<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Comparative transcriptome analysis of oil palm flowers reveals an EAR-motif-containing R2R3-MYB that modulates phenylpropene biosynthesis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Li, Ran; Reddy, Vaishnavi Amarr; Jin, Jingjing; Rajan, Chakaravarthy; Wang, Qian; Yue, Genhua; Lim, Chin Huat; Chua, Nam-Hai; Ye, Jian; Sarojam, Rajani</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>Li, R., Reddy, V. A., Jin, J., Rajan, C., Wang, Q., Yue, G., et al. (2017). Comparative transcriptome analysis of oil palm flowers reveals an EAR-motif-containing R2R3-MYB that modulates phenylpropene biosynthesis. BMC Plant Biology, 17(1), 219-.</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2017</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/45360">http://hdl.handle.net/10220/45360</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>© The Author(s). 2017 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<a href="http://creativecommons.org/publicdomain/zero/1.0/">http://creativecommons.org/publicdomain/zero/1.0/</a>) applies to the data made available in this article, unless otherwise stated.</td>
</tr>
</tbody>
</table>
Comparative transcriptome analysis of oil palm flowers reveals an EAR-motif-containing R2R3-MYB that modulates phenylpropene biosynthesis

Ran Li¹,6†, Vaishnavi Amarr Reddy¹,2†, Jingjing Jin¹, Chakaravarthy Rajan¹,7, Qian Wang¹,8, Genhua Yue¹, Chinh Huat Lim³, Nam-Hai Chua⁴, Jian Ye¹,5* and Rajani Sarojam¹*

Abstract

Background: Oil palm is the most productive oil crop and the efficiency of pollination has a direct impact on the yield of oil. Pollination by wind can occur but maximal pollination is mediated by the weevil E. kamerunicus. These weevils complete their life cycle by feeding on male flowers. Attraction of weevils to oil palm flowers is due to the emission of methylchavicol by both male and female flowers. In search for male flowers, the weevils visit female flowers by accident due to methylchavicol fragrance and deposit pollen. Given the importance of methylchavicol emission on pollination, we performed comparative transcriptome analysis of oil palm flowers and leaves to identify candidate genes involved in methylchavicol production in flowers.

Results: RNA sequencing (RNA-Seq) of male open flowers, female open flowers and leaves was performed using Illumina HiSeq 2000 platform. Analysis of the transcriptome data revealed that the transcripts of methylchavicol biosynthesis genes were strongly up-regulated whereas transcripts encoding genes involved in lignin production such as, caffeic acid O-methyltransferase (COMT) and Ferulate-5-hydroxylase (F5H) were found to be suppressed in oil palm flowers. Among the transcripts encoding transcription factors, an EAR-motif-containing R2R3-MYB transcription factor (EgMYB4) was found to be enriched in oil palm flowers. We determined that EgMYB4 can suppress the expression of a monolignol pathway gene, EgCOMT, in vivo by binding to the AC elements present in the promoter region. EgMYB4 was further functionally characterized in sweet basil which also produces phenylpropenes like oil palm. Transgenic sweet basil plants showed significant reduction in lignin content but produced more phenylpropenes.

Conclusions: Our results suggest that EgMYB4 possibly restrains lignin biosynthesis in oil palm flowers thus allowing enhanced carbon flux into the phenylpropene pathway. This study augments our understanding of the diverse roles that EAR-motif-containing MYBs play to fine tune the metabolic flux along the various branches of core phenylpropanoid pathway. This will aid in metabolic engineering of plant aromatic compounds.

Keywords: MYB transcription factor, Phenylpropene, Lignin, Oil palm, Basil, Phenylpropanoid pathway
Background

Oil palm (Elaeis guineensis) is one of the most important monocot cash crop of Southeast Asia [1, 2]. Palm oil derived from its fruits is the largest source of edible vegetable oil in the world [3]. It was first introduced in Southeast Asia in 1848 and was planted on a commercial scale around 1917. The subsequent accelerated planting of palm trees and expansion of oil palm plantations boosted the economies of the developing Southeast Asian countries [4]. Oil palm is monoecious, producing male and female flowers on the same tree but at different times. To produce fruits, the plant needs to attract efficient pollinators to its flowers. Elaeidobius kamerunicus, a type of weevil which originated from West Africa, is considered the most competent and dominant insect pollinator species of oil palm. Introduction of this weevil into Malaysia during the 1980s enhanced the pollination rate of oil palm trees leading to 20–30% increase in fruit production [5]. To attract the weevil for pollination, both male and female flowers release a volatile phenylpropene compound called methylchavicol (also known as estragole) [6].

Phenylpropanes (C6-C3 carbon skeleton) are a class of volatile organic compounds (VOCs) produced by plants that serve as pollinator attractors and aid in pathogen defense [7, 8]. Commonly produced phenylpropanes include chavicol and eugenol, and their derivatives methylchavicol, methyl Eugenol and isoeugenol. Phenylpropanes are produced by the general phenylpropanoid pathway [9]. This pathway is also responsible for the production of lignins, flavonoids, phenolic acids and stilbenes ([10]; Fig. 1). Among these phenylpropanoid derived metabolites, lignin is vital for plant growth and development. It enhances cell wall strength and is required for water transport and mechanical strength [11]. Lignin is a polymer derived from three monolignols, p-coumaryl, coniferyl and sinapyl alcohol. These three precursors produce the hydroxycinnamates (H), guaiacyl (G) and syringyl (S) units of lignin respectively (Fig. 1; [12]). Phenylpropanes are also produced from monolignols; for instance, chavicol is made from p-coumaryl alcohol, whereas, eugenol is derived from coniferyl alcohol (Fig. 1; [13]). Hence, plant tissues producing phenylpropanes, such as flowers, need to prudently regulate carbon flux more into scent production than towards lignin formation.

