***New Phytologist* Supporting Information Figs S1-S11, Methods S1 and Tables S1-S5**

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

Authors: Chu Wang, Ying Wang, Jing Chen, Lang Liu, Mingxia Yang, Zhengguo Li, Chengyuan Wang, Eran Pichersky and Haiyang Xu

Article acceptance date: 19 April 2022

The following Supporting Information is available for this article:

Fig. S1 LC-QTOF-MS analysis of stable isotope incorporation into pogostone from *Pogostemon cablin* leaves fed with 4-methylvaleric-d_{11} acid.

Fig. S2 RT-qPCR 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 LC-QTOF-MS analyses of relative abundances of leucine in *N. 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 Modeling structures of PcIBMS1 and PcIPMS1.

Fig. S9 Structure modeling and substrate docking of PcIPMS1-L135M and PcIBMS1-M132L.

Fig. S10 Sequence alignment of C-terminal domains of PcIBMS1, PcIPMS1, MdIPMS1, MdIPMS2, AtIPMS1, AtIPMS2 and the three IPMS-like proteins from *N. benthamiana*.

Fig. S11 Sequence alignment of N-terminal catalytic regions of PcIBMS1, PcIPMS1, MdIPMS1, MdIPMS2, AtIPMS1, AtIPMS2 and the three IPMS-like proteins from *N. benthamiana*.

Table S1 All primers used in this present study.

Table S2 Bioinformatic analysis of gene candidates involved in 4-methylvaleric acid biosynthesis screened from *P*. *cablin* RNAseq database.

Table S3 Average normalized counts of gene candidates involved in 4-methylvaleric acid biosynthesis and *PcAAE2* in *P. cablin* RNA-seq database.

Table S4 Ranking of the top 10 unique genes in the *P. cablin* RNAseq database by coexpression analysis with the *PcAAE2* gene.

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

Methods S1 Details on experimental procedures.

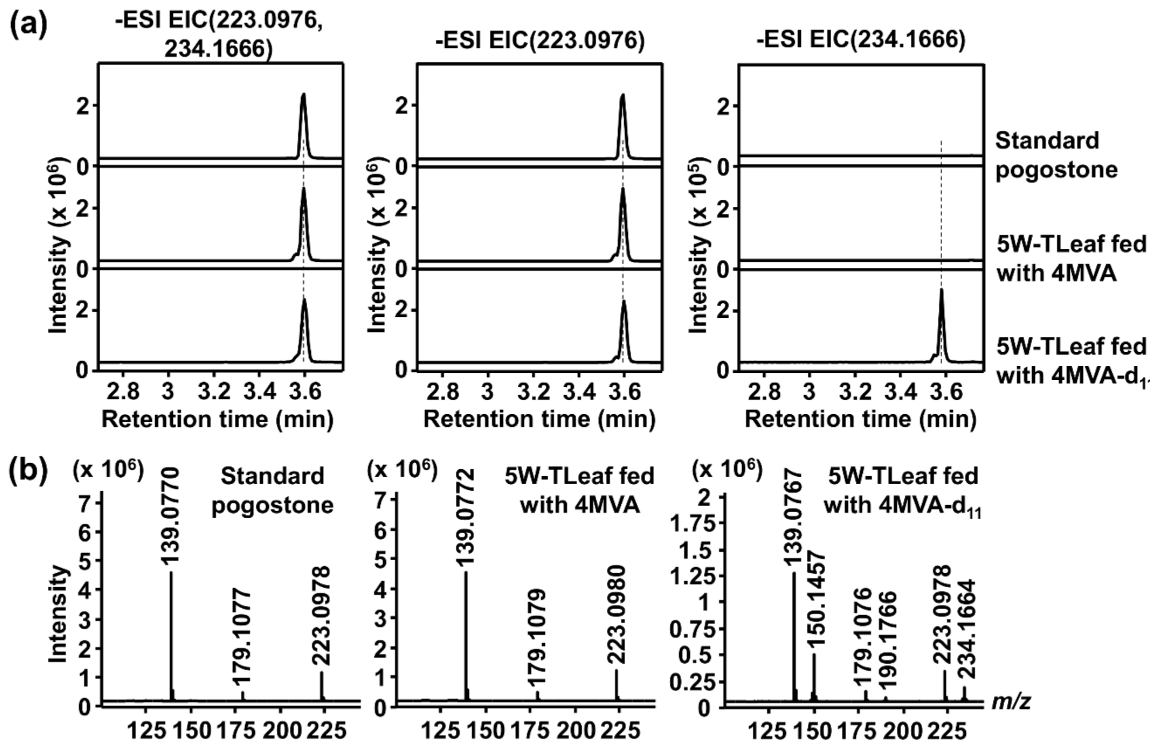
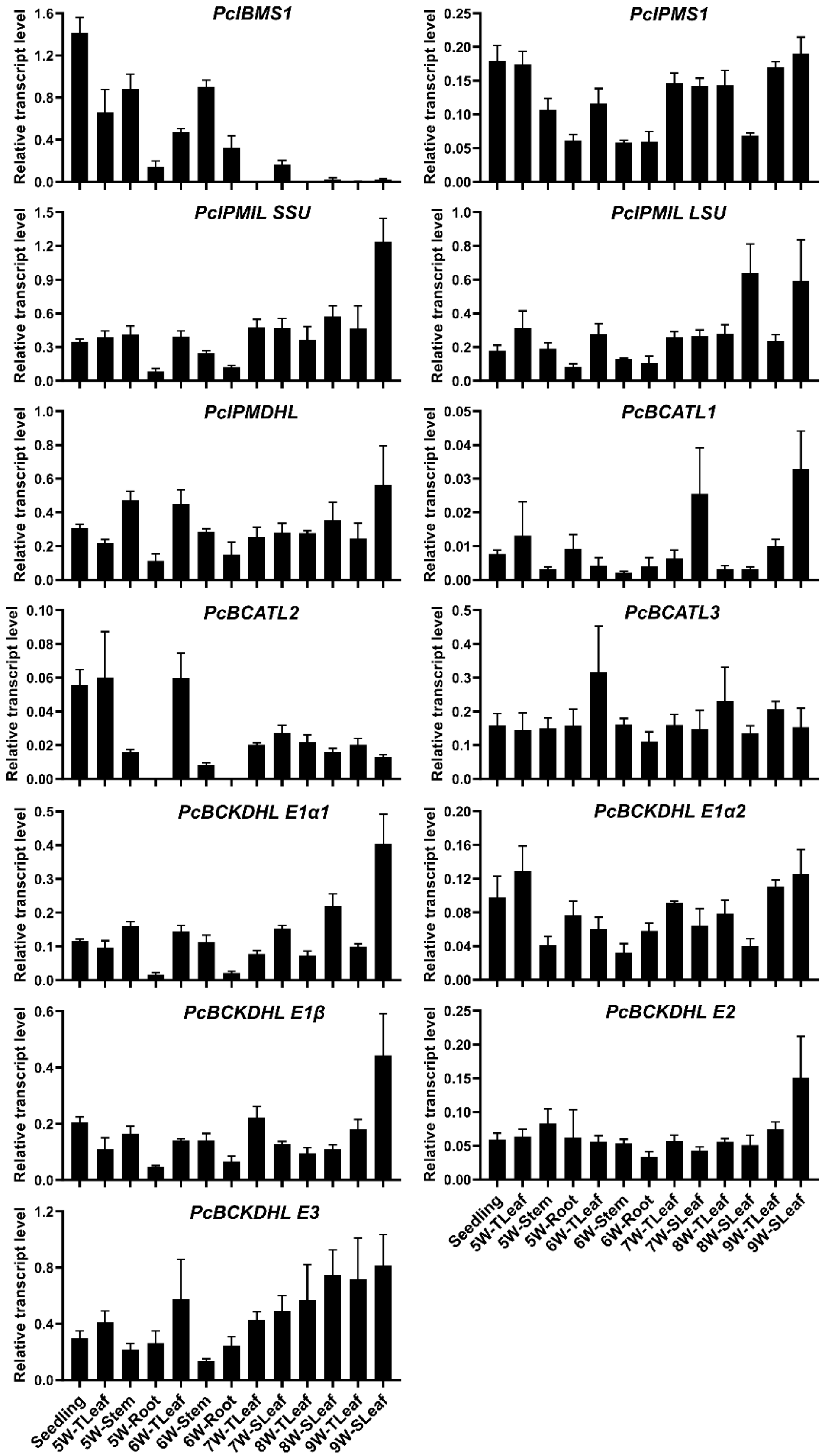


Fig. S1 LC-QTOF-MS analysis of stable isotope incorporation into pogostone from *Pogostemon cablin* leaves fed with 4-methylvaleric-d_{11} acid. (a) LC-QTOF-MS analysis of *P. cablin* five-week-old top leaf (5W-TLeaf) fed with 4MVA or 4MVA-d_{11}. The extracted ion chromatograms in negative mode of *m/z* 223.0976 and 234.1666 (left) for both pogostone and pogostone-d_{11}, 223.0976 (middle) for pogostone and 234.1666 (right) for pogostone-d_{11} are shown. (b) The mass spectrums in negative mode in LC-QTOF-MS analysis of standard pogostone (left), pogostone from *P. cablin* 5W-TLeaf fed with 4MVA (middle) and pogostone-d_{11} from *P. cablin* 5W-TLeaf fed with 4MVA-d_{11} (right). Note the 11-mass-unit shift of the parent ions of pogostone (223.0976) to pogostone-d_{11}(234.1666) upon feeding with 4MVA-d_{11}. The retention times of pogostone and pogostone-d_{11} are slightly different, most likely due to the slight differences in physical and chemical properties between these two compounds.



PcIBMS1
PcIPMS1
PcIPMIL SSU
PcIPMIL LSU
PcIPMDHL
PcBCATL1
PcBCATL2
PcBCATL3
PcBCKDHL E1α1
PcBCKDHL E1α2
PcBCKDHL E1β
PcBCKDHL E2
PcBCKDHL E3
Relative transcript level
Seedling
5W-TLeaf
5W-Stem
5W-Root
6W-TLeaf
6W-Stem
6W-Root
7W-TLeaf
7W-SLeaf
8W-TLeaf
8W-SLeaf
9W-TLeaf
9W-SLeaf

Fig. S2 RT-qPCR analysis of transcript levels of gene candidates involved in 4-methylvaleric acid biosynthesis in different tissues of *Pogostemon cablin* at different stages of development. The RT-qPCR data were presented as relative expression of these genes relative to *P. cablin GAPDH* (glyceraldehyde-3-phosphate dehydrogenase), *Actin 7* and *Tubulin 3* in each tissue sample. Data are means ± SD of four independent biological replicates.

