

Synthesis of 4-methylvaleric acid, a precursor of pogostone, involves a 2-isobutylmalate synthase related to 2-isopropylmalate synthase of leucine biosynthesis

Chu Wang^{1,2*}, Ying Wang^{1,2*}, Jing Chen^{1,2*}, Lang Liu^{1,2}, Mingxia Yang³, Zhengguo Li^{1,2} ,
Chengyuan Wang³ , Eran Pichersky⁴  and Haiyang Xu^{1,2} 

¹School of Life Sciences, Chongqing University, Chongqing 401331, China; ²Center of Plant Functional Genomics, Institute of Advanced Interdisciplinary Studies, Chongqing University, Chongqing 401331, China; ³The Center for Microbes, Development and Health, Institute Pasteur of Shanghai, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai 200031, China; ⁴Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA

Summary

Author correspondence:
Haiyang Xu
Email: hyxu@cqu.edu.cn

Received: 2 March 2022
Accepted: 19 April 2022

New Phytologist (2022) 235: 1129–1145
doi: 10.1111/nph.18186

Key words: 2-isobutylmalate, 2-isobutylmalate synthase, 2-isopropylmalate synthase, 4-methylvaleric acid, branched-chain fatty acid, *Pogostemon cablin*, pogostone, α -ketoacid elongation pathway.

- We show here that the side chain of pogostone, one of the major components of patchouli oil obtained from *Pogostemon cablin* and possessing a variety of pharmacological activities, is derived from 4-methylvaleric acid.
- We also show that 4-methylvaleric acid is produced through the one-carbon α -ketoacid elongation pathway with the involvement of the key enzyme 2-isobutylmalate synthase (IBMS), a newly identified enzyme related to isopropylmalate synthase (IPMS) of leucine (Leu) biosynthesis.
- Site-directed mutagenesis identified Met¹³² in the N-terminal catalytic region as affecting the substrate specificity of PclBMS1. Even though PclBMS1 possesses the C-terminal domain that in IPMS serves to mediate Leu inhibition, it is insensitive to Leu.
- The observation of the evolution of IBMS from IPMS, as well as previously reported examples of IPMS-related genes involved in making glucosinolates in Brassicaceae, acylsugars in Solanaceae, and flavour compounds in apple, indicate that IPMS genes represent an important pool for the independent evolution of genes for specialised metabolism.

Introduction

Pogostemon cablin is a traditional Chinese medicinal plant in the Lamiaceae family, from whose aerial parts *Pogostemonis Herba*, a herbal preparation, is obtained. *Pogostemonis Herba* or its essential oil component (known as patchouli oil) have been widely used in many Asian countries for the treatment of many ailments since ancient times due to a variety of pharmacological activities (Li *et al.*, 2013; Wang *et al.*, 2016; Hu *et al.*, 2017; Junren *et al.*, 2021). Pogostone, one of the major chemical components of patchouli oil, has recently been demonstrated to possess various bioactivities including antimicrobial activities as well as pharmacological activities (Li *et al.*, 2012, 2014; Huang *et al.*, 2014; Peng *et al.*, 2014; Chen *et al.*, 2015; Cao *et al.*, 2017; Luchesi *et al.*, 2020; Ye *et al.*, 2021).

The biosynthetic pathway of pogostone is still unclear, but an outline of the pathway has been proposed that involves the branched-chain 4-methylvaleric acid (Fig. 1; Chen *et al.*, 2021). Furthermore, PCAA2, a cytosolic acyl-activating enzyme that catalyses the conversion of 4-methylvaleric acid to 4-

methylvaleryl-CoA as part of the proposed pogostone biosynthetic pathway, has been identified and characterised (Chen *et al.*, 2021).

The formation of the branched C6 fatty 4-methylvaleric acid has not yet been elucidated in any species, although evidence for the involvement of either α -ketoacid elongation (α KAE) pathway or the fatty acid synthase (FAS)-mediated elongation pathway in the elongation of other branched-chain fatty acids in various species has been presented (Ohlrogge & Browse, 1995; Kroumova & Wagner, 2003; Slocombe *et al.*, 2008; Li-Beisson *et al.*, 2013). In the FAS route, both carbons from acetyl-ACP are retained per elongation cycle. In the α KAE pathway, acetyl-CoA is added but only one carbon is retained per elongation cycle. The α KAE pathway plays a role in the biosynthesis of some primary and specialised metabolites such as leucine (Leu) and glucosinolates (Melendez-Hevia *et al.*, 1996; de Quiros *et al.*, 2000; Halkier & Gershenzon, 2006; Binder, 2010).

The α KAE steps involved in leucine biosynthesis have been well documented in various biological organisms (Binder, 2010) and are illustrative of the α KAE route in general (Fig. 1). First, 2-isopropylmalate synthase (IPMS; EC 2.3.3.13) catalyses an aldol-type condensation between acetyl-CoA and 2-oxoisovalerate to

*These authors contributed equally to this work.

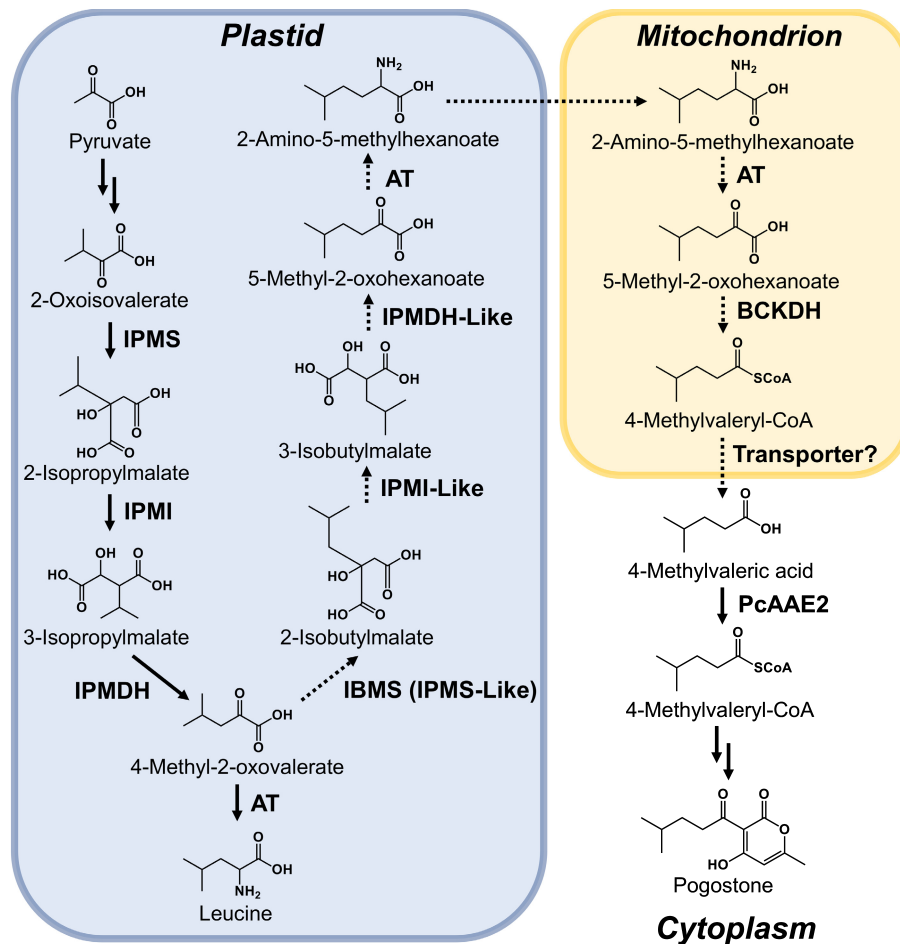


Fig. 1 The proposed biosynthetic pathway of 4-methylvaleric acid for pogostone biosynthesis in *Pogostemon cablin*. The anabolism and catabolism of branched chain amino acid leucine are well documented and occur in plastids and mitochondria, respectively, in land plants (Binder *et al.*, 2007; Binder, 2010). The anabolism and catabolism of 2-amino-5-oxohexanoate, the leucine analogue with one more carbon than leucine, most likely occurs in the same way as that of leucine. 5-Methyl-2-oxohexanoate, the precursor of 2-amino-5-oxohexanoate biosynthesis, is elongated by one carbon atom in plastids through a α -keto acid elongation (α KAE) pathway from 4-methyl-2-oxovalerate, which is derived from the anabolism of leucine through the α KAE pathway (Binder, 2010). The one-carbon elongation process for 5-methyl-2-oxohexanoate through the α KAE route is performed in three sequential chemical reactions: 2-oxo acid substrate 4-methyl-2-oxovalerate first undergoes an aldol-type condensation with acetyl co-enzyme A (acetyl-CoA) by 2-isobutylmalate synthase (IBMS) to give 2-isobutylmalate, which then undergoes isomerisation and oxidative decarboxylation by isopropylmalate isomerase (IPMI) and 3-isopropylmalate dehydrogenase (IPMDH) or their homologies to yield 5-methyl-2-oxohexanoate in the plastid. 5-Methyl-2-oxohexanoate is regenerated from 2-amino-5-oxohexanoate through transamination catalysed by aminotransferase (AT) in mitochondria. 4-Methylvaleryl-CoA is formed from 5-methyl-2-oxohexanoate in a reaction catalysed by branched-chain keto acid dehydrogenase (BCKDH) in mitochondria, and then transported into cytosol by a transporter along with losing the CoA group to give free 4-methylvaleric acid, which is subsequently activated again to CoA derivative by the cytosol acyl-activating enzyme 2 from *P. cablin* (PcAAE2) for pogostone biosynthesis (Chen *et al.*, 2021). The dashed arrows represent the proposed steps remaining to be elucidated. IPMS, 2-isopropylmalate synthase.

yield 2-isopropylmalate. Next, 2-isopropylmalate is converted using isopropylmalate isomerase (IPMI; EC 4.2.1.33) to 3-isopropylmalate. The final reaction, the oxidative decarboxylation of 3-isopropylmalate to 4-methyl-2-oxovalerate, is catalysed by the isopropylmalate dehydrogenase (IPMDH; EC 1.1.1.85). The one-carbon elongated product 4-methyl-2-oxovalerate then undergoes a reversible transamination catalysed by branched-chain aminotransferase (BCAT; EC 2.6.1.42) to yield the branched-chain amino acid (BCAA) leucine in plastid. 4-Methyl-2-oxovalerate can be regenerated using transamination reactions using BCAT from leucine in mitochondria and converted into isovaleryl-CoA through irreversible oxidative decarboxylation catalysed using a branched-chain keto acid dehydrogenase

(BCKDH, EC 1.2.4.4) (Binder, 2010). The isovaleryl-CoA generated in mitochondria from 4-methyl-2-oxovalerate might be transported into cytosol using a transporter along with losing the CoA group to give free 3-methylbutanoic (Xu *et al.*, 2013).

In the Solanaceae, many species produce specialised metabolites called acyl sugars, in which at least some of the acyls attached to the sugar moiety are elongated branched chains (Kroumova & Wagner, 2003; Slocombe *et al.*, 2008; Ning *et al.*, 2015). Feeding studies with labelled amino acids have shown that the branched short chain fatty acid (2-methylpropanoic acid, 3-methylbutanoic acid and 2-methylbutanoic acid) are directly derived from Val, Leu and Ile, respectively (Kroumova & Wagner, 2003), while the branched acyl chains of C6 to C12 are derived from these BCAAs

using elongation through the α KAE pathway in tobaccos and petunia, but from the FAS pathway in *Lycopersicon pennellii* and *Datura metel* (Kroumova & Wagner, 2003).

As the branched C5 fatty 3-methylbutanoic acid is derived from leucine metabolism involving the one-carbon elongation α KAE pathway (Kroumova & Wagner, 2003; Binder, 2010), 4-methylvaleric acid, the branched C6 fatty acid with one more carbon than 3-methylbutanoic acid, is most likely to be derived from leucine metabolism by one additional elongation cycle of α KAE involving three reactions catalysed using IPMS, IPMI and IPMDH, or analogous enzymes (Fig. 1). We show here that the first reaction, in which 4-methyl-2-oxovalerate is elongated to 2-isobutylmalate, is catalysed using a novel enzyme, 2-isobutylmalate synthase (IBMS), that is closely related to IPMS.