Lignin biosynthesis is extensively controlled by R2R3-MYB transcription factors. AC elements, which serve as the binding sites for R2R3-MYBs, have been found to be enriched in the promoter regions of many lignin biosynthesis genes [14–16]. Several R2R3-MYBs function as activators of lignin synthesis. However, members of the R2R3-MYB subgroup 4 that contain an EAR motif act as transcriptional repressors and some have been identified as negative regulators of lignin production [10, 17, 18]. Perturbations in the expression levels of many of these repressors not only affected the final lignin content but also changed the flux along the various phenylpropanoid pathways. The first R2R3-MYBs characterized to down regulate lignification were AmMYB308 and AmMYB330 from Antirrhinum majus. Overexpression of AmMYB308 or AmMYB330 in tobacco repressed both lignin and phenolic acid metabolism [19]. From Arabidopsis, AtMYB4 and AtMYB32 were identified as repressors of lignin biosynthesis genes [20, 21]. The main target of AtMYB4 was shown to be cinnamate 4-hydroxylase (C4H), a core enzyme of the phenylpropanoid pathway, and AtMYB4 mutants showed enhanced accumulation of sinapate esters. AtMYB32 has been suggested to be a repressor of Arabidopsis caffeic acid O-methyltransferase (COMT) gene. Eucalyptus gunnii EgMYB1 affected lignin formation and repressed the expression of two key lignin genes EgCCR (cinnamoyl-CoA reductase) and EgCAD2 (cinnamyl alcohol dehydrogenase) [22]. Further, heterologous expression of chrysanthemum CmMYB1, reduced lignin and flavonoid production in Arabidopsis [23].

From monocots, a few subgroup 4 R2R3-MYBs have been characterized as repressors of lignin genes. Maize ZmMYB31 and ZmMYB42 downregulate both maize and Arabidopsis COMT genes [24]. Ectopic expression of maize ZmMYB42 in Arabidopsis decreased lignin content and suppressed flavonoid production as well [25]. But ectopic expression of maize ZmMYB31 in Arabidopsis led to decreased lignin content and increased anthocyanin levels due to the redirection of carbon flux towards anthocyanin pathway. Apart from the COMT promoter, it was also shown that ZmMYB31 could bind to the promoter of maize Ferulate-5-hydroxylase (F5H) gene in vivo [26]. Recently it was found that COMT is a common target of MYB31 and MYB42 in the mature leaves of maize, sorghum and rice. MYB31 and MYB42 were shown to target other genes involved in lignin pathway as well but in a more species-specific manner [27]. ZmMYB11 from maize was identified based on the sequence similarity with ZmMYB31 and ZmMYB42, and shown to regulate lignin metabolism by binding to COMT promoter [18]. A switchgrass (Panicum virgatum) ortholog of AtMYB4, PvMYB4 was shown to negatively regulate lignin formation in both tobacco and switchgrass [16]. Recently from banana, MusaMYB31 was identified as a repressor of lignin and polyphenols. It was able to down regulate many genes involved in lignin and general phenylpropanoid pathway [28].

In comparison to lignin biosynthesis, very little is known about the regulation of volatile phenylpropanes in plants. Only few TFs have been reported to regulate the production of phenylpropanes. In Petunia, R2R3-MYBs, ODORANT1 (ODO1), EMISSION OF BENZENOIDS II (EOBII)
and PhMYB4, were identified as regulators of volatile benzenoid/phenylpropanoid compounds in petals [29–33]. Both PhODO1 and PhEOBII function as positive regulators of various shikimate and phenylpropanoid pathway genes while PhMYB4 acts as a repressor of C4H gene and indirectly affects the formation of petunia floral volatiles. Production of Anthocyanin Pigment (PAPI) MYB transcription factor from Arabidopsis when ectopically expressed in Petunia or rose increased the production of phenylpropenes [34, 35]. Recently, FaEOBII was identified to control the production of eugenol in ripe strawberry fruits. In FaEOBII-silenced plants, the expression of CAD and Eugenol synthase 2 (FaEGS2) was down-regulated [36]. However, regulation of biosynthesis of volatile phenylpropenes in monocot plants have not been investigated.

In this study, we performed transcriptome sequencing of oil palm leaves, male flowers and female flowers and did a comparative study to identify the genes involved in methylchavicol production. From the differentially expressed TFs, we identified an EAR-motif-containing R2R3-MYB gene, EgMYB4 (Elaeis guineensis MYB4), which was highly enriched in open flowers. Expression of EgMYB4 coincided with the spatial and developmental production of methylchavicol. Promoter studies showed that EgMYB4 was able to bind to the promoter of oil palm EgCOMT (caffeic acid O-methyltransferase) and suppress its expression. To further characterize the function of EgMYB4, it was ectopically expressed in sweet basil (Ocimum basilicum) plants which produce high quantities of phenylpropenes similar to oil palm. Overexpression of EgMYB4 decreased lignin content and enhanced the production of phenylpropenes in transgenic sweet basil plants. This suggests that, EgMYB4 presumably suppresses lignin production in oil palm flowers and redirects the carbon flux to phenylpropene production to promote successful pollination.