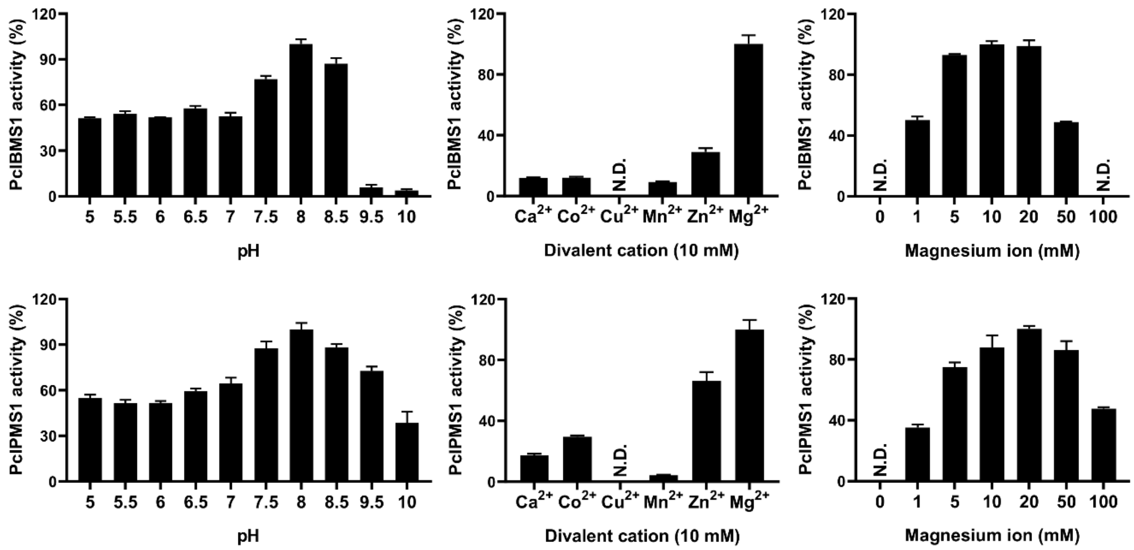


Fig. S3 Dependency of PcIBMS1 and PcIPMS1 enzymatic activity on incubation conditions and components. These enzymatic assays were performed using end-point enzyme assay (DTNB). These activities were measured with 0.5 mM 4-methyl-2-oxovalerate for PcIBMS1 and 0.5 mM 2-oxoisovalerate for PcIPMS1 in the presence of 0.5 mM acetyl-CoA and different concentration of divalent ions. To ensure optimum buffering capacity, 100 mM potassium phosphate buffer and 100 mM Tris-HCl buffer were used from pH 5-7 and pH 7.5-10, respectively. Divalent cations (Ca^{2+}, Co^{2+}, Cu^{2+}, Mn^{2+}, Zn^{2+}, Mg^{2+}) were used as chloride salts at a final concentration of 10 mM. To ensure optimum concentration of magnesium ion, the concentrations of magnesium ion ranging from 1mM to 100 mM were tested. The highest activity was set as 100%. The data are presented as relative mean percentages (means ± SD) of three independent biological replicates. N.D. represents “not detected or relative activities < 5%”.

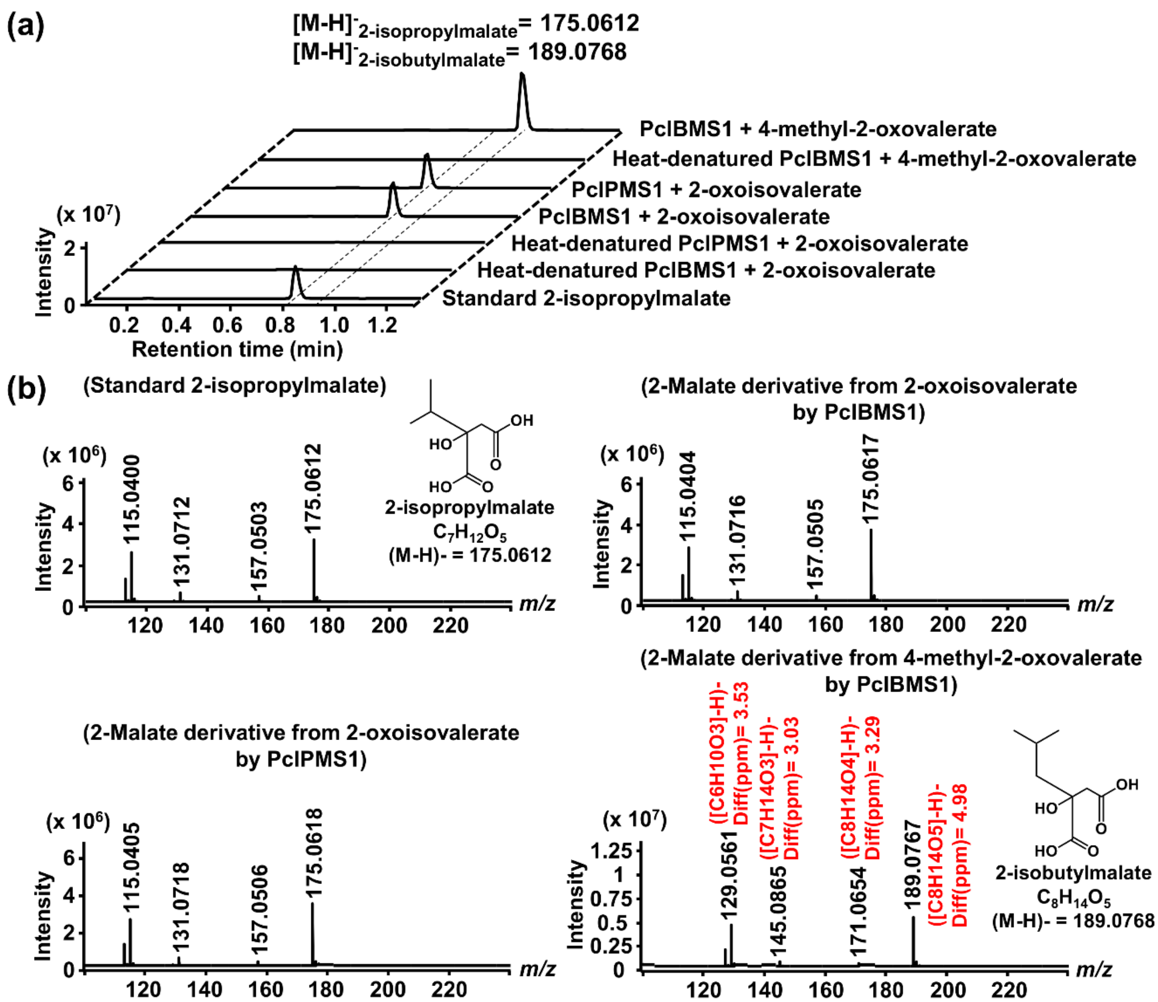


Fig. S4 *In vitro* analyses of PcIBMS1 and PcIPMS1 activities. (a) LC-QTOF-MS analysis of 2-malate derivatives formed in reactions containing purified PcIBMS1 and PcIPMS1, acetyl-CoA, $MgCl_2$, 2-oxoisovalerate or 4-methyl-2-oxovalerate after 30 min of incubation. The extracted ion chromatograms in negative mode of *m/z* 175.0612 and 189.0768 are shown for 2-isopropylmalate and 2-isobutylmalate respectively. (b) Mass spectrums in negative mode of standard 2-isopropylmalate, and 2-malate derivatives (predicted 2-isopropylmalate and 2-isobutylmalate) generated *in vitro* by the action of PcIBMS1 and PcIPMS1 with the substrates 2-oxoisovalerate or 4-methyl-2-oxovalerate. The 2-malate derivatives from 2-oxoisovalerate by the action of PcIBMS1 and PcIPMS1 were identified as 2-isopropylmalate through the comparison of its retention time and mass spectrums with that of standard 2-isopropylmalate. The 2-malate derivative from 4-methyl-2-oxovalerate by the action of PcIBMS1 was identified as 2-isobutylmalate based on the exact mass and fragment pattern. The chemical structures of 2-isopropylmalate and 2-isobutylmalate were placed along with the mass spectrums of standard 2-isopropylmalate and predicted 2-isobutylmalate respectively. The predicted molecular formula corresponding to the mass fragment of 2-malate derivative generated *in vitro* by the action of PcIBMS1 with the substrate 4-methyl-2-oxovalerate was marked in red color.

