




Pathway elucidation and heterologous reconstitution of the long-chain alkane pentadecane biosynthesis from *Pogostemon cablin*

Jing Wen^{1,2,†}, Wanxian Xia^{1,2,†}, Ying Wang^{1,2,†}, Juan Li^{1,2}, Ruihao Guo^{1,2}, Yue Zhao^{1,2}, Jing Fen^{1,2}, Xinyu Duan^{1,2}, Guo Wei³, Guodong Wang⁴ , Zhengguo Li^{1,2}  and Haiyang Xu^{1,2,*} 

¹Key Laboratory of Plant Hormones and Molecular Breeding of Chongqing, School of Life Sciences, Chongqing University, Chongqing, China

²Center of Plant Functional Genomics, Institute of Advanced Interdisciplinary Studies, Chongqing University, Chongqing, China

³College of Horticulture and Landscape Architecture, Yangzhou University, Yangzhou, China

⁴State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

Received 19 September 2024;

revised 4 November 2024;

accepted 5 November 2024.

*Correspondence (Tel +86-18716380892;

fax +86-023-65678491; email hyxu@cqu.edu.cn).

[edu.cn](mailto:hyxu@cqu.edu.cn))

[†]These authors contributed equally to this work.

Summary

Very-long-chain (VLC) alkanes are major components of hydrophobic cuticular waxes that cover the aerial epidermis of land plants, serving as a waterproofing barrier to protect the plant against environmental stresses. The mechanism of VLC-alkane biosynthesis has been extensively elucidated in plants. However, little is known about the biosynthesis of long-chain alkanes (LC, C₁₃ ~ C₁₉) such as pentadecane in plants. Alkanes with different chain lengths are also major constituents of fossil fuels and thus the discovery of the alkane biosynthetic machinery in plants would provide a toolbox of enzymes for the production of renewable hydrocarbon sources and next generations of biofuels. The top leaves of *Pogostemon cablin* at young stage accumulate large amounts of LC-alkane pentadecane, making this plant an excellent system for the elucidation of LC-alkane biosynthetic machinery in plant. We show here that LC-alkane pentadecane biosynthesis in *P. cablin* involves an endoplasmic reticulum (ER)-localized complex made of PcCER1-LIKE3 and PcCER3, homologues of Arabidopsis ECERIFERUM1 (AtCER1) and AtCER3 proteins that are involved in Arabidopsis VLC-alkane biosynthesis. We reconstitute the biosynthesis of pentadecane in *Nicotiana benthamiana* by co-expression of PcCER1-LIKE3 and PcCER3 and further improve its production by silencing multifunctional acetyl-CoA carboxylases involved in fatty acid elongation pathway. Taken together, we uncovered the key biosynthetic machinery of LC-alkane pentadecane in *P. cablin* and demonstrated that using these newly identified enzymes to engineer this LC-alkane for liquid biofuel production in a heterologous plant host is possible.

Keywords: long-chain alkanes, pentadecane, biofuel, CER1-LIKE enzyme, aldehyde decarboxylase, *Pogostemon cablin*.

Introduction

Alkanes consisting mainly of carbon (C) and hydrogen (H) atoms are acyclic saturated hydrocarbons and have the general chemical formula C_nH_{2n+2}. While very-long-chain (VLC; >C₁₉) alkanes are abundant in plants as part of the cuticular waxes (Bernard and Joubès, 2013; Lee and Suh, 2015), long-chain (LC; C₁₃ ~ C₁₉) and medium-chain (C₆ ~ C₁₂) alkanes also occur in land plants, but they are rarely reported, and their mode of synthesis has not been elucidated (Cheesbrough and Kolattukudy, 1984; Debnath et al., 2021; Savage et al., 1996a,b).

In cyanobacteria that produce medium- to long-chain alkanes (mainly C₁₅ and C₁₇), of which the biological function remains elusive, acyl-acyl carrier proteins (ACPs) are converted into alkanes through the acyl-ACP reductase (AAR)/aldehyde deformylating oxygenase (ADO) pathway with coproduction of formate instead of CO (Figure S1a; Li et al., 2012; Schirmer et al., 2010). Many microalgae also synthesize large amounts of LC-alka(e)nes (mainly C₁₅, C₁₇, C_{17:1}) that accumulate in cell walls and can be utilized to provide energy when photosynthesis is not possible (Sorigué et al., 2016). Their biosynthesis depends on the

activity of a unique fatty acid photodecarboxylase catalysing the decarboxylation of free fatty acid to alka(e)nes (Figure S1b; Sorigué et al., 2017).

Land plants have a protective barrier called cuticle, a hydrophobic layer consisting of cutin polyester and cuticular waxes that protects them against desiccation, ultraviolet radiation, pathogen infection and other abiotic and biotic environmental stress (Bernard and Joubès, 2013; Bowman et al., 2017; Lee and Suh, 2015; Riederer and Schreiber, 2001). In addition to its physiological functions, lipid components derived from the plant cuticle have been used in industrial applications including the production of paints, lubricants, adhesives, pharmaceuticals, cosmetics and biofuels (Lee and Suh, 2015). Cuticular waxes consist of a complex mixture of multiple homologous series of very-long-chain aliphatic compounds including fatty acids, primary and secondary alcohols, aldehydes, esters, ketones and alkanes, as well as nonacyl lipid cyclic components including terpenoids and flavonoids (Bernard and Joubès, 2013; Lee and Suh, 2015; Samuels et al., 2008). Aliphatic acyl chains of cuticular waxes are derived from VLC-fatty acids (VLCFAs) with chain lengths ranging from 22 to 38 carbons, resulting from the activity

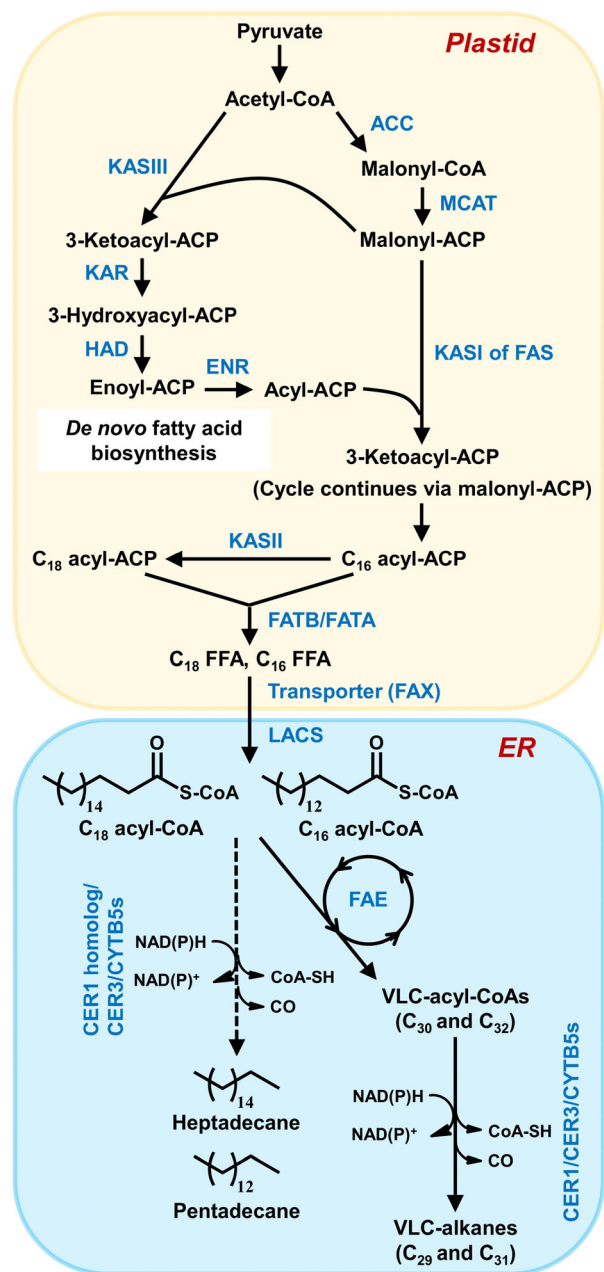


Figure 1 The proposed biosynthetic pathway of LC-alkanes pentadecane and heptadecane in *Pogostemon cablin*. The schematic view of the proposed biosynthetic pathway of LC-alkanes pentadecane and heptadecane in *Pogostemon cablin* is based on the fatty acid and VLC-alkane biosynthetic pathways in *Arabidopsis* (Bernard *et al.*, 2012; Bernard and Joubès, 2013; Lee and Suh, 2015; Pascal *et al.*, 2019). The C₁₆ and C₁₈ free fatty acids (FFAs) that are synthesized through the plastidial de novo synthesis pathway are exported by fatty acid exporters (FAX) (Bonaventure *et al.*, 2003) from plastid into cytosol and activated to C₁₆ acyl-CoA (palmitoyl-CoA) and C₁₈ acyl-CoA (stearoyl-CoA) by long-chain acyl-CoA synthetase (LACS) (Schnurr *et al.*, 2004; Weng *et al.*, 2010; Zheng *et al.*, 2019), respectively. The enzymatic alkane-forming complex in the endoplasmic reticulum (ER) that contains ECERIFERUM1 (CER1) homologue, CER3 and CYTB5s from *P. cablin* catalyses the conversion of palmitoyl-CoA and stearoyl-CoA to long-chain alkanes pentadecane and heptadecane, respectively. Meanwhile C₁₆ acyl-CoA and C₁₈ acyl-CoA are elongated to very-long-chain (VLC) acyl-CoAs by ER-associated multienzyme fatty acyl-CoA elongase (FAE) complexes. The VLC-acyl-CoAs are converted to VLC-alkanes (C₂₉ and C₃₁) by the alkane-forming complex formed by CER1, CER3 and CYTB5s from *P. cablin*. The dashed arrow represents proposed steps remaining to be elucidated in this study. Enzymes catalysing each step are indicated in blue. ACC, acetyl-CoA carboxylase (multi-subunit); ACP, acyl carrier protein; CER, eceriferum; CYTB5, cytochromes b5 isoforms; ENR, enoyl-ACP reductase; FAS, fatty acid synthase; FATA(B), fatty acyl-ACP thioesterase A(B); HAD, hydroxyacyl-ACP dehydratase; KAR, ketoacyl-ACP reductase; KAS, ketoacyl-ACP synthase; MCAT, malonyl-CoA: ACP acyltransferase.

of endoplasmic reticulum (ER)-associated multienzyme fatty acyl-CoA elongase (FAE) complexes (Lee and Suh, 2015, 2022). In *Arabidopsis*, VLC-alkanes represent up to 70% of the total wax content in leaves (Bernard and Joubès, 2013).

VLCFAs in *Arabidopsis* are processed via the alkane-forming pathway, also called the decarbonylation pathway, to form the wax components VLC-alkanes, aldehydes, secondary alcohols and ketones. The elongation of C₁₆ and C₁₈ fatty acids, which are synthesized through the plastidial de novo synthesis pathway (Figure 1; Bonaventure *et al.*, 2003), to VLCFAs is required prior to the formation of waxes. It initially uses C₁₆ and C₁₈ fatty acids that are esterified to Co-enzyme A (CoA) before entering the ER-associated multienzyme FAE complexes (Figure 1; Schnurr *et al.*, 2004; Weng *et al.*, 2010; Zheng *et al.*, 2019). In the alkane-forming pathway, VLC-alkanes biosynthesis in *Arabidopsis*

involves two different alkane-forming complexes, the ECERIFERUM1(CER1)/CER3/CYTOCHROME B5(CYT5) complex catalysing the conversion of acyl-CoAs to VLC-alkanes with strict substrate specificity for compounds containing more than 29 carbons and the CER1-LIKE1/CER3/CYT5 complex preferring to catalyse the formation of VLC-alkanes with 25 and 27 carbon atoms (Figure S1b; Bernard *et al.*, 2012; Pascal *et al.*, 2019). Besides VLC-alkanes widely synthesized in land plants, a variety of other organisms including animals and microorganisms can convert fatty acids into alkanes, mainly to protect themselves from environmental threats (Jaroensuk *et al.*, 2020; Marsh and Waugh, 2013). Insects produce a large spectrum of hydrocarbons (from C₂₁ to C₃₇) that serve as essential waterproofing components on their epicuticle to prevent moisture loss and are also important for their social behaviours and insect-to-insect recognition (Qiu *et al.*, 2012). Insects convert VLC-acyl-CoA into VLC-alka(e)nes by the sequential activities of two unique enzymes, an acyl-CoA reductase and an insect-specific P450 oxidative decarbonylase (CYP4G) (Figure S1d; Qiu *et al.*, 2012).