![Diagram of the phenylpropanoid pathway in plant](https://example.com/diagram.png)

**Fig. 1** The phenylpropanoid pathway in plant. CAAT, coumaryl alcohol acetyl transferase; CAD, cinnamyl alcohol dehydrogenase; CFAT, coniferyl alcohol acetyl transferase; CCOMT, caffeoyl-CoA O-methyltransferase; CCR, cinnamoyl-CoA reductase; C3H, p-coumaroylshikimate 3′-hydroxylase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumaroyl CoA ligase; COMT, caffeic acid O-methyltransferase; CvOMT, chavicol O-methyltransferase; CVS, chavicol synthase; EGS, eugenol synthase; EOMT, eugenol O-methyltransferase; F5H, ferulate-5-hydroxylase; HCT, hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase; PAL, phenylalanine ammonia lyase
Results

Methylchavicol is produced at open flower stage of oil palm

In oil palm plantations, open flowers give a strong odor like aniseed which is attributed to the emission of methylchavicol [37, 38]. Analysis of male flower, female flower and leaf samples by GC-MS confirmed that the strong odor from open flowers was due to methylchavicol emission (Additional file 1). In leaves, only one green leaf volatile, 2-hexenal, was detected. To investigate methylchavicol emission during different stages of flower development, volatile compounds from three developmental stages of both male and female flowers were extracted and analyzed by GC-MS. The developmental stages analyzed were; before anthesis stage (closed bud), at anthesis stage (newly opened) and after anthesis stage (old flowers). Our results showed that methylchavicol is strongly emitted in male and female flowers at the open flower stage (Fig. 2a and b).

Sequencing, de novo assembly and annotation of transcriptome

To elucidate methylchavicol biosynthesis in oil palm, we performed RNA-sequencing (RNA-Seq) of male open flowers, female open flowers and leaves using Illumina HiSeq 2000. More than 22 million high quality reads of 101 base pairs from different tissues were generated. The quality of Illumina sequencing outputs was high, as evaluated by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) (Additional file 2). Using the Trinity method (default parameter) [39], the sequence reads were finally assembled into 59,078 non-redundant unigenes with 96,062 isoforms (with N50 = 1884 bp and mean size = 1104 bp). All unigenes were longer than 200 bp. The unigenes were annotated by performing a BLASTX search against various protein databases including TAIR and UniProt (e-value bigger than 1e-3). Functions were assigned by top hit candidates. Expression levels for assembled transcripts were calculated by mapping the reads onto them using Bowtie [40]. Later, by using RSEM (RNA-Seq by Expectation-Maximization), the abundance of assembled transcripts was estimated and the expression levels were measured [41].

EgMYB4 is an EAR-motif-containing R2R3-MYB TF that is enriched in flowers

Transcription factors (TFs) play an important role in regulating plant secondary metabolism [10, 42]. Seventy-nine upregulated and 219 downregulated TF transcripts were found in male and female open flowers when compared to leaves (Additional file 3A). The top 10 differentially expressed TFs between leaves and flowers are listed in Additional file 3B. Among them, the four MADS box TFs have been proposed to be involved in flower development [43]. The MYB-like TF was shown to be involved in regulating flowering time [44]. Out of the four MYB TFs identified, two were similar to Arabidopsis MYB21 and one each to Arabidopsis MYB108 and MYB4 respectively. Studies on AtMYB21 showed that it functions in floral organ development particularly stamen development [45, 46], and MYB108 was shown to regulate jasmonate-mediated stamen development [47]. The other differentially expressed MYB TF identified was AtMYB4 which is implicated in regulating lignin pathway [20]. Analysis of AtMYB4 and other R2R3-MYBs belonging to subgroup 4 revealed that perturbations in lignin production was able to affect flux along the various phenylpropanoid pathways altering the formation of secondary metabolites [20, 26]. Since the production of both lignin and volatile phenylpropenes share common precursors of phenylpropanoid pathway, we decided to focus on this TF. Full length coding sequence of this MYB was obtained by reverse transcription PCR and named as EgMYB4. Amino acid sequence analysis indicated that it encodes a typical R2R3 domain and has an EAR motif (C2 motif) at the C-terminal similar to other members of R2R3-MYB subgroup 4 (Fig. 3a). Phylogenetic analysis based on amino acid sequences with other characterized R2R3-MYBs of subgroup 4 showed that EgMYB4 is closely related to Eucalyptus gunnii MYB1 (Fig. 3b). Expression profile analysis revealed high expression of EgMYB4 in oil palm female and male open flowers when compared to leaves (Fig. 3c). Among the three developmental stages of flowers analyzed, EgMYB4 expression increased in open flowers as compared to before anthesis stage flowers correlating with the emission of methylchavicol. In female flowers, the expression of EgMYB4 decreased in after anthesis stage flowers while in male flowers it increased (Fig. 3c). To investigate the subcellular localization of EgMYB4, 35S: EgMYB4-YFP was transiently expressed in N. benthamiana leaves. Fluorescence analysis showed that EgMYB4 exclusively localize in the nuclear bodies as confirmed by 4, 6-diamidino-2-phenylindole (DAPI) staining (Fig. 3d).

