SI Appendix, Materials and Methods

Isotope feeding study with ^{13}C acetic acid.

Application of stable acetic acid isotopes. Five filter paper discs (11.5-mm dia.) were placed on a glass side and each was wetted with approximately 20 µL of a solution (adjusted to pH 7 with 1 M KOH) with or without 0.1 M MES, with and without methanol (20 mM), and containing unlabeled acetic acid (20 mM) or 1-[^{13}C]- or 2-[^{13}C] or 1,2-[$^{13}C_2$] acetic acid (20 mM) having 99 atom % ^{13}C enrichment (Sigma-Aldrich, St. Louis, MO, USA). Peel plugs (11.5-mm dia.) were removed randomly from the surface of fruit and trimmed to 1- to 2-mm thickness. The resulting apple discs were placed skin-side-up onto the wetted filter paper discs and approximately 40 µL of additional incubation solution was added to each of the apple five discs on the slide. The slide was then placed into a horizontally-oriented 30-mL glass vial sealed with a cap housing a valved septum (Mininert valve 20/400mm; Sigma-Aldrich). The samples were incubated for 24 to 26 h at 22 °C. The incorporation of ^{13}C into headspace volatiles was analyzed by gas chromatography (GC) coupled with time-of-flight mass spectrometry (TOFMS) as described below. Four biological replicates were performed with each replicate being derived from a single fruit. Two of the four biological replicates were incubated in solutions containing MES buffer and in the other two biological replicates, discs were incubated in pH-adjusted water. The presence of the buffer did not affect the volatile profile or incorporation of label into volatiles. Following analysis of volatiles, apple discs were freeze-dried and derivatized acids were quantified using GC/MS as described below. The data for headspace volatiles were from apple discs incubated without MES buffer and the data for soluble metabolites were from apple discs incubated with MES.

Headspace volatile analysis. Headspace ester, aldehyde, and alcohol contents were measured by GC/TOFMS. The GC (HP-6890, Hewlett Packard Co., Wilmington, DE) was equipped with a mass selective detector (Pegasus II, LECO Corp., St. Joseph, MI) operated using electron ionization (70 eV). Headspace volatiles were sampled using a 1-cm long, solid-phase microextraction (SPME) fiber (65 μm PDMS-DVB, Supelco Co., Bellefonte, PA). Following a 3-min sorption time, the SPME fiber was immediately transferred to the GC injection port (230 °C) and desorbed for 2 min. Conditions of GC separation and TOFMS analysis were as previously described (1). In brief, desorbed volatiles were trapped on-column using a liquid nitrogen cryofocusing trap. Separation of volatiles was by capillary column (HP-5MS, 29 m x 0.25 mm i.d., 0.25 μm film thickness, Agilent Technologies, Santa Clara, CA). The temperature of the GC was ramped from 40 to 240 °C at a rate of 40 $°C \cdot min^{-1}$, the flow rate of the helium carrier gas was 1 $mL \cdot min^{-1}$, and the GC was operated in splitless mode. Identification of compounds was by comparison of the mass spectrum with authenticated reference standards and/or with spectra in the National Institute for Standard and Technology (NIST) mass spectrum library (version 05) when no standard was available. Volatile compound concentrations ($nmol \cdot L^{-1}$) in the chamber headspace were quantified by comparison of instrument response to that of a 72-component standard mixture of authenticated, high-purity, esters and alcohols (Sigma-Aldrich Corp., St. Louis, MO, USA).

Soluble metabolite analysis. Apple discs were freeze-dried and placed into 15-mL polypropylene tubes containing three 4-mm dia. stainless steel balls and ground using a vibrating grinder (GenoGrinder 2000, SPEX CertiPrep, Inc., Metuchen, NJ) operated at 700 strokes per min. Approximately 0.2 to 0.25 mL of solution containing acetonitrile, 2-propanol, and water in

a ratio of 3:3:2 (v/v/v) was added to a 1.5-mL microfuge tube containing 0.020 to 0.025 g of ground tissue and held for 5 min at 22 °C for extraction. The extract was centrifuged at 14,000 × g for 5 min and 100 µL of the supernatant was transferred to a clean 1.5-mL microfuge tube and vacuum dried. After drying, 50 µL of methoxyamine hydrochloride in pyridine (10 mg·mL^{-1}) was added and the solution incubated at 50 °C overnight. The incubated solution was derivatized by adding 100 µL of N-methyl-N-(*tert*-butyldimethylsilyl)trifluoroacetamide (MTBSTFA) and incubated overnight at 22 °C. Metabolites were quantified by comparison of instrument response to that of an authenticated, high purity derivatized standard mixture containing 50 µM each of valine, threonine, isoleucine, leucine, aspartic acid, citramalic acid, α-ketobutyric acid, α-keto-β-methylvaleric acid, 2-isopropylmalic acid (Sigma-Aldrich, St. Louis, MO), and citraconic acid (Fisher Scientific USA, Pittsburgh, PA). In addition, a standard mixture containing 50 µM each of 2-hydroxyglutarate and citramalic acid was prepared, to assure these two compounds having identical mass and similar mass spectra could be separated.

Derivatized acid analysis was performed using a mass spectrometer (Agilent 5973, Agilent Technologies, Santa Clara, CA or Pegasus II, Leco, Inc., Saint Joseph, MI) coupled to a GC (Agilent 6890). The injection port temperature was 280 °C. Separation of volatiles was by capillary column (DB-5MS, 30 m x 0.25 mm i.d., 0.25 µm film thickness, Agilent, Santa Clara, CA). The GC was programmed as follows: 80 °C for 2 min, ramped 30 °C·min^{-1} to 130 °C, then 15 °C·min^{-1} to 300 °C, and held at 300 °C for 3 min. The flow rate of the helium carrier gas was 1.2 mL·min^{-1}. The GC was operated in either split (1:20 ratio) or splitless mode, depending on the need for sensitivity. Electron ionization (70 eV) was employed and selected ion monitoring (SIM) was used when needed to achieve low levels of detection. Specifically, SIM *m/z* 188, 216, 288, 301,302, 404, and 461 were used to detect α-ketobutyrate, α-keto-β-methylvalerate, valine,

citraconic acid, isoleucine, threonine, and 2-isopropylmalate, respectively. Injection volumes were 0.2 μL for split mode and either 0.2 μL or 1 μL for splitless mode. Identification of compounds was by comparison of the mass spectrum and GC retention time with those of derivatized standards and comparison to the NIST 05 mass spectrum library using commercial software (ChemStation, Agilent, Santa Clara, CA). The quantification was by comparison of peak areas with a known amount of derivatized standards as described previously. 2-Ethylmalic acid (no standard available) was quantified by estimation of the instrument response factor for adjacent eluting compounds (citramalic acid and 2-isopropylmalic acid) in the standard mixture for which standards were available.

Isotopic analysis. Isotopologs (mass isotopomers) of headspace volatiles and soluble acids were quantified as described under the section "Headspace volatile analysis." Propanol, propanal, methyl propanoate, methyl butanoate, and butanal were quantified by integrating extracted ion chromatogram peaks for the molecular ions (M^+) *m/z* 60, 58, 88, 102, and 72, respectively. 2-Methylbutanol and methyl 2-methylbutanoate were quantified by integrating extracted ion chromatogram peaks for fragment ions *m/z* 70 and 88, respectively. Incorporation of label into carbon 1 and 2 of methyl butanoate was quantified by integrating the chromatographic peak for *m/z* 74, a fragment corresponding to $\bullet CH_2C(=OH+)OCH_3$ generated by the McLafferty rearrangement. Abundances of the ion at this mass and its heavier isotopologs were used to distinguish the extent of ^{13}C incorporation into C-1 and/or C-2 of this ester. This was done to determine whether the alkanoate portion of the ester is synthesized via one- or two-carbon fatty acid biosynthesis (FAB). The acids isoleucine, citramalate, citraconate, 2-ethylmalate, threonine, and the mixture of 2-isopropylmalate and 2-propylmalate were quantified by integrating peaks at

m/z 302, 433, 301, 447, 404 and 461, respectively. These masses correspond to the molecular masses of the *tert*-butyldimethylsilyl derivatives minus 57 Da, corresponding to loss of the *tert*-butyl group from the derivative. The isotope enrichment was assessed by integrating peaks corresponding to the heavier isotopologs, and mole percent enrichments were calculated by correcting for natural isotope content by unlabeled acetate control on the mass distribution as described by Biemann (2). The mass isotopolog distribution is represented as the unlabeled mass fraction (M), the one ^{13}C-labeled mass fraction (M+1), the two ^{13}C-labeled mass fraction (M+2), and so on, up to the five ^{13}C-labeled mass fraction (M+5). The expected position of the isotopic carbon from labeled acetate in the various compounds of interest is described in the proposed pathway (Fig. 1). The amounts of threonine, α-ketobutyrate, and α-keto-β-methylvalerate were in low abundance or poorly resolved; however, the data for threonine are included to help settle the question of label in isoleucine arising via threonine deaminase.

Developmental changes in *MdCMS*, *MdIPMS1* and *MdIPMS2* expression and citramalate content.

To determine the developmental pattern of gene expression and citramalate content, eight developmental stages were selected based on physiological changes during ripening. These stages are: stage 1 (day 0), early preclimacteric; stage 2 (day 11), late preclimacteric and onset of trace ester biosynthesis; stage 3 (day 25), onset of autocatalytic ethylene and rapid increase of ester biosynthesis; stage 4 (day 32), half-maximal ester biosynthesis and engagement of the respiratory climacteric; stage 5 (day 39), near maximal ester biosynthesis, peak in respiratory activity, and onset of rapid tissue softening; stage 6 (day 49), end of maximal ester biosynthesis, conclusion of the respiratory climacteric, and completion of tissue softening; stage 7 (day 60),

midpoint in the decline in ester biosynthesis, maximal ethylene production, and onset of senescence; and stage 8 (day 70), postclimacteric minimum in ester production and extensive fruit senescence.

Plant material. 'Jonagold' apples were harvested for examination every three to four days from research plots at the Michigan State University Clarksville Horticultural Experiment Station, Clarksville, MI, from 2 Sept. 2004 (day 0) until ripening was fully engaged on 7 Oct. 2004 (day 35). After it was apparent that ripening was underway, the approximately 200 fruit remaining on the trees were harvested and transported to the laboratory on 7 Oct. 2004 (day 35). This was done to avoid damage in the field due to freezing and fruit drop. After every harvest, fruit were held overnight in the laboratory to equilibrate to laboratory temperature (21±1 °C). After 7 Oct., harvested fruit were covered with ventilated, black, 0.1-mm thick plastic bags to avoid desiccation and responses to intermittent laboratory light before analysis. Fruit were subsequently examined every three to four days until the conclusion of the study on 23 Nov. 2004 (day 81).