Alkanes are the major constituents of fossil fuels and exist in the form of a gas, liquid or wax depending on the length and structure of the carbon backbone. Alkanes of various chain lengths are important targets for biotechnology for the next generations of biofuels that can effectively function as “drop-in” replacements for gasoline (mainly C₅–C₉ hydrocarbons), jet fuel (C₅–C₁₆) and diesel (C₁₂–C₂₀), leading to renewed interest, especially in the area how alkanes are biosynthesized and in the area of mechanistic enzymology (Choi *et al.*, 2020; Marsh and Waugh, 2013). Although the biosynthetic pathway of plant VLC-alkanes has been largely elucidated (Bernard and Joubès, 2013; Lee and Suh, 2015, 2022; Pascal *et al.*, 2019),

the use of plant hydrocarbons as a renewable source of liquid fuels is hampered by predominance of constituents with high carbon numbers ($>C_{25}$), which are in solid state at ambient temperature (Jetter and Kunst, 2008). Therefore, there is a great interest in finding unique plant species that can accumulate shorter chain hydrocarbons ($<C_{19}$), of which the discovery of the biosynthetic machinery would provide a toolbox of enzymes for generation of renewable plant sources of liquid fuels and large production of the next generations of biofuels in a heterologous host.

The top leaves of *Pogostemon cablin* at young stage could accumulate large amounts of the LC-alkane pentadecane, making the plant an excellent system for the elucidation of LC-alkane biosynthetic machinery in plant. Here, we show that *P. cablin* plants synthesize the LC-alkane pentadecane and VLC alkanes using distinct homologues of the Arabidopsis VLC alkane biosynthesis machinery, with no overlap between the two types of complexes. By silencing multifunctional acetyl-CoA carboxylase (ACC) that catalyses the carboxylation of acetyl-CoA to form malonyl-CoA, which is used in the cytosol in various biosynthetic pathways, especially including fatty acyl-CoA elongation process, the yield of LC-alkane pentadecane was dramatically increased in *N. benthamiana* leaves transiently co-expressing the newly identified LC-alkane biosynthesis gene *PcCER1-LIKE3* and *PcCER3*. Therefore, we provide the possibility that the newly discovered proteins involved in LC-alkane biosynthesis could be utilized in a heterologous plant host to produce LC-alkane pentadecane, an important liquid biofuel component.

Results

The top leaves of *P. cablin* at young stage accumulate large amounts of pentadecane

Pogostemon cablin plant, a traditional Chinese medical plant in the Lamiaceae family, is the natural source of *Pogostemonis Herba* and patchouli oil (its essential oil), which have been widely used in pharmaceutical and cosmetic industries (Chen et al., 2021b). Pogostone, one of the major constituents of patchouli oil, has been shown to exert various bioactive activities including antimicrobial and pharmacological activities (Chen et al., 2021b). During our previous analysis of pogostone distribution by gas chromatography mass spectrometry (GC–MS) in different tissues of *P. cablin* plant at young developmental stage (Chen et al., 2021a), we found that leaves of *P. cablin* also produce substantial amounts of the LC-alkane pentadecane (Figure 2a), suggesting this plant as a potential material for elucidation of LC-alkane biosynthetic machinery in plant. The levels of LC-alkane pentadecane in different tissues of *P. cablin* plant at different developmental stages were further determined by GC–MS analysis of methyl tert-butyl ether (MTBE)-extracted macerated tissues. This analysis showed that pentadecane is largely produced in *P. cablin* seedlings and in the top leaves of the main stem of plants younger than 7 weeks (Figure 2b). The top leaves of the main stem of plants older than 7 weeks contained much smaller amounts of pentadecane compared to that of younger plants, indicating that the distribution of pentadecane in *P. cablin* newly generated leaves from the main stem is developmentally controlled. Meanwhile, the levels of pentadecane in the older leaf tissue 7W-SLeaf, 8W-SLeaf and 9W-SLeaf steadily decreased (Figure 2b), suggesting the old leaves are losing pentadecane.

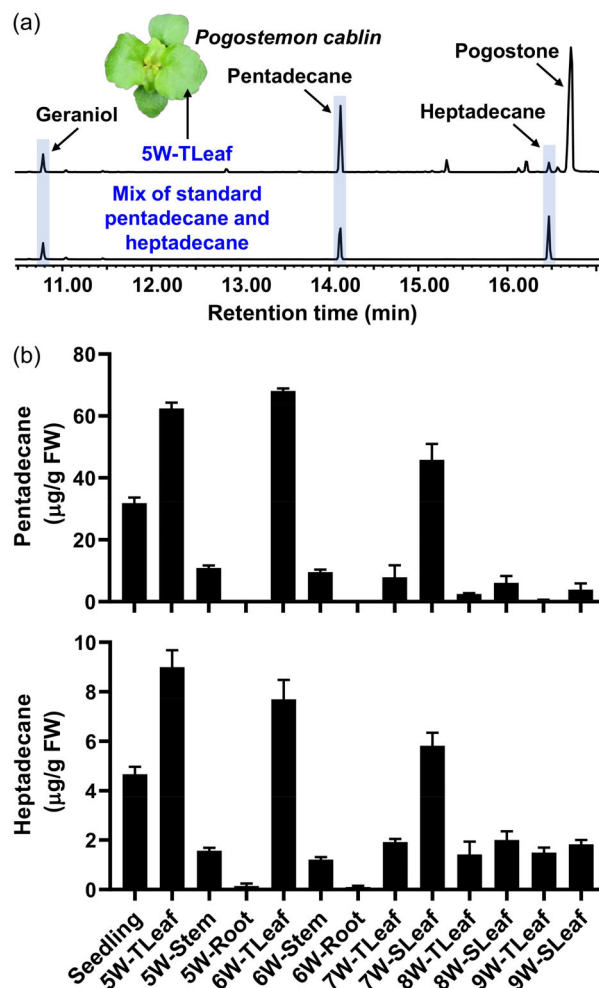


Figure 2 GC–MS analysis of LC-alkane pentadecane and heptadecane in different tissues of *Pogostemon cablin* at different development stages. (a) GC–MS chromatograms in total ion mode of MTBE (methyl tert-butyl ether) extracts from 5 W-TLeaf (the top leaves from five-week-old *P. cablin* plants). Geraniol was used as internal standard. The peaks of long-chain alkanes pentadecane and heptadecane as well as geraniol and pogostone were labelled. (b) Levels of pentadecane and heptadecane in different tissues of *P. cablin* at different development stages. The organic extracts were analysed by GC–MS and quantification was achieved by normalization of the peaks to the internal standard geraniol and comparison with standard curves of pentadecane and heptadecane. The data represent means \pm SD of four independent biological replicates. FW, fresh weight. The *P. cablin* tissues used in this study were designated in detail as previously described (Chen et al., 2021a), and the older leaf tissue 7W-SLeaf, 8W-SLeaf and 9W-SLeaf are the same tissues in different developmental stages with the younger leaf tissue 6W-TLeaf, 7W-TLeaf and 8W-TLeaf, respectively.

Besides pentadecane, another odd-number LC-alkane heptadecane was also detected in *P. cablin* plants and showed a similar distribution pattern to that of pentadecane in the tested tissues, albeit present at much lower levels (Figure 2). In addition, the VLC-alkanes such as nonacosane (C_{29}) and hentriacontane (C_{31}) that are generally the main components of plant cuticular wax alkanes remain at a relatively stable level in tested aerial parts of *P. cablin* (Figure S2).

Screening the candidate decarbonylases for LC-alkane biosynthesis in *P. cablin*

In general, the naturally synthesized alkanes comprise an odd number of carbons, suggesting that they are derived by loss of the carboxyl carbon from fatty acid with even numbers of carbons (Marsh and Waugh, 2013). The formation of LC-alkanes involves decarbonylation processes, the removal of the aldehyde carbonyl group from the intermediacy of the corresponding fatty aldehydes catalysed by a group of enzymes called aldehyde decarbonylases (AD) (Marsh and Waugh, 2013). The decarbonylation of the intermediate fatty aldehyde proceeds by different mechanisms depending on the organisms. The insect AD (CYP4G) is a membrane-bound P450-type enzyme while the cyanobacterial AD (ADO) is a soluble non-heme di-iron oxygenase (Khara et al., 2013; Qiu et al., 2012). The plant AD (CER1 or its homologues) is thought to be similar to the non-heme iron membrane proteins represented by stearoyl desaturase and fatty acid hydroxylase (Wang et al., 2019).

To identify candidate alkane-forming enzymes involved in LC-alkane pentadecane biosynthesis in *P. cablin* plant, we first performed the screening of candidate ADs by homology searches of our previously constructed *P. cablin* RNAseq database using aforementioned AD genes identified from different organisms including insects, cyanobacteria and plants through TBLASTN software. Our search identified four AtCER1 homologues and no other potential AD (Tables S1 and S2). The one with the highest amino acid identity with AtCER1 was designed as PcCER1, and the other three were in turn designed as PcCER1-LIKE1 ~ 3 according to the degree of their amino acid identities with PcCER1 (Figure S3). Sequence alignment showed that all four candidate ADs possess a putative fatty acid hydroxylase domain and WAX2 domain at the N- and C-terminal regions, respectively, and the three conserved CER1 His-rich clusters that have been shown to be essential for alkane synthesis (Figure S3; Bernard et al., 2012). These His-rich clusters are part of the catalytic site of CER1 that serve as ligands for a di-iron binding site mediating electron transfer from CYTB5s (Bernard et al., 2012). In addition, Microalgae fatty acid photodecarboxylase catalysing the decarboxylation of free fatty acid to alka(e)nes (Sorigué et al., 2017) was also utilized as the query gene, but no homologue was identified from *P. cablin* RNAseq database.

Since in Arabidopsis CER1 and its homologue CER1-LIKE1 physically interact with CER3 to form different alkane-forming complexes that produce a series of VLC-alkanes with different chain lengths (C_{25} ~ C_{37}) (Bernard et al., 2012; Pascal et al., 2019), we further screened the *P. cablin* RNAseq database for homologues of AtCER3. One gene, designed as PcCER3, was identified.

To test the hypothesis that pentadecane is produced in young *P. cablin* leaves by a combination of PcCER3 and one of the PcCER1 homologues, an expression correlation analysis to identify the best PcCER1 homologue candidate was first performed (Figure S4). To more accurately test whether the transcript profiles of these screened five genes correlated with the distribution of pentadecane in the 13 selected *P. cablin* tissues as previously described (Chen et al., 2021a) using Pearson correlation analysis, their transcript levels were tested using RT-qPCR in the same 13 selected tissues samples used for pentadecane measurement. This analysis showed that the expression patterns of both *PcCER1-LIKE2* and *PcCER1-LIKE3* in the tested tissue samples showed a strong positive correlation

with the biosynthesis of pentadecane (Table S3), displaying a similar developmental pattern as the distribution pattern of pentadecane (Figures 2b and 3). However, *PcCER1-LIKE3* showed much higher transcript levels than *PcCER1-LIKE2* in tested seedling tissue and leaf tissues of plants younger than 7 weeks (Figure 3), implying it is the best candidate gene.

PcCER1 and PcCER1-LIKE1 ~ 3 physically interact with PcCER3

In Arabidopsis both CER1 and its homologue CER1-LIKE1 can physically interact with CER3 to form an ER-localized complex that catalyses VLC-alkane formation in plant cells (Bernard et al., 2012; Pascal et al., 2019). To test the hypothesis that PcCER1 or its homologues can physically interact with PcCER3 to form a complex in plant cell, we first studied their subcellular localizations by transiently coexpressing ER-localized AtWAK2-mCherry fluorescent fusion protein with PcCER1-GFP, PcCER1-LIKE1-GFP, PcCER1-LIKE2-GFP, PcCER1-LIKE3-GFP or PcCER3-GFP fluorescent fusion proteins in Arabidopsis leaf protoplasts. Confocal microscopy analyses showed that all these five proteins colocalized with AtWAK2 at the ER (Figure S5). We next tested the physical interactions of PcCER1 and PcCER1-LIKE1 ~ 3 with PcCER3 using both split-ubiquitin yeast two-hybrid (SUY2H) assay and *Nicotiana benthamiana* transient split-luciferase assay. The physical interactions between PcCER3 and PcCER1 or its homologues PcCER1-LIKE1 ~ 3 were verified in SUY2H assay by the observation that the yeast cells coexpressing NubG-PcCER3 with Cub-PcCER1 or Cub-PcCER1-LIKE1 ~ 3 were able to grow on different selective media and show β -galactosidase activities unlike yeast cells coexpressing Cub and NubG-PcCER3 as a negative control (Figure 4a). The physical interaction between PcCER1 and its homologues PcCER1-LIKE1 ~ 3 with PcCER3 was further tested by split-luciferase assay. LUC activities detected in infiltrated *N. benthamiana* leaves further indicated that both PcCER1 and its homologues PcCER1-LIKE1 ~ 3 can not only interact with PcCER3 to form a complex but also interact with themselves in plant cells (Figure 4b).

