes in *Arabidopsis*. Further evidence for differential regulation of ‘early’ and ‘late’ genes. *Plant Physiol.* 113, 1437–1445 (1997).
27. Petroni, K. & Tonelli, C. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci.* 181, 219–229 (2011).
28. Rahimah, A. B., Cheah, S. C. & Rajinder, S. Freeze-drying of oil palm (Elaeis guineensis) leaf and its effect on the quality of extractable DNA. *J. Oil Palm Res.* 18, 296–304 (2006).
29. Corley, R. H. & Tinker, P. B. in: *The Oil Palm* 4th edn 287–325 (Blackwell Science, 2003).
30. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425 (1987).

**Acknowledgements**

We thank the Genome Institute at Washington University for genome and transcriptome sequencing. We appreciate the assistance given by Noh Ahmad and Norziha Abdullah of the MPOB research station at Kluang, Johore in sampling and phenotyping of palms. We thank United Plantations, FELDA Agricultural Services, Kulim and Sime Darby for providing materials and phenotype information of individual palms. The Africa map figure was generated using free World Map PowerPoint Slides (www.m62.net/powerpoint-slides/logistics-presentations/world-map-powerpoint-slides/). The project was endorsed by the Ministry of Plantation Industries and Commodities (MPIC), Malaysia and funded by MPOB. We appreciate the unswerving support from Datuk Dr Choo Yuen May, Director General of MPOB. R.A.M. is supported by a grant from NSF 0421604 ‘Genomics of Comparative Seed Plant Evolution’.

**Author contributions**

R.S. initiated the preliminary work on the fruit colour marker/gene. R.S., E.-T.L.L., M.O.-A. and R.S. conceptualized the research programme. R.S., E.-T.L.L., M.O.-A., R.N., M.A.A.M., N.L., S.W.S., J.M.O., R.A.M. and R.S. developed the overall strategy, designed the experiments and coordinated the project. R.S., M.M., M.I., N.L., R.A.M. and R.S. identified appropriate samples for transcriptome sequencing. R.S., L.C.-L.O. and M.M. coordinated tagging and collection of fruit bunches for transcriptome sequencing. R.S., L.C.-L.O., M.O.-A., N.-C.T, P.-L.C., J.N., M.A.B., N.L., B.B., A.V.B., C.W., D.H., J.D.M., M.H., J.M.O., S.W.S. and R.S. conducted laboratory experiments and were involved in data analysis. R.S. and L.C.-L.O. constructed the genetic map. E.-T.L.L., K.-L.C., R.R., M.A.H., N.A., M.H., D.H. and S.W.S. performed bioinformatic analyses.

**Additional information**

**Accession codes:** Transcriptome data has been deposited in GenBank/EMBL/DDBJ sequence read archive (SRA) under the accession code SUB497076. Annotated genomic sequence of the VIRESCENS gene from the reference *E. guineensis* genome has been deposited in GenBank/EMBL/DDBJ nucleotide core database under the accession code KJ789862.

**Supplementary Information** accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: R.A.M. is a consultant for Orion Genomics, LLC. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Singh, R. et al. *The oil palm VIRESCENS gene controls fruit colour and encodes a R2R3-MYB*. Nat. Commun. 5:4106 doi: 10.1038/ncomms5106 (2014).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/