On each evaluation date, 20 apples were randomly chosen and their internal ethylene content measured (3). Of these, the fourteen fruit with an internal ethylene content nearest to the median were selected for further analysis. The four fruit having ethylene levels closest to the median were chosen for analysis of CO_2 production and ester emission. Of the remaining 10 fruit, maturity analysis (percent red coloration, background color, °Brix, and starch index) was performed as described by Mir et al. (4). From these fruit, the skin and 2 to 3 mm of underlying cortex tissue were removed, immediately frozen in liquid nitrogen, and stored at -80 °C. Two

replicates were created, each consisting of pooled tissue samples from five fruit. The pooled tissue samples were used for soluble metabolite analyses and gene expression studies.

***Isolation of RNA*.** Approximately 5 g of frozen 'Jonagold' apple tissue was ground using liquid nitrogen chilled mortar and pestle. Ground tissue was extracted using hot borate/phenol followed by LiCl precipitation (5). To compare gene expression, samples of leaf, root, and stem tissues from 'Jonagold' apple trees grown in a commercial orchard were collected. Root tissue was collected from a tree with scion rooting. Tissue samples (3 g) were subsequently ground and extracted as described for fruit tissue.

Microarray printing, design, labeling, and statistical analysis. Custom cDNA microarray slides were created using a robotic printing device in the Genomics Technology Support Facility (GTSF) Genomics Core in Michigan State University, East Lansing, MI. Approximately 10,000 unsequenced cDNA gene fragments were generated from the lambda phage cDNA library from 'Mutsu' apple fruit using a mass excision kit and protocols as described by the manufacturer (ZAP-cDNA synthesis kit, Stratagene, LaJolla, CA). In addition to the unsequenced cDNA fragments, 116 apple ESTs available in GenBank (kindly donated by Dr. Schuyler Korban, University of Illinois, Urbana-Champaign) were also included on the array. The microarray printing procedure, experimental design, microarray protocol used for labeling, hybridization and washing, image scanning, and statistical analyses were described previously (3, 6). Gene fragments undergoing greater than a four-fold change in expression and/or identified as undergoing significant changes in expression ($P<0.00025$) relative to day 0 were sequenced. Assigned tentative identities for sequenced gene fragments was based on BLAST analysis of the

predicted amino acid sequence against the NCBI non-redundant protein and Arabidopsis protein databases. The initial tentative identity for citramalate synthase (*MdCMS*) was *IPMS*. Following protein characterization, the designation *MdCMS* was adopted and is used uniformly hereafter for improved clarity.

Determination of mRNA transcript levels by reverse transcription polymerase chain reaction (RT-PCR). The expression of a mixture of two *MdCMS* alleles, *MdIPMS1*, *MdIPMS2*, and *18S ribosomal RNA* (*18s rRNA*) was measured using semiquantitative RT-PCR analysis. For each analysis, there were two biological and two technical replications. cDNA synthesis and PCR were performed using commercially available kits according to manufacturer directions (Invitrogen, Carlsbad, CA). Before creating cDNA, total RNA was treated with DNase using an RNase-free DNase kit according to the manufacturer (Qiagen Inc, Valencia, CA). DNase-treated total RNA (1.0 μg) was reverse transcribed using oligo(dT)12-18 primer or random hexamer and SuperScript II as described by the manufacturer (Invitrogen). cDNA created with oligo (dT)12-18 primer was used for 18s rRNA as well as *MdCMS* expression, cDNA created with random hexamer was used for 18s rRNA and *MdIPMS* expression analysis. The cDNA (1.0 μL) was used as a template in a 50-μL PCR cocktail containing 10 μM of the forward and reverse gene-specific primers (GSP). Primer sequences, the expected size of the PCR product, optimum cycle number, and optimum temperature for primer binding are listed in *SI Appendix, Table S1*. The PCR was performed as follows: 1) 5 min at 95 °C, 2) 30 s at 95 °C, 3) 30 s at 57–59 °C, 4) 30 s at 72 °C, repeating 18–33 cycles from steps 2–4, and final elongation 5 min at 72 °C. The PCR products were separated by electrophoresis on a 1.5% (w/v) agarose gel, visualized with UV fluorescence of ethidium bromide, and photographed. Relative light density of the bands was

quantified by a digital imaging system (EagleEye II, Stratagene, La Jolla, CA). To identify the optimum cycle number, the gene products amplified by PCR had to be visible on the gel electrophoresis and be quantifiable by light density measurement without saturation of pixels. For each PCR run, the target cycle number was bracketed by 2 to 3 cycles to ensure that transcript synthesis had not plateaued. PCR products were cleaned using a PCR purification kit (QIAquick, Qiagen) and sequenced at the GTSF to verify identity. All the PCR generated sequences were 98% to 100% identical to the original sequence. A partial sequence of *18s rRNA* (gi: 85717895) was used as an internal control for PCR analyses. The spot density for the *18s rRNA* varied approximately ±10% across the eight developmental stages.

MdCMS and MdIPMS cloning, identification, and sequencing.

Total RNA extracted from ripe 'Jonagold' fruit skin as described previously was used to identify nucleotide sequences. All primers used for identifying genomic and coding sequences are listed in *SI Appendix, Table S2*. First-strand cDNA was synthesized using 4 µg of total RNA, oligo(dT)12-18 primer, and SuperScript II (Invitrogen) according to manufacturer instructions. Gene fragments were amplified by PCR containing cDNA, 10 µM of the forward and reverse gene-specific primers, and a high-fidelity DNA polymerase (Phusion, New England Biolabs Inc., Ipswich, MA) in a 50-µL PCR cocktail buffer as described by the manufacturer. The PCR products were gel purified (QIAquick gel extraction kit, Qiagen) and the purified PCR products were used to add dATP in the 50-µL reaction buffer containing 0.2 mM dATP and PCR solutions as provided by the manufacturer ($MgCl_2$, PCR 10x buffer, and Taq DNA polymerase) (Invitrogen) and incubated for 15 min at 72 °C. The mixture was ligated into a modified PCR

2.1-TOPO vector (Invitrogen) and used to transform TOP10 *E. coli* cells (Invitrogen). The transformed colonies were screened by restriction analysis and sequenced at the GTSF.

MdCMS mRNA sequence. An *MdCMS* clone was obtained from the 'Mutsu' cDNA library created as described previously and sequenced at the GTSF to obtain a full-length sequence of the mRNA, which was about 2000 nucleotides long. The predicted open reading frame (ORF) and the translated protein size of 473 amino acids were obtained by the NCBI ORF finder program. However, the anticipated protein size was much shorter than Arabidopsis IPMS and MAM proteins that 3' RACE was performed to verify the accurate coding region.

3' RACE of MdCMS. 3' RACE was performed using a commercially available kit (3' RACE system for rapid amplification of cDNA ends, Invitrogen). In brief, first-strand cDNA was synthesized using 4 µg of total RNA isolated from Jonagold apple fruit, SuperScript II, and an adapted primer. cDNA (2.0 µL) was used as templates in a 50-µL PCR cocktail containing CMS GSP (*SI Appendix,* Table S2, no. 3) and an abridged universal amplification primer provided by the manufacturer. PCR products were gel purified and cloned into a modified pCR 2.1-TOPO vector, which was used to transform DH5α *E. coli* cells. Sequencing of the insert was performed by GTFS.

MdCMS gDNA sequence. To determine the genomic sequence of *MdCMS*, approximately 100 mg of gDNA was isolated from young buds of 'Jonagold' and 'Mutsu' using a commercial kit (DNeasy Plant Mini Kit, Qiagen). gDNA was used as a template in a 50-µL PCR cocktail containing 10 µM of the ORF forward and reverse primer (*SI Appendix, Table S2*, nos. 1 and 2).

PCR products were gel purified and cloned directly into the vector using a commercial kit (Zero Blunt TOPO PCR cloning kit, Invitrogen). Inserts in the transformed colonies were verified by sequencing at GTSF.

MdIPMS mRNA sequences. Clones of the five apple ESTs (GenBank nos. CN488984, DT000542, DT001339, DT002187, DT001294) annotated as *IPMS* in the GenBank database were kindly donated by Dr. Schuyler S. Korban. Of these, only DT001294 and DT001339 were of sufficient length to be used for further analysis. However, the 5' end of the sequence was missing in both clones. Therefore, we performed 5' RACE as described below.

5' RACE of MdIPMS. 5' RACE was performed as described by Scotto-Lavino et al. (7) with modification. In brief, first-strand cDNA was synthesized using 5 µg of total RNA, SuperScript II, and IPMS GSP (*SI Appendix, Table S2*, no. 19). A 5' end tailed cDNA was created from the cDNA by incubating with dATP and terminal deoxynucleotidyl transferase (Invitrogen). Two PCRs were performed using high-fidelity DNA polymerase (Phusion, New England Biolabs): the first PCR cocktail contained primers IPMS GSP1 (no. 20), Qt (no. 16), and Qo (no. 17), and the second nested PCR contained the first PCR cocktail as a template and primers IPMS GSP2 (no. 21) and Qi (no. 18). The nested PCR products were gel purified, cloned into a modified pCR 2.1-TOPO vector, transformed to DH5α *E. coli* cells, and the insert sequence was verified at the GSTF.

Once the tentative cDNA sequence was identified from 5' RACE, the ORF was obtained by PCR using forward and reverse primers nos. 12 to 15 (*SI Appendix, Table S2*). The gel purified PCR products were cloned into a modified pCR 2.1-TOPO vector, transformed to DH5α

E. coli cells, and sequenced. We isolated two IPMS sequences, which were designated *MdIPMS*1 and *MdIPMS*2, with sequences identical to clones DT001339 and DT001294, respectively.

MdCMS allelic composition and gene expression in association with volatile profile.