Coexpression of PcCER1-LIKE3 and PcCER3 in *N. benthamiana* leaves largely increases LC-alkane pentadecane and heptadecane production

To functionally characterize the alkane-forming complexes formed by PcCER3 and each of PcCER1 and its homologues PcCER1-LIKE1 ~ 3 *in planta*, we transiently expressed these complexes and their individual components in *N. benthamiana* leaves. No significant changes in levels of odd-number LC and VLC-alkanes in *N. benthamiana* leaves transiently expressing *PcCER3* alone were observed compared with that in *N. benthamiana* leaves expressing empty vector as a control (Figure 5). Transient coexpression of PcCER1-LIKE3 and PcCER3 in *N. benthamiana* leaves led to the largest increase in pentadecane and heptadecane production among the four complexes and their individual components, representing 52.5-fold increase in pentadecane level and 7.0-fold increase in heptadecane level than that in control *N. benthamiana* leaves (Figure 5). Coexpression of PcCER3 and PcCER1-LIKE1 or PcCER1-LIKE2 also led to an increase in LC-alkane pentadecane and heptadecane production compared with that in control *N. benthamiana* leaves, albeit to a much lower extent than expression of the complex formed by PcCER1-LIKE3 and PcCER3 (Figure 5). These results indicated that the complex formed by PcCER1-LIKE3 and PcCER3 is the predominant catalyst responsible for the formation of

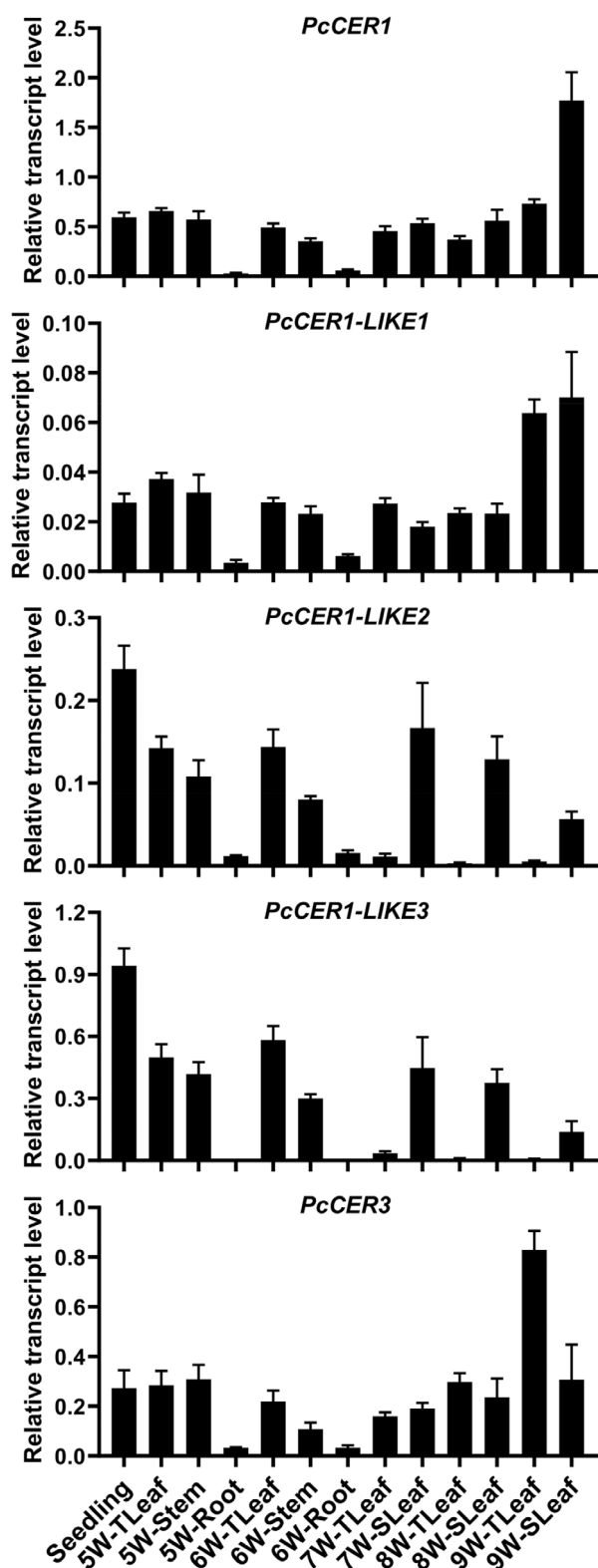


Figure 3 RT-qPCR analysis of transcript levels of *PcCER1*, *PcCER1-LIKE1* ~ 3 and *PcCER3* in several *Pogostemon cablin* tissues at different developmental stages. The RT-qPCR data were expressed as relative transcript levels of these indicated genes relative to *P. cablin* *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase), *Actin7* and *Tubulin3* in each tissue sample. Data are means \pm SD of four independent biological replicates.

pentadecane and heptadecane compared with the complexes formed by *PcCER3* and each of *PcCER1*, *PcCER1-LIKE1* and *PcCER1-LIKE2* (Figure 5).

Besides the LC-alkanes pentadecane and heptadecane, some other LC and VLC-alkane production in *N. benthamiana* leaves was also altered when the complex formed by *PcCER3* and each of *PcCER1* and *PcCER1-LIKE1* ~ 3 was transiently expressed (Figure S6). The nonadecane (C_{19}) level in *N. benthamiana* leaves was greatly elevated when all four possible complexes (*PcCER3* and each of *PcCER1-LIKE1* ~ 3) were tested, but the complexes of *PcCER3* and *PcCER1-LIKE2* or *PcCER3* and *PcCER1-LIKE3* led to much larger increase compared with *PcCER3* and *PcCER1-LIKE1* (52.6-fold, 46.9-fold and 13.8-fold compared with that in control *N. benthamiana* leaves (Figure S6)). On the other hand, transient expression of the complex formed by *PcCER3* and *PcCER1-LIKE1* or *PcCER1-LIKE2* in *N. benthamiana* leaves led to a much larger increase in VLC-alkane heneicosane (C_{21}), tricosane (C_{23}) and pentacosane (C_{25}) production compared with that in control *N. benthamiana* leaves than coexpression of *PcCER3* and *PcCER1-LIKE3*, which led to a relatively lower increase in those VLC alkenes (39.0-fold, 33.5-fold and 9.2-fold increase, respectively) (Figure S6). These results indicated that the complex formed by *PcCER3* and *PcCER1-LIKE1* or *PcCER1-LIKE2* is more efficient in catalysing the formation of VLC-alkanes with carbon numbers ranging from 21 to 25 compared with the complex formed by *PcCER3* and each of *PcCER1* and *PcCER1-LIKE3*. The level of heptacosane (C_{27}) and nonacosane (C_{29}) in *N. benthamiana* leaves expressing the complex formed by *PcCER3* and each of *PcCER1*, *PcCER1-LIKE1* and *PcCER1-LIKE2* also increased significantly compared with that in control *N. benthamiana* leaves, while no significant change in the level of hentriacontane (C_{31}) and tritriacontane (C_{33}) was observed in transgenic *N. benthamiana* leaves, which yielded almost the same content of these two chemicals as the control *N. benthamiana* leaves (Figure S6). These results indicate that *P. cablin* plant contains multiple alkane-forming complexes that can produce a series of *n*-alkanes with different chain lengths.

Reconstitution of LC-alkane pentadecane biosynthesis in yeast

In yeast, LC-acyl-CoAs including C_{16} acyl-CoA and C_{18} acyl-CoA are generated by the type I fatty acid synthase (FAS) system and elongated up to C_{26} acyl-CoA in endoplasmic reticulum (Runguphan and Keasling, 2014; Yu et al., 2017), probably supplying substrates for production of alkanes by heterogeneous expression of alkane-forming complex. To further verify the catalytic activities of the complex formed by *PcCER3* and each of *PcCER1* and its three homologues from *P. cablin*, we first overexpressed each of *PcCER3*, *PcCER1* and *PcCER1-LIKE1* ~ 3 driven by GAL promoters in DSY-1 yeast strains, and found that pentadecane and heptadecane can be detected in small amounts in the hexane extract from the lysate of different transgenic yeast cells and control yeast cells harbouring empty vectors after 96 h of galactose incubation, and their yields did not alter significantly among transgenic and control yeast cells (Figure 6), while other alkanes with more than 17 carbon atoms cannot be detected. We then overexpressed each of the complexes formed by *PcCER3* and each of *PcCER1* and *PcCER1-LIKE1* ~ 3 driven by GAL promoters in DSY-1 yeast strains. Similar cases about the alkane production as above were observed in these transgenic yeast cells and control yeast cells, except that the yield of pentadecane in the yeast cells

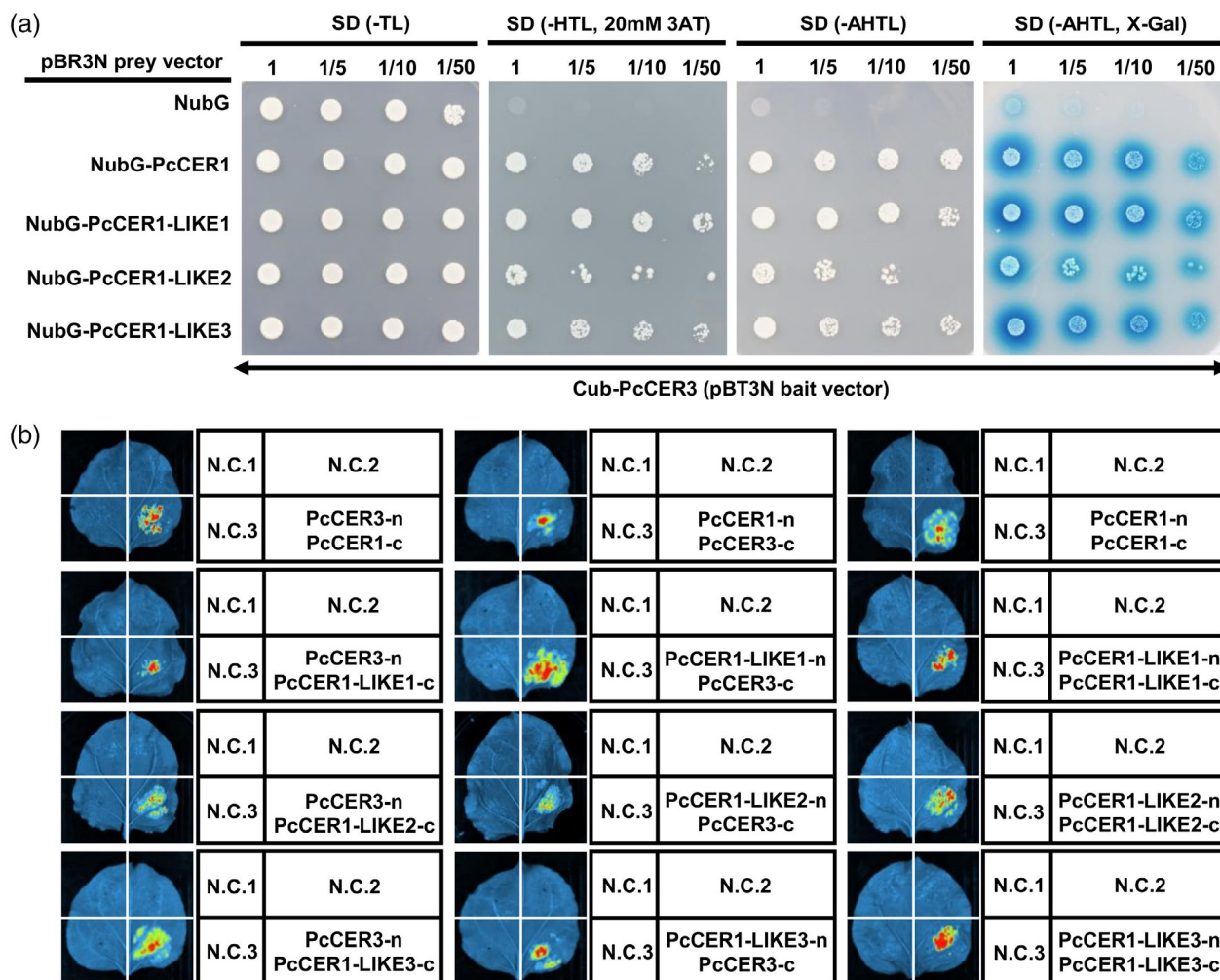


Figure 4 Direct interactions between PcCER3 and members of PcCER1 subfamily. (a) Direct interactions of PcCER3 with each of PcCER1, PcCER1-LIKE1, PcCER1-LIKE2 and PcCER1-LIKE3 in the SUY2H system. Yeast cells transformed with Cub-PcCER3 (in pBT3N bait vector) and NubG-PcCER1, NubG-PcCER1-LIKE1, NubG-PcCER1-LIKE2 or NubG-PcCER1-LIKE3 (all in pBR3N prey vector) are able to grow on the selective SD medium lacking His, Leu and Trp and containing 20 mM 3-amino-1,2,4-triazole (3AT) (-HTL, 20 mM 3AT) and the selective medium lacking His, Leu, Trp and adenine (-AHTL), and show β -galactosidase activity (-AHTL, X-Gal). Yeast cells cotransformed with Cub-PcCER3 and empty pBR3N prey vector (NubG) were used as negative control and cannot grow on any aforementioned selective SD mediums. (b) Direct interactions between PcCER3 and members of PcCER1 subfamily containing PcCER1, PcCER1-LIKE1, PcCER1-LIKE2 and PcCER1-LIKE3 in *N. benthamiana*. Luciferase image of *N. benthamiana* leaves coinfiltrated with different combinations of the agrobacteria containing indicated vectors. N.C.1, negative control 1 (nLuc + cLuc); N.C.2, negative control 2 (Gene-nLuc + cLuc); N.C.3, negative control 3 (nLuc + Gene-cLuc).

overexpressing the complex formed by PcCER3 and PcCER1-LIKE3 was increased by 3.6-fold compared with that in control yeast cells (Figure 6). These results suggested that the complex formed by PcCER3 and PcCER1-LIKE3 prefer to utilize the native C₁₆ acyl-CoA to generate pentadecane in yeast, which is consistent with its catalytic preferences in *N. benthamiana* leaves.