Materials and Methods

Plant materials and chemicals

The plants of *P. cablin* and *Nicotiana benthamiana* were grown in soil under conditions described previously (Chen *et al.*, 2021). The harvested tissues were flash frozen in liquid nitrogen and stored at -80°C until use.

As standard 2-isobutylmalate is not available commercially, the substituted 2-malate derivative generated by the condensation of acetyl-CoA to 4-methyl-2-oxovalerate by PclBMS1 was identified as 2-isobutylmalate based on the exact mass and fragment pattern by liquid chromatography–quadrupole time-of-flight–mass spectrometry (LC-QTOF-MS). All other commercial chemicals were purchased from Sigma-Aldrich.

Isotope feeding study with deuterium-labelled [$^2\text{H}_{11}$] 4-methylvaleric acid (4MVA- d_{11})

The 5-wk-old top leaves (5W-TLeaf) of *P. cablin* plants were placed in a 10-ml feeding solution immediately before the onset of darkness and transferred to a vacuum desiccator equipped with a vacuum pump. After incubation for 4 min at the atmospheric pressure of 0.04 MPa, the leaves in the feeding solution were transferred to a growth chamber and incubated under darkness for 12 h at 25°C . Feeding for both 4-methylvaleric acid (4MVA) and 4MVA- d_{11} were performed at a concentration of 1 mM. The detailed procedures for analysis of metabolites of *P. cablin* leaves fed with 4MVA or 4MVA- d_{11} can be found in Supporting Information Methods S1.

Gene identification and coexpression analysis

Identification of gene candidates involved in 4-methylvaleric acid biosynthesis from our previously constructed *P. cablin* transcriptome database (bioproject accession no. PRJNA713906) was performed by querying our local nucleotide database with Arabidopsis representative enzymes involved in BCAA Leu metabolism using the tBLASTn function as previously described (Chen *et al.*, 2021). Coexpression analysis was performed by comparing the transcript abundance of *PcAAE2* and the identified

gene candidates involved in 4-methylvaleric acid biosynthesis using the Pearson correlation method to quantitate the similarity of their expression profiles as described previously (Li *et al.*, 2018).

Quantitative RT-PCR analysis of the identified gene candidates involved in 4-methylvaleric acid biosynthesis

For quantitative real-time PCR (RT-qPCR) analysis of transcripts in different tissues, RNA was extracted using the Total RNA Isolation Kit from Omega (Norcross, GA, USA) with a DNA digestion step. The reverse transcription reaction was performed using the High-Capacity cDNA Reverse Transcription Kit from Thermo Fisher Scientific (Waltham, MA, USA), following the manufacturer's instructions using random primers. The RT-qPCR analysis was performed using the Power SYBR Green PCR master mix as previously described (Chen *et al.*, 2021). The relative expression levels of these genes relative to the three reference genes (*GAPDH*, *Actin7* and *Tubulin3*) in each tissue were calculated as previously described (Chen *et al.*, 2021). Primers used in this study are listed in the Table S1.

Isolation of His-tagged recombinant proteins of PclBMS1, PclPMS1, PclBMS1-M132L and PclPMS1-L135M

The constructs for expression of His-tagged recombinant proteins of PclBMS1, PclPMS1, PclBMS1-M132L and PclPMS1-L135M in *Escherichia coli* were constructed based on expression vector pET-28a(+). The details of vector construction can be found in Methods S1. Recombinant proteins of PclBMS1, PclPMS1, PclBMS1-M132L and PclPMS1-L135M were expressed in *E. coli* strain BL21(BE3) and purified using Ni-NTA agarose chromatography (Qiagen) as previously described (Xu *et al.*, 2018). The primer pairs used for construction of these pET-28a(+) expression vectors fused with HIS₆ residues at the N-terminus for expression in *E. coli* are listed in Table S1.

Enzymatic assays of recombinant PclBMS1, PclPMS1, PclBMS1-M132L and PclPMS1-L135M

The enzyme assay for the condensation reaction between acetyl-CoA and different 2-oxo acids was performed with a spectrophotometric end-point assay with DTNB as previously described (de Kraker *et al.*, 2007) with minor modifications and the enzymatic products of PclBMS1 and PclPMS1 towards 2-oxoisovalerate and 4-methyl-2-oxovalerate were further measured using an LC-QTOF-MS-based method. Details of these enzymatic assays for the condensation reaction can be found in Methods S1.

Modelling structures of PclBMS1, PclPMS1, PclPMS1-L135M and PclBMS1-M132L

The target sequences were first split into domains by FUPRED based on the deep-learning predicted contact map (Zheng *et al.*, 2020). Then the model of each domain and the full-chain sequence were generated through ALPHA FOLD2 (Jumper *et al.*,

2021). Finally, all individual domains were assembled together to create the final full-length model by DEMO (Zhou *et al.*, 2019) based on the domain orientations derived from the ALPHA FOLD2 full-length model.

Subcellular localisation of PcIBMS1 and PcIPMS1

The vector used for subcellular localisation of analysis was pTF486. The construction of these vectors expressing fusion proteins of PcIBMS1 and PcIPMS1 with green fluorescent protein (GFP) fused at C-terminus was performed as previously described (Chen *et al.*, 2021). The full-length open reading frames of PcIBMS1 and PcIPMS1 were cloned into pTF486 between the *SalI* and *BamHI* sites using the In-Fusion[®] HD Cloning Kit (Takara, Kusatsu, Japan) and following the protocol. 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 S1.

Transient expression in *N. benthamiana* leaves

The pEAQ-HT constructs carrying each of *PcIBMS1* and *PcIPMS1* used for transformation were constructed as described previously (Xu *et al.*, 2019). The procedure used for transient expression in *N. benthamiana* leaves was performed as previously described (Xu *et al.*, 2019). *N. benthamiana* plants transformed with empty pEAQ-HT vector were used as the control. The infiltrated leaves were collected 7 d after infiltration. The detailed procedures for analysis of metabolites in *N. benthamiana* leaves can be found in Methods S1.

GC-MS analysis and LC-QTOF-MS analysis

For GC-MS analysis, 1 µl aliquot of the sample was injected into the Agilent GC-MSD (Agilent 7890B-5977B, Santa Clara, CA, USA) system equipped with the Rxi-5Sil MS column (30 m × 0.25 mm × 0.25 µm film thickness; Restek, Bellefonte, PA, USA). The oven temperature was programmed as previously described (Chen *et al.*, 2021). For LC-QTOF-MS analysis, 2 µl aliquot of the sample was injected into the LC-QTOF-MS system (6545 LC/QTOF-MS; Agilent) coupled with a C₁₈ column (Zorbax RRHD Plus C₁₈, Φ2.1 × 50 mm, 1.8 µm) at 35°C. The gradient (solvent A, water + 0.1% formic acid; solvent B, acetonitrile) program and operating parameters were set as previously described (Chen *et al.*, 2021).

Results

4-Methylvaleric acid is a precursor of pogostone

Chen *et al.* (2021) hypothesised that the side chain of pogostone is derived from 4-methylvaleric acid, and furthered showed that the expression of PcAAE2, a cytosolic acyl-activating enzyme that catalyses the conversion of 4-methylvaleric acid to 4-methylvaleryl-CoA, is correlated with pogostone biosynthesis.

However, a direct proof that 4-methylvaleric acid is a precursor of pogostone has not yet been presented. To determine if the side chain of pogostone is derived from 4-methylvaleric acid, 5-wk-old top leaves (5W-TLeaf) of *P. cablin* plants were placed in a solution containing the deuterium-labelled [²H₁₁] 4-methylvaleric acid (4-methylvaleric-d₁₁ acid or 4MVA-d₁₁) at the concentration of 1 mM. Following an overnight treatment with the labelled 4MVA-d₁₁, volatiles were extracted with methyl tert-butyl ether (MTBE) and aliquots were analysed using gas chromatography–mass spectrometry (GC-MS) (Fig. 2). Another portion of each MTBE extract was further dried and dissolved in 90% methanol for LC-QTOF-MS (Fig. S1). We observed that 7.8% of the pogostone molecules extracted had an 11-D shift (from 223.0976 for pogostone to 234.1666 for pogostone-d₁₁ in negative mode in LC-QTOF-MS analysis and from 224.1 for pogostone to 235.2 for pogostone-d₁₁ in GC-MS analysis) in the parent ion mass (Figs 2, S1), indicating that 4-methylvaleric acid is a genuine precursor of pogostone biosynthesis.

Identification of the gene candidates involved in 4-methylvaleric acid biosynthesis using coexpression analysis

To identify the candidate genes involved in 4-methylvaleric acid biosynthesis, we performed homology searches on our previously constructed *P. cablin* RNA-seq database (Chen *et al.*, 2021), using genes involved in leucine metabolism in *Arabidopsis thaliana* using tBLASTn software (Fig. 1). We obtained a total of 13 unique genes (Table S2): two genes with similarity to AtIPMS (named *PcIBMS1* and *PcIPMS1*, please refer to the following lines); two genes (*PcIPMIL SSU* and *PcIPMIL LSU*; ‘L’ for ‘-like’) related to the small subunit and large subunit of Arabidopsis IPMI, respectively; one gene (*PcIPMDHL*) related to AtIPMDH; three genes putatively encoding branched-chain aminotransferases (*PcBCATL1*, *PcBCATL2* and *PcBCATL3*); and five genes putatively encoding the three subunits of branched-chain α-keto acid dehydrogenase (*PcBCKDHL E1α1*, *PcBCKDHL E1α2*, *PcBCKDHL E1β*, *PcBCKDHL E2* and *PcBCKDHL E3*).

We next performed coexpression analysis of these 13 unique genes through Pearson correlation analysis to look for genes whose expression was best correlated with that of *PcAAE2*, whose expression was previously shown to be correlated with pogostone biosynthesis (Chen *et al.*, 2021), in all six *P. cablin* tissues including Seedling, 5W-Root, 5W-Stem, 5W-TLeaf, 8W-TLeaf and 8W-SLeaf (Table S3). Two of the 13 genes, *PcIBMS1* and *PcIPMDHL*, showed a statistically significant correlation ($P < 0.01$) (Table 1). In addition, in the coexpression analysis of all the unique genes in the *P. cablin* RNA-seq database with the *PcAAE2* gene through Pearson correlation analysis, *PcIBMS1* ranked the top third unique gene (Table S4).

To more accurately test whether the transcript profiles of these 13 genes correlated with that of *PcAAE2* in different *P. cablin* tissues at different developmental stages as previously described (Chen *et al.*, 2021), their expression patterns were further confirmed using RT-qPCR in the same 13 tissue samples used for *PcAAE2* (Fig. S2). Again, the expression patterns of *PcIBMS1* in the tested tissue samples showed a strong positive correlation with