Ninety-nine accessions of apple previously characterized for volatile profiles by Sugimoto et al. (8), were genotyped. Sixty-five were listed as *Malus* × *domestica*, thirty-four as hybrid (*SI Appendix*, germplasm_genotype_BCester.xls). Leaf tissue was collected from the USDA Plant Genetic Resources Unit's *Malus* Germplasm Repository, Geneva, NY in summer and fall of 2020 and DNA was extracted using the DNeasy Plant Mini Kit (Qiagen). Amplicons for sequencing were produced by PCR with primers CMS_SNPF and CMS_SNPR as listed in *SI Appendix,* Table S3. The 25 µL reaction mixture consisted of 1X of 10X ThermoPol Reaction Buffer (New England BioLabs), 200 µM dNTPs (Promega), 0.2 µM forward and reverse primer (Integrated DNA Technologies), 0.625 units *Taq* DNA polymerase (New England BioLabs) and 1 µL ~50 ng/µL DNA. Reaction conditions consisted of 30 s/95 °C, 30 cycles of 15 s/95 °C, 15 s/51 °C, 45 s/68 °C, followed by 5 min/68 °C. The 251 bp amplicon was purified with Monarch PCR & DNA Cleanup Kit (New England BioLabs). Purified amplicons were then sent for sequencing by the RTSF with primer CMS_SNPF. Analysis was performed with the software Chromas (Technelysium Pty Ltd).

In addition, gene expression of *MdCMS* in ripe and unripe fruit was determined for fruit having high or low emission of esters derived from 2-methylbutanol or 2-methylbutanoate (2MB) relative to straight-chain (SC) esters as described by Sugimoto et al (8). Five high 2MB/SC ratio and five low 2MB/SC ratio lines were selected. There were two replicates per line, each a pooled sample from four fruit. RNA was isolated from skin tissue and the expression of a

mixture of the two MdCMS alleles and 18s rRNA was measured using semiquatitative RT-PCR analysis as described above.

Expression of MdCMS and MdIPMS in *E. coli* and protein purification

All the primers used for expressing proteins are listed in *SI Appendix, Table S2*. The cloning procedures were described previously in "*MdCMS* and *MdIPMS* cloning, identification, and sequencing".

Isolation of RNA and cDNA cloning into expression vector. Total RNA from ripe 'Jonagold' apple skin was used to create cDNA. PCR was performed using forward and reverse primer nos. 4 to 7. To generate full length ORF and truncated ORF lacking a putative chloroplast transit peptide (9), forward primers nos. 4 and 5 were used with the combinations of ORF reverse primers nos. 6 and 7, which, respectively, do and do not contain a stop codon. dATP was added to the gel purified PCR product, cloned into an expression vector (pBAD-TOPO, Invitrogen) to create *MdCMS_1*- and *MdCMS_2*-pBAD-TOPO, which were used to transform TOP10 *E. coli* cells (Invitrogen). The insert was verified by sequencing at the GTSF.

The protein expression (see below) of the pBAD-TOPO construct was very poor so the PET 101/D-TOPO (Invitrogen) expression vector containing C-terminal 6xHis tag was used. The *MdCMS* ORF was generated using *MdCMS*-pBAD-TOPO as a template in a PCR cocktail using forward and reverse primer nos. 8 to 11. The ORF for the *MdIPMS* genes was generated using *MdIPMS*-pCR 2.1 as a template in a PCR cocktail using primer nos. 22 to 27. The forward primer nos. 8 to 10, 22, 24, 25, and 27 were used to generate full length ORF and truncated ORFs (lacking a putative chloroplast transit peptide) as previously described. For CMS, only the

full-length protein was expressed, and the truncated protein failed to express despite using two different forward primers (nos. 9 and 10). For IPMS, both full-length and truncated proteins were expressed. The gel purified PCR products were cloned directly into the PET 101/D-TOPO vector, generating full length *MdCMS*-PET 101/D and full and truncated length *MdIPMS*-PET 101/D. Vectors were used to transform TOP10 *E. coli* cells (Invitrogen) and the insert was verified by sequencing at the GTSF. The screened constructs were used to transform BL21(DE3) *E. coli* cells (Invitrogen) for protein expression analysis.

Protein expression in E. coli. *MdCMS*- or *MdIPMS*-PET 101/D-TOPO in BL21(DE3) cells were grown in 250 mL of Luria-Bertani medium containing antibiotic carbenicillin (50 $\mu g \cdot mL^{-1}$) at 37 °C for 1-2 h until the OD_{600} reached 0.6. The expression of the target gene was induced by adding isopropyl β-D-1-thiogalactopyranoside optimized to final concentrations of 0.25 mM and 1 mM for CMS and IPMS, respectively, and was further incubated for 6 to 7 h at 22 °C. After incubation, the cells were harvested by centrifugation at 5000 × g for 10 min at 4 °C and the bacterial pellets were stored at –80 °C.

Expressed His-tag protein purification. Bacterial pellets were homogenized with 10 mL lysis buffer (50 mM sodium phosphate buffer pH 8.0, 300 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, 0.1% TritonX-100, 20 mM imidazole), 200 µL lysozyme (50 $mg \cdot mL^{-1}$, Hoffmann-La Roche Ltd., Basel, Switzerland), 5 µL benzonase (Sigma-Aldrich), and 1 tablet of EDTA-free protease inhibitor (Hoffmann-La Roche Ltd.) for 20 min at 4 °C. The cells were disrupted using a sonicator (Misonix Sonicator S-4000 Ultrasonic Processor, Qsonica, LLC., Newtown, CT) fitted with a Microtip probe (Qsonica, LLC.). The cells were lysed using pulsed sonication (10 s)

followed by 10 s with no sonication, repeated for a total process time of 1.5 min. The sonicator power setting was 4 to 5 W. The disrupted cells were precipitated by centrifugation at 12,000 × g for 15 min at 4 °C. The bacterial lysate was added to a chromatography column (Econo-Pac, Bio-Rad Laboratories, Inc., Hercules, CA) packed with 1.5 mL Ni-NTA agarose (Qiagen) previously rinsed with lysis buffer. Binding of the His-tag protein to the nickel resin was facilitated by mixing for 1 to 2 min. The column was washed with a washing buffer (lysis buffer containing 35 mM imidazole). The His-tag protein was eluted with 3 mL elution buffer (lysis buffer containing 300 mM imidazole). The eluent was transferred to a desalting column (Econo-Pac 10 DG column, Bio-Rad Laboratories) and 4 mL of desalting buffer (50 mM Tris buffer pH 8.0, 1 mM $MgCl_2$, and 10% glycerol) was applied as described by de Kraker et al. (10). The desalted eluent was used for enzyme assays.

The protein concentration was determined by Bradford assay using bovine serum albumin as a standard and the protein concentration ranged from 0.2 to 0.4 $\mu g \cdot \mu L^{-1}$. The His-tag protein was separated by electrophoresis on a 10% SDS-PAGE gel, stained using Coomassie Brilliant Blue R-250, photographed, and the sequence verified at the Research Technology Support Facility of the Proteomics Core at Michigan State University using LC/MS/MS.

MdCMS and MdIPMS protein characterization.

End-point enzyme assay (DTNB) for CMS and IPMS. Activity and substrate preference assays for CMS_1, CMS_2 and IPMS1 and 2 were performed as described by de Kraker et al. (9) with modifications for α-ketobutyrate, α-ketoisovalerate, α-keto-β-methylvalerate, α-ketoisocaproate, α-ketohexanoic acid, α-ketooctanoic acid, oxaloacetate, glyoxylate, α-ketoglutarate, α-keto-γ-(methylthio)butyric acid, pyruvate, and α-ketovaleric acid. An aliquot of 5 to 20 µL of the

protein preparation was added to an enzyme assay mixture (100 mM Tris buffer pH 8.0, 4 mM $MgCl_2$, 0.25 mM acetyl-CoA, 10 mM α-ketoacids) for a final volume of 150 µL and incubated for 20 min at 22 °C. The reaction was stopped by the addition of 200 µL ethanol and color was developed by adding 200 µL of a 1 mM solution of Ellman's reagent [5,5'-dithiobis-(2-nitrobenzoic acid) or DTNB] (Sigma-Aldrich) in 100 mM Tris buffer 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-colored TNB^{2-} dianion product. After full color development, the absorbance of the mixture was measured at 412 nm. The absorbance was adjusted by subtracting the background of the identical enzyme assay mixture without α-ketoacids. For α-keto-γ-(methylthio)butyric acid, the results were further corrected for the slight reactivity between its thiol group and DTNB using reactions containing the α-ketoacid, but not acetyl-CoA. The molar extinction coefficient (ε_{412}=14140 $M^{-1}\cdot cm^{-1}$) was used to calculate enzyme activity (11). The enzyme product was derivatized and analyzed with GC/MS as previously described and confirmed with an authenticated standard.

Enzyme kinetics. Enzyme assays for kinetic analyses were DTNB end-point assays run at 22 °C using 5 to 15 µL aliquots of protein eluent. Absorbance of the reaction was found to increase linearly for at least 30 min. For individual assays, absorbance was recorded every 5 min for 20 min. Correction for background absorbance was as previously described. K_m and V_{max} were determined by regression analysis of the Lineweaver-Burke plot. Each assay was repeated 2 to 3 times. The concentration of acetyl-CoA was fixed at 250 µM while the concentration of α-ketoacids (pyruvate, α -ketobutyrate, α -ketoisovalerate) ranged from 0.33 to 16 mM. For MdCMS_1 and 2 and MdIPMS1, the concentration of the α-ketoacids was fixed at 10 mM while

the acetyl-CoA concentrations ranged from 16.6 to 200 μM. For MdIPMS2, the concentration of α-ketobutyrate was fixed at 10 mM and α-ketoisovalerate at 5 mM while the acetyl-CoA concentration ranged from 16.6 to 200 μM. The concentrations of the substrates were selected so that the reciprocal substrate concentrations were evenly distributed on the Lineweaver-Burk plot to prevent uneven weighing during curve-fitting.

Determination of pH and amino acid feedback regulation. The optimum pH range for , MdIPMS1 and MdIPMS2 was determined using several buffers spanning pH 5.5 to 10.5. Buffers included potassium or sodium phosphate with citrate (pH 5.5 to 7.0), potassium or sodium phosphate (pH 6.0 to 8.0), Tris (pH 7.0 to 9.0), 2-amino-2-methyl-propane (pH 9.0 to 10.5). The rate of reaction was determined by end-point DTNB assay as previously described.

To determine inhibition by BCAAs, activity was measured following addition of valine, leucine, isoleucine, and threonine at concentrations of 0, 0.05, 0.1, 0.15, 0.3, 0.5, 1, 2.5, and 10 mM in the previously described enzyme assay mixture. Pyruvate (10 mM) was used as the α-keto substrate for MdCMS and α-ketoisovalerate (10 mM) was used for MdIPMS1 and 2. The rate of reaction was determined by end-point enzyme assay as previously described. Data are expressed as a percentage of the activity of control reaction containing no inhibitor. Each reaction was repeated twice.