Phenylpropene synthases are enriched while COMT and FSH transcripts are repressed in oil palm flowers

Among all the metabolic derivatives of phenylpropanoid pathway, lignin and phenylpropene biosynthesis share the most common precursors. To investigate the biosynthesis of lignin and methylchavicol in flowers, we first analyzed the expression of shared upstream biosynthesis genes from the RNA-Seq data. Many phenylpropanoid-pathway genes like Phenylalanine ammonia-lyase (PAL) 4-coumarate:CoA ligase (4CL), C4H, Cinnamoyl-CoA reductase (CCR1) and CAD exist as a multigene family in various plant species. Studies have shown that they can have distinct or overlapping functions in phenylpropanoid metabolism depending on their substrate specificity.
We indeed identified more than one transcript encoding these genes from our RNA-Seq data of leaves and flowers. Few transcripts showed constitutive expression while few other transcripts had tissue specific expression (Additional file 4). These genes can act in different branches of phenylpropanoid biosynthesis pathway, hence their function in the pathway cannot be confirmed without the support of experimental data. To determine a clear expression pattern of the active metabolite pathway in flowers, genes specifically involved in latter steps of lignin and methylchavicol synthesis were investigated. Most differentially expressed transcripts between leaves and flowers were the genes encoding enzymes involved in the final two steps of methylchavicol.
biosynthesis; chavicol synthase (CVS) and chavicol O-
methyltransferase (CvOMT). They were significantly
up-regulated in both male and female open flowers
(Fig. 4b–f). CVS belongs to PIP family of NADPH-
dependent reductases which also include eugenol
synthases (EGS) and isoeugenol synthases (IGS). EGS
and IGS have been identified and characterized from
basil, Petunia, Clarkia, anise and recently from straw-
berry [7, 49–52]. A chavicol synthase (LtCES1) was
also characterized from Larrea tridentata that can act
on both coniferyl acetate and coumaryl acetate
substrates to form eugenol and chavicol, respectively
[53]. The oil palm EgCVS showed high sequence
similarity to these identified enzymes (Fig. 4c). Simi-
larly, phenylpropene O-methyltransferases have been
characterized from sweet basil and apple which are
required for generating methylated phenylpropenes
like, methyleugenol and methylchavicol [54, 55]. Two
transcripts similar to chavicol O-methyltransferases
(EgCvOMT1 and EgCvOMT2) were found in our
RNA-Seq data, which share high amino acid identity
with sweet basil and apple CvOMTs (Fig. 4a). On the
Fig. 4 (See legend on next page.)
other hand, transcripts encoding genes involved in lignin biosynthesis genes such as caffeic acid O-methyltransferase (COMT) and Ferulate-5-hydroxylase (F5H) were highly down-regulated in oil palm male and female open flowers (Fig. 4h and i). COMT and F5H are key genes mainly committed to the formation of S-lignin subunit [12]. EgCOMT and EgF5H showed high similarity to the well-characterized genes from other species (Fig. 4e and g). Protein accession numbers and nucleotide sequences of the genes in Fig. 4 are listed out in Additional files 5 and 6 respectively.

**EgMYB4 interacts with EgCOMT promoter**

AC elements are present in the promoters of many lignin biosynthesis genes and are known to be bound by MYBs which regulate their expression. Additionally, COMT and F5H genes are known to be down-regulated by EAR motif-containing MYBs in monocots. Maize ZmMYB11 was found to bind to the promoter of ZmCOMT in vivo; ZmMYB31 and ZmMYB42 was found to bind to the promoters of both ZmCOMT and ZmF5H genes and repress their expression [18, 26]. To investigate if EgF5H and EgCOMT are regulated by EgMYB4, we cloned and screened a 2 kb-promoter region of both genes. One AC-IV element (ACAAAC) was found in both the promoters (Fig. 5a; [16]). Within 0.25 kb of promoter region of EgCOMT, we identified two new AC elements (ACACC) and named it as AC-V. To test the interaction of EgMYB4 to these AC elements, an electrophoretic mobility shift assay (EMSA) was performed using the purified recombinant His-EgMYB4 protein. His-EgMYB4 bound to both AC-IV and AC-V elements but failed to interact with mutated versions of AC-IV (GAAGGGA) and AC-V (GGAGGAT) (Fig. 5b). DNA binding specificity was further confirmed by a competition experiment using 250-fold excess unlabeled cold probe which led to the disappearance of labeled DNA/protein complex.

Further, the transcriptional repression activity of EgMYB4 was determined by using N. benthamiana as a transient expression system [56]. Based on the distribution of AC elements, 1 kb promoter region of EgF5H and 2 kb promoter region of EgCOMT was introduced into pCAMBIA1391 vector. EgF5H promoter: GLU or EgCOMT promoter: GLIS were used as reporters and 35S:EgMYB4-YFP was used as an effector (Fig. 5c). The promoter activity of EgF5H had no difference between EgMYB4-YFP and only YFP expressing leaves (Fig. 5d and g). However, the promoter activity of EgCOMT was significantly suppressed in EgMYB4-YFP expressing leaves when compared to leaves expressing only YFP (Fig. 5e and f). These results indicate that EgCOMT is repressed by EgMYB4. Suppression of COMT activity is known to reduce lignin formation [57–60]. From the above results, we can postulate that increased expression of EgMYB4 in oil palm flowers reduces lignin synthesis by repressing EgCOMT, which might indirectly affect the metabolic flux into methylchavicol synthesis.