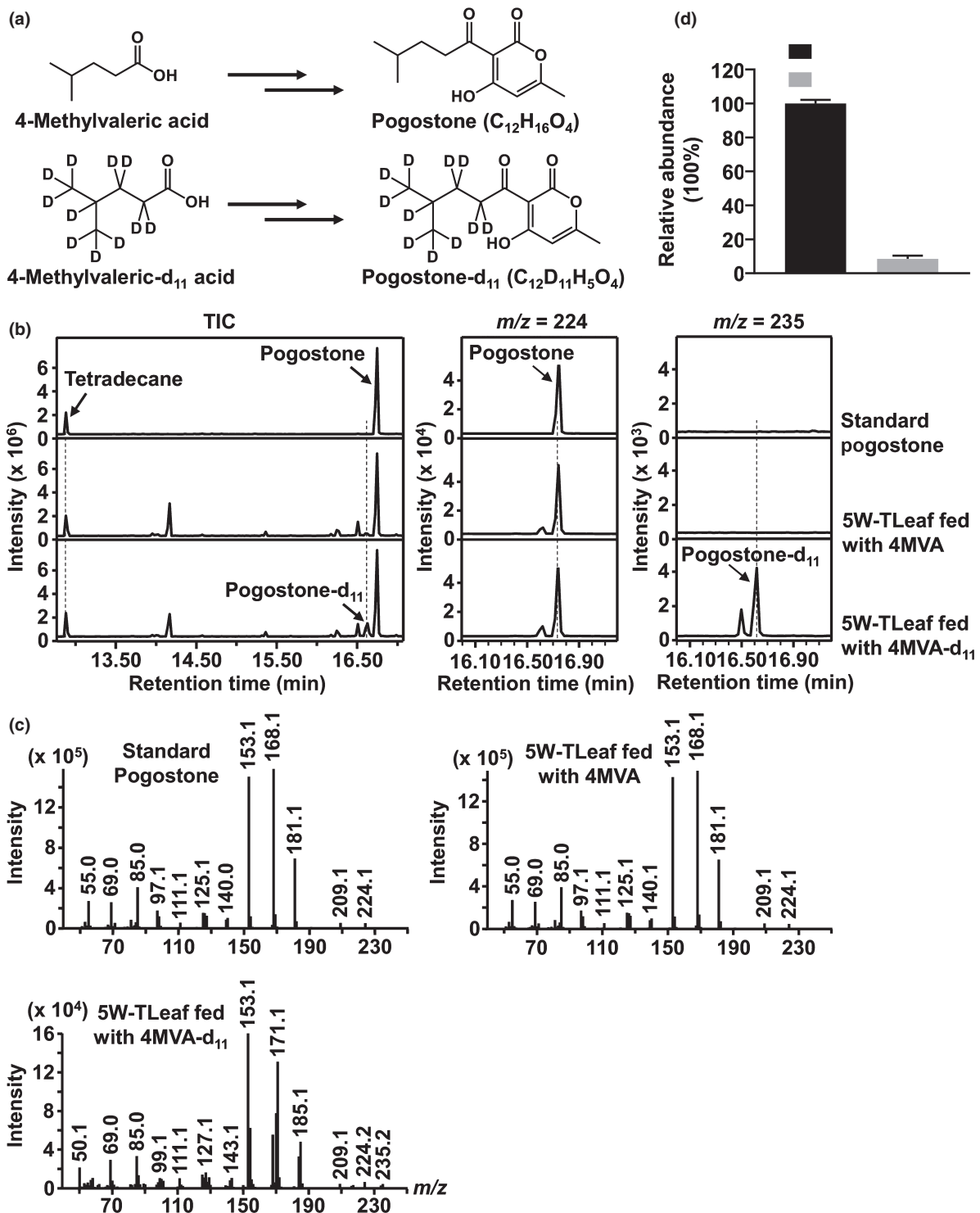


Fig. 2 Gas chromatography–mass spectrometry (GC-MS) analysis of stable isotope incorporation into pogostone from *Pogostemon cablin* leaves fed with 4-methylvaleric- d_{11} acid. (a) Chemical structures of pogostone from 4-methylvaleric acid (4MVA) and 4-methylvaleric- d_{11} acid (4MVA- d_{11}). (b) GC-MS analysis of *P. cablin* 5-wk-old top leaf (5W-TLeaf) fed with 4MVA or 4MVA- d_{11} . The total ion chromatograms (TIC) (left) and extracted ion chromatograms of m/z 224 for pogostone (middle) and of m/z 235 for pogostone- d_{11} (right) are shown. Tetradecane was used as an internal standard. The retention times of pogostone and pogostone- d_{11} are slightly different, most probably due to the slight differences in physical and chemical properties between these two compounds. (c) Mass spectra in GC-MS analysis of standard pogostone (upper left), pogostone from *P. cablin* 5W-TLeaf fed with 4MVA (upper right) and pogostone- d_{11} from *P. cablin* 5W-TLeaf fed with 4MVA- d_{11} (lower left). Note the 11-mass-unit shift of the parent ions of pogostone (224) to pogostone- d_{11} (235) upon feeding with 4MVA- d_{11} . (d) Relative abundances of pogostone and pogostone- d_{11} in *P. cablin* 5W-TLeaf fed with 4MVA- d_{11} . Relative abundances in GC-MS analysis of pogostone and pogostone- d_{11} in *P. cablin* 5W-TLeaf fed with 4MVA- d_{11} were calculated by peak area normalisation to the internal standard tetradecane and the relative abundance of pogostone was set as 100%. Data are represented as mean \pm SD of three independent biological replicates.

Table 1 Ranking of 4-methylvaleric acid biosynthetic pathway gene candidates using coexpression correlation analysis with the *PcAAE2* gene.

Genes	Pearson correlation coefficient	<i>P</i> (two tailed)
<i>PcIBMS1</i>	0.969	0.001
<i>PcIPMDHL</i>	0.95	0.004
<i>PcIPMIL SSU</i>	0.732	0.098
<i>PcBCATL3</i>	0.528	0.281
<i>PcIPMIL LSU</i>	0.455	0.365
<i>PcIPMS1</i>	0.451	0.369
<i>PcBCATL2</i>	0.387	0.449
<i>PcBCKDHL E3</i>	0.228	0.664
<i>PcBCKDHL E1β</i>	-0.066	0.9
<i>PcBCKDHL E1α2</i>	-0.143	0.788
<i>PcBCKDHL E2</i>	-0.203	0.699
<i>PcBCKDHL E1α1</i>	-0.385	0.452
<i>PcBCATL1</i>	-0.605	0.203

that of *PcAAE2*, displaying a similar developmental pattern as that of *PcAAE2* and the distribution pattern of pogostone (Chen *et al.*, 2021).

Subcellular localisation of *PcIBMS1* and *PcIPMS1*

To investigate the subcellular compartmentation of *PcIBMS1*, subcellular localisation prediction software TARGETP 2.0 was first utilised for signal peptide prediction of *PcIBMS1* as well as for *PcIPMS1*. Both *PcIBMS1* and *PcIPMS1* were predicted to harbour chloroplast transit peptides in their N-termini with a high score (Table S2). The subcellular localisations of these two proteins were further analysed experimentally by transient expression in Arabidopsis leaf protoplasts *PcIBMS1* and *PcIPMS1* each fused to the GFP at their C-termini. In these experiments, complete co-localisation of both *PcIBMS1*-GFP and *PcIPMS1*-GFP signals with red chlorophyll autofluorescence was observed (Fig. 3), indicating that both proteins are targeted to the plastids.

In vitro characterisation of isopropylmalate and isobutylmalate synthase activities of *PcIBMS1*

The combination of the positive correlation of expression patterns of *PcIBMS1* with that of *PcAAE2* and its further confirmation by RT-qPCR analysis made it the most likely candidate gene encoding the key enzyme in the biosynthetic pathway of 4-methylvaleric acid compared with the other 12 unique genes. To test this hypothesis, we performed *in vitro* biochemical assays with *PcIBMS1* and, for comparison, *PcIPMS1*. Truncated *PcIBMS1* and *PcIPMS1* proteins, both lacking the first 40 amino acids (the estimated length of the transit peptide), were produced in *E. coli* with a fused N-terminal His6 tag, and purified as soluble proteins. Both purified recombinant proteins were initially tested in a spectrophotometric end-point enzyme assay (5,5'-dithiobis(2-nitrobenzoic acid) (DTNB)) with a variety of 2-oxo acid substrates at a 0.3 mM concentration of the 2-oxo acid substrate and 1 mM of acetyl-CoA.

Recombinant *PcIBMS1* showed strong activity with 4-methyl-2-oxovalerate as a substrate, and also had significant activity with 2-oxoisovalerate (55% of its activity with 4-methyl-2-oxovalerate), but lower or no activity with the other substrates tested (Table 2). Recombinant *PcIPMS1* utilised 2-oxoisovalerate as the best substrate, with only low levels of activity or no activity with all other substrates tested (Table 2). In particular, *PcIPMS1* showed no detectable activity with 4-methyl-2-oxovalerate (Table 2).

To further characterise *PcIBMS1* and *PcIPMS1*, we first tested the dependency of their enzymatic activities on pH and cations. The highest levels of activities of *PcIBMS1* and *PcIPMS1* with their respective preferred substrates were observed at pH 8.0 (Fig. S3). The activities of *PcIBMS1* and *PcIPMS1* were dependent on divalent cations with Mg^{2+} being the most preferred cation at optimal concentration of 10 mM (Fig. S3). The kinetic properties of *PcIBMS1* with 4-methyl-2-oxovalerate and 2-oxoisovalerate as substrates, and *PcIPMS1* with 2-oxoisovalerate as the substrate were

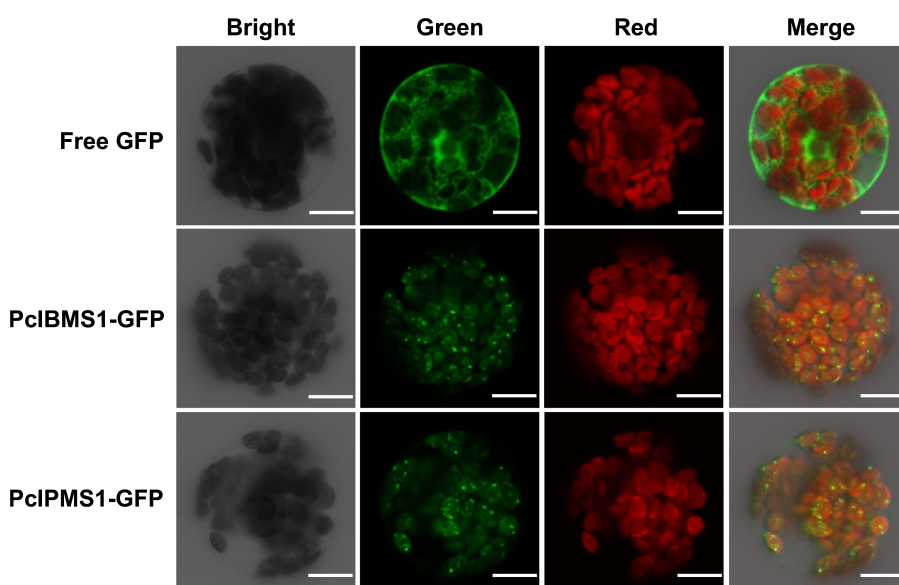


Fig. 3 Subcellular localisation of *PcIBMS1* and *PcIPMS1* in Arabidopsis leaf mesophyll protoplasts revealed using laser confocal microscopy. The signal of free GFP was used as a control and chloroplasts are revealed by red chlorophyll autofluorescence. Bar, 10 μ m.

Table 2 Substrate specificity of recombinant PclBMS1 and PclPMS1.

Substrate	Structure	PclBMS1	PclPMS1
Pyruvate		nd	nd
2-Oxobutyrate		18	10
4-Methylthio-2-oxobutanoic acid		15	nd
2-Oxoisovalerate		55	100 ^b
2-Oxovalerate		34	20
2-Oxoglutaric acid		nd	nd
3-Methyl-2-oxovalerate		8	nd
4-Methyl-2-oxovalerate		100 ^a	nd
2-Oxohexanoic acid		7	nd

These activities were measured with substrate and acetyl-CoA at concentration of 0.3 and 1 mM, respectively, in the end-point enzyme assay (DTNB). Data are expressed as relative mean percentages from triplicate independent assays. nd, not detected or relative activities < 5%.

^a100% relative activity of PclBMS1 corresponds to 1.60 $\mu\text{mol min}^{-1} \text{mg}^{-1}$ on 4-methyl-2-oxovalerate.

^b100% relative activity of PclPMS1 corresponds to 1.42 $\mu\text{mol min}^{-1} \text{mg}^{-1}$ on 2-oxoisovalerate.

measured in a spectrophotometric end-point enzyme assay. The kinetic analysis revealed that PclBMS1 had a K_m value of $139.8 \pm 39.4 \mu\text{M}$ and a catalytic efficiency of $1.72 \times 10^4 \text{s}^{-1} \text{M}^{-1}$

for 4-methyl-2-oxovalerate and a K_m value of $1303.3 \pm 127.7 \mu\text{M}$ and a catalytic efficiency of $3.337 \times 10^3 \text{s}^{-1} \text{M}^{-1}$ for 2-oxoisovalerate, and that PclPMS1 had a K_m value of $288.3 \pm 8.2 \mu\text{M}$ and a catalytic efficiency of $9.94 \times 10^3 \text{s}^{-1} \text{M}^{-1}$ for 2-oxoisovalerate (Table 3). The enzymatic products of PclBMS1 and PclPMS1 with 2-oxoisovalerate and 4-methyl-2-oxovalerate as substrates were further verified with LC-QTOF-MS (Fig. S4).