Subcellular localization of MdCMS and MdIPMS. The ORFs of *MdCMS_1*, *MdCMS_2*, *MdIPMS1* and *MdIPMS2* were generated from the *MdCMS*-pBAD-TOPO and *MdIPMS*-PCR 2.1 constructs using forward and reverse primer nos. 28 to 36 in *SI Appendix,* Table S2. The gel purified PCR products were cloned directly into the plasmid pDONR207 by BP clonase

recombination reactions as described by the manufacturer (Invitrogen). The product of the BP recombination reactions was used to transform DH5α *E. coli* cells and the inserts were verified by sequencing at the GTSF. *MdCMS_1* and *2* and *MdIPMS1* and *2* in the pDONR207 construct were transferred to the destination vector pEarleyGate 101 (35S-Gateway-YFP-HA tag-OCS 3', ABRC stock no. CD3-683) (12) using LR clonase recombination reactions as described by the manufacturer (Invitrogen). The product of LR recombination reactions were used to transform DH5α *E. coli* cells, generating *MdCMS*- or *MdIPMS*-pEarleyGate 101, and the insert was verified by sequencing at the GTSF. *MdCMS*- or *MdIPMS*-pEarleyGate 101 constructs were used to transform *Agrobacterium tumefaciens* strain EHA105 as previously described.

Approximately 4- to 5-week-old tobacco (*Nicotiana tabacum* cv. Samson) plants grown at 22 °C in a growth chamber were used for transient expression assays. *MdCMS*- or *MdIPMS*-pEarleyGate 101 in *Agrobacterium* was syringe-infiltrated into leaves and the infiltrated areas were analyzed after three days by confocal microscopy as described by Reumann et al. (13).

Yeast complementation.

To delete TD gene, primers 1-4 (*SI appendix,* Table S4) were used to amplify the *Kluyveromyces lactis* TRP1 selective marker from plasmid pBS1479, which were kindly donated by Dr. Min-Hao Kuo at Michigan State University, USA (14). The amplified PCR product was transformed into YMRX-3B (Matα trp1-Δ, his3-Δ200, ura3-52, lys2-801, ade2-1 leu4::kanMX4, yor108w(leu9)::hIS3Mx6), which we received from Dr. Enrico Casalone at Università di Chieti, Italy (15), to replace the TD gene, thereby creating a triple knockout YMRX-3B-TD. The presence of the insert was screened by tryptophan prototroph selection. Yeast growth media and

conditions were based on standard procedures (16) and transformation was done with the lithium acetate method (17).

Transient expression in *Nicotiana benthamiana.*

Nicotiana benthamiana seeds treated with 10% w/v trisodium orthophosphate for 10 min, rinsed with distilled water and sown on a covered tray of Redi-Earth. After the second true leaf plants were transplanted to four-inch square pots of Redi-Earth and grown in a growth room at 23 °C under fluorescent lights (145 µmol m^{-2} s^{-1}) under a 16-h photoperiod and were supplemented with half-strength Hoagland's solution. The open reading frames lacking stop-codons of *MdCMS_1, MdCMS_2* and *MdIPMS2* from pDONR207, as described previously, were recombined into pEAQ-HT DEST3 vector (18) with the addition of a 6x His-Tag. The constructs were first transformed into *E. coli* strain DH5α and subsequently purified and transformed into *Agrobacterium tumefaciens* strain LBA4404 with selection for kanamycin and rifampicin. Preparation and infiltration of *A. tumefaciens* was as previously described (19) for five-week-old *N. benthamiana* plants. After five days, infiltrated leaf tissue was collected, immediately frozen in liquid nitrogen, and held at -80 °C.

Extraction, derivatization, and analysis of metabolites from *N. benthamiana* transient expression

Collected infiltrated tissue was ground by mortar and pestle in liquid nitrogen and 100 mg was aliquoted for extraction. A solution of internal standard consisting of a cell-free 2 mM amino acid mixture - ^{13}C, ^{15}N, (MilliporeSigma) was dissolved in 1:1 water:acetonitrile and added to samples and calibration standards at 100 µL per mL extraction buffer (1:1

water:acetonitrile) or calibration standard. Ground tissue was added to 2 mL of extraction buffer with internal standard, vortexed for 15 s, and incubated at 65 °C for 10 min. Samples were then placed on ice for 2 min and then centrifuged at 2300×*g* for 15 min. From the extract, 500 µL was transferred to a 0.2 µm nylon centrifugal filter (Costar, Corning) and centrifuged at 20000×*g* for 2 min. This was repeated to produce 1 mL of filtered extract. To prepare for derivatization, 75 µL of 1% w/v 4-nitrophenol (MilliporeSigma) was added followed by 50 µL 1M NaOH (MilliporeSigma) to each extract and calibration standard. Samples were then dried by SpeedVac for 54 hr at RT. For derivatization, 100 µL of 40 mg mL^{-1} methoxyamine hydrochloride (MilliporeSigma) dissolved in pyridine (MilliporeSigma) was added to each sample and incubated at 60 °C for 12 h. 100 µL of *N*-methyl-*N*--*tert*--butyldimethylsilyltrifluoroacetamide containing 1% *tert*-butyldimethyldilyl chloride was then added to each sample and incubated at 60 °C for 12 h. Samples were centrifuged at 20000×*g* for 10 min prior to GC-MS analysis. GC-MS analysis was as described above for "*Soluble metabolite analysis"* at a 1:10 split ratio with 1 µL injections on an HP-6890 GC (Hewlett Packard Co., Wilmington, DE) coupled to a Pegasus II TOFMS (LECO Corp., St. Joseph, MI). Quantification was performed by comparison of ratios of peak areas of metabolites to internal standard. Identification of compounds was by comparison of the mass spectrum and GC retention time with those of derivatized standards and comparison to the NIST 05 mass spectrum library. Valine (*m/z* 186) was compared to labeled valine (*m/z* 191). Leucine (*m/z* 200) was compared to labeled leucine (*m/z* 206). Isoleucine (*m/z* 200), threonine (*m/z* 303), and citramalic acid (*m/z* 433) were compared to labeled isoleucine (*m/z* 206). Calibration standards consisted of 0.001 to 1.0 µM of the following compounds: valine, leucine, isoleucine, threonine, and citramalic acid (MilliporeSigma).

Identification MdCMS protein in situ.

To confirm the presence of MdCMS protein in apple fruit, we first identified fragment peptides from purified preparation of MdCMS_1 and then determined whether those peptides, or others having sequence identity to MdCMS_1, were present in apple fruit. To identify peptide fragments of MdCMS, we extracted protein from *E. coli* BL21(DE3) cells transformed to contain the coding region of *MdCMS_1* as previously described. A 2-D electrophoresis (2-DE) gel of the *E. coli* MdCMS protein was used to estimate mass and isoelectric point using 2-D SDS-PAGE standards (Bio-Rad Laboratories, Hercules, CA, USA) and 18 cm Immobiline DryStrip gels (GE Healthcare, Baie-d'Urfé, QC, Canada) with non-linear pH gradients (pH 3–11 NL) under conditions and procedures described by Zheng, et al. (20). Digital images of 2-DE gels were processed and analyzed using PDQuest 2-D Image Analysis Software (Version 7.40, Bio-Rad Laboratories, Hercules, CA, USA). Selected spots with an estimated mass of 52 kDa were manually excised from a 2-DE gel under the exposure of UV (*SI Appendix,* Fig. S10). Excised gel spots were then treated with DTT and alkylated with iodoacetamide, then digested with trypsin. An LC-MS protocol reported previously (20) was used to identify characteristic fragments. Briefly, mass spectrometry was performed on a hybrid triple-quadrupole linear ion trap (Q-TRAP 4000, Applied Biosystems, Foster City, CA) equipped with a nanospray ion source. Spectra were acquired using Information Dependent Acquisition mode in Analyst 1.4.1 software (Applied Biosystems, Foster City, CA). For each MS scan (*m/z* 400-1400), the four most intense ions were selected for an enhanced resolution scan in order to determine charge states. The ion spray voltage was 2100 V, curtain gas was set to 20 (arbitrary units), and declustering potential was 60 V. MS/MS peak lists were generated via MASCOT v2.3 software (Matrix Science, London, UK). Product ion scans for precursors within 1.0 Da were summed if

collected within 30 IDA cycles of each other. All spectra were centroided and de-isotoped. The raw MS/MS data were searched against NCBI viridiplantae entries (278,115 sequences), updated 31 Dec., 2010 (NIH, Bethesda, MD, USA) using the MASCOT algorithm (Matrix Science, London, UK). The digested MdCMS yielded 13 unique peptides identical to the known CMS sequence, collectively representing 27% coverage (*SI Appendix*, Table S6). The MS and MS/MS mass tolerance were 1.2 and 0.5 Da, respectively, and one missed cleavage was allowed.

To verify that MdCMS protein was present in ripening apple tissue, a non-gel proteomic approach using LC-MS was used. For fruit analysis, 'Golden Delicious' (*Malus* × *domestica* Borkh.) apples were harvested from the Kentville Research and Development Center, Agriculture and Agri-Food Canada, Kentville, Nova Scotia, CA on 2 Oct., 2005. Fruit were determined to be pre-climacteric based on the internal ethylene concentration and starch index (21). Fruit were divided into two lots, one lot was treated with 36 $\mu L \cdot L^{-1}$ ethylene (ET) for 36 hours and the other was left untreated. Chambers were vented at the flow rate of 30 mL min^{-1}. Fruit were allowed to ripen 0, 7, 14, and 21 d at room temperature. Three sample groups were created: 1) day 0, day 7 and ET day 7; 2) day 0, day 14, and ET day 14; and 3) day 0, day 21, and ET day 21. Protein preparation and analysis was essentially as described by Zheng et al. (22). Protein was extracted from five individual fruits for each ripening period and pooled. Samples were taken from the apple peel (2-5 mm thick) via a phenol extraction followed by ammonium acetate-methanol precipitation (23). Four hundred (400) µg of protein extract were dissolved in an aqueous solution of 25 mM dithiothreitol (DTT) for 60 min at 37 °C, alkylated with 12.5 mM iodoacetamide for 30 min at room temperature in the dark, and then digested with trypsin in a 1:5 trypsin:protein ratio (w/w) (Promega, Madison, WI) at 37 °C for 12 hrs. The resulting peptides were labeled by reductive methylation of primary amino groups (N-term and K). Samples of fruit