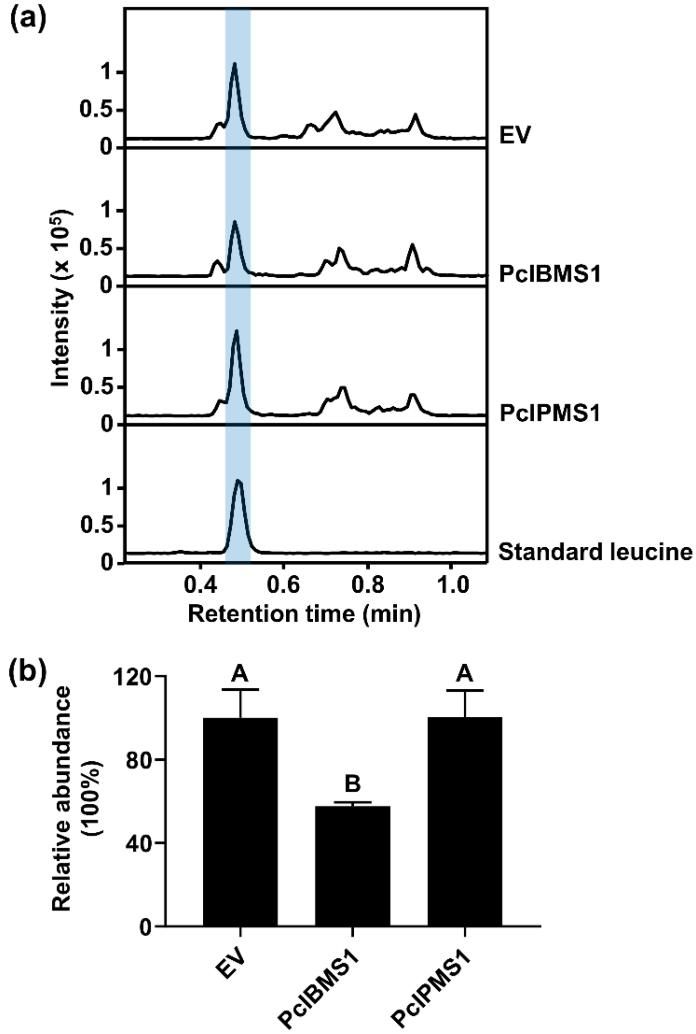


Fig. S5 LC-QTOF-MS analyses of relative abundances of leucine in *N. benthamiana* leaves expressing *PcIBMS1* and *PcIPMS1*. (a) LC-QTOF-MS analyses of *N. benthamiana* leaves expressing the two enzymes PcIBMS1 and PcIPMS1. *N. benthamiana* leaves transformed with empty vector were used as negative control. Extracted ion chromatograms in negative mode of *m/z* 130.0874 for leucine are shown. Leucine was identified based on the comparison of retention time and mass spectrums with the authentic standard. (b) The relative abundance of leucine in *N. benthamiana* leaves expressing *PcIBMS1* or *PcIPMS1* and *N. benthamiana* leaves transformed with empty vector. The relative abundance of leucine in transgenic *N. benthamiana* leaves was calculated by normalization of peak area of leucine to that of internal standard naringenin (*m/z*, 271.0612). The relative abundance of leucine in *N. benthamiana* leaves transformed with empty vector was set as 100%. The data are presented as relative mean percentages (means ± SD) of three independent biological replicates. Significant difference among the relative abundance of leucine in transgenic *N. benthamiana* leaves was tested using independent *t*-tests $P < 0.01$ (Two-tailed distribution; Two-sample equal variance (homoscedastic)). Different letters above the bars mark the statistically significant groups after *t*-tests. EV, empty vector.