Heterologous expression of PclBMS1 and PclPMS1 in *N. benthamiana*

To further test the activities of PclBMS1 and PclPMS1 *in planta*, *N. benthamiana* leaves were infiltrated with *Agrobacterium tumefaciens* strains harbouring plasmids containing either PclBMS1 or PclPMS1. For control, *A. tumefaciens* strains harbouring empty vector (pEAQ-HT) were infiltrated into *N. benthamiana* leaves. In these experiments, the complete open reading frames of PclBMS1 and PclPMS1 were used, including the corresponding transit peptides that were shown to direct the protein to the plastid. Transformed leaves were harvested 7 d after infiltration and the products were extracted and analysed using LC-QTOF-MS. As expected, transient expression of PclBMS1 in *N. benthamiana* leaves resulted in the overproduction of 2-isobutylmalate and 2-isopropylmalate, which were verified by the comparison of their retention times and mass spectra with corresponding standards (Fig. 4a). By contrast, transient expression of PclPMS1 in *N. benthamiana* leaves resulted in negligible amounts of 2-isopropylmalate and 2-isobutylmalate compared with that produced in *N. benthamiana* leaves transiently expressing PclBMS1 (Fig. 4a).

To test whether the introduction of PclBMS1 or PclPMS1 can lead to the production of free 4-methylvaleric acid or an increase in its content, *N. benthamiana* leaf samples were extracted with MTBE, and the MTBE extracts were further analysed using GC-MS. No free 4-methylvaleric acid was

Table 3 Kinetic properties of recombinant PclBMS1, PclPMS1, PclBMS1-M132L and PclPMS1-L135M.

Enzyme	Substrate	K_m (μM)	K_{cat} (s^{-1})	K_{cat}/K_m ($\text{s}^{-1} \text{M}^{-1}$)
PclBMS1	2-Oxoisovalerate ^a	1303.3 ± 127.7	4.5 ± 0.21	3337 ± 156
	Acetyl-CoA ^b	42.0 ± 5.8	2.3 ± 0.07	26361 ± 761
	4-Methyl-2-oxovalerate ^c	139.8 ± 39.4	2.4 ± 0.09	17246 ± 653
	Acetyl-CoA ^d	145.4 ± 21.2	2.5 ± 0.02	17124 ± 106
PclBMS1-M132L	2-Oxoisovalerate ^a	443.5 ± 44.4	3.6 ± 0.16	8007 ± 352
	Acetyl-CoA ^b	57.6 ± 11.7	3.1 ± 0.01	53336 ± 220
	4-Methyl-2-oxovalerate ^c	95.0 ± 17.9	0.95 ± 0.01	9978 ± 87
	Acetyl-CoA ^d	131.9 ± 6.5	0.95 ± 0.03	7185 ± 189
PclPMS1	2-Oxoisovalerate ^a	288.3 ± 8.2	2.9 ± 0.04	9940 ± 126
	Acetyl-CoA ^b	48.1 ± 2.8	2.7 ± 0.02	56094 ± 429
PclPMS1-L135M	2-Oxoisovalerate ^a	621.5 ± 58.3	1.2 ± 0.03	1985 ± 49
	Acetyl-CoA ^b	22.7 ± 1.9	1.1 ± 0.02	48828 ± 1072

Data are presented as mean \pm SD from triplicate independent assays using end-point enzyme assay (DTNB) method.

^aKinetic parameters were determined with 1 mM acetyl-CoA.

^bKinetic parameters were determined with 2 mM 2-oxoisovalerate.

^cKinetic parameters were determined with 1 mM acetyl-CoA.

^dKinetic parameters were determined with 2 mM 4-methyl-2-oxovalerate.

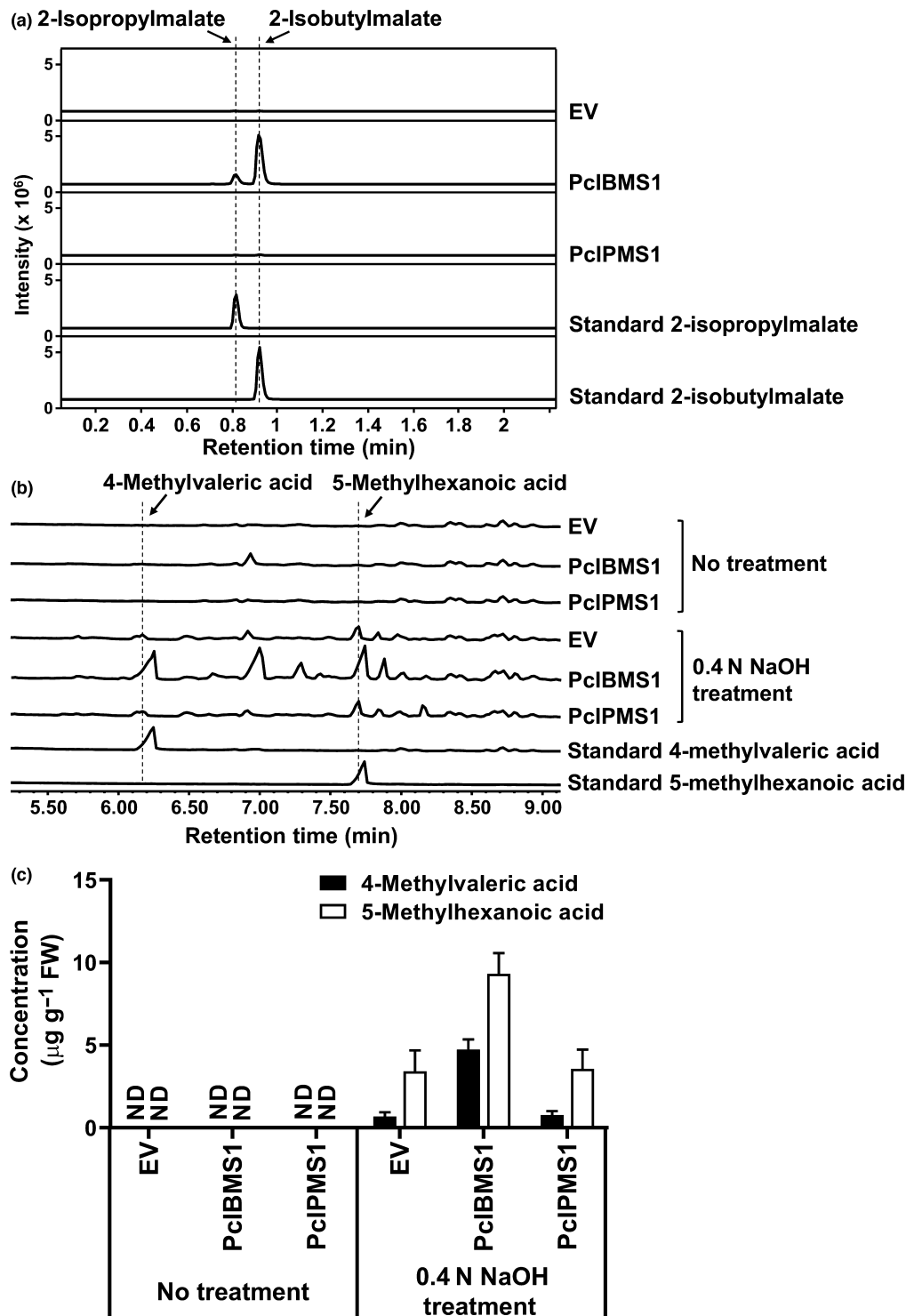


Fig. 4 Liquid chromatography–quadrupole time-of-flight–mass spectrometry (LC-QTOF-MS) and gas chromatography–mass spectrometry (GC-MS) analyses of *Nicotiana benthamiana* leaves expressing *PclBMS1* and *PclPMS1*. (a) LC-QTOF-MS analyses of *N. benthamiana* leaves expressing the two enzymes *PclBMS1* and *PclPMS1*. *Nicotiana benthamiana* leaves transformed with empty vector (EV) were used as a negative control. Extracted ion chromatograms in negative mode of both m/z 175.0612 for 2-isopropylmalate and 189.0768 for 2-isobutylmalate are shown. Standard 2-isopropylmalate is available commercially, whereas standard 2-butylmalate was generated by the *in vitro* enzymatic reactions of purified *PclBMS1* with 4-methyl-2-oxovalerate as the substrate in the presence of acetyl co-enzyme A (acetyl-CoA). (b) GC-MS analyses of *N. benthamiana* leaves expressing the two enzymes *PclBMS1* and *PclPMS1* with or without base treatment. The methyl tert-butyl ether (MTBE) extracts of *N. benthamiana* leaf samples with or without base treatment were analysed using GC-MS and total ion chromatograms are shown. (c) Concentration of 4-methylvaleric acid and 5-methylhexanoic acid in *N. benthamiana* leaves expressing the two enzymes *PclBMS1* and *PclPMS1* with or without base treatment. The quantifications were achieved by normalisation of the peaks to the internal standard geraniol and comparison with the standard curves of authentic 4-methylvaleric acid and 5-methylhexanoic acid ($n = 3$; means \pm SD). FW, fresh weight; nd, not detected.

detected in control *N. benthamiana* leaves and in *N. benthamiana* leaves transiently expressing *PcIBMS1* or *PcIPMS1* (Fig. 4b). As it had been previously observed that compounds with carboxylic groups generated in *N. benthamiana* leaves by the introduction of heterologous genes tended to be modified by endogenous glycosylases (Yang *et al.*, 2011; Xu *et al.*, 2018), we searched for glycosylated 4-methylvaleric acid by performing hydrolysis on leaf samples with 0.4 N NaOH at 80°C for 20 min. Surprisingly, 4-methylvaleric acid and 5-methylhexanoic acid (containing one more carbon than 4-methylvaleric acid) were detected at low levels in the hydrolysed MTBE extracts of control *N. benthamiana* leaves. However, these two compounds in *N. benthamiana* leaves expressing *PcIBMS1* showed 6.9-fold and 2.7-fold increases in their contents, respectively, compared with that in the control *N. benthamiana* leaves (Fig. 4b,c). Transient expression of *PcIPMS1* did not result in any observable changes in the contents of these two branched-chain fatty acids in the hydrolysed MTBE extracts compared with that in the control *N. benthamiana* leaves (Fig. 4b,c). In addition, the relative content of Leu in *N. benthamiana* leaves expressing *PcIBMS1* declined to *c.* 58% compared with that in the control *N. benthamiana* leaves (Fig. S5), while no change in Leu concentrations was observed in *N. benthamiana* leaves expressing *PcIPMS1*, which yielded almost the same content of Leu as the control *N. benthamiana* leaves (Fig. S5).

Leu feedback inhibition assay of PcIBMS1 and PcIPMS1

Isopropylmalate synthases from plants and bacteria generally form a homodimeric protein (Koon *et al.*, 2004; de Carvalho *et al.*, 2005; de Kraker *et al.*, 2007; de Kraker & Gershenzon, 2011). Each monomer contains two core domains including an N-terminal catalytic region consisting mainly of a (β/α)₈ barrel (TIM barrel) with a divalent metal cofactor necessary for substrate binding (Koon *et al.*, 2004), and a C-terminal allosteric regulatory domain responsible for Leu feedback inhibition in microbes and plants (de Carvalho *et al.*, 2005; de Kraker & Gershenzon, 2011). The two major domains are separated by two small subdomains that have a flexible hinge in between them (Koon *et al.*, 2004). The full-length *PcIBMS1* gene encodes a protein containing 591 amino acids, showing 74%, 66% and 64% amino acid identities with PcIPMS1, AtIPMS1 and OsIPMS1, respectively (Fig. S6), the latter two proteins were functionally identified as typical IPMSs in *A. thaliana* and *Oryza sativa* (de Kraker *et al.*, 2007; Xing & Last, 2017; He *et al.*, 2019). Sequence comparison with other typical IPMSs showed that the protein encoded by the *PcIBMS1* gene contains the N-terminal catalytic region and the likely C-terminal allosteric regulatory domain (Fig. S6). However, many changes in residues occur in the PcIBMS1 protein at positions corresponding to the conserved sites of the other typical IPMS proteins in both the N-terminal catalytic region (Fig. 5) and the C-terminal allosteric regulatory domain, notably the loss of five amino acids in the C-terminal domain of the PcIBMS1 protein (Fig. S7). The sequence