at the three stages within each group were distinguished as ‘light’ (L), ‘intermediate’ (M), and ‘heavy’ (H) labeled by incubating with CH_2O (light formaldehyde), CD_2O (intermediate formaldehyde), and $^{13}CD_2O$ (heavy formaldehyde), respectively, in the presence of sodium cyanoborohydride ($NaBH_3CN$, light and intermediate) and sodium cyanoborodeuteride ($NaBD_3CN$, heavy). After labeling, three differentially labeled samples were pooled in a ratio of 1:1:1 (w/w) and desalted prior to fractionation. Peptides were then fractionated by strong cation exchange using an Agilent 1100 LC equipped with a PolySulfoethyl A column (100 mm x 2.1 mm x 3 µm particles with 300-Å pores, PolyLC, Columbia, MD) at a flow rate of 0.2 $mL{\cdot}min^{-1}$ using a linear gradient from 10 to 500 mM ammonium formate over 45 min. A total of 20 fractions were collected and dried in a vacuum centrifuge. The resultant peptides were extracted in washes of ammonium bicarbonate solution, acetonitrile and 1.0% formic acid. Extraction solvent was removed under vacuum and the peptides were resuspended with 30 µL of an aqueous solution containing 5% methanol and 0.5% formic acid. Peptides were analyzed on a quadrupole TOF LC/MS instrument (XeVo, Waters, Milford, MA, USA) as previously described (24). Only peptides that matched the identified the MdCMS_1 protein better than any other protein in the database with MASCOT scores of 25 or greater were selected for further analysis (*SI Appendix*, Table S7). Protein preparations from apple fruit yielded 4 peptides identical to the known CMS sequence, representing 11% coverage. False positive peptide identification rates were also calculated using the decoy option provided by MASCOT and estimated as below 1.5%.

Protein sequence alignment and phylogenetic tree. Protein sequences were aligned using the MUSCLE algorithm (25) and formatted with ESPript (26). Phylogenies were created using the Neighbor-Joining method (27). The optimal tree with the sum of branch length = 5.39 is shown.

The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches (28). All sequences were obtained from Genbank, Sol Genomics Network (https://solgenomics.net), and Rosaceae database (https://www.rosaceae.org). For pear (*Pyrus communis*) and strawberry (*Fragaria ananassa*), amino acid sequences were constructed by adding all exon nucleotides from the Rosaceae database and translation. NCBI accessions, UniProt accessions, and/or gene IDs are listed in *SI Appendix*, Table S5. The evolutionary distances were computed using the Poisson correction method (29) in MEGA X software (30) and are in the units of the number of amino acid substitutions per site. Signal sequences, regulatory chloroplast targeting regions, alignment gaps, and missing data were eliminated.

SI Appendix, Table S1. GenBank gi number, forward and reverse primer sequence (5'→3'), expected PCR gene fragment size (bp), optimum cycle number, and optimum annealing temperature ('Temp') for isopropylmalate synthase (*MdIPMS1* and *2*) and citramalate synthase (*MdCMS_1*) genes and the 18S ribosomal RNA (18s rRNA) gene from apple for semi-quantitative RT-PCR.

Target	GenBank gi no.	Forward primer sequence	Reverse primer sequence	Size (bp)	Cycle No.	Temp (° C)
MdIPMS1	48408575	agggtttaggaggcgactgt	cggagagtggtgtcgaagat	322	29	59
MdIPMS2	71823902	cgcatctccgaccctaacta	acctgctccttgctcttcct	386	33	59
MdCMS_1	XM_008348555.3	caccgtgaaggcactatgaa	ggctcggagacaatccttct	456	27	57
18s rRNA	85717895	gagaaacggctaccacatcc	gagcgtaggcttgctttgag	390	18	59

SI Appendix, Table S2. Primer list for identified procedures used to clone genomic and coding sequences and to express proteins of 2-isopropylmalate synthase (*MdIPMS1* and *2*) and citramalate synthase (*MdCMS_1* and *_2*) genes from 'Jonagold' apple.

Primer no.	Primer name	Primer sequence (5'->3')
		Genomic DNA
1	CMS forward	caccatggccttctcagcagaaaat
2	CMS reverse	caccgtgaaggcactatgaa
		3' RACE
3	CMS GSP	agctgaacaagactgccgatca
		MdCMS-pBAD-TOPO
4	CMS full forward	atggccttctcagcagaaaatc
5	CMS short forward	tgctccgctctctcccaatc
6	CMS full reverse (-stop)	aatggccacctgataaaccaatg
7	CMS full reverse (+stop)	ttaaatggccacctgataaacca
		MdCMS-PET 101/D-TOPO
8	CMS full forward	caccatggccttctcagcagaaaat
9	CMS short forward	caccatgtgctccgctctct
10	CMS short forward2	caccatggcctgctccgctctctcccaat
11	CMS full reverse	aatggccacctgataaaccaatg
		ORF/mRNA
12	IPMS1 forward	aaccatggcggctgtctgcacaa
13	IPMS1 reverse	tcatgcagacaccggcgttagtt
14	IPMS2 forward	cttctgctcaaaccaggaacca
15	IPMS2 reverse	tcatgcagacactttgtttcgttca
		5' RACE
16	Qt	ccagtgagcagagtgacgaggactcgagctcaagcttttttttttttttttt
17	Qo	ccagtgagcagagtgacg
18	Qi	gaggactcgagctcaagc
19	IPMS GSP	ctgatggataccactttcatgtgc
20	IPMS GSP1	ccaacaatagccttgtgtggct
21	IPMS GSP2	tgcaacccagtgtactcttcca
		MdIPMS-PET 101/D-TOPO

22	IPMS1 full forward	caccatggcggctgtctgcacaaa
23	IPMS1 full reverse	tgcagacaccggcgttagttcc
24	IPMS1 short forward	caccatggcgtgctctcaaactgacaaccccaaa
25	IPMS2 full forward	caccatggcaactgtctgcacgca
26	IPMS2 full reverse	tgcagacactttgtttcgttca
27	IPMS2 short forward	caccatggcgtgctctcaaactgacaaccccaaa
	MdCMS- and *MdIPMS*-pDONR207/YFP	
28	CMS full forward	ggggacaagtttgtacaaaaaagcaggcttcaccatggccttctcagcagaaaatc
29	CMS short forward	ggggacaagtttgtacaaaaaagcaggcttcaccatggcctgctccgctctctcccaatc
30	CMS reverse	ggggaccactttgtacaagaaagctgggtcaatggccacctgataaaccaat
31	IPMS1 full forward	ggggacaagtttgtacaaaaaagcaggcttcaccatggcggctgtctgcacaaacc
32	IPMS1 short forward	ggggacaagtttgtacaaaaaagcaggcttcaccatggcgtgctctcaaactgacaaccc
33	IPMS1 reverse	ggggaccactttgtacaagaaagctgggtctgcagacaccggcgttagttcc
34	IPMS2 full forward	ggggacaagtttgtacaaaaaagcaggcttcaccatggcaactgtctgcacgcacc
35	IPMS2 short forward	ggggacaagtttgtacaaaaaagcaggcttcaccatggcgtgctctcaaactgacaaccc
36	IPMS2 reverse	ggggaccactttgtacaagaaagctgggtctgcagacactttgtttcgttca

SI Appendix, Table S3. Primer list used to clone coding sequences containing the non-synonymous SNP (AA 387) of citramalate synthase (*MdCMS_1* and *_2*) genes from the genomic DNA of 99 apple lines from the USDA *Malus* Germplasm Repository, Geneva, NY, USA.

Primer no.	Experiment	Primer name	Primer sequence (5'->3')
1	Genomic	MdCMS_SNP forward	gtggaagagtacagcggatt
2	DNA	MdCMS_SNP reverse	gccaaaataatctcataggtgctc

SI Appendix, Table S4. Primers used to delete threonine deaminase in yeast strain YMRX-3B (LEU4 and LEU9 mutant) to create a triple knockout in an attempt to create a strain lacking the ability to make isoleucine.

Primer No.	Primer name	Primer sequence (5'->3')
1	Replace trp1 forward 1	atgtcagctactctactaaagcaaccattatgtacggttgtttacgactcactatagggc
2	Replace trp1 forward 2	caagccacatttaaactaagtcaattacacaaagttagtgatgtcagctactctactaaa
3	Replace trp1 reverse 1	ttaatatttcaagaatttttgataaacagtgttatcagtttctgaagcttgatatcgaat
4	Replace trp1 reverse 2	acaagttgttgcgtaaatttataaagtaaattgtcggttttaatatttcaagaatttttg

SI Appendix, Table S5. Amino acid similarity of putative IPMS and CMS genes in Rosaceae.

Species	Search database	Tentative ID	Target ID	eValue to MdCMS	eValue to MdIPMS1	eValue to MdIPMS2
Pyrus communis	Pyrus communis Bartlett DH Genome v2.0 transcripts	CMS	pycom05g14850	0E+00	2E-169	1.7E-167
Pyrus communis	Pyrus communis Bartlett DH Genome v2.0 transcripts	IPMS2	pycom01g04920	4E-73	0E+00	0.0E+00
Pyrus communis	Pyrus communis Bartlett DH Genome v2.0 transcripts	IPMS1	pycom15g29280	2E-42	3E-175	1.4E-163
Pyrus bretschneideri	Pyrus bretschneideri 'DangshanSuli' Genome Assembly v1.1 chromosomes	CMS	chr5-v1.1-pbr	4E-105	4E-60	1.3E-59
Pyrus bretschneideri	Pyrus bretschneideri 'DangshanSuli' Genome Assembly v1.1 chromosomes	IPMS2	NW_008988196.1-v1.1-pbr	1E-58	3E-107	1.1E-105
Pyrus bretschneideri	Pyrus bretschneideri 'DangshanSuli' Genome Assembly v1.1 chromosomes	IPMS1	chr15-v1.1-pbr	3E-58	4E-108	1.7E-106
Pyrus betulifolia	Pyrus betulifolia Genome v1.0 chromosomes	CMS	GWHAAYT00000005	6E-105	9.70E-60	2.7E-59
Pyrus betulifolia	Pyrus betulifolia Genome v1.0 chromosomes	IPMS2	GWHAAYT00000001	3.4E-91	2.7E-106	1.4E-108
Pyrus betulifolia	Pyrus betulifolia Genome v1.0 chromosomes	IPMS1	GWHAAYT00000015	2.3E-58	NA	6.4E-107
Pyrus ussuriensis x communis	Pyrus ussuriensis x communis Zhongai Genome v1.0	CMS	Pdr0g066290.1	0.0E+00	4.6E-170	1.0E-167
Pyrus ussuriensis x communis	Pyrus ussuriensis x communis Zhongai Genome v1.0	IPMS1	Pdr15g005160.1	3.7E-169	0.0E+00	0.0E+00
Pyrus ussuriensis x communis	Pyrus ussuriensis x communis Zhongai Genome v1.0	IPMS2	Pdr1g002310.1	3.1E-168	0.0E+00	0.0E+00
Prunus avium	Prunus avium Tieton Genome v1.1	IPMS	chr6	7.4E-58	1.5E-94	9.7E-96
Prunus armeniaca	Prunus armeniaca Genome v1.0	IPMS	PARG02544m01	5.8E-165	0.0E+00	0.0E+00
Prunus persica	Peach Genome v2.0.a1 primary transcripts	IPMS	Prupe.6G172500.1	6.5E-165	0.0E+00	0.0E+00
Fragaria x ananassa	Fragaria x ananassa Camarosa Genome Assembly v1.0 & Annotation v1.0.a1	IPMS	augustus_masked-Fvb1-2-processed-gene-162.3-mRNA-1	3.2E-162	0.0E+00	0.0E+00
Rubus occidentalis	Rubus occidentalis Whole Genome v3.0 Assembly & Annotation	IPMS	Ro01	2.6E-59	2.7E-95	9.1E-97
Rosa chinensis	Rosa chinensis Old Blush homozygous genome v2.0	IPMS	RcHm_v2.0_Chr2	2.6E-58	3.2E-95	2.6E-96