In Arabidopsis, CER1 was found to interact with the cytochrome b5 isoforms (CYTB5s) serving as ER-reducing components, and overexpression of cytochrome b5 isoform B (CYTB5-B) in the yeast cells coexpressing CER1 and CER3 were found to increase alkane production (Bernard *et al.*, 2012). We screened the closest homologue of Arabidopsis CYTB5-B from *P. cablin* RNAseq database and designated it as PcCYTB5-B. To check if we can further improve the yield of pentadecane in yeast cells overexpressing the complex formed by PcCER3 and PcCER1-LIKE3, we further coexpressed PcCYTB5-B in such transgenic

yeast cells. As expected, overexpression of PcCYTB5-B in the yeast cells coexpressing PcCER3 and PcCER1-LIKE3 further led to improvement of pentadecane production by 2.2-fold and even heptadecane production by 2.4-fold (Figure 6).

Silencing acetyl-CoA carboxylase dramatically increases pentadecane production in *N. benthamiana* leaves co-expressing PcCER3 and PcCER1-LIKE3

Our results show that the ER-localized complex formed by PcCER3 and PcCER1-LIKE3 catalyses the formation of LC-alkane pentadecane and heptadecane from C₁₆ and C₁₈ acyl-CoA, respectively. In plant, the saturated C₁₆ and C₁₈ acyl-CoA can be further elongated to VLC acyl-CoA by FAE complexes, providing precursors for the formation of a series of compounds including VLC-alkanes (Figure 7a; Bernard and Joubès, 2013; Lee and Suh, 2022). Theoretically, the biosynthesis of LC-alkane

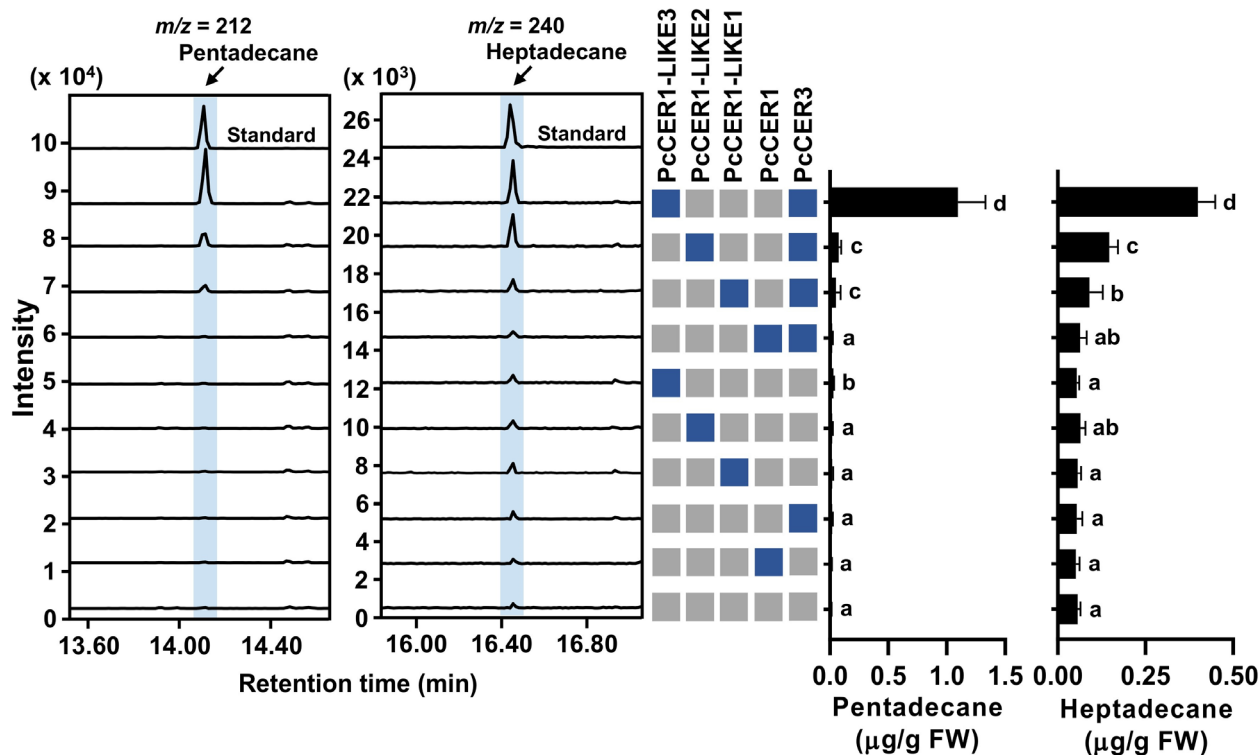


Figure 5 GC-MS analysis of the odd-numbered LC-alkane pentadecane and heptadecane in *Nicotiana benthamiana* leaves expressing indicated genes or gene combinations. Extracted ion chromatograms of both m/z 212 for pentadecane and m/z 240 for heptadecane of MTBE extracts of *N. benthamiana* leaves transiently expressing indicated genes or gene combination are shown as graphs on the left. Blue box indicates the gene that is transiently expressed in *N. benthamiana* leaves. *Nicotiana benthamiana* leaves transformed with empty vector (EV) were used as a negative control (all grey boxes). Levels of the odd-numbered LC-alkane pentadecane and heptadecane in *N. benthamiana* leaves expressing indicated genes and gene combinations are also shown. The quantification was achieved by normalization of the peaks to the internal standard geraniol and comparison with the standard curves of authentic standards. Data are expressed as means \pm SD of eight independent biological replicates. Significant differences among the levels of LC and VLC-alkanes in *N. benthamiana* leaves expressing indicated genes and gene combinations were tested using independent t -tests $P < 0.05$ (Two-tailed distribution; Two-sample equal variance). Different letters to the right of the bars mark the statistically significant groups after t -tests. FW, fresh weight.

pentadecane and heptadecane competes with fatty acyl-CoA elongation (FAE) process.

A multifunctional acetyl-CoA carboxylase (ACC) that catalyses the carboxylation of acetyl-CoA in the cytosol to form malonyl-CoA, an essential building block for FAE process to create the VLC-acyl-CoA precursors for VLC-alkane biosynthesis, has been described in several plants (Baud *et al.*, 2003; Lu *et al.*, 2011). In Arabidopsis, partial blockage of the ER-based FAE process through mutation of AtACC1 (At1g36160), resulting in the reduced production of malonyl-CoA, led to the accumulation of free C₁₆ and C₁₈ acyl-CoAs, although some of it was shunted into the cutin, suberin and TAG biosynthetic pathways (Lu *et al.*, 2011). This blockage of malonyl-CoA biosynthesis through the AtACC1 mutation also resulted in a large reduction of cuticular wax constituents such as VLC-alkanes (Lu *et al.*, 2011). This observation allowed us to postulate that silencing the multifunctional NbACC to partially block the ER-based FAE process may provide more C₁₆ and C₁₈ acyl-CoA precursors for LC-alkane pentadecane and heptadecane biosynthesis in *N. benthamiana* leaves co-expressing PcCER3 and PcCER1-LIKE3.

To test this hypothesis, we first screened candidate multifunctional ACC from *N. benthamiana* genome by homology searches using AtACC1 protein (Figure S7). Our search identified

four isoforms of NbACCs which are designated as NbACC1 ~ 4. The four NbACC isoforms show all around 80% identities with AtACC1 and 89% ~ 97% identities among themselves at the protein level (Figure S7). Virus-induced gene silencing (VIGS) of all four NbACC isoform genes in *N. benthamiana* plants was performed by infiltration with pTRV2-NbACC construct inserted with the NbACC4 gene fragment within their highly conserved regions (Figure S8a). The NbACC-VIGS leaves of *N. benthamiana* plants show a wrinkled and folded phenotype compared with that of control *N. benthamiana* plants infiltrated with pTRV2 empty vector (EV) (Figure 7b), similar to the observation in AtACC1-mutated Arabidopsis plant (Lu *et al.*, 2011). The total relative transcript levels of NbACC1 ~ 4 in NbACC-VIGS leaves decreased by 83.4% compared with that in control *N. benthamiana* leaves (Figure 7c; Figure S8b). Accordingly, the levels of VLC-alkanes nonacosane (C₂₉), hentriacontane (C₃₁) and tritriacontane (C₃₃) decreased by 65%, 72% and 80% in NbACC-VIGS leaves, respectively, compared with that in control leaves (Figure 7d). As predicted, the yields of LC-alkane pentadecane and heptadecane in NbACC-VIGS leaves co-expressing PcCER3 and PcCER1-LIKE3 increased by 3.7-fold and 1.5-fold, respectively, compared with that in control leaves co-expressing PcCER3 and PcCER1-LIKE3 (Figure 7e).

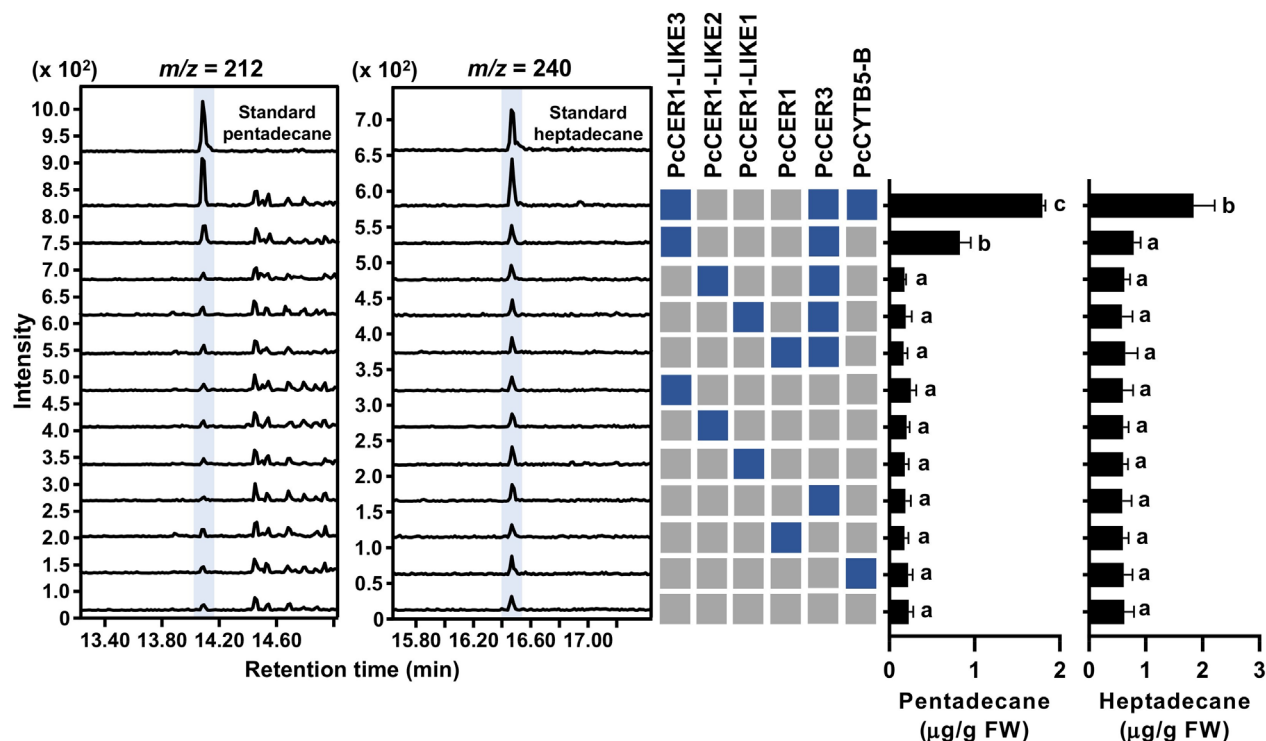


Figure 6 GC–MS analysis of LC-alkane pentadecane and heptadecane in yeast strain (DSY-1) expressing indicated genes or gene combinations. Extracted ion chromatograms of m/z 212 for pentadecane and m/z 240 for heptadecane of MTBE extract of DSY-1 yeast strains expressing indicated genes or gene combinations are shown as graphs in left. Blue box indicates the gene that is expressed in DSY-1 yeast cells. DSY-1 yeast strains transformed with empty vector (EV) were used as a negative control (all grey boxes). Level of pentadecane and heptadecane in the DSY-1 yeast strains expressing indicated genes and gene combinations are also shown in right. The quantifications were achieved by normalization of the peaks to the internal standard geraniol and comparison with the standard curves of authentic standards. Data are expressed as means \pm SD of four independent biological replicates. Significant differences among the levels of pentadecane and heptadecane in DSY-1 yeast strains expressing indicated genes and gene combinations were tested using independent t -tests $P < 0.05$ (Two-tailed distribution; Two-sample equal variance). Different letters to the right of the bars mark the statistically significant groups after t -tests. FW, fresh weight.