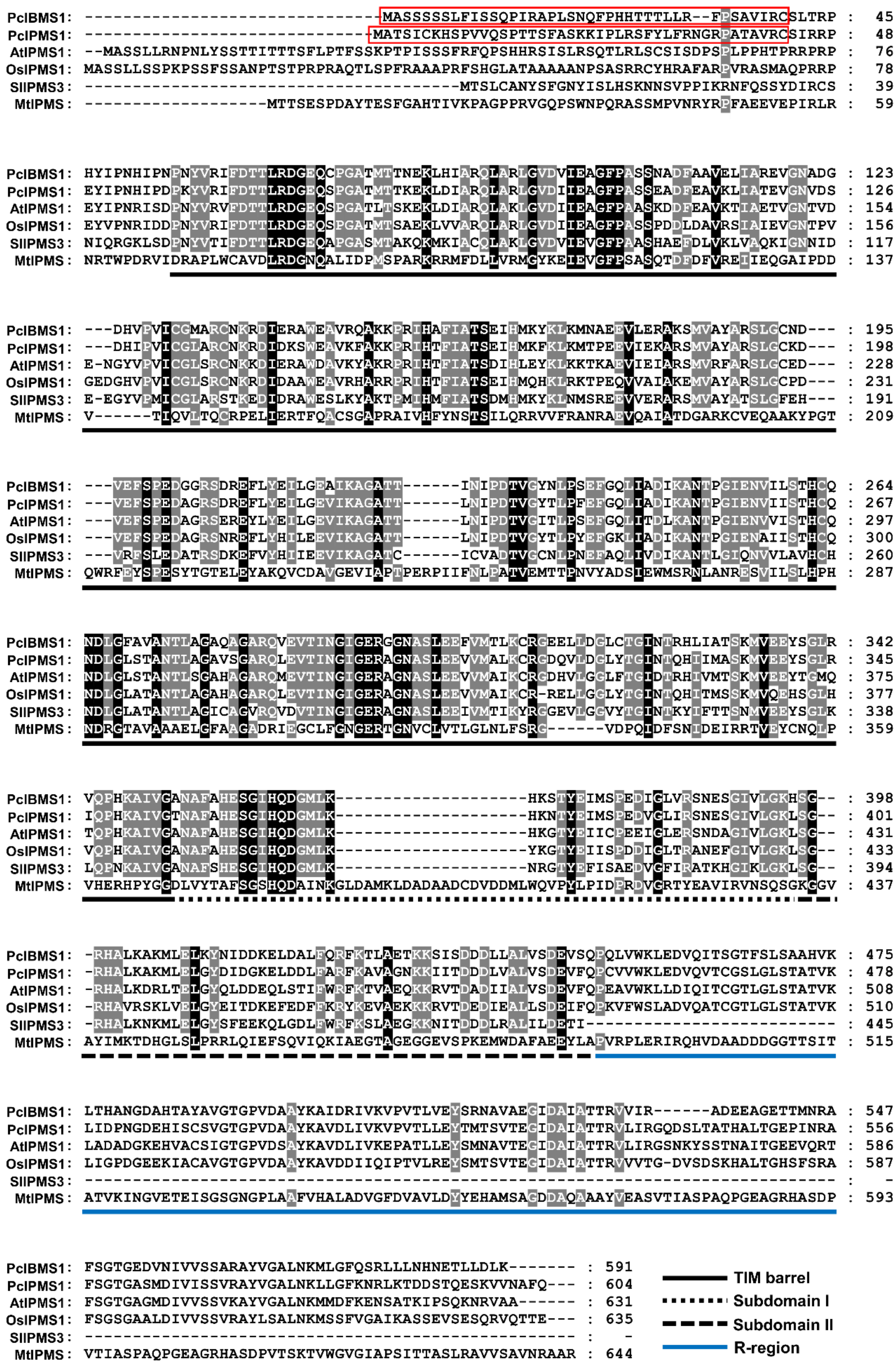

PcIBMS1: ------------------------------MASSSSSLFISSQPIRAPLSNQFPHHTTTLLR--FPSAVIRCSLTRP : 45
PcIPMS1: -----------------------------MATSICKHSPVVQSPTTSFASKKIPLRSFYLFRNGRPATAVRCSIRRP : 48
AtIPMS1: --MASSLLRNPNLYSSTTITTTSFLPTFSSKPTPISSSFRFQPSHHRSISLRSQTLRLSCSISDPSPLPPHTPRRPRP : 76
OsIPMS1: MASSLLSSPKPSSFSSANPTSTPRPRAQTLSPFRAAAPRFSHGLATAAAAANPSASRRCYHRAFARPVRASMAQPRRP : 78
SlIPMS3: --------------------------------------MTSLCANYSFGNYISLHSKNNSVPPIKRNFQSSYDIRCS : 39
MtIPMS: ------------------MTTSESPDAYTESFGAHTIVKPAGPPRVGQPSWNPQRASSMPVNRYRPFAEEVEPIRLR : 59
PcIBMS1: HYIPNHIPNPNYVRIFDTTLRDGEQCPGATMTTNEKLHIARQLARLGVDVIEAGFPASSNADFAAVELIAREVGNADG : 123
PcIPMS1: EYIPNHIPDPKYVRIFDTTLRDGEQSPGATMTTKEKLDIARQLARLGVDIIEAGFPASSEADFEAVKLIATEVGNVDS : 126
AtIPMS1: EYIPNRISDPNYVRVFDTTLRDGEQSPGATLTSKEKLDIARQLAKLGVDIIEAGFPAASKDDFEAVKTIAETVGNTVD : 154
OsIPMS1: EYVPNRIDDPNYVRIFDTTLRDGEQSPGATMTSAEKLVVARQLARLGVDIIEAGFPASSPDDLDAVRSIAIEVGNTPV : 156
SlIPMS3: NIQRGKLSDPNYVTIFDTTLRDGEQAPGASMTAKQKMKIACQLAKLGVDVIEVGFPAASHAEFDLVKLVAQKIGNNID : 117
MtIPMS: NRTWPDRVIDRAPLWCAVDLRDGNQALIDPYSPARKRRMFDLLVRMGYKEIEVGFPSASQTDFDFVREIIEQGAIPDD : 137
PcIBMS1: ---DHVPVICGMARCNKRDIERAWEAVRQAKKPRIHAFIATSEIHMKYKLKMNAEEVLERAKSMVAYARSLGCND--- : 195
PcIPMS1: ---DHIPVICGLARCNKRDIDKSWEAVKFAKKPRIHTFIATSEIHMKFKLKMTPEEVIEKARSMVAYARSLGCKD--- : 198
AtIPMS1: E-NGYVPVICGLSRCNKKDIERAWDAVKYAKRPRIHTFIATSDIHLEYKLKKTKAEVIEIARSMVRFARSLGCED--- : 228
OsIPMS1: GEDGHVPVICGLSRCNKRDIDAAWEAVRHARRPRIHTFIATSEIHMQHKLRKTPEQVVAIAKEMVAYARSLGCPD--- : 231
SlIPMS3: E-EGYVPMICGLARSTKEDIDRAWESLKYAKTPMIHMFIATSDMHMKYKLNMSREEVVERARSMVAYATSLGFEH--- : 191
MtIPMS: V------TIQVLTQCRPELIERTFQACSGAPRAIVHFYNSTSILQRRVVFRANRAEVQAIATDGARKCVEQAAKYPGT : 209
PcIBMS1: ---VEFSPEDGGRSDREFLYEILGEAIKAGATT------INIPDTVGYNLPSEFGQLIADIKANTPGIENVILSTHCQ : 264
PcIPMS1: ---VEFSPEDAGRSDREFLYEILGEVIKAGATT------LNIPDTVGYTLPSEFGQLIADIKANTPGIENVIISTHCQ : 267
AtIPMS1: ---VEFSPEDAGRSEREYLYEILGEVIKAGATT------LNIPDTVGITLPSEFGQLITDLKANTPGIENVVISTHCQ : 297
OsIPMS1: ---VEFSPEDAGRSNREFLYHILEEVIKAGATT------LNIPDTVGYTLPYEFGKLIADIKANTPGIENAIISTHCQ : 300
SlIPMS3: ---VRFSLEDATRSDKEFVYHIIEEVIKAGATC------ICVADTVGCNLPNEFAQLIVDIKANTLGIQNVVLAVHCH : 260
MtIPMS: QWRFEYSPESYTGTELEYAKQVCDAVGEVIAPTPERPIIFNLPATVEMTTPNVYADSIEWMSRNLANRESVILSLHPH : 287
PcIBMS1: NDLGFAVANTLAGAQAGARQVEVTINGIGERGGNASLEEFVMTLKCRGEELLDGLCTGINTRHLIATSKMVEEYSGLR : 342
PcIPMS1: NDLGLSTANTLAGAVSGARQLEVTINGIGERAGNASLEEVVMALKCRGDQVLDGLYTGINTQHIIMASKMVEEYSGLR : 345
AtIPMS1: NDLGLSTANTLSGAHAGARQMEVTINGIGERAGNASLEEVVMAIKCRGDHVLGGLFTGIDTRHIVMTSKMVEEYTGMQ : 375
OsIPMS1: NDLGLATANTLAGAHAGARQLEVTINGIGERAGNASLEEVVMAIKCR-RELLGGLYTGINTQHITMTSKMVQEHSGLH : 377
SlIPMS3: NDLGLATANTLAGICAGVRQVDVTINGIGERAGNASLEEIVMTIKYRGGEVLGGVYTGINTKYIFTTSNMVEEYSGLK : 338
MtIPMS: NDRGTAVAAAELGFAAGADRIEGCLFGNGERTGNVCLVTLGLNLFSRG------VDPQIDFSNIDEIRRTVEYCNQLP : 359
PcIBMS1: VQPHKAIVGANAFAHESGIHQDGMLK--------------------HKSTYEIMSPEDIGLVRSNESGIVLGKHSG-- : 398
PcIPMS1: IQPHKAIVGTNAFAHESGIHQDGMLK--------------------HKNTYEIMSPEDVGLIRSNESGIVLGKLSG-- : 401
AtIPMS1: TQPHKAIVGANAFAHESGIHQDGMLK--------------------HKGTYEIICPEEIGLERSNDAGIVLGKLSG-- : 431
OsIPMS1: VQPHKAIVGANAFAHESGIHQDGMLK--------------------YKGTYEIISPDDIGLTRANEFGIVLGKLSG-- : 433
SlIPMS3: LQPNKAIVGANAFSHESGIHQDGMLK--------------------NRGTYEFISAEDVGFIRATKHGIKLGKLSG-- : 394
MtIPMS: VHERHPYGGDLVYTAFSGSHQDAINKGLDAMKLDADAADCDVDDMLWQVPYLPIDPRDVGRTYEAVIRVNSQSGKGGV : 437
PcIBMS1: -RHALKAKMLELKYNIDDKELDALFQRFKTLAETKKSISDDDLLALVSDEVSQEQLVWKLEDVQITSGTFSLSAAHVK : 475
PcIPMS1: -RHALKAKMLELGYDIDGKELDDLFARFKAVAGNKKIITDDDLVALVSDEVFQECVVWKLEDVQVTCGSLGLSTATVK : 478
AtIPMS1: -RHALKDRLTELGYQLDDEQLSTIFWRFKTVAEQKKRVTDADIIALVSDEVFQPEAVWKLLDIQITCGTLGLSTATVK : 508
OsIPMS1: -RHAVRSKLVELGYEITDKEFEDFFKRYKEVAEKKKRVTDEDIEALLSDEIFQEKVFWSLADVQATCGTLGLSTATVK : 510
SlIPMS3: -RHALKNKMLELGYSFEEKQLGDLFWRFKSLAEGKKNITDDDLRALILDETI------------------------- : 445
MtIPMS: AYIMKTDHGLSLPRRLQIEFSQVIQKIAEGTAGEGGEVSPKEMWDAFAEEYLAEVRPLERIRQHVDAADDDGGTTSIT : 515
PcIBMS1: LTHANGDAHTAYAVGTGPVDAAYKAIDRIVKVPVTLVEYSRNAVAEGIDAIATTRVVIR------ADEEAGETTMNRA : 547
PcIPMS1: LIDPNGDEHISCSVGTGPVDAAYKAVDLIVKVPVTLLEYTMTSVTEGIDAIATTRVLIRGQDSLTATHALTGEPINRA : 556
AtIPMS1: LADADGKEHVACSIGTGPVDSAYKAVDLIVKEPATLLEYSMNAVTEGIDAIATTRVLIRGSNKYSSTNAITGEEVQRT : 586
OsIPMS1: LIGPDGEEKIACAVGTGPVDAAYKAVDDIIQIPTVLREYSMTSVTEGIDAIATTRVVVTG-DVSDSKHALTGHSFSRA : 587
SlIPMS3: -- : -
MtIPMS: ATVKINGVETEISGSGNGPLAAFVHALADVGFDVAVLDYYEHAMSAGDDAQAAAYVEASVTIASPAQPGEAGRHASDP : 593
PcIBMS1: FSGTGEDVNIVVSSARAYVGALNKMLGFQSRLLLNHNETLLDLK------- : 591
PcIPMS1: FSGTGASMDIVISSVRAYVGALNKLLGFKNRLKTDDSTQESKVVNAFQ--- : 604
AtIPMS1: FSGTGAGMDIVVSSVKAYVGALNKMMDFKENSATKIPSQKNRVAA------ : 631
OsIPMS1: FSGSGAALDIVVSSVRAYLSALNKMSSFVGAIKASSEVSESQRVQTTE--- : 635
SlIPMS3: --- : -
MtIPMS: VTIASPAQPGEAGRHASDPVTSKTVWGVGIAPSITTASLRAVVSAVNRAAR : 644
TIM barrel
Subdomain I
Subdomain II
R-region

Fig. S6 Amino acid sequence alignment of PcIBMS1, PcIPMS1, AtIPMS1, OsIPMS1, SlIPMS3 and MtIPMS. Black background shows perfectly conserved sequences across the IPMS proteins, while the light grey background shows the amino acid identity higher than 80% across the IPMS proteins. Red frames indicate the predicted chloroplast transit peptides of PcIBMS1 and PcIPMS1. Protein domain designations are derived from MtIPMS crystal structure (Koon *et al.*, 2004). PcIBMS1 and PcIPMS1 were directly obtained from *P. cablin* transcriptome database. The information of other proteins used in this sequence alignment is shown in Table S5.

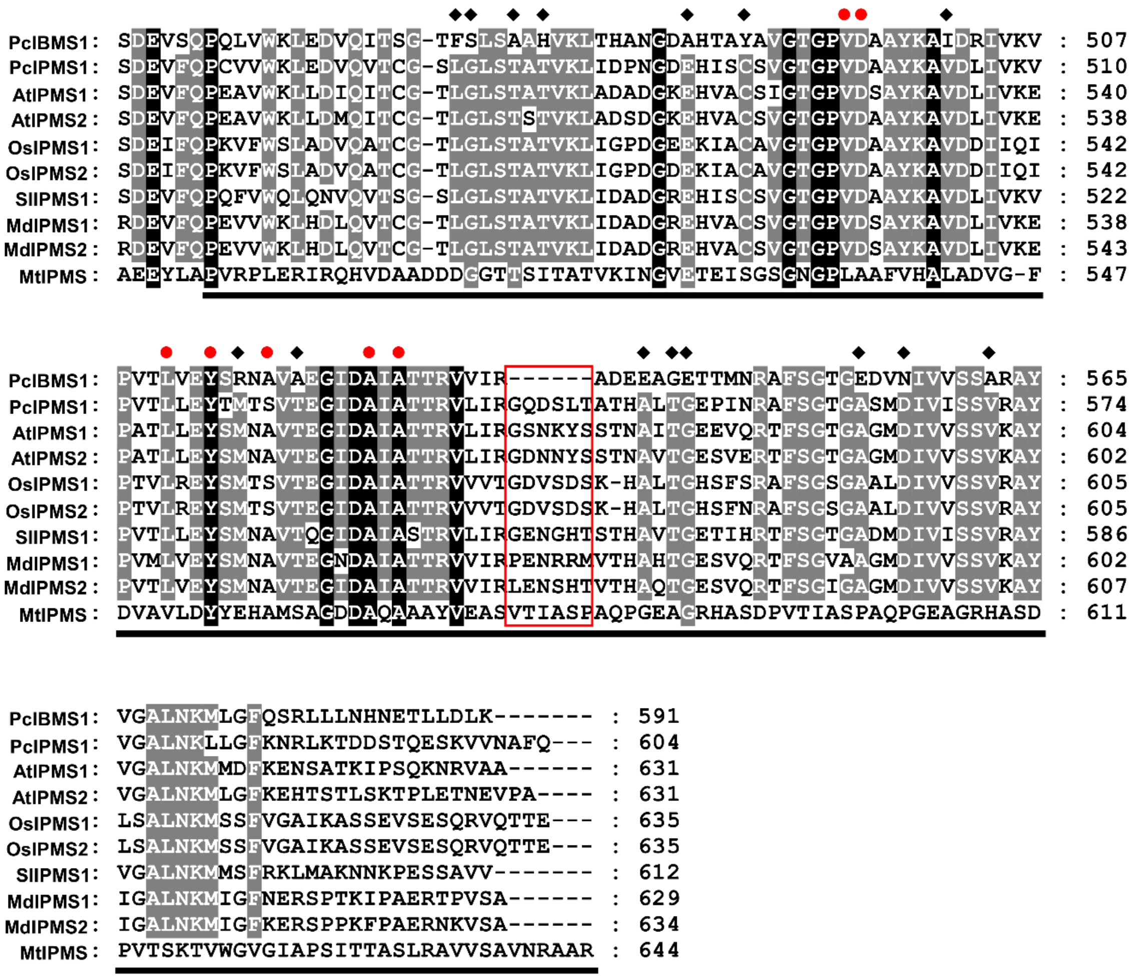


Fig. S7 Sequence alignment of C-terminal domain of PcIBMS1, PcIPMS1, MtIPMS and other typical plant IPMSs that are subject to Leu feedback inhibition. The IPMSs used in this sequence alignment except PcIBMS1 have been biochemically verified to be subject to Leu feedback inhibition (Koon *et al.*, 2004; de Carvalho *et al.*, 2005; de Kraker *et al.*, 2007; Ning *et al.*, 2015; He *et al.*, 2019; Sugimoto *et al.*, 2021). The solid diamond indicates amino acid divergence in the site of PcIBMS1 protein corresponding to the conserved amino acid site in the plant typical IPMSs occurs. Solid red circle marks the Leu binding sites that are derived from *Mycobacterium tuberculosis* IPMS (MtIPMS) crystal structure (Koon *et al.*, 2004). Red frame indicates the sites where PcIBMS1 loss five amino acids. Blank lines under the sequence alignment depict the C-terminal allosteric regulatory domain based on the MtIPMS crystal structure (Koon *et al.*, 2004). Black background shows perfectly conserved sequences across the IPMS proteins, while the light gray background shows the amino acid identity higher than 80% across the IPMS proteins. PcIBMS1 and PcIPMS1 were directly obtained from *P. cablin* transcriptome database. The information of other proteins used in this sequence alignment is shown in Table S5.