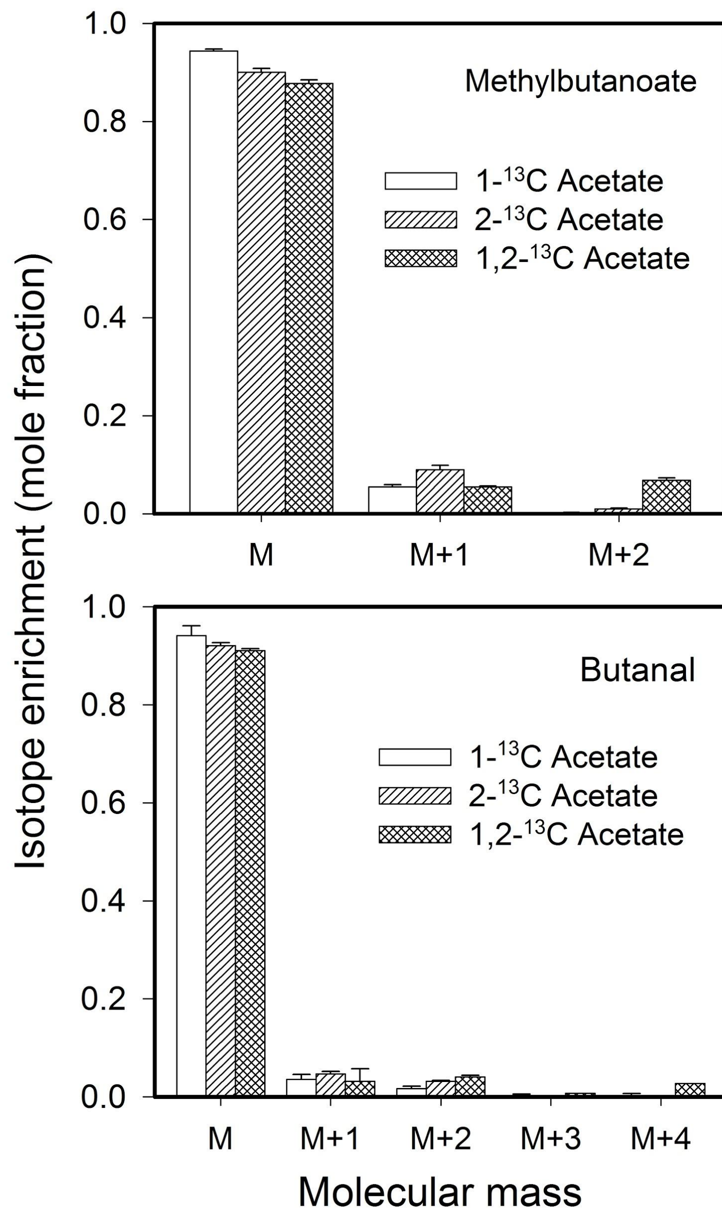
SI Appendix, Table S6. Identification of protein fragments extracted from purified CMS protein. Protein spots excised from gels stained with Sypro Ruby were subjected to digestion with trypsin and identified following mass spectrometry analysis (LC/MS/MS). Protein search was against MdCMS_1.

Spot	Mascot Score	Match % coverage	pI	
1	143	12	7.1	IPNENYV TADQVLELAK VDVIEAGFPASSK AGNASLEEFVMAVK.T + Oxidation SLTNEDLESLVYQVAI
2	106	5	7.0	TADQVLELAK AGNASLEEFVMAVK.T + Oxidation
3	506	25	6.95	IPNENYV RVVDTTLRDG EQAAGASMTR LRVDVIEA GFPASSK SDVDAAWESVK TADQVLELAK SLGAQDITFVCEDAGRSEK QVEVTINGIGERA AGNASLEEFVMAVK
4	406	24	6.9	IPNENYV RVVDTTLR LRVDVIEA GFPASSK SDVDAAWESVK TADQVLELAK SLGAQDITFVCEDAGRS EKEFLYR QVEVTINGIGERA AGNASLEEFVMAVK SLTNEDLESLVYQVAI
5	355	17	6.85	IPNENYVR VDVIEAGFPASSK

				SDVDAAWESVK TADQVLELAK SLGAQDITFVCEDAGR QVE VTINGIGERA AGNASLEEFVMAVK
6	281	17	6.8	IPNENYV VDVIEAGFPASSK SDVDAAWESVK TADQVLELAK SLGAQDITFVCEDAGR QVEVTINGIGERA AGNASLEEFVMAVK

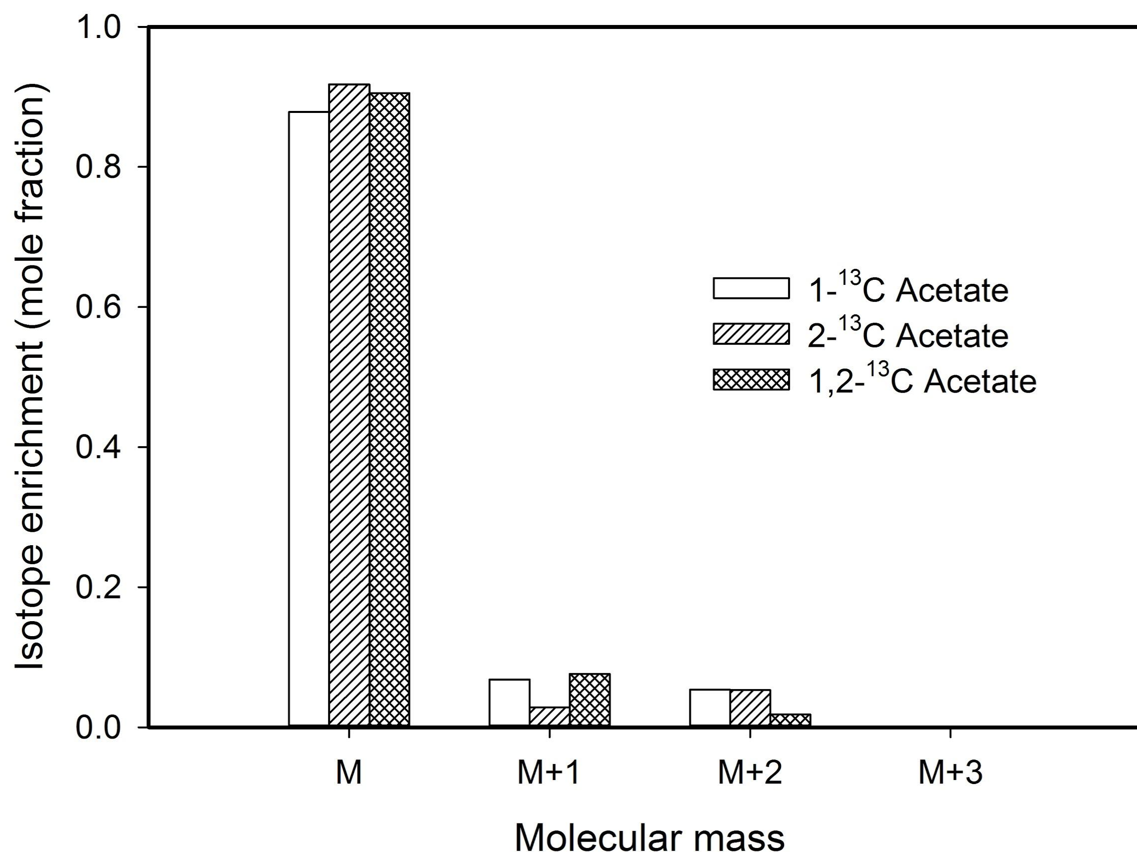
SI Appendix, Table S7. MdCMS peptides detected, identity (mascot score), and abundance ratio relative to day 0 in ethylene-treated (ET) and untreated (UT) apple fruit after 14 (D14) and 21 (D21) days holding at 20 °C. MdCMS was not detected in apple extracts from fruit held for 7 days. Protein search was against MdCMS_1.