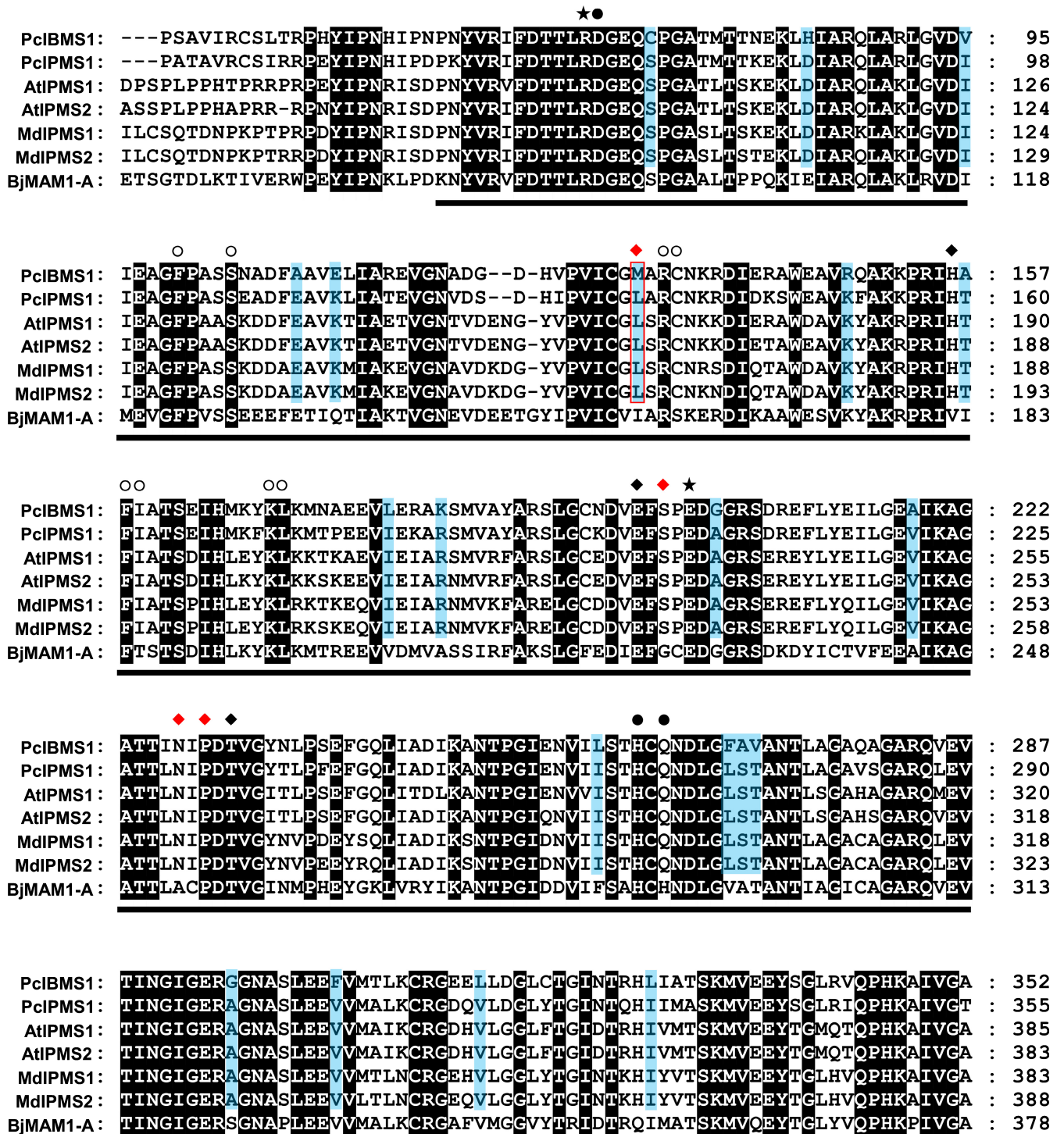
divergence in the C-terminal allosteric regulatory domain of PcIBMS1 protein compared with other typical IPMSs that are sensitive to Leu feedback inhibition led us to hypothesise that *PcIBMS1* encoded an IPMS family protein that was not subject to Leu feedback inhibition.

To test this hypothesis, both recombinant PcIBMS1 and PcIPMS1 were assayed for aldol-type condensation activities between acetyl-CoA and 4-methyl-2-oxovalerate and 2-oxoisovalerate, in the end-point enzyme assay (DTNB) in the presence of Leu at a series of concentrations from 0.05 to 10 mM. PcIBMS1 activity was insensitive to Leu concentrations up to the highest concentration measured (10 mM) no matter whether the 2-oxo acid substrate was 4-methyl-2-oxovalerate or 2-oxoisovalerate (Fig. 6a,b). Conversely, the activity of PcIPMS1 with the substrate 2-oxoisovalerate declined to 57.8% at 50 μ M Leu and was reduced to 34.7% at the highest Leu concentration measured (10 mM) (Fig. 6b). The observations of substrate specificity and kinetic properties of PcIPMS1 enzyme towards 2-oxoisovalerate and acetyl-CoA and its sensitivity to Leu feedback inhibition indicated that PcIPMS1 encoded a genuine IPMS enzyme responsible for Leu biosynthesis.

Determination of amino acid sites in the N-terminal domain that may affect the substrate specificity of PcIBMS1 and PcIPMS1

Our data indicate that PcIBMS1 is an IPMS-like protein with an altered preferred substrate (4-methyl-2-oxovalerate rather than 2-oxoisovalerate). Previous studies on the 3D structure of MtIPMS from *Mycobacterium tuberculosis* and BjMAM1-A from *Brassica juncea* identified seven amino acid sites in the N-terminal catalytic domain that may affect the 2-oxo acid substrate chain length specificities of IPMS and methylthioalkylmalate synthase (MAMS) (de Kraker & Gershenzon, 2011; Kumar *et al.*, 2019), the latter is a neofunctionalised form of IPMS involved in the biosynthesis of methionine-derived glucosinolates (de Kraker *et al.*, 2007; de Kraker & Gershenzon, 2011). The overall fold of MAMs shares with IPMSs the N-terminal catalytic α/β -barrel domain and the C-terminal α -helical region that forms part of the CoA binding site (Kumar *et al.*, 2019). The two major structural differences between MAMs and IPMSs are the loss of the N-terminal extension and the C-terminal Leu binding regulatory domain found in the Leu biosynthesis enzyme. In both MAMs and IPMSs, the catalytic machinery and reaction chemistry are conserved (Kumar *et al.*, 2019).

Sequence comparison of the N-terminal domain of PcIBMS1 that catalyses the extension of the longer chain substrate 4-methyl-2-oxovalerate with those of the five IPMS proteins, including PcIPMS1, AtIPMS1, AtIPMS2, MdIPMS1 and MdIPMS2, all of which show no or negligible activities towards the longer chain substrate 4-methyl-2-oxovalerate compared with that towards the classic IPMS substrate 2-oxoisovalerate (de Kraker *et al.*, 2007; Sugimoto *et al.*, 2021), highlights the sequence divergence in PcIBMS1 (Fig. 5). Twenty conserved



★ Catalytic sites ● Metal binding sites ○ CoA binding sites ◆ ◆ α-ketoacid binding sites — TIM barrel

Fig. 5 Multiple sequence alignment of N-terminal catalytic regions of *Pogostemon cablin*, *Malus domestica* and *Arabidopsis* isopropylmalate synthases (IPMS) and *Brassica juncea* MAM1-A (BjMAM1-A). Multiple sequence alignment of the N-terminal catalytic regions of *P. cablin*, *M. domestica* and *Arabidopsis* IPMS and BjMAM1-A is shown. Sequence features corresponding to the characterised BjMAM1-A crystal structure are shown (Kumar *et al.*, 2019). Residues in catalytic sites, metal binding sites, CoA binding sites and α-ketoacid binding sites are marked by an asterisk, solid circle symbol, hollow circle symbol and solid diamond symbol, respectively. The black solid diamond symbol depicts α-ketoacid binding sites implicated in substrate selectivity, while red solid diamond symbol are those shown to impact substrate size discrimination (de Kraker & Gershenzon, 2011; Kumar *et al.*, 2019). Protein domain designations are derived from the BjMAM1-A crystal structure (Kumar *et al.*, 2019). Black background shows perfectly conserved sequences across the IPMS, while the light blue background shows that amino acids in the conserved sites in the selected five typical IPMSs diverge from that in the corresponding sites in PcIBMS1. PcIBMS1 and PcIPMS1 were directly obtained from the *P. cablin* transcriptome database. The information for other proteins used in this sequence alignment is shown in Supporting Information Table S5.

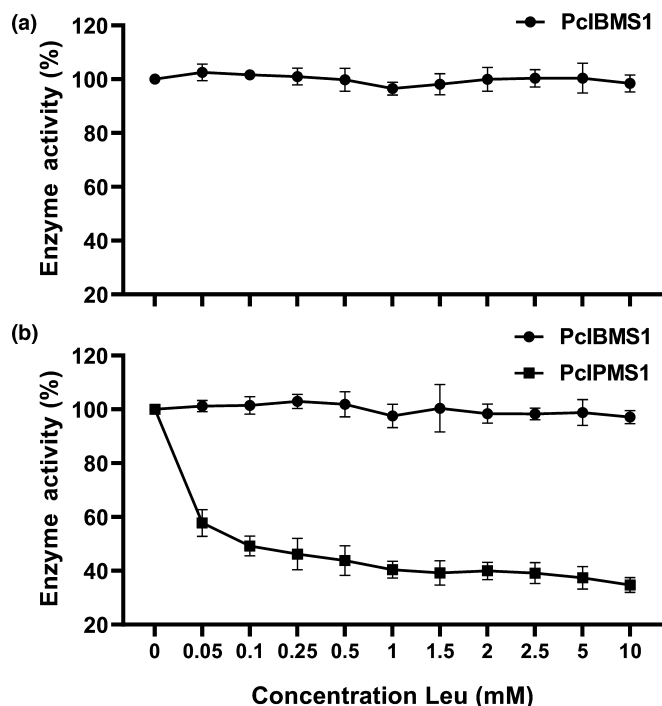


Fig. 6 Effect of leucine (Leu) on the enzymatic activity of PcIBMS1 and PcIPMS1. (a) Effect of Leu on the enzymatic activities of PcIBMS1 towards 4-methyl-2-oxovalerate. (b) Effect of Leu on the enzymatic activities of both PcIBMS1 and PcIPMS1 towards 2-oxoisovalerate. Activities are expressed as a percentage of the activity in the absence of Leu. The 100% activity was set in the absence of Leu. Data represent means \pm SD of three independent biological replicates.

amino acid sites of N-terminal catalytic domains of these five typical IPMS proteins diverged from the corresponding sites in PcIBMS1 protein (Fig. 5). In particular, the conserved Leu located in one of the seven sites in these IPMSs that were identified as accounting for the differences in substrate binding between IPMSs and MAMs (de Kraker & Gershenzon, 2011; Kumar *et al.*, 2019), was replaced by Met¹³² in the same position in PcIBMS1 (Fig. 5), suggesting that the substituted Met¹³² may play a vital role in the alteration of substrate specificity of PcIBMS1 from typical IPMSs. To further explore the implication of the Leu-to-Met substitution, we performed structure modelling of PcIBMS1 and PcIPMS1. The overall structures of PcIBMS1 and PcIPMS1 obtained by modelling were similar and shared a conserved substrate binding pocket (Fig. S8). The Leu-to-Met substitution in this pocket resulted in the distance between 4-methyl-2-oxovalerate and the residues changing from 4.9 to 3.1 Å (Figs S8, S9). Therefore, the Met¹³² in PcIBMS1 had a stronger binding affinity to 4-methyl-2-oxovalerate than Leu¹³⁵ in PcIPMS1.

To test the effect of Met¹³² in determining substrate specificity of PcIBMS1, we mutated this residue in PcIBMS1 to the corresponding residue Leu found in the typical IPMS, and we did the reciprocal experiment by changing the corresponding Leu¹³⁵ in PcIPMS1 to Met. We assayed the resulting PcIBMS1-M132L and PcIPMS1-L135M point mutants using 4-methyl-2-oxovalerate and 2-oxoisovalerate and determined their kinetic

parameters (Table 3). The PcIBMS1-M132L mutant exhibited *c.* a three-fold decrease in K_m value for 2-oxoisovalerate with a corresponding *c.* 2.4-fold increase in catalytic efficiency (k_{cat}/K_m), while showing only a modest difference in the K_m value for 4-methyl-2-oxovalerate and *c.* a two-fold decrease in catalytic efficiency vs the wild-type. The PcIPMS1-L135M point mutant exhibited *c.* a 2.2-fold increase in K_m value for 2-oxoisovalerate with a corresponding *c.* five-fold decrease in catalytic efficiency (k_{cat}/K_m) vs the wild-type, while still showing no detectable activities with 4-methyl-2-oxovalerate. These results suggested that this amino acid site indeed contributed to the ability of the enzyme to accept the longer chain 2-oxo acid 4-methyl-2-oxovalerate but that other amino acids played a crucial role as well.