Discussion

Biosynthesis of the LC-alkane pentadecane in *P. cablin* involves the complex formed by PcCER3 and PcCER1-LIKE3

The aerial surfaces of land plants are covered with the cuticle layer consisting of two hydrophobic components: cutin and cuticular waxes, which act as protective barriers to protect them from various environmental stresses such as severe water deficiency and excessive solar and UV radiation (Bernard and Joubès, 2013; Bowman *et al.*, 2017; Lee *et al.*, 2021; Samuels *et al.*, 2008). VLC-alkanes are a major component of cuticular waxes (Lee and Suh, 2015), but plants also produce shorter alkanes, which could be used as biofuels. Therefore, the complete elucidation of the alkane biosynthetic pathways in plants would also provide a toolbox of enzymes for the production of renewable hydrocarbon sources and next generations of biofuels (Joyce and Stewart, 2012). The VLC-alkane biosynthetic pathway of plants has been largely elucidated (Bernard *et al.*, 2012; Lee and Suh, 2015; Pascal *et al.*, 2019). Biosynthesis of VLC-alkanes in *Arabidopsis* involves two coexisting alkane-forming complexes with distinct chain-length specificities. The complex made of *Arabidopsis* CER1 and CER3 catalyses the conversion of acyl-CoAs to VLC-alkanes mainly with 29 carbons in engineered yeast

INVSUR4# which allowed production of VLCFAs (C_{28} and C_{30}) that are not otherwise synthesized by yeast (Bernard *et al.*, 2012), while the complex made of *Arabidopsis* CER1-LIKE1 and CER3 catalyses the formation of VLC-alkanes with 25 and 27 carbon atoms in such engineered yeast (Pascal *et al.*, 2019).

In addition, CER1 and CER1-LIKE1 were found to interact with the cytochrome b5 isoforms (CYTB5s) serving as ER-localized reducing components (Bernard *et al.*, 2012; Pascal *et al.*, 2019), suggesting that cytochrome b5 may be a cofactor for CER1 and CER1-LIKE1. A set of data collected on *Arabidopsis* ER-localized CER1/CER3 complex suggests that CER3 catalyses the reduction of VLC-acyl-CoA into aldehydes, which, remaining potentially bound to the complex, are decarbonylated by CER1 into n -alkanes and carbon monoxide (Pascal *et al.*, 2019). As the plant AD enzyme, CER1 is an integral membrane protein that has some sequence similarity to the fatty acid hydroxylase superfamily and stearoyl-CoA desaturase (Aarts *et al.*, 1995). The difficulties associated with expressing and purifying eukaryotic membrane proteins have hindered the mechanistic analysis of the reaction catalysed by the plant AD enzymes, making it the most enigmatic (Marsh and Waugh, 2013).

The LC-alkane pentadecane, a liquid hydrocarbon at ambient temperature, is a major component of transportation fuels such as diesel. We found that pentadecane is largely produced in

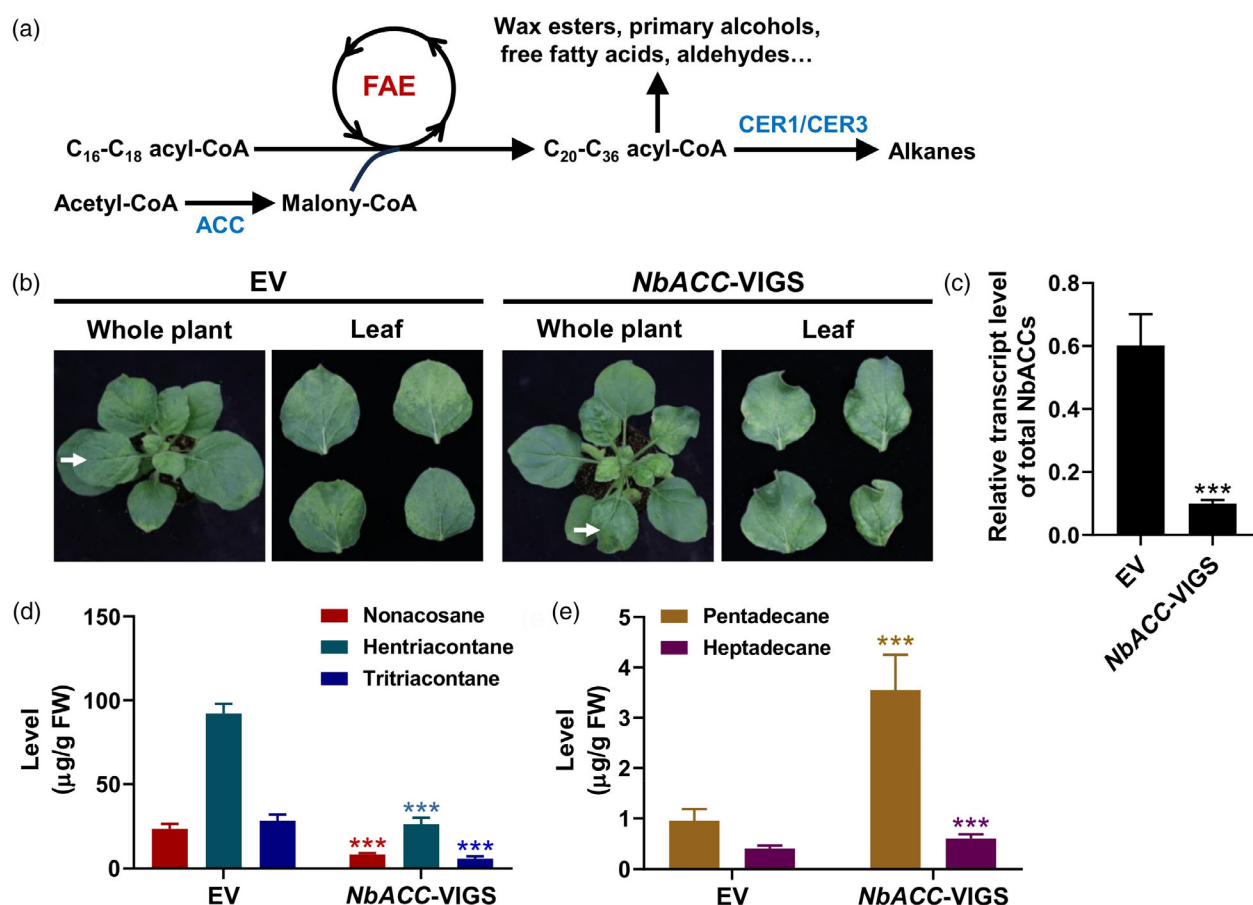


Figure 7 Improvement of pentadecane and heptadecane biosynthesis in *Nicotiana benthamiana* leaves coexpressing PcCER3 and PcCER1-LIKE3. (a) A multifunctional acetyl-CoA carboxylase (ACC) is involved in the fatty acyl-CoA elongation (FAE) process for alkane biosynthesis in plant (Bernard and Joubès, 2013; Lu et al., 2011). The first step of the elongation cycle is the condensation of two malonyl-CoA to the acyl-CoA chain. (b) Phenotypes of *N. benthamiana* plant infiltrated with empty TRV (EV, control) and NbACC-VIGS construct at 2 weeks after infiltration. The whole plant (left) and leaves (right) that are indicated in location by white arrow in whole plant (left) are shown. The shown leaves correspond in location and developmental stages to that of *N. benthamiana* plant infiltrated with NbPDS-VIGS construct which shows strong photobleaching phenotype under the same growth condition. (c) Relative transcript levels of total NbACCs (NbACC1 ~ 4) in the indicated leaves shown in (a) of *N. benthamiana* plant infiltrated with empty TRV (EV) and NbACC-VIGS construct at 2 weeks after infiltration. The RT-qPCR data were expressed as relative transcript levels of total NbACCs relative to *N. benthamiana* GAPDH (Niben101Scf05634g01008.1). Data are means \pm SD of nine independent biological replicates. (d) The levels of VLC-alkane nonacosane (C₂₉), hentriacontane (C₃₁) and tritriacontane (C₃₃) in the indicated leaves shown in (a) of control and NbACC-VIGS *N. benthamiana* plants. Data are means \pm SD of 10 independent biological replicates. (e) The level of LC-alkane pentadecane and heptadecane in the indicated leaves shown in (a) that were harvested at 7 days after co-infiltration with pEAQ-PcCER3 and pEAQ-PcCER1-LIKE3. Data are means \pm SD of 10 independent biological replicates. Significant differences between data from control (EV) and NbACC-VIGS *N. benthamiana* leaves were tested using independent *t*-tests (****P* < 0.001).

P. cablin seedlings and the top leaves of the main stem of the plants younger than 7 weeks, and its production is under strict developmental control (Figure 2b), which makes the plant an excellent system for discovery of plant LC-alkane biosynthetic machinery. Through homology searches on our previously constructed *P. cablin* RNAseq database using the identified AD genes from different lineages including insect, cyanobacteria, green algae and plant, only four candidate *P. cablin* ADs that are homologues of Arabidopsis CER1 were screened and designated as PcCER1 and PcCER1-LIKE1 ~ 3. Accordingly, only one homologue of Arabidopsis CER3 that catalyses the formation of aldehyde precursors for Arabidopsis CER1 was also screened and named as PcCER3. The expression patterns of PcCER1-LIKE2 and PcCER1-LIKE3 among the screened four *P. cablin* ADs display a similar developmental pattern as the distribution pattern of pentadecane in tested *P. cablin* tissues (Figure 3; Table S3). The

much higher expression levels of PcCER1-LIKE3 in tested seedling tissue and young leaf tissues than PcCER1-LIKE2 make it the best candidate AD for pentadecane biosynthesis (Figure 3).

The catalytic activities of the ER-localized complexes formed by PcCER3 and each of the candidate *P. cablin* ADs were investigated in *N. benthamiana* leaves by transient co-expression and indicated that *P. cablin* plant contains several alkane-forming complexes with distinct chain-length specificities (Figure 5; Figure S6). Among these four complexes, the complex formed by PcCER3 and PcCER1-LIKE3 is more efficient in catalysing the formation of LC-alkane pentadecane and heptadecane compared with the other three complexes (Figure 5), suggesting that PcCER1-LIKE3 is the *P. cablin* AD that catalyse the decarbonylation of hexadecanal that are formed by reduction of C₁₆ acyl-CoA catalysed by PcCER3 to generate pentadecane. Failure to establish a reliable *P. cablin* transformation system, including gene

silencing (VIGS) using tobacco rattle virus, prevents us from testing the biological functions of these *P. cablin* ADs and PcCER3 directly in *P. cablin* plant. Therefore, we further tested the catalytic activities of these complexes formed by PcCER3 and each of the candidate *P. cablin* ADs in the yeast system (Figure 6). The obvious differences with the observations in tobacco system are that the complex formed by PcCER3 and PcCER1-LIKE3 increased the LC-alkane pentadecane and heptadecane production in yeast to a much lower extent than in *N. benthamiana* leaves, and no other longer-chain alkanes are detected in yeast even under the expression of the complex formed by PcCER3 and each of PcCER1 and its homologues (Figure 6), probably because of the lower endogenous LC and VLC-acyl-CoA pools in yeast than in *N. benthamiana* leaves. Overall, these results strengthen the conclusion that the ER-localized complex formed by PcCER3 and PcCER1-LIKE3 is involved in the biosynthesis of LC-alkane pentadecane in *P. cablin* plant. In addition, the highly elevated production of heptadecane in *N. benthamiana* leaves coexpressing *PcCER3* and *PcCER1-LIKE3* and yeast cells coexpressing *PcCER3*, *PcCER1-LIKE3* and *PcCYTB5-B* suggested that the biosynthesis of heptadecane at relatively low levels in *P. cablin* plant is most likely the result of the action of PcCER1-LIKE3.