			Abundance Ratio			
Target Gene	Peptides identified	Mascot score	$D14_{UT}/D0$	$D14_{ET}/D0$	$D21_{ET}/D0$	$D21_{UT}/D0$
MdCMS_1	GKDILGGLHTGINTK KFHEVFEHFK SLTNEDLESLVYQVAI SLGAQDITFV	102	0.90	0.90	0.69	0.46

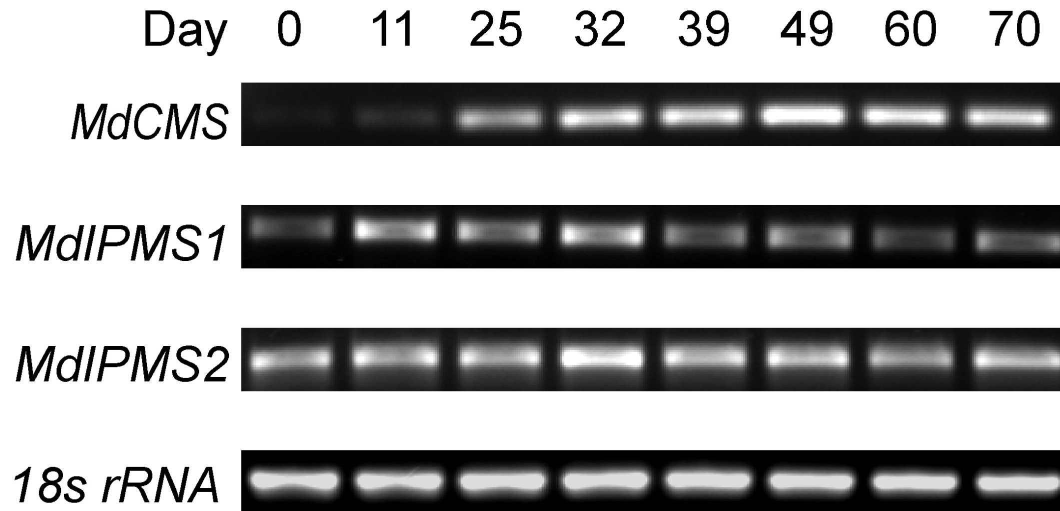


SI Appendix Figure S1. Mass isotopolog distribution of butanal and methylbutanoate from apple discs fed with 1-[^{13}C] acetate, 2-[^{13}C] acetate, and 1,2-[$^{13}C_2$] acetate. The isotope distribution (in mole fraction) is expressed as unlabeled mass (M) and one mass unit heavier than the unlabeled mass (M+1) up to 4 mass units heavier (M+4) than the unlabeled compound. For methylbutanoate, *m/z* 74, a product of the McLafferty rearrangement containing carbons C-1 and

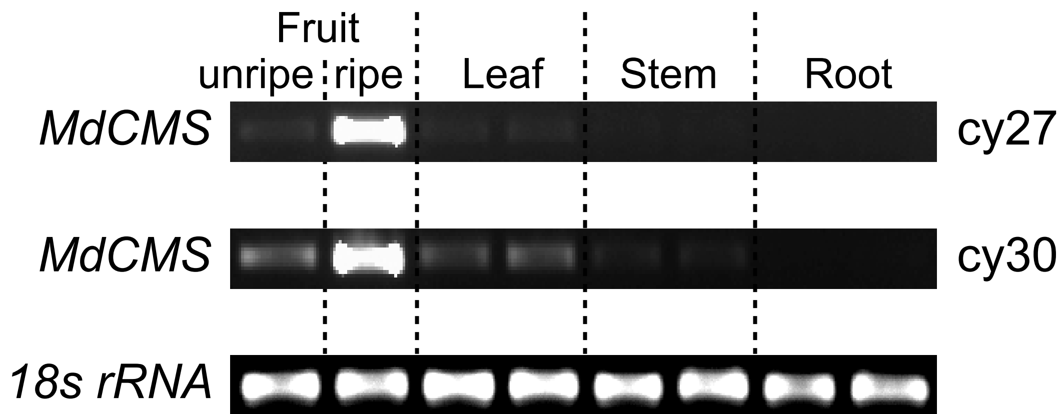
C-2, was used as M. For butanal, the molecular ion *m/z* 72 was used. Vertical bars represent standard deviation of the mean of two biological replications.



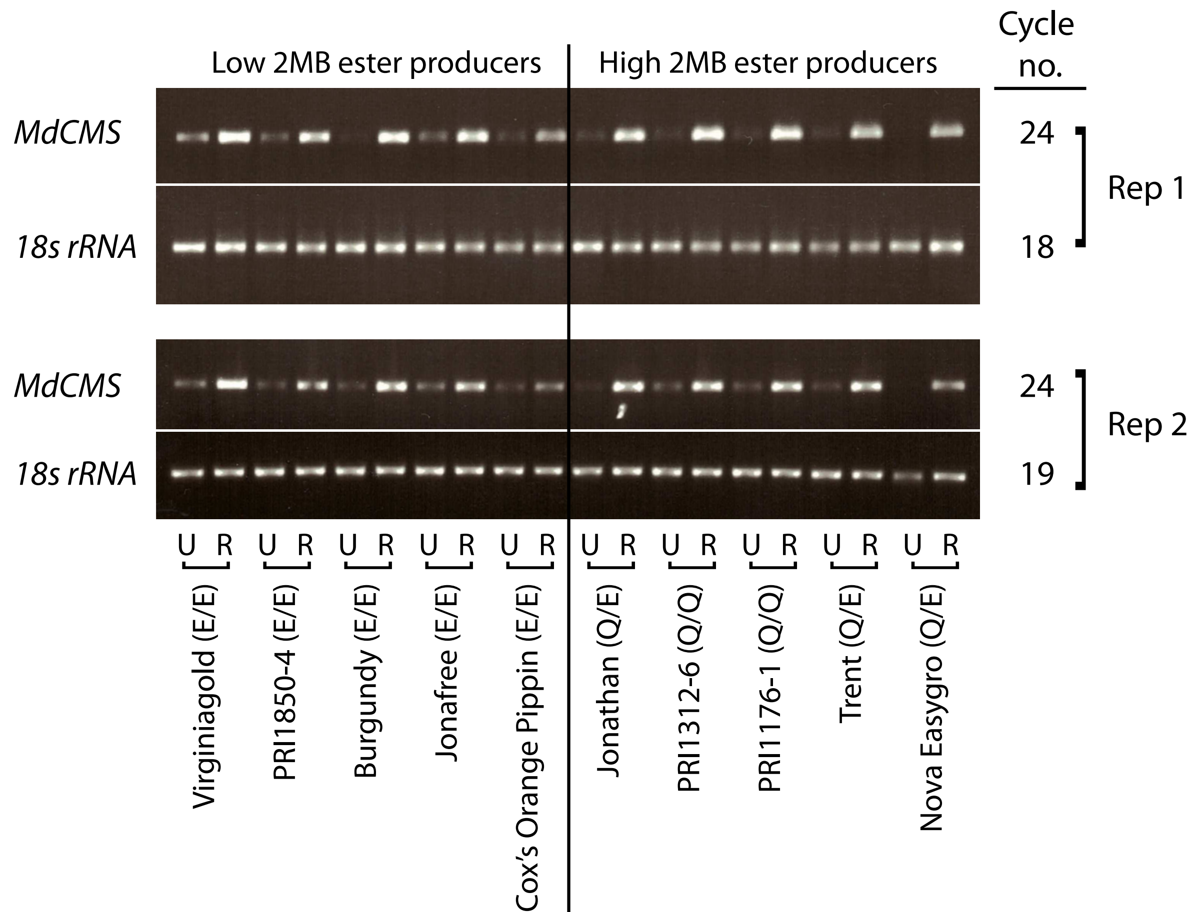
SI Appendix Figure S2. Mass isotopolog distribution of threonine from apple discs fed with 1-[^{13}C] acetate, 2-[^{13}C] acetate, and 1,2-[$^{13}C_2$] acetate. The isotope distribution (in mole fraction) is expressed as unlabeled mass (M) and one mass unit heavier than the unlabeled mass (M+1) up to 3 mass units heavier (M+3) than the unlabeled compound. Fragment ion *m/z* 404 was used as M, the molecular mass of the *tert*-butyldimethylsilyl derivatives minus 57 Da, corresponding to loss of the *tert*-butyl group from the derivative. Due to low instrument response for threonine, only data for replicate 1 are given.



SI Appendix, Figure S3. Semiquantitative RT-PCR analysis of the apple citramalate synthase (designated *MdCMS*, but likely a combination of *MdCMS_1* and *MdCMS_2*) and 2-isopropylmalate synthase (*MdIPMS1* and *MdIPMS2*) expression for 'Jonagold' apple fruit during ripening. Eight time points (day 0, 11, 25, 32, 39, 49, 60, and 70) were selected based on distinct physiological stages as fruit ripened from 2 Sept. 2004 (day 0) to 12 Nov. 2004 (day 70). Total RNA was isolated from fruit at each time point and 18s rRNA was used as a control (20 PCR cycles). The optimum PCR cycles were 27, 29, and 33 for *MdCMS*, *MdIPMS1*, and *MdIPMS2* respectively. RT-PCR analysis was repeated and essentially identical results were obtained.



SI Appendix, Figure S4. Semiquantitative RT-PCR analysis of apple citramalate synthase (designated *MdCMS*, but likely a combination of *MdCMS_1* and *MdCMS_2*) gene expression in different organs of 'Jonagold' apple. Samples were from unripe (day 0) and ripe (day 49) fruit peel tissues and leaf, stem, and root tissues. *18s rRNA* was used as a control (20 PCR cycles). Two biological replicates were performed for leaf stem and root tissue, shown as two lanes. Data are presented for 27 and 30 PCR cycles (cy27 and cy30, respectively) to better reveal low levels of expression for unripe fruit and leaf tissue.