Phylogenetic relationships of PcIBMS1 homologues

To gain further insight into the evolution of PcIBMS1, we expanded the phylogenetic analysis of PcIBMS1 and PcIPMS1 to include their closest homologues from plant species and microorganisms, including genuine IPMSs, MAMs from *A. thaliana* and *B. juncea*, and citramalate synthase (CMS) from *Malus domestica* (Fig. 7). PcIBMS1, MdCMS1 and MAMs, as well as SIIPMS3, a tomato true IPMS that is nonetheless lacking the regulatory C-terminal regulatory domain and is involved in synthesising the C5 acyl precursor for acylsugar biosynthesis rather than leucine, split off from typical plant IPMSs before the split between monocot and dicot IPMSs (Fig. 7a). However, these leucine-insensitive enzymes represented neofunctionalised forms of typical plant IPMS (Ning *et al.*, 2015; Kumar *et al.*, 2019; Sugimoto *et al.*, 2021) that either lack the regulatory C-terminus domain altogether or possess the C-terminal domain but show high rates of change in it. We therefore redid the phylogenetic analysis with the protein sequences from which the C-terminal regulatory domain was removed (Fig. 7b). The phylogenetic tree obtained in this analysis showed that PcIBMS1 was most closely related to PcIPMS1.

Discussion

4-Methylvaleric acid is a precursor of pogostone

Pogostone possesses many pharmaceutical activities and therefore is considered as a potential therapeutic agent for treatment of many diseases. In our previous study, we constructed a *P. cablin* RNA-seq database, from which we isolated and characterised PcAAE2, an enzyme that catalyses the formation of 4-methylvaleryl-CoA from 4-methylvaleric acid, a branched-chain fatty acid (Chen *et al.*, 2021). The expression of PcAAE2 correlated spatially and temporally with pogostone biosynthesis and, based on these data as well as the structure of pogostone, 4-methylvaleric acid was hypothesised as a precursor of pogostone (via 4-methylvaleryl-CoA). Here, by feeding *P. cablin* young leaves with deuterium-labelled [²H₁₁] 4-methylvaleric acid (4MVA-d₁₁) and following the incorporation of this label into pogostone, we were able to show that 4-methylvalerate was indeed a precursor in pogostone biosynthesis (Figs 2, S1).

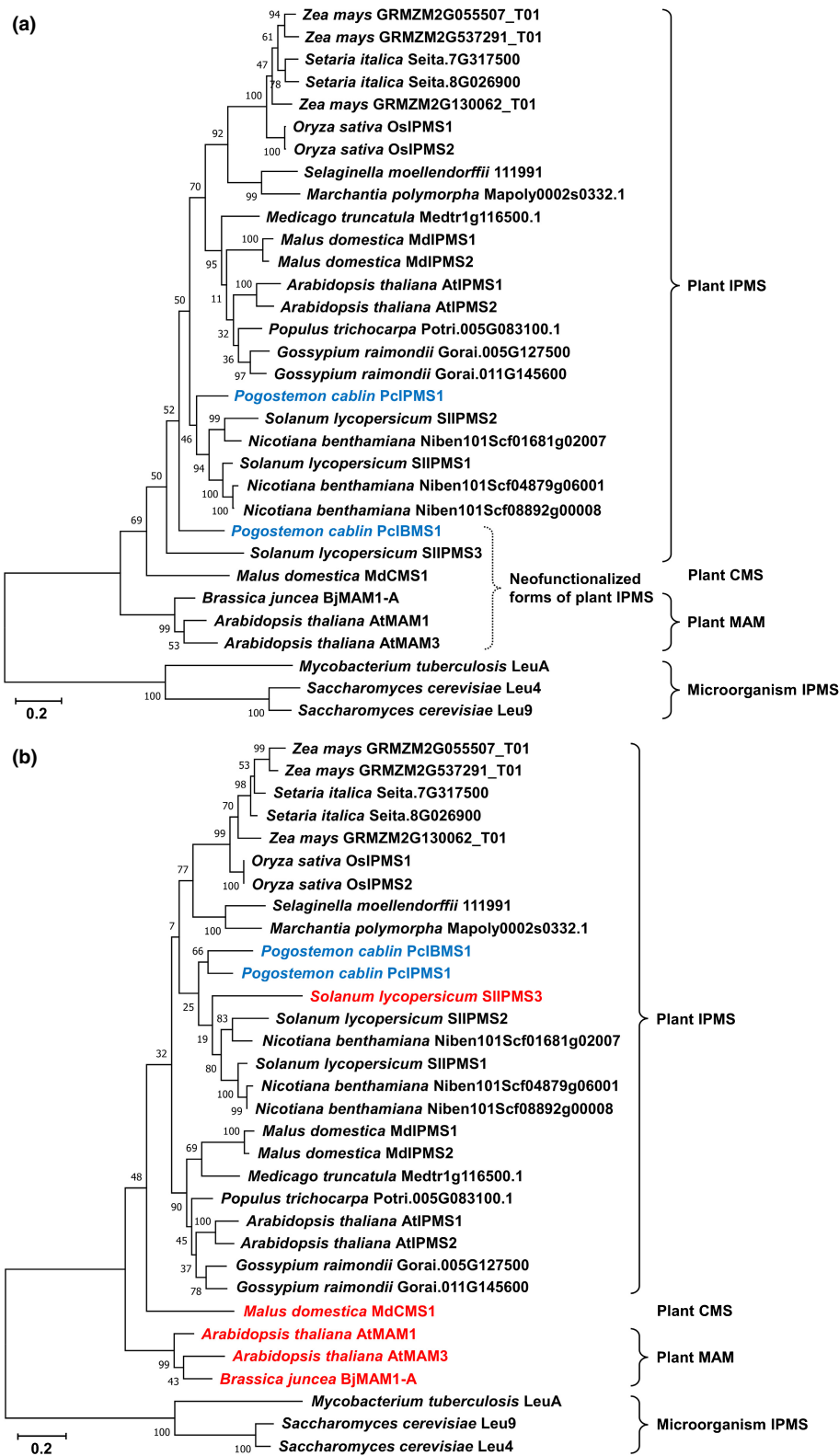


Fig. 7 Phylogenetic analyses of PclBMS1 homologues, methylthioalkylmalate synthases (MAMs) and citramalate synthase (CMS) from selected species.

(a) Phylogenetic analyses of whole PclBMS1 homologues, MAMs and CMS from selected species. The whole protein sequences of PclBMS1 homologues, MAMs and CMS were used in this phylogenetic analysis. (b) Phylogenetic analyses of the N-terminal domains of PclBMS1 homologues, MAMs and CMS from selected species. PclBMS1 and all the 2-isopropylmalate synthases (IPMS) except SIIPMS3, from which the predicted C-terminal allosteric regulatory domains were removed, were used in this phylogenetic analysis. The proteins naturally lacking the C-terminal allosteric regulatory domain that is presented in typical IPMS were marked in red colour. For the construction of phylogenetic trees in (a) and (b), all proteins except PclBMS1 and PclPMS1 were obtained from the PHYTOZONE v.12.1 (<https://phytozone.jgi.doe.gov/pz/portal.html>), National Center for Biotechnology Information (NCBI) website (<https://www.ncbi.nlm.nih.gov/>), The Arabidopsis Information Resource (TAIR) website (<https://www.arabidopsis.org/>) and UniProt website (<https://www.uniprot.org/>) (Supporting Information Table S5). PclBMS1 and PclPMS1, the protein sequences of which were directly obtained from *Pogostemon cablin* transcriptome database, are marked in blue colour. Sequences were aligned using MUSCLE (Edgar, 2004), and the trees were conducted 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 trees 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.

PclBMS1 is a key enzyme involved in the formation of 4-methylvaleric acid

Branched and straight fatty acids have been hypothesised as the precursors of many specialised metabolites such as bitter acids in

Humulus lupulus, acylsugars in Solanaceae and cannabinoids in *Cannabis sativa* (Kroumova & Wagner, 2003; Stout *et al.*, 2012; Xu *et al.*, 2013; Ning *et al.*, 2015), although direct evidence for the activity of key enzymes in their biosynthesis has seldom been presented. To further elucidate the route leading to biosynthesis

of pogostone, we set out to characterise the key enzyme involved in the biosynthesis of 4-methylvaleric acid. This branched C6 fatty acid is longer by one carbon than 3-methylbutanoic acid, a derivative of Leu whose own synthesis involves the α -ketoacid elongation (α KAE) pathway (Kroumova & Wagner, 2003; Binder, 2010). Therefore, 4-methylvaleric acid is most likely to have been derived from leucine metabolism by one additional elongation cycle through the α KAE pathway.

Among the 13 gene candidates for 4-methylvaleric acid biosynthesis that were identified by utilising Arabidopsis representative genes involved in leucine metabolism to query the assembled *P. cablin* transcriptome data through homology search by tBLASTn software, PcIBMS1 showed the highest correlation with pogostone biosynthesis markers (Tables 1, S4; Fig. S2) and was homologous to IPMS, the key enzyme in the one-carbon elongation of 2-oxoisovalerate to 4-methyl-2-oxovalerate (via 2-isopropylmalate) in Leu biosynthesis, making it the most likely candidate encoding the key enzyme for 4-methylvaleric acid biosynthesis. *In vitro* biochemical assays of the recombinant PcIBMS1 protein showed that PcIBMS1 not only effectively catalysed the formation of 2-isobutylmalate from 4-methyl-2-oxovalerate and acetyl-CoA, but was also able to catalyse the formation of 2-isopropylmalate from 2-oxoisovalerate and acetyl-CoA (Table 2, 3; Fig. S4). A second IPMS-like gene found in the *P. cablin* transcriptome, but whose expression pattern did not correlate with pogostone biosynthesis, did not have activity with 4-methyl-2-oxovalerate (Table 3). This gene, which was designated as PcIPMS1, was shown to be a *bona fide* isopropylmalate synthase that effectively used 2-oxoisovalerate as a substrate with kinetic properties similar to other plant typical IPMSs (Table 3; de Kraker *et al.*, 2007; Ning *et al.*, 2015). The K_m value of PcIBMS1 for 2-oxoisovalerate was *c.* five-fold higher than that of PcIPMS1, but its K_m value for 4-methyl-2-oxovalerate was three-fold lower than the K_m value of PcIPMS1 with 2-oxoisovalerate. However, PcIBMS1 has a somewhat high K_m value for acetyl-CoA when acting on its preferred substrate, compared with IPMSs (Table 3; de Kraker *et al.*, 2007; Ning *et al.*, 2015). These enzymatic properties of PcIBMS1 suggested that it may perform distinct biological functions from typical plant IPMSs that are responsible for Leu biosynthesis.

Attempts to establish a reliable *P. cablin* transformation system, including gene silencing (VIGS) using tobacco rattle virus, were not successful, preventing us from testing the biological functions of PcIBMS1 directly in *P. cablin* plants. Therefore, the activities of PcIBMS1 *in planta* were tested in *N. benthamiana* leaves by transient expression. Analysis of compounds produced in *N. benthamiana* leaves when *PcIBMS1* was transiently expressed lent additional evidence for its function in 4-methylvaleric acid biosynthesis. We were able to directly detect free 2-isopropylmalate and 2-isobutylmalate, which were not detected in the control, as well as a large increase in the content of glycosylated 4-methylvaleric acid and glycosylated 5-methylhexanoic acid (Fig. 4). These results indicated that when *PcIBMS1* was overexpressed in *N. benthamiana*, the resulting protein could also act with 2-oxoisovalerate, despite its high K_m value for this substrate, and also with 5-methyl-2-oxohexanoate.

This further indicated that endogenous *N. benthamiana* proteins could complete the elongation cycle with 2-isobutylmalate and longer intermediates. *N. benthamiana* is known to synthesise these longer branched-chain fatty acids as part of its acylsugar biosynthesis in glandular trichomes (Kroumova & Wagner, 2003; Slocombe *et al.*, 2008), which is a possible explanation why we detected low levels of glycosylated 4-methylvaleric acid and glycosylated 5-methylhexanoic acid in control *N. benthamiana* leaves.