**Ectopic expression of EgMYB4 in sweet basil decreases total lignin content and increases phenylpropene production**

As oil palm is a non-model plant and transformation studies are not feasible, we decided to characterize EgMYB4 in sweet basil plants. Similar to oil palm flowers, sweet basil also produces phenylpropanes in leaf glandular trichomes which form the main components of sweet basil essential oil (Additional file 7; [9]). To validate our hypothesis that EgMYB4 mediated lignin suppression in oil palm flowers affects flux into phenylpropene synthesis, EgMYB4 was ectopically expressed in sweet basil plants. The sweet basil variety used in our study produces two types of phenylpropanes, majorly eugenol/methyleugenol and small amounts of methylchavicol. Five independent transgenic lines were initially selected and of them two lines, MYB4-2 and MYB4-5, were advanced for further characterization. Both of these lines showed high expression of EgMYB4 transgene (Fig. 6a). Plants overexpressing EgMYB4 showed many phenotypic alterations. These plants were dwarfed, exhibited weak stem, had reduced leaf size and delayed flowering (Fig. 6b and c). These phenotypes are characteristic of lignin-deficient mutants [26, 61]. In plants, the cell walls of sclerenchyma and vascular tissues are lignified providing strength and rigidity. Transverse sections of mature stems from transgenic and wild type (WT) sweet basil plants were analyzed by toluidine blue and phloroglucin staining to look for possible changes in these lignified tissues. There was considerable amount of reduction in the number of lignified cell layers forming the sclerenchyma and vascular tissues in the transgenic plants (Fig. 7a-d). The amount of total lignin in mature stems was measured by thioglycolic acid method and was found to be reduced by 34.2% and 31.4%
overexpression of *EgMYB4* in sweet basil could reduce lignin biosynthesis in sweet basil plants. We have performed RNA sequencing of sweet basil leaf glandular trichomes from which the orthologs of *F5H* and *COMT* genes were identified. Expression levels of these genes were measured in *EgMYB4* overexpressing and wild type sweet basil plants. The expression levels of *ObCOMT1*, *ObCOMT2* and *ObF5H* were significantly decreased in the transgenic lines when compared to WT plants (Fig. 7f-h). These results suggest that ectopic expression of *EgMYB4* down regulates the expression of both *ObCOMT* and *ObF5H* genes thereby affecting lignin biosynthesis in sweet basil plants. Although in in vivo studies *EgMYB4* did not suppress *EgF5H*, we did observe a decrease in *ObF5H* expression when *EgMYB4* was ectopically expressed. The direct interaction of *EgMYB4* with the promoters of *ObCOMTs* and *ObF5H* remains to be confirmed. However similar results have been observed.
with other R2R3-MYBs like ZmMYB31 and AtMYB4 where overexpression leads to dose dependent selection of additional target genes in transgenic plants [20, 26].

To determine the effect of EgMYB4 overexpression on phenylpropenes, volatile compounds were extracted and analyzed from transgenic lines and WT plants. The transgenic plants exhibited a significant increase in total phenylpropene levels (Fig. 8d). The amount of methylchavicol in MYB4–2 and MYB4–5 was ~15- and 10-folds higher than WT plants (Fig. 8a). Meanwhile, the amount of methyleugenol in MYB4–2 and MYB4–5 was ~ 4- and 3-folds higher than WT plants (Fig. 8c). However, we did not see any changes in eugenol levels (Fig. 8b). This result suggests that ectopic expression of EgMYB4 can affect flux into phenylpropene pathway.

**Discussion**

Efficient pollination is vital for plant’s survival. The interaction between plant and its pollinator is considered to be mutualistic and the result of coevolution between plant and pollinator species. A stringent one-to-one coevolution is generally found between one plant species and its insect partner, for example fig plants and its specific pollinator [62]. In case of oil palm, average yields in Southeast Asia increased considerably only after introduction of Elaeodobius kamarunicus from Africa. This strongly suggests that oil palm is highly dependent on just a single species of insect for pollination suggesting a specialized mutualistic palm–pollinator relationship [63]. Phenylpropanes form one of the main components of floral scents that help attract pollinators. Oil palm flowers emit methylchavicol to attract insect pollinators. Phenylpropanes are derived from the general phenylpropanoid pathway. Pathway studies propose that both lignin and phenylpropene production share common substrates [13]. Hence within a cell, tight regulation of metabolic pathway enzymes is required to ensure that substrates used by competing pathways are regulated aptly to produce the desired metabolic outcome. The biosynthesis of lignin consumes large amounts of carbon resources. The biomass of the oil palm is lignocellulosic rich with lignin contents ranging from 18 to 23% among different varieties and tissues [64]. From a plant’s perspective, suppression of lignin biosynthesis in flowers allows the plant to reallocate carbon sources towards a vital floral phenylpropene production, which will aid in attracting pollinators. Transcriptome data analysis of open oil palm flowers revealed that EgCVS and EgCrOMT are highly expressed in flowers whereas expression levels of EgCOMT and EgF5H are significantly reduced when compared to leaves (Additional file 4). By comparative transcriptome analysis and protein-DNA interaction assays, we identified a R2R3 EAR-motif-containing EgMYB4 from oil palm which can bind to AC elements of EgCOMT and suppress its activity in vivo. No interaction with EgF5H was observed (Fig. 5). Although one AC element exists in the promoter of EgF5H which was cloned, we cannot completely rule out the possibility that the length
of the promoter tested might not have been enough to observe the binding of EgMYB4 to EgF5H. Additionally, there might be other EAR-motif-containing MYBs that represses EgF5H in oil palm flowers. Monocot transcription factors like ZmMYB11, MusaMYB31, ZmMYB31, ZmMYB42 and PvMYB4 which show high sequence identity with EgMYB4 are reported to repress COMT along with other genes in the lignin biosynthesis pathway [25–28]. In our study, we tested the interaction of EgMYB4 with the promoters of EgCOMT and EgF5H only, we cannot dismiss the possibility that it might regulate other genes in phenylpropanoid pathway. qRT-PCR analysis of EgMYB4 at different developmental stages of flowers showed high levels of EgMYB4 in after anthesis stage male flowers. At this stage, flowers do not emit phenylpropenes suggesting additional role for EgMYB4 in flowers. Perturbation in one branch of phenylpropanoid pathway is known to affect other branches as well. After examining the
transcripts of genes involved in flavonoid pathway from the RNA-Seq data of oil palm flowers, we were not able to observe a clear pattern that would suggest either up or downregulation of flavonoid pathway in flowers (Additional file 8). Further work is required to analyze other phenylpropanoid derived metabolites in oil palm flowers apart from methylchavicol to determine the effect of EgMYB4 on their production.