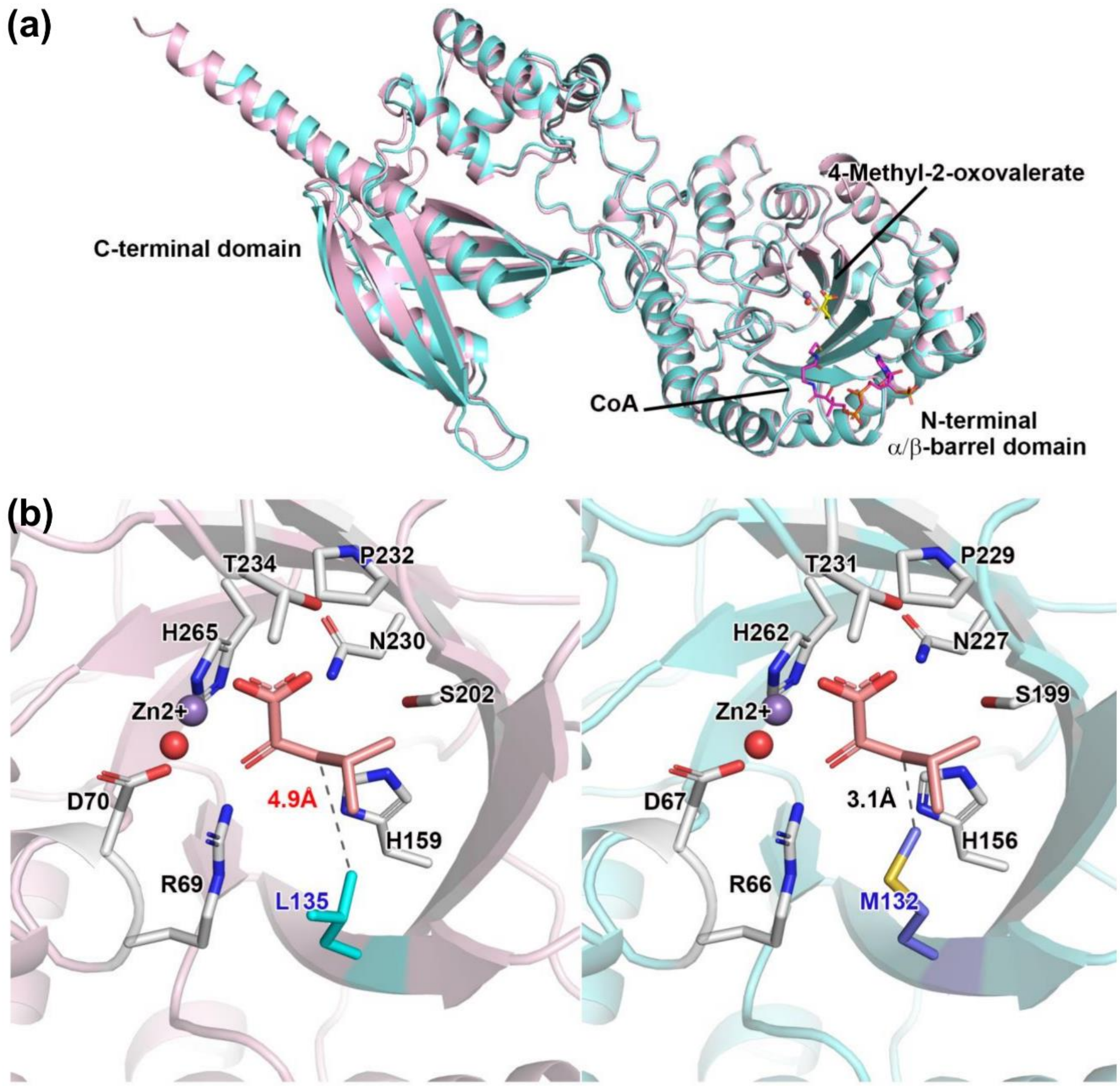


Fig. S8 Modeling structures of PcIBMS1 and PcIPMS1. (a) Superimposed structures of PcIBMS1 and PcIPMS1 bound with CoA (pink stick) and 4-Methyl-2-oxovalerate (yellow stick). PcIBMS1 and PcIPMS1 are shown with cyan and light-pink cartoon, respectively. (b) Zoom-in view shows the residues of PcIPMS1 (left) and PcIBMS1 (right) interacting with 4-Methyl-2-oxovalerate (salmon stick). The non-conserved residues in the binding pocket are highlighted in cyan (L135, PcIPMS1) and blue (M132, PcIBMS1), respectively.

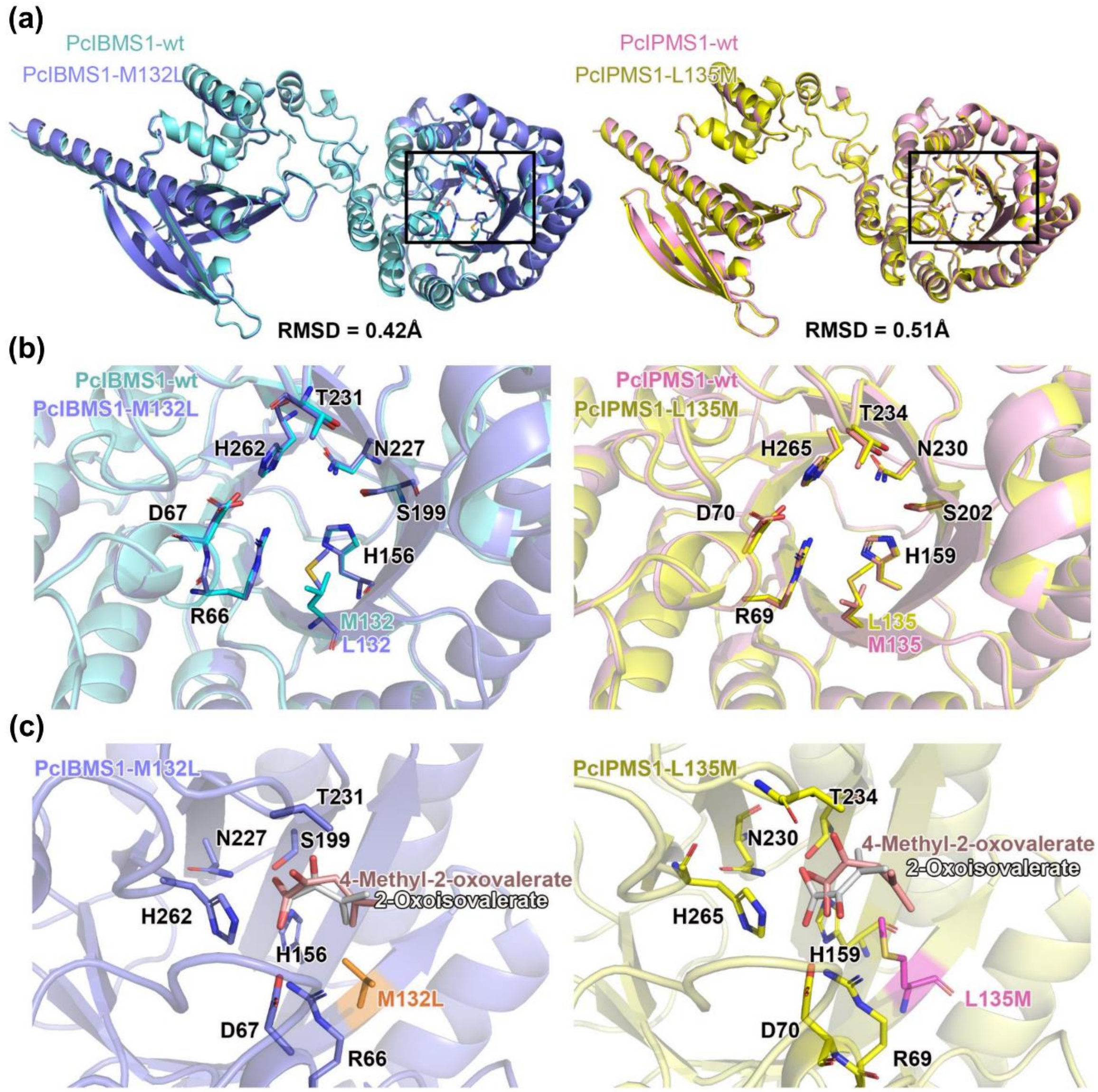


Fig. S9 Structure modeling and substrate docking of PcIPMS1-L135M and PcIBMS1-M1132L. (a) Superimposition of modeling structures including PcIBMS1-wt (wild-type) /PcIBMS1-M132L with RMSD 0.42Å and PcIPMS1-wt/PcIPMS1-L135M with RMSD 0.51Å. Structures are shown with cartoon in cyan (PcIBMS1-wt), light blue (PcIBMS1-M1132L), pink (PcIPMS1-wt) and yellow (PcIPMS1-L135M). (b) A zoom-in view of modeling structures' activity catalytic center shows the detail structure orientation of key residues. The overall structures are shown with cartoon and key residues in catalytic center are shown with stick. (c) Substrate docking results. The docked substrates are shown with stick in white (2-oxoisovalerate) and pink (4-methyl-2-oxovalerate). The mutated residues

M132L (PcIBMS1-M1132L) and L135M (PcIPMS1-L135M) are highlighted and labelled in orange and magenta, respectively.