Met¹³² amino acid site in PcIBMS1 affects its substrate preference

The MAM structure of a MAMS from *B. juncea*, designated as BjMAM1-A, was determined using X-ray crystallography (Kumar *et al.*, 2019). Seven amino acid sites that may affect substrate preference of IPMS and MAMS were identified. Our sequence comparisons showed that one of the seven corresponding sites was substituted by Met¹³² in PcIBMS1 (Fig. 5). This site corresponded to Leu¹⁴³ in Mt-IPMS that was shown to have a hydrophobic interaction with one of the two methyl groups of 2-oxoisovalerate by RCSB PDB Ligand Explorer 3.5 (Koon *et al.*, 2004; de Kraker & Gershenzon, 2011), suggesting that Met¹³² may play a vital role in substrate preference of PcIBMS1. The PcIBMS1-M132L mutant obtained using site-directed mutagenesis showed much higher affinity and catalytic efficiency towards 2-oxoisovalerate and lower catalytic efficiency towards 4-methyl-2-oxovalerate than wild-type PcIBMS1 (Table 3), which was consistent with the prediction of the effect of the Leu-to-Met substitution in the binding pocket in the structural modelling of PcIBMS1 (Figs S8, S9). The reciprocal PcIPMS1 mutant, PcIPMS1-L135M, showed a strong decrease in affinity and catalytic efficiency towards 2-oxoisovalerate compared with the wild-type enzyme, but it still did not have the capability of utilising the larger substrate 4-methyl-2-oxovalerate. Therefore, while Met¹³² plays a critical role in PcIBMS1 substrate specificity, simply substituting Leu in this position in IPMS, is clearly not sufficient to convert it into IBMS.

PcIBMS1 has altered feedback regulation of enzyme activity

PcIPMS1 is a typical IPMS enzyme that exhibits Leu feedback inhibition both *in vitro* (Fig. 6) and *in planta* (overexpression of *PcIPMS1* in *N. benthamiana* did not lead to an increase in Leu concentrations; Fig. S5). By contrast, our sequence comparisons showed that PcIBMS1 had a wide sequence divergence from typical plant IPMS proteins in both the catalytic N-terminal (Fig. 5) and the C-terminal regulatory domain (Fig. S7). In the latter domain, a deletion of five amino acids in PcIBMS1 was particularly notable (Fig. S7). The feedback inhibition assay indicated that *PcIBMS1* encoded an enzyme that was not subject to Leu inhibition (Fig. 6). The loss of Leu regulation by the sequence divergence and the loss of five amino acids in PcIBMS1 were therefore obtained by a different mechanism of the complete loss of the C-terminal domain, which is how Brassica MAMS and

Solanaceae SIIPMS3 acquired their insensitivity to Leu inhibition. In addition, the loss of Leu feedback inhibition in PclBMS1 was consistent with the observation that transient expression of *PcIBMS1* in *N. benthamiana* led to the overproduction of 2-isopropylmalate, but transient expression of *PcIPMS1* did not. Further structure analysis and gain-of-function mutations in the C-terminal regulatory domain of PclBMS1 might provide a better understanding of the mechanism leading to the insensitivity to feedback inhibition of this enzyme.

IPMSs have been recruited and neofunctionalised to specialised metabolism pathways multiple times throughout plant evolution

Several instances of IPMS neofunctionalisation and recruitment to specialised metabolism within eudicots have been documented, including Brassicaceae MAMs involved in the biosynthesis of Met-related glucosinolate compounds, apple CMS involved in flavour esters, and wild tomato *IPMS3* involved in acylsugar biosynthesis (Ning *et al.*, 2015; Kumar *et al.*, 2019; Sugimoto *et al.*, 2021). In these three cases, a loss of the C-terminal allosteric regulatory domain had occurred, which freed these enzymes from Leu inhibition. When the phylogeny of these enzymes was analysed using standard methodology, they appeared to have split from the genuine IPMS clade before the split of monocot and dicot IPMSs (Fig. 7a). However, this analysis is probably misleading, as the genes encoding these enzymes have undergone a major change involving the loss of a domain and possibly accelerated substitutions elsewhere. When only the catalytic domain was used in the phylogenetic analysis, MAMs and CMS still split early from the IPMS clade, but PclBMS1 now clustered with PcIPMS1, and wild tomato IPMS3 groups with other Solanaceae IPMSs (Fig. 7b). This analysis indicated that at least for PclBMS1, its recruitment and neofunctionalisation into the pogostone pathway occurred relatively recently.

In addition to the changes in the catalytic domain that were obviously necessary to evolve a new substrate specificity, the uniform loss of Leu inhibition is of interest. As the end product of the pathways in which these enzymes participate is not Leu, a Leu sensitivity will serve no clear function. But perhaps, more importantly, specialised metabolites such as glucosinolates, flavour esters, acylsugars, and pogostone are typically made in large quantities in specific tissues or organs, and inhibition by related metabolites would be detrimental to achieving such high levels of production.

Based on our results with PclBMS and the previous research on the origin of MAMs and CMS, it is likely that many more IPMS-derived enzymes responsible for synthesising substrates for the biosynthesis of specialised metabolism are still to be discovered. In particular, one anticipates these genes in the Solanaceae, in which many species synthesise acyl sugars using acyls up to 12 carbons long, and for which the α KAE pathway was shown to be involved in *Nicotiana* and *Petunia*. In fact, our detection in *N. benthamiana* leaves (not expressing PclBMS) of basal levels of 4-methylvaleric acid and 5-methylhexanoic acid (Fig. 4) suggested that *N. benthamiana* has endogenous enzymes related to IBMS.

However, the three IPMS-like genes that can be presently identified in the *N. benthamiana* genome all clustered together with genuine IPMS sequences, whereas PclBMS1 was in a separate clade (Fig. 7). Furthermore, all three shared the motif that was responsible for Leu feedback inhibition (Fig. S10), and the IPMS-specific Leu residue corresponding to Met¹³² in PclBMS1 (Fig. S11). Therefore, at this point it cannot be determined if any of these three IPMS-like genes from *N. benthamiana* encoded an enzyme with IBMS activity, or whether the observed IBMS activity in *N. benthamiana* was due to an enzyme encoded by another gene.

PclBMS1 catalyses the committed step in pogostone biosynthesis

The initial elongation step of 4-methylvaleric acid biosynthesis catalysed by PclBMS1 was critical for connecting primary and specialised metabolism (Fig. 1). It is of interest that PclBMS1 has evolved to use 4-methyl-2-oxovalerate as a more efficient substrate than 2-oxoisovalerate, the preferred substrate of typical plant IPMSs, but still retained high enzymatic activities towards the latter substrate (Table 3). Both 2-isobutylmalate and 2-isopropylmalate were detected in *N. benthamiana* leaves transiently expressing PclBMS1 (Fig. 4), suggesting that 2-oxoisovalerate can also be a natural substrate *in planta*. It is plausible that the initial elongation steps from both 2-oxoisovalerate and 4-methyl-2-oxovalerate in the α KAE route for 4-methylvaleric acid biosynthesis are catalysed by only PclBMS1, allowing ultimately high levels of pogostone production from Leu inhibition. As only one homologue of each of IPMI and IPMDH was screened from *P. cablin* transcriptome database (Table S2), the elongation cycles of the α KAE route for 4-methylvaleric acid biosynthesis probably shared both enzymes with Leu biosynthetic pathway in the plastid.

Four out of five screened BCKDH subunits were predicted to locate in mitochondria (Table S2), suggesting that the proposed BCKDH-like reactions in 4-methylvaleric acid formation occurred in mitochondria such as leucine catabolism (Fig. 1; Binder, 2010). A recent characterisation of the intrinsic acyl-CoA thioesterase activity of a peroxisomal ATP binding cassette transporter that was required for transport and metabolism of fatty acids (Lousa *et al.*, 2013) raised the possibility of the involvement of this type of transporter in the transport of 4-methylvaleryl-CoA generated by the action of BCKDH-like reactions in mitochondria into the cytosol, along with losing the CoA group to yield free 4-methylvaleric acid (Fig. 1). Although the complete pathway leading to pogostone biosynthesis is still unresolved, the characterisation of PclBMS1 will allow the breeding of high-yield-pogostone *P. cablin* varieties and reconstitution of the pathway leading to biosynthesis of the branched side of pogostone combined with PcAAE2 in a heterologous system (e.g. *N. benthamiana*).

Acknowledgements

We would like to thank Dr Deng at the Analytical and Testing Center of Chongqing University for their assistance with LC-

QTOF-MS analysis and laser scanning confocal microscopy analysis. This work was supported by the National Natural Sciences Foundation of China (grant no. 31970319 to HX), Fundamental Research Funds for the Central Universities (2020CDJQY-A077), the National Key Research and Development Program (2018YFD1000800), and 100 Talent Program (Chongqing University, 0304001104433 to HX).





Competing interests

The authors have no conflicts of interest to declare.

Author contributions

HX designed the experiments; Chu Wang, YW, JC, LL and ZL conducted the experiments, analysed data or provided material; MY and Chengyuan Wang conducted and analysed the protein modelling. EP and HX wrote the article; all authors edited the article. CW, YW and JC contributed equally to this work.

ORCID

Zhengguo Li  <https://orcid.org/0000-0001-7019-2560>
 Eran Pichersky  <https://orcid.org/0000-0002-4343-1535>
 Chengyuan Wang  <https://orcid.org/0000-0002-8605-0716>
 Haiyang Xu  <https://orcid.org/0000-0002-6954-0000>

Data availability

Data supporting the finding of this study are available within the article and its supplemental files. The sequence data used in this study can be obtained from the National Center for Biotechnology Information (NCBI, USA) with the following GenBank accession nos.: PcIBMS1, MW413958, PcIPMS1 and MW413959.