Overexpression of EgMYB4 in sweet basil resulted in increased phenylpropene levels and decreased lignin content (Figs. 7 and 8). Sweet basil EgMYB4 overexpressing plants exhibited growth phenotypes typical to lignin deficient plants, including dwarfism, weakened stem, reduced leaf size and delayed flowering [26]. These phenotypes of decreased lignin imply that ectopic expression of EgMYB4 can affect lignin biosynthesis genes in sweet basil. Downregulation of ObCOMTs and ObF5H was observed in transgenic plants. However, in the in vivo studies EgMYB4 did not suppress EgF5H. There is a possibility that EgMYB4 might have different targets in different plants or the resulting perturbation in the lignin pathway might affect the steady state mRNA level of ObF5H. Additionally, this might also be due to the high level of EgMYB4 expression in the transgenics. Similiar results have been observed with AtMYB4 and ZmMYB31. AtMYB4 mainly regulates cinnamate 4-hydroxylase (C4H) to control hydroxycinnamic acid metabolism. But when overexpressed it was able to down regulate the expression of additional genes (CHS, 4CL1 and 4CL3) in the pathway. It was found that there is a dose dependent selection of target genes by AtMYB4 [20]. Similarly, ZmMYB31 interacts with ZmCOMT and ZmF5H promoters in vivo. However, it can repress the expression of Arabidopsis 4CL and C3H when overexpressed in Arabidopsis [26]. The direct binding of EgMYB4 to the promoters of sweet basil COMTs, FSH and other lignin pathway genes needs to be confirmed. Downregulation of COMT and FSH by transgenic approaches has been reported in many plants like Arabidopsis, tobacco, poplar, alfalfa, maize, switchgrass and fescue to manipulate lignin polymer. They mainly affect the formation of S-lignin subunits [57–60, 65–69]. We observed that both methylchavicol and methyleugenol was increased in EgMYB4 overexpressing sweet basil lines which are derived from p-coumaryl and coniferyl alcohol respectively. Probably the downregulation of sinapyl alcohol (S-lignin) pathway increases flux towards the formation of p-coumaryl and coniferyl alcohol thus enhancing the formation of both methylchavicol and methyleugenol which are derived from these precursors in sweet basil.

In oil palm, transcripts encoding enzymes directly responsible for methylchavicol biosynthesis like EgCvOMT
and EgCVS show transcriptional activation only in flowers (Additional file 4). However, in addition to direct activation of enzymes that are responsible for metabolite production, activation and suppression of enzymes involved in flux leading to its formation also help to fine tune the amount of metabolite formation, in a particular tissue at a developmentally relevant time. For example, in Petunia R2R3-MYBs ODORANT1 (ODO1), EMISSION OF BENZENOIDS I, EMISSION OF BENZENOIDS II (EOBII), are all identified as positive regulators of various genes involved in the production of floral volatiles [29–33]. But PhMYB4 belonging to R2R3-MYB subgroup 4 acts as a repressor of C4H gene of phenylpropanoid pathway and indirectly affects the formation of petunia floral volatiles [33]. Similarly, production of methylchavicol in oil palm flowers can be regulated at various levels by different genes acting as positive or negative regulators. Multiple or single TFs might be involved in stage specific activation of EgCvOMT and EgCVS. Our work presented here shows the regulation imparted by one of the probably many transcription factors involved in temporal and spatial regulation of methyl chavicol biosynthesis in oil palm flowers. Emission of methylchavicol forms an important agronomic trait in oil palm as it affects pollination. In our study, we identified a transcription factor potentially regulating methylchavicol emissions and enzymes catalyzing its production. These genes may be used as potential candidates in molecular breeding strategies of oil palm. Discovering genetic variations of these genes in different oil palm accessions and its effect on emission/yield can help design breeding of high yielding varieties.

Conclusions
A volatile phenylpropene, methylchavicol is released from oil palm flowers to attract the pollinator weevil. RNA-Seq analysis of flowers showed increased expression of transcripts coding for enzymes involved in methylchavicol production but decreased expression of COMT and F5H transcripts which are specifically involved in lignin formation. We identified a R2R3-MYB enriched in flowers, EgMYB4 that can bind to EgCOMT promoter and suppress it. Functional analysis of EgMYB4 in sweet basil demonstrated the ability of EgMYB4 to reduce lignin formation and enhance the formation of phenylpropene suggesting a similar role for EgMYB4 in oil palm flowers. This study on methylchavicol emission in oil palm flowers will further enrich our understanding of transcriptional regulation of the complex phenylpropanoid pathways which produces an array of compounds in plants. Moreover, it will help in oil palm breeding and metabolic engineering of phenylpropanoid pathway to produce aromatic compounds of interest.