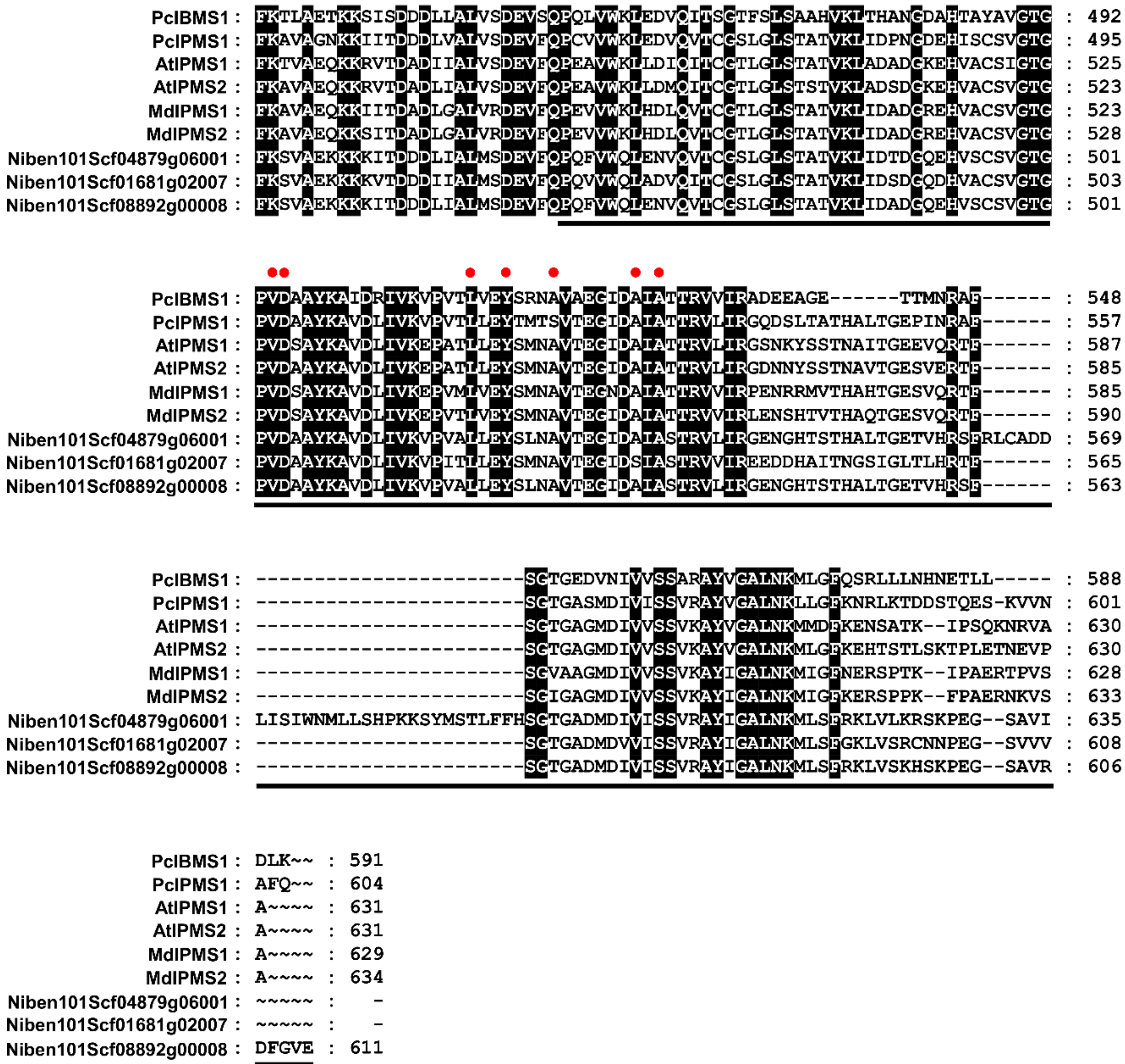


Fig. S10 Sequence alignment of C-terminal domains of PcIBMS1, PcIPMS1, MdIPMS1, MdIPMS2, AtIPMS1, AtIPMS2 and the three IPMS-like proteins from *N. benthamiana*. The IPMSs including MdIPMS1, MdIPMS2, AtIPMS1 and AtIPMS2 have been biochemically verified to be subject to Leu feedback inhibition (de Kraker *et al.*, 2007; Sugimoto *et al.*, 2021). Solid red circle marks the Leu binding sites that are derived from *Mycobacterium tuberculosis* IPMS (MtIPMS) crystal structure (Koon *et al.*, 2004). Blank lines under the sequence alignment depict the C-terminal allosteric regulatory domain based on the MtIPMS crystal structure (Koon *et al.*, 2004). Black background shows perfectly conserved sequences across the IPMS proteins.

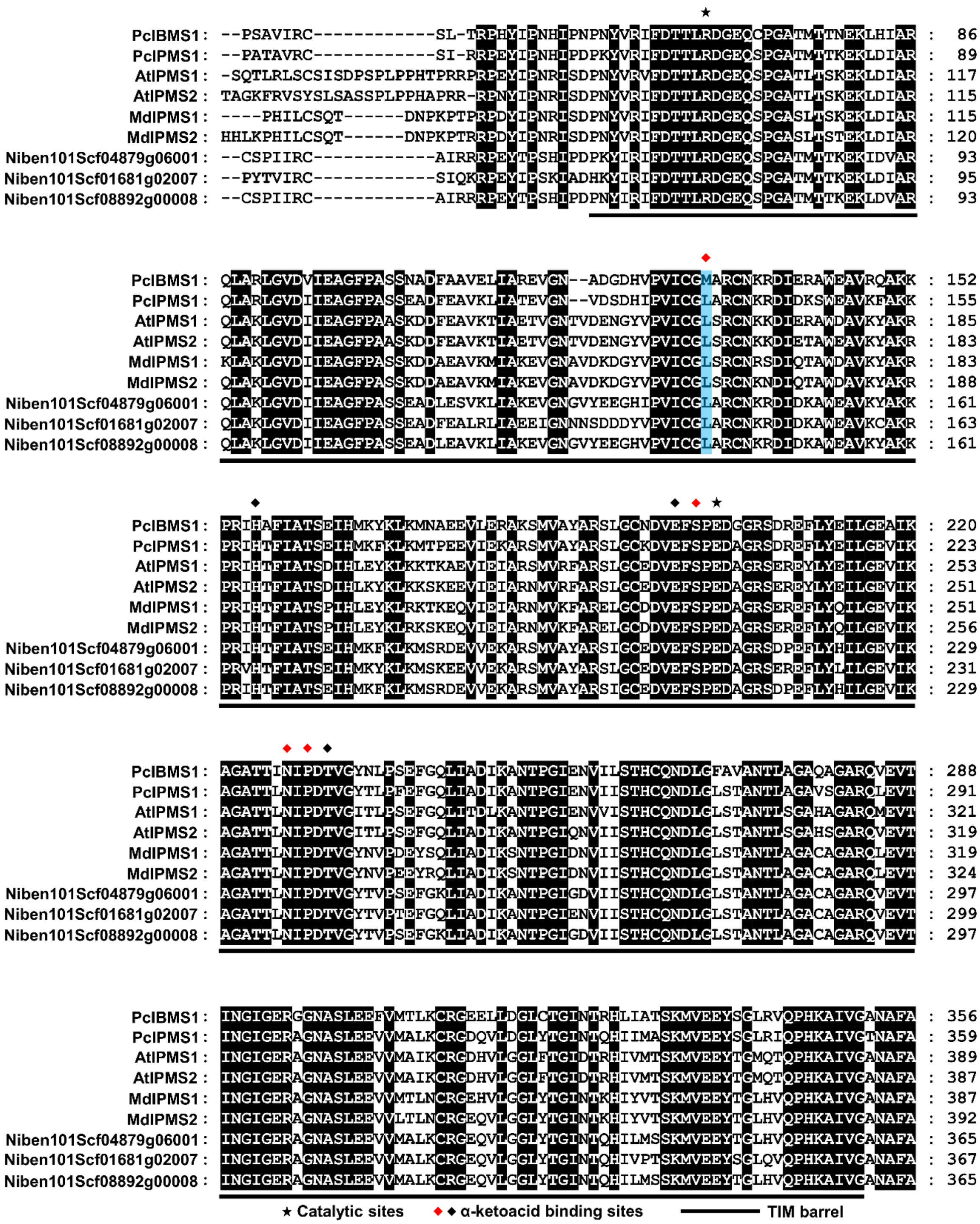


Fig. S11 Sequence alignment of N-terminal catalytic regions of PcIBMS1, PcIPMS1, MdIPMS1, MdIPMS2, AtIPMS1, AtIPMS2 and the three IPMS-like proteins from *N. benthamiana*. Multiple sequence alignment of the N-terminal catalytic regions of PcIBMS1, PcIPMS1, MdIPMS1, MdIPMS2, AtIPMS1, AtIPMS2 and the three IPMS-like proteins including Niben101Scf04879g06001, Niben101Scf01681g02007 and Niben101Scf08892g00008 from *N. benthamiana* is shown. Sequence features corresponding to the characterized BjMAM1-A crystal structure are shown (Kumar *et al.*, 2019). Residues in catalytic sites and α-ketoacid binding sites are marked by asterisk 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 BjMAM1-A crystal structure (Kumar *et al.*, 2019). Black background shows perfectly conserved sequences across the IPMSs, while the light blue background shows amino acid Met132 in PcIBMS1 and its corresponding amino acids in the other eight IPMSs.

Table S1 All primers used in this present study.

Name of primers	**DNA sequences (5'- 3')**
Vector construction for heterologous expression in *E. coli*	
Construction of pET-28a-PcIBMS1 and pET-28a-PcIPMS1	
PcIBMS1-TRTN-F	TGGTGCCGCGCGGCAGCCATATGTCCCTCACTCGCCCTCACTACATC
PcIBMS1-PETN-R	GTCGACGGAGCTCGAATTCGTCATTTCAAATCAAGCAACGTCTCATTG
PcIPMS1-TRTN-F	TGGTGCCGCGCGGCAGCCATATGGTCCGCTGTTCCATCCG
PcIPMS1-PETN-R	GTCGACGGAGCTCGAATTCGTTACTGGAATGCATTTACCACCTTACTTTCTTG
Construction of pET-28a-PcIBMS1-M132L	
PcIBMS1-M132L-F1	TGGTGCCGCGCGGCAGCCATATGTCCCTCACTCGCCCTCAC
PcIBMS1-M132L-R1	GAGGCCGCAGATCACCGGC
PcIBMS1-M132L-F2	TGCCGGTGATCTGCGGCCTC**GCG**AGGTGCAACAAGAGGG
PcIBMS1-M132L-R2	GTCGACGGAGCTCGAATTCGTCATTTCAAATCAAGCAACGTCTC
Construction of pET-28a-PcIPMS1-L135M	
PcIPMS1-L135M-F1	TGGTGCCGCGCGGCAGCCATATGGTCCGCTGTTCCATCCG
PcIPMS1-L135M-R1	TCCACAGATCACCGGGATATG
PcIPMS1-L135M-F2	CATATCCCGGTGATCTGTGGA**ATG**GCGAGGTGTAATAAGAGGGATATTG
PcIPMS1-L135M-R2	GTCGACGGAGCTCGAATTCGTTACTGGAATGCATTTACCACCTTAC
Vector construction for transient heterologous expression in *Nicotiana benthamiana*	
PcIBMS1-AQ-F	ATATTCTGCCCAAATTCGCGAATGGCGTCTTCCTCCTCTTCC
PcIBMS1-AQ-R	ATGAAACCAGAGTTAAAGGCCTCATTTCAAATCAAGCAACGTCTC
PcIPMS1-AQ-F	ATATTCTGCCCAAATTCGCGAATGGCGACATCCATTTGCAAG
PcIPMS1-AQ-R	ATGAAACCAGAGTTAAAGGCCTTACTGGAATGCATTTACCACCTTAC
For RT-qPCR analysis	
PcIBMS1-RT-F	CGAAGAAGCCGAGGATTCAC
PcIBMS1-RT-R	CTCTTAGCCCTCTCCAACACC
PcIPMS1-RT-F	GTCTCCTGAAGATGTTGGACTTC
PcIPMS1-RT-R	AATGGCCTTGAAGCGTGTG
PcIPMILSSU1-RT-F	GGATTGCCCTCCACATACCA
PcIPMILSSU1-RT-R	GAGCCGCACCCAAAGTTATC
PcIPMILLSU1-RT-F	CGCAAACCTTCGACTACTGG