References

- Binder S. 2010. Branched-chain amino acid metabolism in *Arabidopsis thaliana*. *The Arabidopsis Book* 8: e0137.
- Binder S, Knill T, Schuster J. 2007. Branched-chain amino acid metabolism in higher plants. *Physiologia Plantarum* 129: 68–78.
- Cao ZX, Yang YT, Yu S, Li YZ, Wang WW, Huang J, Xie XF, Xiong L, Lei S, Peng C. 2017. Pogostone induces autophagy and apoptosis involving PI3K/Akt/mTOR axis in human colorectal carcinoma HCT116 cells. *Journal of Ethnopharmacology* 202: 20–27.
- de Carvalho LP, Argyrou A, Blanchard JS. 2005. Slow-onset feedback inhibition: inhibition of *Mycobacterium tuberculosis* alpha-isopropylmalate synthase by L-leucine. *Journal of the American Chemical Society* 127: 10004–10005.
- Chen H, Liao H, Liu Y, Zheng Y, Wu X, Su Z, Zhang X, Lai Z, Lai X, Lin Z-X *et al.* 2015. Protective effects of pogostone from *Pogostemonis herba* against ethanol-induced gastric ulcer in rats. *Fitoterapia* 100: 110–117.
- Chen J, Liu L, Wang Y, Li Z, Wang G, Kraus GA, Pichersky E, Xu H. 2021. Characterization of a cytosolic acyl-activating enzyme catalyzing the formation of 4-methylvaleryl-CoA for pogostone biosynthesis in *Pogostemon cablin*. *Plant and Cell Physiology* 62: 1556–1571.
- Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
- Halkier BA, Gershenzon J. 2006. Biology and biochemistry of glucosinolates. *Annual Review of Plant Biology* 57: 303–333.
- He Y, Cheng J, He Y, Yang B, Cheng Y, Yang C, Zhang H, Wang Z. 2019. Influence of isopropylmalate synthase OsIPMS1 on seed vigour associated with amino acid and energy metabolism in rice. *Plant Biotechnology Journal* 17: 322–337.
- Hu G, Peng C, Xie X, Zhang S, Cao X. 2017. Availability, pharmacetics, security, pharmacokinetics, and pharmacological activities of patchouli alcohol. *Evidence-Based Complementary and Alternative Medicine: eCAM* 2017: 4850612.
- Huang SH, Xian JD, Kong SZ, Li YC, Xie JH, Lin J, Chen JN, Wang HF, Su ZR. 2014. Insecticidal activity of pogostone against *Spodoptera litura* and *Spodoptera exigua* (Lepidoptera: noctuidae). *Pest Management Science* 70: 510–516.
- Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* 8: 275–282.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A *et al.* 2021. Highly accurate protein structure prediction with ALPHAFOLD. *Nature* 596: 583–589.
- Junren C, Xiaofang X, Mengting LI, Qiuyun X, Gangmin LI, Huiqiong Z, Guanru C, Xin XU, Yanpeng Y, Fu P *et al.* 2021. Pharmacological activities and mechanisms of action of *Pogostemon cablin* Benth: a review. *Chinese Medicine* 16: 5.
- Koon N, Squire CJ, Baker EN. 2004. Crystal structure of LeuA from *Mycobacterium tuberculosis*, a key enzyme in leucine biosynthesis. *Proceedings of the National Academy of Sciences, USA* 101: 8295–8300.
- de Kraker JW, Gershenzon J. 2011. From amino acid to glucosinolate biosynthesis: protein sequence changes in the evolution of methylthioalkylmalate synthase in Arabidopsis. *Plant Cell* 23: 38–53.
- de Kraker JW, Luck K, Textor S, Tokuhisa JG, Gershenzon J. 2007. Two Arabidopsis genes (*IPMS1* and *IPMS2*) encode isopropylmalate synthase, the branchpoint step in the biosynthesis of leucine. *Plant Physiology* 143: 970–986.
- Kroumova AB, Wagner GJ. 2003. Different elongation pathways in the biosynthesis of acyl groups of trichome exudate sugar esters from various solanaceous plants. *Planta* 216: 1013–1021.
- Kumar R, Lee SG, Augustine R, Reichelt M, Vassao DG, Palavalli MH, Allen A, Gershenzon J, Jez JM, Bisht NC. 2019. Molecular basis of the evolution of methylthioalkylmalate synthase and the diversity of methionine-derived glucosinolates. *Plant Cell* 31: 1633–1647.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* 35: 1547–1549.
- Li CW, Wu XL, Zhao XN, Su ZQ, Chen HM, Wang XF, Zhang XJ, Zeng HF, Chen JN, Li YC *et al.* 2013. Anti-inflammatory property of the ethanol extract of the root and rhizome of *Pogostemon cablin* (Blanco) benth. *Scientific World Journal* 2013: 434151.
- Li W, Zhou F, Pichersky E. 2018. Jasmone hydroxylase, a key enzyme in the synthesis of the alcohol moiety of pyrethrin insecticides. *Plant Physiology* 177: 1498–1509.
- Li Y-C, Liang H-C, Chen H-M, Tan L-R, Yi Y-Y, Qin Z, Zhang W-M, Wu D-W, Li C-W, Lin R-F *et al.* 2012. Anti-*Candida albicans* activity and pharmacokinetics of pogostone isolated from *Pogostemonis herba*. *Phytomedicine* 20: 77–83.
- Li Y-C, Xian Y-F, Su Z-R, Ip S-P, Xie J-H, Liao J-B, Wu D-W, Li C-W, Chen J-N, Lin Z-X *et al.* 2014. Pogostone suppresses proinflammatory mediator production and protects against endotoxic shock in mice. *Journal of Ethnopharmacology* 157: 212–221.
- Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, DeBono A, Durrett TP *et al.* 2013. Acyl-lipid metabolism. *The Arabidopsis Book* 11: e0161.
- Lousa CD, van Roermund CWT, Postis VLG, Dietrich D, Kerr ID, Wanders RJA, Baldwin SA, Baker A, Theodoulou FL. 2013. Intrinsic acyl-CoA thioesterase activity of a peroxisomal ATP binding cassette transporter is required for transport and metabolism of fatty acids. *Proceedings of the National Academy of Sciences, USA* 110: 1279–1284.
- Luchesi LA, Paulus D, Busso C, Frata MT, Oliveira JB. 2020. Chemical composition, antifungal and antioxidant activity of essential oils from *Baccharis dracunculifolia* and *Pogostemon cablin* against *Fusarium graminearum*. *Natural Product Research* 34: 1–4.

- Melendez-Hevia E, Waddell TG, Cascante M. 1996. The puzzle of the Krebs citric acid cycle: assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution. *Journal of Molecular Evolution* 43: 293–303.
- Ning J, Moghe GD, Leong B, Kim J, Ofner I, Wang Z, Adams C, Jones AD, Zamir D, Last RL. 2015. A feedback-insensitive isopropylmalate synthase affects acylsugar composition in cultivated and wild tomato. *Plant Physiology* 169: 1821–1835.
- Ohlrogge J, Browse J. 1995. Lipid biosynthesis. *Plant Cell* 7: 957–970.
- Peng F, Wan F, Xiong L, Peng C, Dai M, Chen JP. 2014. *In vitro* and *in vivo* antibacterial activity of Pogostone. *Chinese Medical Journal* 127: 4001–4005.
- de Quiros HC, Magrath R, McCallum D, Kroymann J, Scnabelrauch D, Mitchell-Olds T, Mithen R. 2000. alpha-keto acid elongation and glucosinolate biosynthesis in *Arabidopsis thaliana*. *Theoretical and Applied Genetics* 101: 429–437.
- Slocombe SP, Schauvinhold I, McQuinn RP, Besser K, Welsby NA, Harper A, Aziz N, Li Y, Larson TR, Giovannoni J *et al.* 2008. Transcriptomic and reverse genetic analyses of branched-chain fatty acid and acyl sugar production in *Solanum pennellii* and *Nicotiana benthamiana*. *Plant Physiology* 148: 1830–1846.
- Stout JM, Boubakir Z, Ambrose SJ, Purves RW, Page JE. 2012. The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in *Cannabis sativa* trichomes. *The Plant Journal* 71: 353–365.
- Sugimoto N, Engelgau P, Jones AD, Song J, Beaudry R. 2021. Citramalate synthase yields a biosynthetic pathway for isoleucine and straight- and branched-chain ester formation in ripening apple fruit. *Proceedings of the National Academy of Sciences, USA* 118: e2009988118.
- Wang X-F, Huang Y-F, Wang L, Xu L-Q, Yu X-T, Liu Y-H, Li C-L, Zhan J-X, Su Z-R, Chen J-N *et al.* 2016. Photo-protective activity of pogostone against UV-induced skin premature aging in mice. *Experimental Gerontology* 77: 76–86.
- Xing A, Last RL. 2017. A regulatory hierarchy of the Arabidopsis branched-chain amino acid metabolic network. *Plant Cell* 29: 1480–1499.
- Xu H, Li W, Schillmiller AL, van Eekelen H, de Vos RCH, Jongsma MA, Pichersky E. 2019. Pyrethric acid of natural pyrethrin insecticide: complete pathway elucidation and reconstitution in *Nicotiana benthamiana*. *New Phytologist* 223: 751–765.
- Xu H, Moghe GD, Wiegert-Rininger K, Schillmiller AL, Barry CS, Last RL, Pichersky E. 2018. Coexpression analysis identifies two oxidoreductases involved in the biosynthesis of the monoterpene acid moiety of natural pyrethrin insecticides in *Tanacetum cinerariifolium*. *Plant Physiology* 176: 524–537.
- Xu H, Zhang F, Liu B, Huhman DV, Sumner LW, Dixon RA, Wang G. 2013. Characterization of the formation of branched short-chain fatty acid:coAs for bitter acid biosynthesis in hop glandular trichomes. *Molecular Plant* 6: 1301–1317.
- Yang T, Stoopen G, Yalpani N, Vervoort J, de Vos R, Voster A, Verstappen FWA, Bouwmeester HJ, Jongsma MA. 2011. Metabolic engineering of geranic acid in maize to achieve fungal resistance is compromised by novel glycosylation patterns. *Metabolic Engineering* 13: 414–425.
- Ye Q, Ling Q, Shen J, Shi L, Chen J, Yang T, Hou Z, Zhao J, Zhou H. 2021. Protective effect of pogostone on murine norovirus infected-RAW264.7 macrophages through inhibition of NF- κ B/NLRP3-dependent pyroptosis. *Journal of Ethnopharmacology* 278: 114250.
- Yoo SD, Cho YH, Sheen J. 2007. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols* 2: 1565–1572.
- Zheng W, Zhou XG, Wuyun QQG, Pearce R, Li Y, Zhang Y. 2020. FUPRED: detecting protein domains through deep-learning-based contact map prediction. *Bioinformatics* 36: 3749–3757.
- Zhou XG, Hu J, Zhang CX, Zhang GJ, Zhang Y. 2019. Assembling multidomain protein structures through analogous global structural alignments. *Proceedings of the National Academy of Sciences, USA* 116: 15930–15938.
- Fig. S1** Liquid chromatography–quadrupole time-of-flight–mass spectrometry analysis of stable isotope incorporation into pogostone from *Pogostemon cablin* leaves fed with 4-methylvaleric-d₁₁ acid.
- Fig. S2** Quantitative real-time PCR analysis of transcript levels of gene candidates involved in 4-methylvaleric acid biosynthesis in different tissues of *Pogostemon cablin* at different stages of development.
- Fig. S3** Dependency of PcIBMS1 and PcIPMS1 enzymatic activity on incubation conditions and components.
- Fig. S4** *In vitro* analyses of PcIBMS1 and PcIPMS1 activities.
- Fig. S5** Liquid chromatography–quadrupole time-of-flight–mass spectrometry analyses of relative abundances of leucine in *Nicotiana benthamiana* leaves expressing *PcIBMS1* and *PcIPMS1*.
- Fig. S6** Amino acid sequence alignment of PcIBMS1, PcIPMS1, AtIPMS1, OsIPMS1, SlIPMS3 and MtIPMS.
- Fig. S7** Sequence alignment of C-terminal domain of PcIBMS1, PcIPMS1, MtIPMS and other typical plant IPMSs that are subject to Leu feedback inhibition.
- Fig. S8** Modelling structures of PcIBMS1 and PcIPMS1.
- Fig. S9** Structure modelling and substrate docking of PcIPMS1-L135M and PcIBMS1-M1132L.
- Fig. S10** Sequence alignment of C-terminal domains of PcIBMS1, PcIPMS1, MtIPMS1, MtIPMS2, AtIPMS1, AtIPMS2 and the three IPMS-like proteins from *Nicotiana benthamiana*.
- Fig. S11** Sequence alignment of N-terminal catalytic regions of PcIBMS1, PcIPMS1, MtIPMS1, MtIPMS2, AtIPMS1, AtIPMS2 and the three IPMS-like proteins from *Nicotiana benthamiana*.
- Methods S1** Details on experimental procedures.
- Table S1** All primers used in this present study.
- Table S2** Bioinformatic analysis of gene candidates involved in 4-methylvaleric acid biosynthesis screened from *Pogostemon cablin* RNA-seq database.
- Table S3** Average normalised counts of gene candidates involved in 4-methylvaleric acid biosynthesis and *PcAAE2* in *Pogostemon cablin* RNA-seq database.
- Table S4** Ranking of the top 10 unique genes in the *Pogostemon cablin* RNA-seq database by coexpression analysis with the *PcAAE2* gene.

Supporting Information

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

Table S5 GenBank accession or locus numbers of functional IPMSs, MAMs and MdCMS1 from NCBI (plants), The Arabidopsis Information Resource (TAIR) and UniProt (bacteria and yeast) sites used for phylogenetic reconstruction and sequence comparison.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.



About *New Phytologist*

- *New Phytologist* is an electronic (online-only) journal owned by the New Phytologist Foundation, a **not-for-profit organization** dedicated to the promotion of plant science, facilitating projects from symposia to free access for our Tansley reviews and Tansley insights.
- Regular papers, Letters, Viewpoints, Research reviews, Rapid reports and both Modelling/Theory and Methods papers are encouraged. We are committed to rapid processing, from online submission through to publication 'as ready' via *Early View* – our average time to decision is <23 days. There are **no page or colour charges** and a PDF version will be provided for each article.
- The journal is available online at Wiley Online Library. Visit **www.newphytologist.com** to search the articles and register for table of contents email alerts.
- If you have any questions, do get in touch with Central Office (np-centraloffice@lancaster.ac.uk) or, if it is more convenient, our USA Office (np-usaoffice@lancaster.ac.uk)
- For submission instructions, subscription and all the latest information visit **www.newphytologist.com**