Methods
Plant materials
The oil palm species used in this study was Dura [2]. Samples from different flower stages and leaves were collected from palm trees planted in Temasek Life Sciences Laboratory, Singapore and Wilmar International Plantation, Palembang, Indonesia. Commercial sweet basil (O. basilicum) was tested for its secondary metabolites by GC-MS and grown in green house under natural light conditions. Agrobacterium mediated transformation of sweet basil was performed as previously described by [70]. T0 and T1 transgenic plants were selected using GFP as visual marker. For all experiments, T1 plants were used. Nicotiana benthamiana seeds were germinated on MS plate and transferred into soil. Twenty days after growing in the greenhouse, the seedlings were used for Agro-infiltration.

Compound analysis
1.5 cm of one oil palm male flower bunch or 3 female flowers were placed in a 5 mL glass bottle containing 2 mL hexane and shaken for 10 min at room temperature. After centrifugation, 500 μL of supernatant was transferred into a 2 mL GC vial and analyzed by gas chromatography-mass spectrometry (GC-MS). For sweet basil samples, leaves were ground in liquid nitrogen and 200 mg of powder was transferred to 1.5 mL eppendorf tubes. 500 μL of hexane (containing 20 μg of diethyl sebacate as internal standard) was added and vortexed for 2 min. After centrifugation, 400 μL of supernatant was transferred into a 2 mL GC vial and analyzed by GC-MS. Six plants from each transgenic line were measured.

RNA isolation and RNA sequencing
RNA from oil palm tissues was extracted as described previously [71]. Briefly, about 0.5 g of each sample was ground in liquid nitrogen and the powder was then transferred to a pre-chilled polypropylene (Falcon) tube. 5 mL of pre-heated (65 °C) CTAB extraction buffer (2% (w/v) CTAB, 2% (w/v) polyvinlypyrrolidone (PVP-40), 100 mM Tris-HCl (pH 8.0), 25 mM EDTA, 2 M NaCl, 0.1% spermidine and 2% β-mercaptoethanol) was added to each tube and samples were incubated for 30 min at 65 °C. Later, the samples were extracted with chloroform: isoamylalcohol (24:1) CTAB extraction buffer (2% (w/v) CTAB, 2% (w/v) polyvinlypyrrolidone (PVP-40), 100 mM Tris-HCl (pH 8.0), 25 mM EDTA, 2 M NaCl, 0.1% spermidine and 2% β-mercaptoethanol) was added to each tube and samples were incubated for 30 min at 65 °C. Later, the samples were extracted with chloroform: isoamylalcohol (24:1) for 2 times. The supernatant (1.0 mL) was then transferred to RNase free.5 mL eppendorf tubes and 0.5 mL of 96–100% ethanol was added. The mixture was immediately loaded onto RNA binding columns (Qiagen RNA Mini extraction kit) and RNA was extracted according to manufacturer's protocol. RNA sequencing and assembly was performed as
described previously [72]. RNA from sweet basil samples were isolated by using RNA Mini extraction kit (Qiagen).

**Quantitative real-time PCR (qRT-PCR) and reverse transcription PCR (RT-PCR)**

Eight hundred ng of total RNA for each sample was reverse transcribed using the PrimeScript™ RT-PCR Kit (TaKaRa). qRT-PCR was performed on an ABI 7900 HT fast real time system (Life technologies) using SYBR Green Real-time PCR Master Mixes (Life technologies). For RT-PCR, genes were amplified in T100™ Thermal Cycler (Bio-Rad) by the following program, 95 °C for 2 min; 23 cycles of 95 °C for 40 s, 60 °C for 40 s, 72 °C for 15 s; 72 °C for 5 min. 10 μL of PCR products were analyzed by gel electrophoresis. The primers used for RNA detection of target genes by qRT-PCR and RT-PCR are listed in Additional file 9. Oil palm Ubiquitin gene (EgUBQ) or sweet basil elongation factor (ObEF1a) gene was used as internal controls.

**Constructs**

Full-length open reading frames encoding EgMYB4 without a stop codon were amplified by PCR using Pfu DNA polymerase (Thermo Scientific) with primers listed in Additional file 9. The obtained DNA fragments were cloned either into pBA-YFP vector [56] to generate YFP fused protein or pET28b (Novagene) to generate His-tag fused protein.

**Purification of recombinant protein**

The construct containing His-MYB4 was transformed into E.coli BL21 (DE3). Expression of His-MYB4 was induced by adding 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) followed by an incubation at 37 °C for 3 h. Later, the cells were collected and the recombinant protein was purified using His-Trap (GE healthcare) according to the manufacturer’s instruction.

**Electrophoretic mobility shift assay (EMSA)**

The probes used in EMSA were all labelled by Biotin and are listed in Additional file 9. EMSA was performed using a LightShift Chemiluminescent EMSA Kit (Thermo) according to the manufacturer’s instructions. Competition experiments were performed using unlabeled DNA as a competitor in a 250-fold molar excess.