The identified *P. cablin* LC-alkane forming complex provides a new toolbox of enzymes for production of the biofuel component pentadecane in a heterologous host

The increasing global demand for transportation fuel in conjunction with the rapidly depleting petroleum reservoirs and growing environmental concerns on using non-renewable petroleum feedstock have emphasized the necessity of replacing fossil fuels with clean renewable fuels (Choi *et al.*, 2020; Liu *et al.*, 2021; Zhang *et al.*, 2011). To date, several biofuel compounds such as bioethanol and biodiesel have been commercially produced, yet most of the processes depend on fermenting sugars derived from edible biomass, such as corn, sugar cane and vegetable oils, challenging the world food supply and causing a series of economic problems (Pandiyani *et al.*, 2019; Peralta-Yahya *et al.*, 2012). In addition, the low energy density and hygroscopicity of bioethanol have become an obstacle to its wider application, and the high oxygen content of biodiesel promotes its immiscibility with fossil fuels, increases its viscosity and makes it very reactive and unstable (Mohammad *et al.*, 2013). It is widely accepted that the bio-hydrocarbons, especially the medium- to long-chain alka(e)nes, are the ideal next generation of biofuels because they highly mimic the chemical composition and physical characteristics of petroleum-based fuels (Fatma *et al.*, 2018; Li *et al.*, 2010; Schirmer *et al.*, 2010). Therefore, production of alka(e)nes via microorganisms by applying metabolic engineering strategies is considered as one of the most promising ways to produce drop-in biofuels. Alkanes are widely synthesized by many organisms including microorganisms, animals and plants, mainly for protecting themselves from various environmental stresses (Marsh and Waugh, 2013). Most pathways leading to alkanes are via fatty aldehydes. Aldehyde decarbonylases (ADs) have been discovered in various lineages including plants, insects and cyanobacteria (Bernard *et al.*, 2012; Pascal *et al.*, 2019; Qiu *et al.*, 2012; Schirmer *et al.*, 2010), and can convert fatty aldehydes to alka(e)nes with coproduction of carbon monoxide, carbon dioxide or formate depending on the organisms (Marsh and Waugh, 2013).

There are two routes available for production of alka(e)nes via ADs in microorganisms, and these are from either fatty acyl-CoAs or from free fatty acids by the action of fatty-acyl carrier protein reductase (AAR) or fatty acid reductase (FAR), respectively (Kang and Nielsen, 2017). Cyanobacteria AD, renamed as aldehyde deformylating oxygenase (ADO) due to its actual coproduct formate, has been frequently used for reconstitution of alka(e)nes biosynthetic pathway through coexpression of AAR, FAR or carboxylic acid reductase (CAR) in engineered microorganisms (Basri *et al.*, 2020; Jaroensuk *et al.*, 2020), and the expression of cyanobacteria AAR and ADO together in engineered strains leads to production of predominantly C15 and C17 alkanes (Schirmer *et al.*, 2010; Song *et al.*, 2016; Zhou *et al.*, 2016). The Arabidopsis aldehyde decarbonylase CER1, producing VLC-alkanes mainly with 29 carbons in yeast and plants (Bernard *et al.*, 2012; Pascal *et al.*, 2019), has been used to demonstrate alkane biosynthesis in heterologous microbial systems. This enzyme was even shown to enable production of C8-C14 alkanes up to 580.8 mg/L in an *E. coli* strain that was engineered to produce short-chain fatty acids (Choi and Lee, 2013), which implies the potential advantages of utilizing the newly identified PcCER1-LIKE3, which is shown to produce predominantly LC-alkane pentadecane in yeast and plant (Figures 5 and 6), to reconstitute LC-alkane biosynthesis in an engineered microorganism. Even though many enzymes have been applied to reconstruct the alka(e)ne biosynthesis in microorganisms, they are commonly not very efficient and it is, therefore, difficult to reach high productivities. For instance, the low turnover rate of cyanobacteria ADO results in a deficient amount of alkane conversion (Andre *et al.*, 2013). Besides the efforts to improve enzyme activities through different modification strategies such as directed evolution and immobilization, finding a new metabolic reaction through enzyme discovery is still required to expand the product portfolio and production capacity. For example, a recently discovered fatty acid photodecarboxylase that directly converts free fatty acids (FFAs) to hydrocarbons and prefers acyl-CoAs rather than FFAs as its substrates have replaced the ADO-dependent, inefficient conversion of fatty aldehydes to alka(e)nes (Bruder *et al.*, 2019; Li *et al.*, 2020; Sorigué *et al.*, 2017). Although the requirement of blue light for the catalysis of this enzyme complicates the fermentation process, the finding of this new enzyme implies that novel and efficient enzymes could be discovered by intensive and systematic studies.

The sustainable production of advanced biofuel alka(e)nes by biotechnological means in a world beyond fossil fuels required a diverse enzymatic toolkit. Newly identified enzymes, engineered enzymes with improved performances and new synthetic pathways designed for higher metabolic flux toward target products have presented new possibilities to further enhance biofuel production in an engineered host (Choi *et al.*, 2020; Liu *et al.*, 2021; Monteiro *et al.*, 2022). Engineering plants for accumulation of large amounts of biofuel components in their leaves, roots, seeds, or fruits may enable ultimately sustainable biofuel production (i.e., from CO₂) at costs as low as large-scale cultivation of grains and vegetables (Choi *et al.*, 2020). The catalytic enzymes of plant origin such as PcCER1-LIKE3 enzymatic complexes for LC-alkane pentadecane biosynthesis would be more feasible in plant metabolic engineering for biofuel production. In Arabidopsis, partial blockage of the ER-based FAE process through mutation of *AtACC1* leads to the reduced production of malonyl-CoA and, subsequently, the more accumulation of C₁₆ and C₁₈ acyl-CoA (Lu *et al.*, 2011). Accordingly,

silencing *NbACCs* dramatically improves the LC-alkane pentadecane biosynthesis in *N. benthamiana* leaves co-expressing PcCER1-LIKE3 and PcCER3, suggesting a strategy for future plant engineering to increase the yield of this liquid biofuel component.

Phylogenetic relationships of *P. cablin* ADs and their homologues

Our obtained results showed that the complexes formed by PcCER3 and each of the candidate ADs from *P. cablin* have distinct preferences to catalyse the formation of alkanes with a broad range of chain lengths including long chain and very long chain, suggesting the different substrate specificities of these *P. cablin* ADs. To gain further insight into the evolution of *P. cablin* ADs, we expanded the phylogenetic analysis of *P. cablin* ADs to include their closest homologues from selected plant species including chlorophyte, monocot and eudicot species (Figure 8). It has been shown that chlorophyte is one of the earliest taxa of Viridiplantae in which various alkane-forming pathways/enzymes emerged, and the duplication of *CER1* genes happened in higher land plants to form orthologous and paralogous groups, achieving functional redundancy and divergence (Wang et al., 2019). The *CER1* homologues retrieved from the selected monocots by homology search using *P. cablin* ADs formed a distinct subclade with that from selected eudicots and *P. cablin* ADs (Figure 8). *Pogostemon cablin* ADs were grouped together with their respective homologues from their relative Asterids, suggesting their closest relationships (Figure 8). However, PcCER1-LIKE2 and PcCER1-LIKE3 are grouped in a distinct subclade from PcCER1 and PcCER1-LIKE1 (Figure 8), implying functional divergence, shown by the different catalytic preferences of their corresponding complexes formed with PcCER3.

Materials and methods

Plant materials, growth conditions and chemicals

Pogostemon cablin seeds were directly germinated in soil. The plants of *P. cablin* and *N. benthamiana* were grown under the growth conditions as described previously (Chen et al., 2021a). The various tissues were collected from *P. cablin* and *N. benthamiana* plants, and the partially harvested tissues were flash-frozen in liquid nitrogen and stored at -80°C for RT-qPCR analysis. Arabidopsis plants selected for protoplast isolation for subcellular localization assay were grown under the conditions as described previously (Yoo et al., 2007). Standard alkanes with different chain lengths were all purchased from Sigma-Aldrich.

Gene identification and coexpression analysis

The putative genes involved in C_{16} and C_{18} fatty acid biosynthesis and export as well as candidate long-chain acyl-CoA synthetases (shown in Figure 1) were screened by functional annotations in our *P. cablin* transcriptome database (bioproject accession number, PRJNA713906) and using homology searches with Arabidopsis representative genes as previously described (Wang et al., 2022). Identification of candidate alkane-forming enzyme genes using homology searches from our previously constructed *P. cablin* transcriptome was performed by querying our local nucleotide database with amino acid sequences of the alkane-forming enzymes identified from different lineages including plant, insect, cyanobacteria and microalgae using the TBLASTN. The screened genes from *P. cablin* transcriptome were supplied as Tables S1 and S2, and the alkane-forming enzyme genes used in this screening were supplied as Table S4. Coexpression analysis

was performed by comparing the normalized abundances of pentadecane and heptadecane in different *P. cablin* tissues and the normalized transcript abundance of selected genes in the same selected tissues using the Pearson correlation method to quantitate their correlation (Chen et al., 2020; Li et al., 2018).

Quantitative RT-PCR analysis of the identified gene candidates

Quantitative real-time PCR (RT-qPCR) analysis of transcripts in different tissues of *P. cablin* plant at different developmental stages and in *N. benthamiana* leaves was performed as described previously (Wang et al., 2022). The relative expression of each gene relative to reference genes (*GAPDH*, *Actin7* and/or *Tubulin3*) in each tissue was calculated as previously described (Chen et al., 2021a). Primers used in this study are listed in the Table S5.

Subcellular localization of PcCER3, PcCER1 and PcCER1-LIKE1 ~ 3

The full-length open reading frames of PcCER3, PcCER1 and PcCER1-LIKE1 ~ 3 were obtained by RT-PCR from prepared cDNA of *P. cablin* seedlings and cloned into pTF486 vector between Sall and BamHI restriction sites using In-Fusion® HD Cloning Kit (TaKaRa, Mountain View, CA, USA) following the protocol, in each case generating a construct that expressed the corresponding fusion protein with GFP fused at C-terminus for subcellular localization assays. The construction of the vector expressing ER marker AtWAK2-mCherry fusion protein was performed as previously described (Zhou et al., 2017). Protoplasts were prepared from Arabidopsis leaves, and transformation and confocal microscopy were performed as described previously (Yoo et al., 2007). Primers used in the construction of these vectors are listed in Table S5.

Interaction analysis

Split ubiquitin analysis was performed using the yeast two-hybrid system from DUALmembrane system (Dualsystems Biotech AG, Schlieren, Switzerland) following the protocol. The full-length open reading frames of PcCER3, PcCER1 and PcCER1-LIKE1 ~ 3 were amplified by PCR using *SfiI* restriction site-containing primers, with subsequent orientated cloning of PcCER3 into pBT3N bait vector and PcCER1 and PcCER1-LIKE1 ~ 3 into pPR3N prey vector. DSY-1 yeast cells were cotransformed with pBT3N-PcCER3 and pPR3N empty vector as a negative control. Transformants were selected on -TL, and interactions were assayed on -HTL supplemented with 20 mM 3-amino-1,2,4-triazole and -AHTL. To show β -galactosidase activity of lacZ with X-Gal, the cells were grown on -AHTL containing 0.04 mg/mL X-Gal. For luciferase complementation imaging assay, PcCER3, PcCER1 and PcCER1-LIKE1 ~ 3 genes were fused into the N- or C-terminal part of LUC to generate the corresponding pCAMBIA1300-nLUC and pCAMBIA1300-cLUC fusion constructs. The luciferase complementation imaging assays were performed as previously described (Ban et al., 2018), using three biological replicates, by infiltration of leaves on separate plants. Primers used for the vector construction are shown in Table S5.

Transient expression in *N. benthamiana* leaves and heterologous expression in yeast

The pEAQ-HT constructs carrying each of PcCER3, PcCER1 and PcCER1-LIKE1 ~ 3 used for the transformation of *N. benthamiana* leaves were constructed as described previously (Xu et al., 2019).