PcIPMILLSU1-RT-R	TGTGGTTTCTCTGACGCTCT
PcIPMDHL1-RT-F	TGTTGGAGGCTTCCATGCTA
PcIPMDHL1-RT-R	CAAACTGCTTTGGGTTTCGC
BCKDHL-E1α-1- RT-F	CCTGGTGGAAGGGTCTCATT
BCKDHL-E1α-1- RT-R	TGAAGGGTCACCATGCTTCT
BCKDHL-E1α-2- RT-F	CACTTGCTGTGTTCAGTGCT
BCKDHL-E1α-2- RT-R	AACTGCTCTCCACCATTCCA
BCKDHL-E1β- RT-F	TGCTGAGATTTCTGCCTCCA
BCKDHL-E1β- RT-R	ATGGCTCGAACACAAGAGGA
BCKDHL-E2- RT-F	CAGTCGCTCTCAATCTCGGA
BCKDHL-E2- RT-R	CTTCCCACCAATTGCACCAA
BCKDHL-E3- RT-F	TGGATGGTGAAGTCCGGAAA
BCKDHL-E3- RT-R	CATCTGCCTCTAGGGTGGTC
PcBCATL1- RT-F	ACATTGCACGAGATGAAGGC
PcBCATL1- RT-R	GCTTCCTACAGGAGCAACAC
PcBCATL2- RT-F	CATCGCCTCGTAGCAATCAC
PcBCATL2- RT-R	TGAACACCTCCTCGTCTTCC
PcBCATL3- RT-F	TGAGCCCATCAGCTGGTATT
PcBCATL3- RT-R	CACCCACCCTCAGTCTCATT
PcGAPDH-RT-F	GATTGGAGAGGTGGACGAGC
PcGAPDH-RT-R	CAACAGTTGGGACACGGAAG
PcActin7-RT-F	TACAGAGGCACCACTCAACC
PcActin7-RT-R	CGACCACTAGCGTACAGTGAG
PcTubulin3-RT-F	GCCTGGACAGAGTGAGGAAG
PcTubulin3-RT-R	TCCATAATCAACCGACAACC
Vector construction for subcellular localization analysis	
PTF-PcIBMS1-F	TTACTATTTACAATTACAGTCGACATGGCGTCTTCCTCCTCTTCCCTC
PTF-PcIBMS1-R	CTCGCCCTTGCTCACCATGGATCCTTTCAAATCAAGCAACGTCTCATTGTG
PTF-PcIPMS1-F	TTACTATTTACAATTACAGTCGACATGGCGACATCCATTTGCAAGCAC
PTF-PcIPMS1-R	CTCGCCCTTGCTCACCATGGATCCCTGGAATGCATTTACCACCTTACTTTCTTGG

The overlaps (15–30 bp) in the primer sequences that were used for DNA fragment assembly were underlined. Substituted sequences for construction of PcIBMS1-M132L and PcIPMS1-L135M point mutants are marked in bold and red color.

Table S2 Bioinformatic analysis of gene candidates involved in 4-methylvaleric acid biosynthesis screened from *P. cablin* RNAseq database.

Genes	Closest Arabidopsis homolog, accession no. and amino acid identity	Functional annotation of the closest Arabidopsis homolog	Peptide length (amino acids)	Signal peptide prediction (using Target P-2.0 and PTS1 Predictor)
PcIBMS1	AT1G74040 (71%)	2-Isopropylmalate synthase 1	591	mTP (0.0013), **cTP (0.9729)**, luTP (0.0254), PTS1 (-48.477)
PcIPMS1	AT1G74040 (78%)	2-Isopropylmalate synthase 1	604	mTP (0.0013), **cTP (0.9926)**, luTP (0.0024), PTS1 (-41.119)
PcIPMIL SSU	AT2G43090 （68%）	Isopropylmalate isomerase small subunit 1	255	mTP (0), **cTP (0.9114)**, luTP (0.0869), PTS1 (-50.713)
PcIPMIL LSU	AT4G13430 （81%）	Isopropyl malate isomerase large subunit 1	510	mTP (0.008), **cTP (0.9572)**, luTP (0.0005), PTS1 (-51.337)
PcIPMDHL	AT1G80560 （80%）	Isopropylmalate dehydrogenase 2	404	mTP (0.0107), **cTP (0.9699)**, luTP (0.0142), PTS1 (-16.432)
PcBCATL1	AT1G10070 （70%）	Branched-chain amino acid transferase 2	392	**mTP (0.9813)**, cTP (0.0015), luTP (0), PTS1 (-54.105)
PcBCATL2	AT1G10070 （62%）	Branched-chain amino acid transferase 2	385	**mTP (0.7027)**, cTP (0.0012), luTP (0.0001), PTS1 (-53.954)
PcBCATL3	AT5G65780 （78%）	Branched-chain amino acid aminotransferase 5	418	mTP (0.0001), **cTP (0.7168)**, luTP (0.0341), PTS1 (-41.206)
PcBCKDHL E1α1	AT1G21400 （62%）	Thiamin diphosphate-binding fold (THDP-binding) superfamily protein	482	**mTP (0.7533)**, cTP (0.0068), luTP (0.0002), PTS1 (-51.494)
PcBCKDHL E1α2	AT5G09300 （67%）	Thiamin diphosphate-binding fold (THDP-binding) superfamily protein	468	mTP (0.272), **cTP (0.5969)**, luTP (0.0059), PTS1 (-41.250)
PcBCKDHL E1β	AT1G55510 （88%）	Branched-chain alpha-keto acid decarboxylase E1 beta subunit	347	**mTP (0.9906)**, cTP (0.0001), luTP (0), PTS1 (-12.473)
PcBCKDHL E2	AT3G06850 （61%）	Dihydrolipoamide branched chain acyltransferase	523	**mTP (0.5057)**, cTP (0.0455), luTP (0.0003), PTS1 (-48.029)
PcBCKDHL E3	AT1G48030 （84%）	Mitochondrial lipoamide dehydrogenase 1	506	**mTP (0.9999)**, cTP (0.0001), luTP (0), PTS1 (-25.218)

The most likely organelle that gene is targeted to and the corresponding prediction value were highlighted in bold.

Table S3 Average normalized counts of gene candidates involved in 4-methylvaleric acid biosynthesis and *PcAAE2* in *P. cablin* RNA-seq database.

Unigene ID	**Seeding**	**5W-Root**	**5W-Stem**	**5W-TLeaf**	**8W-TLeaf**	**8W-SLeaf**
TRINITY_DN7984_c0_g1 (PcIBMS1)	1898.7	644.5	1672.67	1334.5	2.3333	4.3333
TRINITY_DN5555_c0_g1 (PcIPMS1)	1722.33	1939	2364.66	1770.3	1563.66	1980.33
TRINITY_DN5796_c0_g1 (PcIPMIL SSU)	3597.67	2497	3380.33	3993.5	2791	2745
TRINITY_DN22473_c0_g3 (PcIPMIL LSU)	4277	2992	4018.67	5021.5	2855	4303
TRINITY_DN28440_c0_g2 (PcIPMDHL)	2537	1982.5	2675.33	2671.5	1638.67	1346.33
TRINITY_DN3284_c1_g1 (PcPcBCATL1)	27.3333	263.5	32.3333	38	121.997	2702.92
TRINITY_DN20752_c0_g2 (PcBCATL2)	274	65	208	405	265	132
TRINITY_DN8015_c0_g3 (PcBCATL3)	1586	1753.49	1701	2119.5	1522	1220.67
TRINITY_DN6515_c0_g1 (PcBCKDHL E1α1)	1162	685	1961.67	1182.5	1766.67	3290.66
TRINITY_DN8285_c0_g2 (PcBCKDHL E1α2)	705	1224.5	552.333	631.5	545	605.333
TRINITY_DN12512_c0_g3 (PcBCKDHL E1β)	1259	1335	1477.33	1234.5	1201.33	1563
TRINITY_DN9874_c0_g2 (PcBCKDHL E2)	1070.67	2207	937.667	738	654	1225.67
TRINITY_DN22140_c0_g2 (PcBCKDHL E3)	5061.33	6519.5	3216.67	6105.5	2386.67	3396.67
TRINITY_DN3370_c0_g1 (PcAAE2)	801.67	350	974	658	152.667	96

Table S4 Ranking of the top 10 unique genes in the *P. cablin* RNAseq database by coexpression analysis with the *PcAAE2* gene.