**Transcriptional repression assay**

The 2 kb promoter region of EgCOMT and 1 kb promoter region of EgFSH were PCR-amplified (primers are listed in Additional file 9) and cloned into pCAM-BIA1391. All constructs were introduced into AGL1 Agrobacterium. Leaves of N. benthamiana were agroinfiltrated with the indicated constructs (Fig. 5) at a ratio of 1:1. Two days after infiltration leaves were harvested and frozen in liquid nitrogen or stained with GUS staining buffer [56]. Each treatment was repeated eight times. GUS quantitative assay was performed as described previously [56].

**Lignin measurement**

Total lignin content was measured using thioglycolic acid method as described previously [61]. 100 mg of fresh sweet basil stems were used and six plants from each transgenic line were measured.

**Histology**

Stems from the basal portion of adult wild type sweet basil and transgenic plants were fixed in historesin and sectioned. Sections were stained either with toluidine blue or phloroglucinol for lignin analysis. Lignin analysis was performed as described previously [73].

**Data analysis**

Differences in total lignin and phenylpropene levels on different lines were determined by analysis of variance (ANOVA). Differences in GUS activities on different treatments were analyzed by using the Student’s t-test. All tests were carried out with Statistica (Statistica, SAS Institute Inc., http://www.sas.com/).

**Additional files**

- **Additional file 1**: Analysis of volatile compounds in different oil palm tissues. (DOCX 115 kb)
- **Additional file 2**: Overview of RNA-seq result. (DOCX 302 kb)
- **Additional file 3**: Transcription factors found in RNA-seq data. (DOCX 209 kb)
- **Additional file 4**: Heat map of phenylpropanoid pathway transcripts expression in oil palm leaves and male and female open flowers generated by Heatmapper. (DOCX 490 kb)
- **Additional file 5**: List of plant genes used in the phylogenetic analysis described in Figs. 3 and 4. (DOCX 25 kb)
- **Additional file 6**: Nucleotide sequence information of phenylpropanoid pathway genes from RNA-seq. (DOCX 22 kb)
- **Additional file 7**: GC-MS analysis of phenylpropenes in sweet basil leaves. (DOCX 74 kb)
- **Additional file 8**: Expression levels of predicted flavonoid biosynthesis genes in oil palm leaves and flowers. (DOCX 20 kb)
- **Additional file 9**: DNA primers used in this study. (DOCX 21 kb)

**Abbreviations**

C4H: cinnamate 4-hydroxylase; CAD: cinnamyl alcohol dehydrogenase; CCR: cinnamoyl-CoA reductase; COMT: caffeic acid O-methyltransferase; EMSA: electrophoretic mobility shift assay; GC-MS: gas chromatography-mass spectrometry; GFP: green fluorescent protein; GUS: β-glucuronidase; RNA-Seq: RNA sequencing; TF: transcription factors; YFP: yellow fluorescent protein

**Acknowledgements**

We thank KharMeng Ng for his invaluable assistance with experiments. This research was supported by funds from Wilmar International Ltd. and Temasek Life Sciences Laboratory.
Funding
This research was supported by funds from Wilmar International Ltd. and Temasek Life Sciences Laboratory.

Availability of data and materials
The RNA-Seq data supporting the result of this article is available in the DNA Data Bank of Japan (DDBJ) http://www.ddbj.nig.ac.jp/ with accession number DRA004319. Nucleotide sequences from this study were submitted to the National Center for Biotechnology Information (NCBI)/GenBank using BankIt with accession numbers: EgCVS (KT778617), EgMYB4 (KT778616), EgCOMT promoter (KT778614), Eg5SH promoter (KT778615). All the phylogenetic trees and associated matrices have been deposited in the TreeBASE (http://treebase.org/treebase-web/home.html) with the study number 21622. The TreeBASE generated URL to access the deposition is http://pubitol.phyto/treebase/phytolows/study/TB2521622.

Authors’ contributions
RL, LY and RS designed experiments. RL, VAR and QW performed the experiments. CHL assisted with oil palm flower sample preparation. JY conducted the experiments related with RNA-Seq. JY analyzed the transcriptome data. CR performed the sweet basil transformation. GY supported oil palm genome information. RL, NH, JY and RS wrote the article, which was reviewed and approved by all authors.

Ethics approval and consent to participate
This research was a collaborative effort between Wilmar International and Temasek Life Sciences Laboratory. Oil palm (Elaeis guineensis) samples were collected from palm trees planted in Temasek Life Sciences Laboratory, Singapore and Wilmar International Plantation, Palembang, Indonesia. Sweet basil seeds were obtained from commercial suppliers and grown in green house at Temasek Life Sciences Laboratory. No specific permits were required.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

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

Author details
1Temasek Life Sciences Laboratory, National University of Singapore, Singapore 117604, Singapore. 2Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore. 3R&D Department, Wilmar International Plantation, Palembang, Indonesia. 4Laboratory of Plant Molecular Biology, Rockefeller University, New York, NY 10065, USA. 5State Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China. 6Present Address: Singapore Centre on Environmental Life Sciences Exploration, Singapore 117543, Singapore. 7Present Address: Singapore Centre on Environmental Life Sciences Exploration, Singapore 117543, Singapore. 3R&D Department, Wilmar International Plantation, Palembang, Indonesia. 4Laboratory of Plant Molecular Biology, Rockefeller University, New York, NY 10065, USA. 5State Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China. 6Present Address: Singapore Centre on Environmental Life Sciences Exploration, Singapore 117543, Singapore. 7Present Address: Singapore Centre on Environmental Life Sciences Exploration, Singapore 117543, Singapore.

Received: 24 May 2017 Accepted: 13 November 2017
Published online: 23 November 2017

References


Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at www.biomedcentral.com/submit