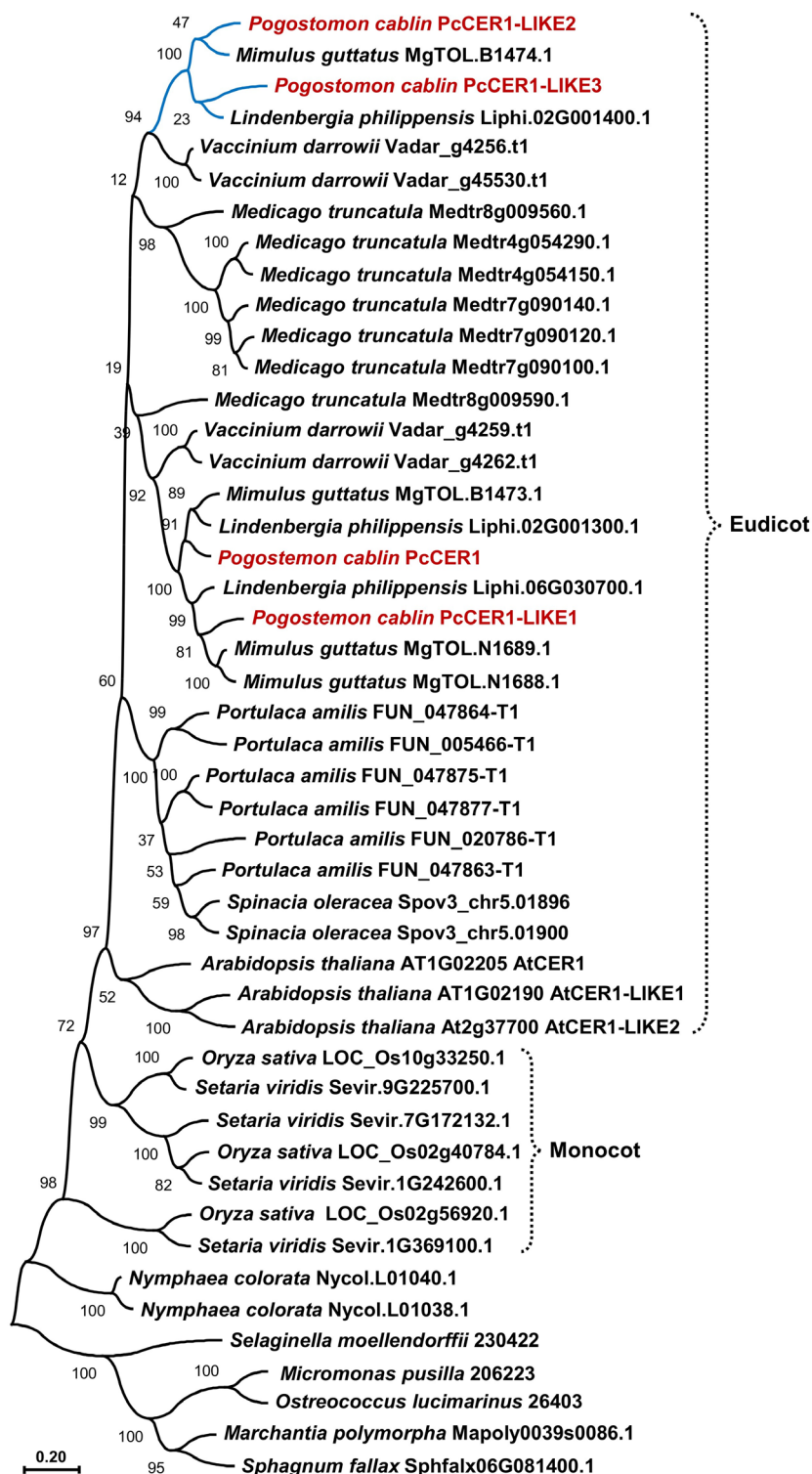


Figure 8 Phylogenetic analyses of PcCER1 homologues from selected plant species. The identification of PcCER1 homologues was conducted in the Phytozome v13 (<https://phytozome-next.jgi.doe.gov/>) and The Arabidopsis Information Resource (TAIR) (<https://www.arabidopsis.org/>) using PcCER1 protein sequences as the Basic Local Alignment Search Tool-protein (BLASTP) queries. The protein sequences of PcCER1 and its homologues in *Pogostemon cablin* were directly obtained from *P. cablin* transcriptome database and are marked in purple colour. Sequences were aligned using Muscle (Edgar, 2004), and the tree was constructed using the maximum likelihood algorithm and the Jones–Taylor–Thornton (JTT) matrix-based model in MEGA X (Jones et al., 1992; Kumar et al., 2018). Branch point bootstrap values were calculated with 1000 replicates. The percentage of tree in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The species are listed along with the protein names. The branches of the subclade in which PcCER1-LIKE2 and PcCER1-LIKE3 are grouped were marked in blue colour.

The procedure used for transient expression in *N. benthamiana* leaves was performed as previously described (Xu *et al.*, 2019). *N. benthamiana* leaves transformed with empty pEAQ-HT vector were used as control. The infiltrated leaves were collected 7 days after infiltration for alkane analysis.

The genes of PcCER3, PcCER1 and PcCER1-LIKE1 ~ 3 were introduced into multiple cloning site 2 region of pESC-LEU vector between BamHI and SalI restriction to generate the corresponding constructs pESC-LEU-PcCER3, pESC-LEU-PcCER1, pESC-LEU-PcCER1-LIKE1 ~ 3 for expression in *Saccharomyces cerevisiae* DSY-1 yeast strains. The genes of PcCER1 and PcCER1-LIKE1 ~ 3 were introduced into multiple cloning site 1 region of the pESC-LEU-PcCER3 construct between NotI and SpeI restriction sites to generate the corresponding constructs pESC-LEU-PcCER3/PcCER1 and pESC-LEU-PcCER3/PcCER1-LIKE1 ~ 3 for coexpression in DSY-1 yeast strains. The PcCYTB5-B gene was introduced into multiple cloning site 1 region of pESC-HIS vector between NotI and SpeI restriction sites to generate the construct pESC-HIS-PcCYTB5-B for expression in DSY-1 yeast strains. DSY-1 cells were transformed with different pESC construct combinations and grown on minimal medium agar plates lacking corresponding amino acids. The resulting positive yeast clones were cultured in 3 mL of synthetic dextrose dropout medium (-Leu, -His) with D-Glc as the carbon source for 2 days. The yeast cells were then harvested and resuspended into synthetic galactose dropout medium (-Leu, -His) with 2% (w/v) Gal. After induction for 4 days, yeast cells from 2.5 mL aliquots of the cultures were collected by centrifugation for alkane analysis after measuring fresh weight. Primers used for the vector construction are shown in Table S5.

VIGS assay

Tobacco rattle virus (TRV)-based VIGS of *NbACCs* in *N. benthamiana* plants was performed using syringe-mediated leaf infiltration method as previously described (Senthil-Kumar and Mysore, 2014), and *N. benthamiana* plants infiltrated with pTRV2-*NbPDS* construct was used as a positive control (Senthil-Kumar and Mysore, 2014). The *NbACC4* gene fragments within the highly homologous regions of the four *NbACC* isoforms were inserted into pTRV2 vector to generate the pTRV2-*NbACC* construct. The lower leaves of two-week-old *N. benthamiana* plants were co-inoculated with *Agrobacterium* strains carrying TRV1 and TRV2 constructs (pTRV2, pTRV2-*NbPDS* or pTRV2-*NbACC*), and the first leaves corresponding in position to pTRV2-*NbPDS* leaves showing photobleaching phenotype in one week after inoculation were used for transient expression of PcCER3 and PcCER1-LIKE3 as described above.

GC–MS analysis of alkanes from *P. cablin* plants, *N. benthamiana* leaves and yeast cells

To analyse the alkane production in different *P. cablin* tissues, *P. cablin* seedlings, roots, main stems and leaf tissues at different developmental stages were collected and cut into small pieces, 100 mg of which were transferred into a tube containing 500 μ L MTBE with 10 ng/ μ L geraniol as internal standard. The tube was vortexed for 3 min at maximum speed and incubated at room temperature for 1 h. The MTBE phase was collected and analysed by GC–MS. To analyse the alkane production in infiltrated *N. benthamiana* leaves, the infiltrated leaves were collected 7 days after infiltration and cut into small pieces, and 500 mg of which were extracted with 1 mL MTBE containing 10 ng/ μ L geraniol as an internal standard as described above for GC–MS analysis. To analyse the alkane production in yeast cells co-transformed with

different pESC construct combinations, the yeast cells collected from 2.5 mL aliquots of the cultures after 4 days of galactose incubation were resuspended in a 2 mL Fastprep tube with 0.5 mL ultrapure water and then added with 0.3 g glass beads (425–600 μ m). The cells were lysed using a Fastprep-24 homogenizer, at 6 m/s for 30 s, put on ice for 2 min and repeated for 10 times. Lysed cells were extracted with 500 μ L hexane containing 10 ng/ μ L geraniol as internal standard as described above for GC–MS analysis. For GC–MS analysis, 1 μ L aliquot of the sample was injected into Agilent GC–MSD (Agilent 7890B-5977B) system equipped with the DB-5MS column (30 m \times 0.25 mm \times 0.25 μ m film thickness, Agilent, Santa Clara, CA, USA). The oven temperature was programmed as follows: initial temperature, 50°C for 2 min, 50–320°C at a rate of 10°C min^{−1} and hold for 5 min. The identification of alkanes was assigned by comparison of their retention times and mass fragmentations with corresponding authentic standards. The measurement of alkanes was performed by normalization of their peak areas to the internal standard geraniol and comparison to their corresponding authentic standard curves.

Acknowledgements

We would like to thank Dr. Pichersky at the University of Michigan for his assistance with editing this manuscript and Dr. Zheng and Dr. Qing at the Analytical and Testing Center of Chongqing University for their assistance with GC–MS analysis and the laser scanning confocal microscopy analysis.

Funding

This work was supported by the Technology Innovation and Application Development Program of Chongqing (cstc2021jscx-cylhX0001), Fundamental Research Funds for the Central Universities (2023CDJXY-009) and the National Natural Sciences Foundation of China (Grant 31970319 to H.X.).

Conflict of interest statement

The authors have no conflicts of interest to declare.

Author contributions

H.X. conceived and supervised the overall research; J.W., W.X., Y.W., R.G., Y.Z., J.F., X.D., G.W. and Z.L. conducted the experiments, analysed data or provided material; H.X. wrote the article; all authors edited the article.

Data availability

Data supporting the findings of this study are available within the article and its supplemental files. The sequence data used in this study can be obtained from NCBI with the following GenBank accession numbers: PcCER1, OQ718413; PcCER1-LIKE1, OQ718414; PcCER1-LIKE2, OQ718415; PcCER1-LIKE3, OQ718416; PcCER3, OQ718417; PcCYTB5-B, OQ718418.

References

- Aarts, M.G.M., Keijzer, C.J., Stiekema, W.J. and Pereira, A. (1995) Molecular characterization of the CER1 gene of *Arabidopsis* involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell* **7**, 2115–2127.