Unigene ID	**Functional annotation by SWISS-PROT database**	**Pearson Correlation Coefficient**	***P* (Two Tailed)**
TRINITY_DN6205_c0_g3	ABC transporter B family member 25	0.979	0.001
TRINITY_DN1882_c0_g1	Transcription factor	0.969	0.001
TRINITY_DN7984_c0_g1 (PcIBMS1)	**2-isopropylmalate synthase A**	**0.969**	**0.001**
TRINITY_DN27812_c0_g1	Unknown protein	0.965	0.002
TRINITY_DN8346_c0_g1	BTB/POZ domain-containing protein	0.955	0.003
TRINITY_DN26973_c0_g1	Receptor like protein 29	0.955	0.003
TRINITY_DN4920_c0_g1	Protein HOTHEAD	0.954	0.003
TRINITY_DN19765_c0_g1	Ferritin-3	0.954	0.003
TRINITY_DN9171_c0_g3	Histidinol dehydrogenase	0.953	0.003
TRINITY_DN7580_c0_g1	Mitogen-activated protein kinase	0.952	0.003

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

Sequence name	**Source**	**GenBank accession or locus no.**	**References**
AtIPMS1	*Arabidopsis thaliana*	AT1G18500	(de Kraker *et al.*, 2007)
AtIPMS2		AT1G74040	
AtMAM1		AT5G23010	(Textor *et al.*, 2004)
AtMAM3		AT5G23020	(Textor *et al.*, 2007)
BjMAM1-A	*Brassica juncea*	CAQ56040	(Kumar *et al.*, 2019)
MdCMS1	*Malus domestica*	MD05G1155100	(Sugimoto *et al.*, 2021)
MdIPMS1		MD15G1332600	
MdIPMS2		MD01G1019700	
OsIPMS1	*Oryza sativa*	Os11g04670.1	(He *et al.*, 2019)
OsIPMS2		Os12g04440.1	
SlIPMS1	*Solanum lycopersicum*	Solyc06g053400.2.1	(Ning *et al.*, 2015)
SlIPMS2		Solyc08g014130.2.1	
SlIPMS3		Solyc08g014230.2.1	
MtIPMS (LeuA)	*Mycobacterium tuberculosis* H37Rv	P9WQB3	(Koon *et al.*, 2004)
Leu4	*Saccharomyces cerevisiae*	YNL104C	(Chang *et al.*, 1984)
Leu9		YOR108W	(Casalone *et al.*, 2000)

Protein sequences from PHYTOZONE v12.1 website were not included in this table and can be obtained directly from relevant websites through the gene names displayed in the phylogenetic tree.

Methods S1 Details on experimental procedures.

Analysis of metabolites of the *P. cablin* leaves fed with 4MVA or 4MVA-d_{11}

The *P. cablin* leaves were cut into pieces and volatiles of 100 mg leaf pieces were extracted with 500 µl MTBE containing 0.01 ng/µl internal standard tetradecane and analyzed by GC-MS. The relative abundances of pogostone and deuterium label incorporated pogostone (pogostone-d_{11}) were calculated by peak area normalization to internal standard tetradecane. 100 µl MTBE extracts were further dried and dissolved in 400 µl 90% methanol for LC-QTOF-MS analysis.

Mass-to-charge ratio (*m/z*) of 224.1 and 235.2 were extracted for pogostone and pogostone-d_{11}, respectively, in GC-MS analysis, while *m/z* of 223.0976 and 234.1666 in negative mode in LC-QTOF-MS analysis were extracted for pogostone and pogostone-d_{11}, respectively. The identification of pogostone-d_{11} were based on the exact mass and fragment pattern by GC-MS analysis and LC-QTOF-MS analysis.

Vector construction for expression in *Escherichia coli*

To obtain soluble proteins for expression in *Escherichia coli*, the truncated open reading frames of *PcIBMS1* and *PcIPMS1*, both missing the first 40 amino acids, were obtained by Reverse transcriptase polymerase chain reaction (RT-PCR) from prepared cDNA of *P. cablin* seedlings and were introduced into the expression vector pET-28a (+) between the restriction sites of NdeI and BamHI using NEBuilder® HiFi DNA Assembly Cloning Kit (NEB, Catalog number: E5520S), in each case generating a fusion gene that encoded a “tag” of HIS_6 residues at the N-terminus for expression in *E. coli*. The corresponding resulted expression vectors are pET-28a-PcIBMS1 and pET-28a-PcIPMS1, respectively. The two fragments for assembly for PcIBMS1 M132L and PcIPMS1 L135M constructs were obtained by RT-PCR from corresponding vectors pET-28a-PcIBMS1 and pET-28a-PcIPMS1 and were assembled into the expression vector pET-28a (+) in the same way as the construction of pET-28a-PcIBMS1 and pET-28a-PcIPMS1. The corresponding resulted expression vectors are pET-28a-PcIBMS1 M132L and pET-28a-PcIPMS1 L135M, respectively.

Enzymatic assay for condensation reaction

The enzyme assay for the condensation reaction between acetyl-CoA and different 2-oxo acids was performed as previously described (de Kraker *et al.*, 2007) with minor modification. Briefly, substrate specificity assays for recombinant PcIBMS1 and PcIPMS1 were determined with a spectrophotometric end-point assay with DTNB (Sigma-Aldrich). An aliquot of 1-5 μL of the protein preparation was added to an enzyme assay mixture (100 mM Tris, pH 8.0, 1 mM acetyl-CoA, 10 mM $MgCl_2$ and 0.3 mM 2-oxo acid) for a final volume of 200 μL and incubated for 10 min at 25 °C. The reaction was stopped by freezing in liquid nitrogen and subsequent addition of 160 μL ethanol and 80 μL of 2 mM DTNB (fresh solution in 100 mM Tris, pH 8.0). The mixture was left for 2 to 3 min for DTNB to react with free thiol groups released from acetyl-CoA to create a yellow-color TNB^{2-} dianion product with an ε_{412} of 14140 M^{-1} cm^{-1} (Kohlhaw, 1988). After full color development, the absorbance of the mixture was measured at 412 nm and then adjusted by subtracting the background of the identical enzyme assay mixture without 2-oxo acid. For 4-Methylthio-2-oxobutanoic acid, the results were further corrected for the slight reactivity between its thiol group and DTNB using reactions containing the 2-oxo acid, but not acetyl-CoA. The assay was generally linear for the first 15 min with 1 μg of protein.

Steady-state kinetic assays were DTNB end-point assays that run at 25 °C using 1 to 5 μL aliquots of protein preparation. Absorbance of the reaction was found to increase linearly for at least 15 min. For individual assays, absorbance was recorded every 5 min for 10 min. Correction for background absorbance was performed as described above. Each assay was repeated 3 times. Kinetic assays for 2-oxo acid were performed with concentration of acetyl-CoA fixed at 1 mM while the concentration of 2-oxo acid ranging from 0.05 to 4 mM. Kinetic assays for acetyl-CoA were performed with the concentration of 2-oxo acid fixed at 2 mM while the concentration of acetyl-CoA ranging from 0.02 mM to 1 mM. Kinetic parameters were determined using the hyperbolic regression analysis method in Hyper32 software (version 1.0.0, http://hyper32.software.informer.com/). The spectrophotometric assay was also used to test the effect of Leu on enzyme activity in the range from 0.05 mM to 10 mM.

In addition, the condensation activities of recombinant PcIBMS1 and PcIPMS1 towards 2-oxoisovalerate and 4-methyl-2-oxovalerate were further measured using an LC-QTOF-MS-based method. Briefly, the reaction mixtures (200 μL) contain 0.1 mM Tris-HCl, pH 8.0, 0.5 mM acetyl-CoA, 10 mM $MgCL_2$, 0.5 mM 2-oxo acid substrate. The reaction was initiated by adding 2 μg purified enzyme protein and stopped after 20 min by addition of 400 μL methanol. The solution was filtered through a 0.45 μm filter prior to LC-QTOF-MS analysis.

Analysis of metabolites in *N. benthamiana* leaves

To analyze the compounds produced in *N. benthamiana* leaves, 500 mg of ground leaf tissues were mixed thoroughly with 100 μl ultrapure water and then centrifuged to remove pellet. 50 μl supernatant were mixed thoroughly with 450 μl methanol containing 0.01mM naringenin and filtered through a 0.45 μm filter for LC-QTOF-MS analysis. 200 μl supernatant were extracted with equal volume of MTBE containing 10 ng/μl geraniol as internal standard. The MTBE extracts were then analyzed by GC-MS.

Base hydrolysis of modified branched chain fatty acids was performed as previously described (Xu *et al.*, 2018). 500 mg of ground leaf tissues were mixed completely with 50 μL 4N NaOH and incubated at 80°C for 20 min, followed by neutralization with 50 μL 4N HCl, and then centrifuged at 10000 g for 15 mins. 200 μl supernatant were extracted with equal volume of MTBE containing 10 ng/μl geraniol for GC-MS analysis.

References：

Casalone E, Barberio C, Cavalieri D, Polsinelli M. 2000. Identification by functional analysis of the gene encoding alpha-isopropylmalate synthase II (LEU9) in *Saccharomyces cerevisiae*. *Yeast* **16**(6): 539-545.

Chang LF, Cunningham TS, Gatzek PR, Chen WJ, Kohlhaw GB. 1984. Cloning and characterization of yeast Leu4, one of two genes responsible for alpha-isopropylmalate synthesis. *Genetics* **108**(1): 91-106.

de Carvalho LP, Argyrou A, Blanchard JS. 2005. Slow-onset feedback inhibition: inhibition of *Mycobacterium tuberculosis* alpha-isopropylmalate synthase by L-leucine. *J Am Chem Soc* **127**(28): 10004-10005.

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**(1): 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 Physiol* **143**(2): 970-986.

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 Biotechnol J* **17**(2): 322-337.

Kohlhaw GB. 1988. Alpha-isopropylmalate synthase from yeast. *Methods Enzymol* **166**: 414-423.

Koon N, Squire CJ, Baker EN. 2004. Crystal structure of LeuA from *Mycobacterium tuberculosis*, a key enzyme in leucine biosynthesis. *Proc Natl Acad Sci U S A* **101**(22): 8295-8300.

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**(7): 1633-1647.

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 Physiol* **169**(3): 1821-1835.

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. *Proc Natl Acad Sci U S A* **118**(3): e2009988118

Textor S, Bartram S, Kroymann J, Falk KL, Hick A, Pickett JA, Gershenzon J. 2004. Biosynthesis of methionine-derived glucosinolates in Arabidopsis thaliana: recombinant expression and characterization of methylthioalkylmalate synthase, the condensing enzyme of the chain-elongation cycle. *Planta* **218**(6): 1026-1035.

Textor S, de Kraker JW, Hause B, Gershenzon J, Tokuhisa JG. 2007. MAM3 catalyzes the formation of all aliphatic glucosinolate chain lengths in arabidopsis. *Plant Physiol* **144**(1): 60-71.

Xu H, Moghe GD, Wiegert-Rininger K, Schilmiller 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 Physiol* **176**(1): 524-537.