- Andre, C., Kim, S.W., Yu, X.H. and Shanklin, J. (2013) Fusing catalase to an alkane-producing enzyme maintains enzymatic activity by converting the inhibitory by product H_2O_2 to the cosubstrate O_2 . *P Natl Acad Sci USA* **110**, 3191–3196.
- Ban, Z.N., Qin, H., Mitchell, A.J., Liu, B.X., Zhang, F.X., Weng, J.K., Dixon, R.A. et al. (2018) Noncatalytic chalcone isomerase-fold proteins in are auxiliary components in prenylated flavonoid biosynthesis. *P Natl Acad Sci USA* **115**, E5223–E5232.
- Basri, R.S., Abd Rahman, R.N.Z.R., Kamarudin, N.H.A. and Ali, M.S.M. (2020) Cyanobacterial aldehyde deformylating oxygenase: Structure, function, and potential in biofuels production. *Int. J. Biol. Macromol.* **164**, 3155–3162.
- Baud, S., Guyon, V., Kronenberger, J., Wuilleme, S., Miquel, M., Caboche, M., Lepiniec, L. et al. (2003) Multifunctional acetyl-CoA carboxylase 1 is essential for very long chain fatty acid elongation and embryo development in *Arabidopsis*. *Plant J.* **33**, 75–86.
- Bernard, A., Domergue, F., Pascal, S., Jetter, R., Renne, C., Faure, J.D., Haslam, R.P. et al. (2012) Reconstitution of plant alkane biosynthesis in yeast demonstrates that *Arabidopsis* ECERIFERUM1 and ECERIFERUM3 are core components of a very-long-chain alkane synthesis complex. *Plant Cell* **24**, 3106–3118.
- Bernard, A. and Joubès, J. (2013) *Arabidopsis* cuticular waxes: advances in synthesis, export and regulation. *Prog. Lipid Res.* **52**, 110–129.
- Bonaventure, G., Salas, J.J., Pollard, M.R. and Ohlrogge, J.B. (2003) Disruption of the FATB gene in *Arabidopsis* demonstrates an essential role of saturated fatty acids in plant growth. *Plant Cell* **15**, 1020–1033.
- Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S.Q., Ishizaki, K., Yamaoka, S. et al. (2017) Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171**, 287–304.
- Bruder, S., Moldenhauer, E.J., Lemke, R.D., Ledesma-Amaro, R. and Kabisch, J. (2019) Drop-in biofuel production using fatty acid photodecarboxylase from *Chlorella variabilis* in the oleaginous yeast *Yarrowia lipolytica*. *Biotechnol. Biofuels* **12**, 202.
- Cheesbrough, T.M. and Kolattukudy, P.E. (1984) Alkane biosynthesis by decarbonylation of aldehydes catalyzed by a particulate preparation from *Pisum sativum*. *P Natl Acad Sci USA* **81**, 6613–6617.
- Chen, C.J., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y.H. and Xia, R. (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **13**, 1194–1202.
- Chen, J., Liu, L., Wang, Y., Li, Z.G., Wang, G.D., Kraus, G.A., Pichersky, E. et al. (2021a) Characterization of a cytosolic acyl-activating enzyme catalyzing the formation of 4-methylvaleryl-coa for pogostone biosynthesis in *Pogostemon cablin*. *Plant Cell Physiol.* **62**, 1556–1571.
- Chen, J.R., Xie, X.F., Li, M.T., Xiong, Q.Y., Li, G.M., Zhang, H.Q., Chen, G.R. et al. (2021b) Pharmacological activities and mechanisms of action of *Pogostemon cablin* Benth: a review. *Chin Med-UK* **16**, 5.
- Choi, K.R., Jiao, S. and Lee, S.Y. (2020) Metabolic engineering strategies toward production of biofuels. *Curr. Opin. Chem. Biol.* **59**, 1–14.
- Choi, Y.J. and Lee, S.Y. (2013) Microbial production of short-chain alkanes. *Nature* **502**, 571–574.
- Debnath, R., Mitra, P., Das, S. and Barik, A. (2021) Leaf surface wax chemicals in *Trichosanthes anguina* (Cucurbitaceae) cultivars mediating short-range attraction and oviposition in *Diaphania indica*. *J. Chem. Ecol.* **47**, 664–679.
- Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 1–19.
- Fatma, Z., Hartman, H., Poolman, M.G., Fell, D.A., Srivastava, S., Shakeel, T. and Yazdani, S.S. (2018) Model-assisted metabolic engineering of *Escherichia coli* for long chain alkane and alcohol production. *Metab. Eng.* **46**, 1–12.
- Jaroensuk, J., Intasian, P., Wattanasuepsin, W., Akeratchatapan, N., Kesornpun, C., Kittipanukul, N. and Chaiyen, P. (2020) Enzymatic reactions and pathway engineering for the production of renewable hydrocarbons. *J. Biotechnol.* **309**, 1–19.
- Jetter, R. and Kunst, L. (2008) Plant surface lipid biosynthetic pathways and their utility for metabolic engineering of waxes and hydrocarbon biofuels. *Plant J.* **54**, 670–683.
- Jones, D.T., Taylor, W.R. and Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–282.
- Joyce, B.L. and Stewart, C.N. (2012) Designing the perfect plant feedstock for biofuel production: using the whole buffalo to diversify fuels and products. *Biotechnol. Adv.* **30**, 1011–1022.
- Kang, M.K. and Nielsen, J. (2017) Biobased production of alkanes and alkenes through metabolic engineering of microorganisms. *J. Ind. Microbiol. Biotechnol.* **44**, 613–622.
- Khara, B., Menon, N., Levy, C., Mansell, D., Das, D., Marsh, E.N.G., Leys, D. et al. (2013) Production of propane and other short-chain alkanes by structure-based engineering of ligand specificity in aldehyde-deformylating oxygenase. *Chembiochem* **14**, 1204–1208.
- Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549.
- Lee, E.J., Kim, K.Y., Zhang, J., Yamaoka, Y., Gao, P., Kim, H., Hwang, J.U. et al. (2021) *Arabidopsis* seedling establishment under waterlogging requires ABCG5-mediated formation of a dense cuticle layer. *New Phytol.* **229**, 156–172.
- Lee, S.B. and Suh, M.C. (2015) Advances in the understanding of cuticular waxes in *Arabidopsis thaliana* and crop species. *Plant Cell Rep.* **34**, 557–572.
- Lee, S.B. and Suh, M.C. (2022) Regulatory mechanisms underlying cuticular wax biosynthesis. *J. Exp. Bot.* **73**, 2799–2816.
- Li, H., Cann, A.F. and Liao, J.C. (2010) Biofuels: biomolecular engineering fundamentals and advances. *Annu Rev Chem Biomol* **1**, 19–36.
- Li, J.B., Ma, Y., Liu, N., Eser, B.E., Guo, Z., Jensen, P.R. and Stephanopoulos, G. (2020) Synthesis of high-titer alka(e)nes in *Yarrowia lipolytica* is enabled by a discovered mechanism. *Nat. Commun.* **11**, 6198.
- Li, N., Chang, W.C., Warui, D.M., Booker, S.J., Krebs, C. and Bollinger, J.M. (2012) Evidence for only oxygenative cleavage of aldehydes to alk(a)enes and formate by cyanobacterial aldehyde decarboxylases. *Biochemistry* **51**, 7908–7916.
- Li, W., Zhou, F. and Pichersky, E. (2018) Jasmonate hydroxylase, a key enzyme in the synthesis of the alcohol moiety of pyrethrin insecticides. *Plant Physiol.* **177**, 1498–1509.
- Liu, Y.Z., Cruz-Morales, P., Zargar, A., Belcher, M.S., Pang, B., Englund, E., Dan, Q.Y. et al. (2021) Biofuels for a sustainable future. *Cell* **184**, 1636–1647.
- Lu, S.Y., Zhao, H.Y., Parsons, E.P., Xu, C.C., Kosma, D.K., Xu, X.J., Chao, D. et al. (2011) The *glossyhead1* allele of ACC1 reveals a principal role for multidomain acetyl-coenzyme A carboxylase in the biosynthesis of cuticular waxes by *Arabidopsis*. *Plant Physiol.* **157**, 1079–1092.
- Marsh, E.N.G. and Waugh, M.W. (2013) Aldehyde decarboxylases: enigmatic enzymes of hydrocarbon biosynthesis. *ACS Catal.* **3**, 2515–2521.
- Mohammad, M., Hari, T.K., Yaakob, Z., Sharma, Y.C. and Sopian, K. (2013) Overview on the production of paraffin based-biofuels via catalytic hydrodeoxygenation. *Renew. Sust. Energ. Rev.* **22**, 121–132.
- Monteiro, R.R.C., da Silva, S.S.O., Cavalcante, C.L., de Luna, F.M.T., Bolivar, J.M., Vieira, R.S. and Fernandez-Lafuente, R. (2022) Biosynthesis of alkanes/alkenes from fatty acids or derivatives (triacylglycerols or fatty aldehydes). *Biotechnol. Adv.* **61**, 108045.
- Pandiyan, K., Singh, A., Singh, S., Saxena, A.K. and Nain, L. (2019) Technological interventions for utilization of crop residues and weedy biomass for second generation bio-ethanol production. *Renew. Energy* **132**, 723–741.
- Pascal, S., Bernard, A., Deslous, P., Gronnier, J., Fournier-Goss, A., Domergue, F., Rowland, O. et al. (2019) *Arabidopsis* CER1-LIKE1 functions in a cuticular very-long-chain alkane-forming complex. *Plant Physiol.* **179**, 415–432.
- Peralta-Yahya, P.P., Zhang, F.Z., del Cardayre, S.B. and Keasling, J.D. (2012) Microbial engineering for the production of advanced biofuels. *Nature* **488**, 320–328.
- Qiu, Y., Tittiger, C., Wicker-Thomas, C., Le Goff, G., Young, S., Wajnberg, E., Fricaux, T. et al. (2012) An insect-specific P450 oxidative decarboxylase for cuticular hydrocarbon biosynthesis. *P Natl Acad Sci USA* **109**, 14858–14863.
- Riederer, M. and Schreiber, L. (2001) Protecting against water loss: analysis of the barrier properties of plant cuticles. *J. Exp. Bot.* **52**, 2023–2032.
- Runguphan, W. and Keasling, J.D. (2014) Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metab. Eng.* **21**, 103–113.
- Samuels, L., Kunst, L. and Jetter, R. (2008) Sealing plant surfaces: cuticular wax formation by epidermal cells. *Annu. Rev. Plant Biol.* **59**, 683–707.

- Savage, T.J., Hamilton, B.S. and Croteau, R. (1996a) Biochemistry of short-chain alkanes (Tissue-specific biosynthesis of n-heptane in *Pinus jeffreyi*). *Plant Physiol.* **110**, 179–186.
- Savage, T.J., Hristova, M.K. and Croteau, R. (1996b) Evidence for an elongation/reduction/C1-elimination pathway in the biosynthesis of n-heptane in xylem of Jeffrey pine. *Plant Physiol.* **111**, 1263–1269.
- Schirmer, A., Rude, M.A., Li, X.Z., Popova, E. and del Cardayre, S.B. (2010) Microbial biosynthesis of alkanes. *Science* **329**, 559–562.
- Schnurr, J., Shockey, J. and Browse, J. (2004) The acyl-CoA synthetase encoded by LACS2 is essential for normal cuticle development in Arabidopsis. *Plant Cell* **16**, 629–642.
- Senthil-Kumar, M. and Mysore, K.S. (2014) Tobacco rattle virus-based virus-induced gene silencing in *Nicotiana benthamiana*. *Nat. Protoc.* **9**, 1549–1562.
- Song, X.J., Yu, H.Y. and Zhu, K. (2016) Improving alkane synthesis in *Escherichia coli* via metabolic engineering. *Appl. Microbiol. Biotechnol.* **100**, 757–767.
- Sorigué, D., Légeret, B., Cuiné, S., Blangy, S., Moulin, S., Billon, E., Richaud, P. et al. (2017) An algal photoenzyme converts fatty acids to hydrocarbons. *Science* **357**, 903–907.
- Sorigué, D., Légeret, B., Cuiné, S., Morales, P., Mirabella, B., Guedéney, G., Li-Beisson, Y. et al. (2016) Microalgae synthesize hydrocarbons from long-chain fatty acids via a light-dependent pathway. *Plant Physiol.* **171**, 2393–2405.
- Wang, C., Wang, Y., Chen, J., Liu, L., Yang, M.X., Li, Z.G., Wang, C.Y. et al. (2022) Synthesis of 4-methylvaleric acid, a precursor of pogostone, involves a 2-isobutylmalate synthase related to 2-isopropylmalate synthase of leucine biosynthesis. *New Phytol.* **235**, 1129–1145.
- Wang, H.L., Ni, X.Z. and Harris-Shultz, K. (2019) Molecular evolution of the plant ECERIFERUM1 and ECERIFERUM3 genes involved in aliphatic hydrocarbon production. *Comput. Biol. Chem.* **80**, 1–9.
- Weng, H., Molina, I., Shockey, J. and Browse, J. (2010) Organ fusion and defective cuticle function in a lacs1 lacs2 double mutant of Arabidopsis. *Planta* **231**, 1089–1100.
- Xu, H.Y., Li, W., Schillmiller, A.L., van Eekelen, H., de Vos, R.C.H., Jongsma, M.A. and Pichersky, E. (2019) Pyrethric acid of natural pyrethrin insecticide: complete pathway elucidation and reconstitution in. *New Phytol.* **223**, 751–765.
- Yoo, S.D., Cho, Y.H. and Sheen, J. (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572.
- Yu, T., Zhou, Y.J.J., Wenning, L., Liu, Q.L., Krivoruchko, A., Siewers, V., Nielsen, J. et al. (2017) Metabolic engineering of *Saccharomyces cerevisiae* for production of very long chain fatty acid-derived chemicals. *Nat. Commun.* **8**, 15587.
- Zhang, F.Z., Rodriguez, S. and Keasling, J.D. (2011) Metabolic engineering of microbial pathways for advanced biofuels production. *Curr. Opin. Biotechnol.* **22**, 775–783.
- Zheng, J., He, C., Qin, Y., Lin, G.F., Park, W.D., Sun, M.H., Li, J. et al. (2019) Co-expression analysis aids in the identification of genes in the cuticular wax pathway in maize. *Plant J.* **97**, 530–542.
- Zhou, F., Wang, C.Y., Gutensohn, M., Jiang, L., Zhang, P., Zhang, D.B., Dudareva, N. et al. (2017) A recruiting protein of geranylgeranyl diphosphate synthase controls metabolic flux toward chlorophyll biosynthesis in rice. *P Natl Acad Sci USA* **114**, 6866–6871.
- Zhou, Y.J.J., Buijs, N.A., Zhu, Z.W., Qin, J.F., Siewers, V. and Nielsen, J. (2016) Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. *Nat. Commun.* **7**, 11709.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 The biochemical reactions leading to alkane formation catalysed by the alkane-forming enzymes identified from plant, microalgae, insects and cyanobacteria.

Figure S2 GC–MS analysis of very-long-chain alkane nonacosane and hentriacontane in different tissues of *P. cablin* at different development stages.

Figure S3 Sequence alignment of PcCER1, PcCER1-LIKE1 ~ 3, AtCER1 and AtCER1-LIKE1.

Figure S4 Heat map of the candidate genes putatively involved in LC and VLC-alkane biosynthesis identified in *P. cablin* transcriptome with distributions of pentadecane and heptadecane in the same 6 selected *P. cablin* tissues for transcriptome construction by co-expression analysis.

Figure S5 Subcellular localization of PcCER1, PcCER1-LIKE1 ~ 3 and PcCER3 in Arabidopsis leaf mesophyll protoplasts.

Figure S6 GC–MS analysis of the odd-numbered LC and VLC-alkanes from C₁₉ to C₃₃ in *Nicotiana benthamiana* leaves expressing indicated genes or gene combinations.

Figure S7 Deduced amino acid sequence alignment of AtACC1 and NbACC1 ~ 4.

Figure S8 Encoding sequence alignment of NbACC1 ~ 4.

Table S1 Unigenes putatively involved in C₁₆ and C₁₈ fatty acids biosynthesis and export and alkane-forming pathway and candidate long-chain acyl-CoA synthetases in *P. cablin* RNAseq database.

Table S2 Average normalized counts of unigenes putatively involved in fatty acid biosynthesis and export and alkane-forming pathway and candidate long-chain acyl-CoA synthetases in *P. cablin* RNAseq database.

Table S3 Correlation analysis of the expression of PcCER3, PcCER1 and PcCER1-LIKE1 ~ 3 with the pentadecane and heptadecane distributions in the same selected 13 *P. cablin* tissues at different developmental stages.

Table S4 GenBank accessions or locus numbers of identified alkane-forming enzymes from plant, microalgae, insects and cyanobacteria and used for screening candidate alkane-forming genes from *P. cablin* RNAseq database.

Table S5 All primers used in this present study.