SI Appendix, Figure S5. Semiquantitative RT-PCR analysis of apple citramalate synthase (designated *MdCMS*, but likely a combination of *MdCMS_1* and *MdCMS_2*) gene expression in unripe (U) and ripe (R) fruit of 10 apple lines characterized as having high or low 2-methylbutanol- and 2-methylbutanoate derived (2MB) ester content. *18s rRNA* was used as a control. Two biological replicates, each a pooled sample of four fruit, were analyzed and the number of PCR cycles is provided. The identity of amino acid 387 for the two *MdCMS* allele translation products is given; glutamine^{387} (Q) is characteristic of the active MdCMS_1 isozyme and glutamate^{387} (E) is indicative of the essentially inactive MdCMS_2 isozyme.

```
          1         10        20        30        40        50        60        70        80        90        100       110       120       130
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  ATGGCCTTCTCAGCAGAAAATCTTCTTTTCATCGCATCAACATCCCCAGTTTCCATAATCAGACATTCAAAACCCACAAACTCCAACATAGACACCACCCTCTTGCTCAAACCCTCATCACTCTTCCCTC
CMS2mRNA  ATGGCCTTCTCAGCAGAAAATCTTCTTTTCATCGCATCAACATCCCCAGTTTCCATAATCAGACATTCAAAACCCACAAACTCCAACATAGACACCACCCTCTTGCCCAAACCCTCATCACTCTTCCCTC
CMS1gDNA  ATGGCCTTCTCAGCAGAAAATCTTCTTTTCATCGCATCAACATCCCCAGTTTCCATAATCAGACATTCAAAACCCACAAACTCCAACATAGACACCACCCTCTTGCTCAAACCCTCATCACTCTTCCCTC
CMS2gDNA  ATGGCCTTCTCAGCAGAAAATCTTCTTTTCATCGCATCAACATCCCCAGTTTCCATAATCAGACATTCAAAACCCACAAACTCCAACATAGACACCACCCTCTTGCCCAAACCCTCATCACTCTTCCCTC
Consensus ATGGCCTTCTCAGCAGAAAATCTTCTTTTCATCGCATCAACATCCCCAGTTTCCATAATCAGACATTCAAAACCCACAAACTCCAACATAGACACCACCCTCTTGCcCAAACCCTCATCACTCTTCCCTC

          131       140       150       160       170       180       190       200       210       220       230       240       250       260
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  TCTCACAACAACGCGCCATCTTGAGGACAAACAAGCGCTGCTCCGCTCTCTCCCAATCCCCCACTCACCCCGAATACATTCCCAACCGAATCCCAAACGAAAACTATGTACGAGTCGTGGACACGACCCT
CMS2mRNA  TCTCACAACAACGCGCCATCTTGAGGACAAACAAGCGCTGCTCCGCTCTCTCCCAATCCCCCACTCACCCCGAATACATTCCCAACCGAATCCCAAACGAAAACTATGTACGAGTCGTGGACACGACCCT
CMS1gDNA  TCTCACAACAACGCGCCATCTTGAGGACAAACAAGCGCTGCTCCGCTCTCTCCCAATCCCCCACTCACCCCGAATACATTCCCAACCGAATCCCAAACGAAAACTATGTACGAGTCGTGGACACGACCCT
CMS2gDNA  TCTCACAACAACGCGCCATCTTGAGGACAAACAAGCGCTGCTCCGCTCTCTCCCAATCCCCCACTCACCCCGAATACATTCCCAACCGAATCCCAAACGAAAACTATGTACGAGTCGTGGACACGACCCT
Consensus TCTCACAACAACGCGCCATCTTGAGGACAAACAAGCGCTGCTCCGCTCTCTCCCAATCCCCCACTCACCCCGAATACATTCCCAACCGAATCCCAAACGAAAACTATGTACGAGTCGTGGACACGACCCT

          261       270       280       290       300       310       320       330       340       350       360       370       380       390
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  ACGGGACGGAGAACAAGCTGCGGGGGCGTCCATGACTCGGCTAGAAAAGCTGGCCATTGCACACCAGCTGGCGAAGCTAAGAGTGGATGTGATCGAGGCGGGGTTCCCGGCTTCGTCCAAATATGATTCG
CMS2mRNA  ACGGGACGGAGAACAAGCTGCGGGGGCGTCCATGACTCGGCTAGAAAAGCTGGCCATTGCACACCAGCTGGCGAAGCTAAGAGTGGATGTGATCGAGGCGGGGTTCCCGGCTTCGTCCAAATATGATTCG
CMS1gDNA  ACGGGACGGAGAACAAGCTGCGGGGGCGTCCATGACTCGGCTAGAAAAGCTGGCCATTGCACACCAGCTGGCGAAGCTAAGAGTGGATGTGATCGAGGCGGGGTTCCCGGCTTCGTCCAAATATGATTCG
CMS2gDNA  ACGGGACGGAGAACAAGCTGCGGGGGCGTCCATGACTCGGCTAGAAAAGCTGGCCATTGCACACCAGCTGGCGAAGCTAAGAGTGGATGTGATCGAGGCGGGGTTCCCGGCTTCGTCCAAATATGATTCG
Consensus ACGGGACGGAGAACAAGCTGCGGGGGCGTCCATGACTCGGCTAGAAAAGCTGGCCATTGCACACCAGCTGGCGAAGCTAAGAGTGGATGTGATCGAGGCGGGGTTCCCGGCTTCGTCCAAATATGATTCG

          391       400       410       420       430       440       450       460       470       480       490       500       510       520
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  GAGACAGTGAAGCGAATAGCAAAAGAGGTTGGAAGCCGTGTTGACGAGTGCGGCTACGTGCCTGTCATATCTGCATATTGTAGGTGCGTTCGTAGCGACGTTGACGCTGCATGGGAGTCGGTGAAGGACG
CMS2mRNA  GAGACAGTGAAGCGAATAGCAAAAGAGGTTGGAAGCCGTGTTGACGAGTGCGGCTACGTGCCTGTCATATCTGCATATTGTAGGTGCGTTCGTAGCGACGTTGACGCTGCATGGGAGTCGGTGAAGGACG
CMS1gDNA  GAGACAGTGAAGCGAATAGCAAAAGAGGTTGGAAGCCGTGTTGACGAGTGCGGCTACGTGCCTGTCATATCTGCATATTGTAGGTGCGTTCGTAGCGACGTTGACGCTGCATGGGAGTCGGTGAAGGACG
CMS2gDNA  GAGACAGTGAAGCGAATAGCAAAAGAGGTTGGAAGCCGTGTTGACGAGTGCGGCTACGTGCCTGTCATATCTGCATATTGTAGGTGCGTTCGTAGCGACGTTGACGCTGCATGGGAGTCGGTGAAGGACG
Consensus GAGACAGTGAAGCGAATAGCAAAAGAGGTTGGAAGCCGTGTTGACGAGTGCGGCTACGTGCCTGTCATATCTGCATATTGTAGGTGCGTTCGTAGCGACGTTGACGCTGCATGGGAGTCGGTGAAGGACG

          521       530       540       550       560       570       580       590       600       610       620       630       640       650
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  CGACGCGCCCGCGCCTATGCATATTCATATCAACCAGCGAAATCCACATGAAGTATAAGCTGAACAAGACTGCCGATCAAGTTCTCGAACTCGCGAAGGAGTCGGTGAGGTATGCCAGAAGCCTTGGAGC
CMS2mRNA  CGACGCGCCCGCGCCTATGCATATTCATATCAACCAGCGAAATCCACATGAAGTATAAGCTGAACAAGACTGCCGATCAAGTTCTCGAACTCGCGAAGGAGTCGGTGAGGTATGCCAGAAGCCTTGGAGC
CMS1gDNA  CGACGCGCCCGCGCCTATGCATATTCATATCAACCAGCGAAATCCACATGAAGTATAAGCTGAACAAGACTGCCGATCAAGTTCTCGAACTCGCGAAGGAGTCGGTGAGGTATGCCAGAAGCCTTGGAGC
CMS2gDNA  CGACGCGCCCGCGCCTATGCATATTCATATCAACCAGCGAAATCCACATGAAGTATAAGCTGAACAAGACTGCCGATCAAGTTCTCGAACTCGCGAAGGAGTCGGTGAGGTATGCCAGAAGCCTTGGAGC
Consensus CGACGCGCCCGCGCCTATGCATATTCATATCAACCAGCGAAATCCACATGAAGTATAAGCTGAACAAGACTGCCGATCAAGTTCTCGAACTCGCGAAGGAGTCGGTGAGGTATGCCAGAAGCCTTGGAGC

          651       660       670       680       690       700       710       720       730       740       750       760       770       780
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  TCAGGACATTACGTTTGTCTGTGAAGACGCTGGAAGGT--------------------------------------------------------------------------------------------
CMS2mRNA  TCAGGACATTACGTTTGTCTGTGAAGACGCTGGAAGGT--------------------------------------------------------------------------------------------
CMS1gDNA  TCAGGACATTACGTTTGTCTGTGAAGACGCTGGAAGGTGGTTAACTCTCTATCTCGTTCTTTCCAAAAACAATTTATATGAAACCGATTTACTATCACATAATGTTCTAATTTAATAAAGTAAAGATATG
CMS2gDNA  TCAGGACATTACGTTTGTCTGTGAAGACGCTGGAAGGTGGTTAACTCTCTATCTCGTTCTTTCCAAAAACAATTTATATGAAACCGATTTACTATCACATAATGTTCTAATTTAATAAAGTAAAGATATG
Consensus TCAGGACATTACGTTTGTCTGTGAAGACGCTGGAAGGT............................................................................................

          781       790       800       810       820       830       840       850       860       870       880       890       900       910
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  ----------------------------------------------------------------------------------------------------------------------------------
CMS2mRNA  ----------------------------------------------------------------------------------------------------------------------------------
CMS1gDNA  TGAAGAGATACCTACACATCACATTATACTAAAAATAAGACTTTAGCCCTGAAATTTAACTTATTTCACATAGAACTTGACATAAATTTTTTAGTCGAATAAAACCCATTGACTAAATTCCACATGAGCA
CMS2gDNA  TGAAGAGATACCTACACATCACATTATACTAAAAATAAGACTTTAGCCCTGAAATTTAACTTATTTCACATAGAACTTGACATAAATTTTTTAGTCGAATAAAACCCATTGACTAAATTCCACATGAGCA
Consensus ..................................................................................................................................

          911       920       930       940       950       960       970       980       990       1000      1010      1020      1030      1040
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  ----------------------------------------------------------------------------------------------------------------------------------
CMS2mRNA  ----------------------------------------------------------------------------------------------------------------------------------
CMS1gDNA  AATTCATTAACACTATTTGGATGAGGAAAATAAACTTAAAATTTGAATAAAAGTCAGAATTTATAAATTGACATACACCAATTTCCTTGTTTGGATTCATAAATATAGAAATTTAGAATTTCGGCATGGA
CMS2gDNA  AATTCATTAACACTATTTGGATGAGGAAAATAAACTTAAAATTTGAATAAAAGTCAGAATTTATAAATTGACATACACCAATTTCCTTGTTTGGATTCATAAATATAGAAATTTAGAATTTCGGCATGGA
Consensus ..................................................................................................................................

          1041 1050      1060      1070      1080      1090      1100      1110      1120      1130      1140      1150      1160      1170
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  ----------------------------------------------------------------------------------------------------------------------------------
CMS2mRNA  ----------------------------------------------------------------------------------------------------------------------------------
CMS1gDNA  AAAAAAAATTGGAATTTGGGACCTTCAATTCCTCAGTTTAAATTTTATGTAAATATGTGTAATTTCCCAATTTCTATGATTGAAAGTTTAAAAATAAAAAATTCCGCATTCAATTTCATTGTTCTTTTCC
CMS2gDNA  AAAAAAAATTGGAATTTGGGACCTTCAATTCCTAAGTTTAAATTTTATGTAAATATGTGTCATTTCCCAATTTCTATGATTGAAAGTTTAAAAATAAAAAATTCCGCATTCAATTTCATTGTTCTTTTCC
Consensus ..................................................................................................................................

          1171 1180      1190      1200      1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  ----------------------------------------------------------------------------------------------------------------------------------
CMS2mRNA  ----------------------------------------------------------------------------------------------------------------------------------
CMS1gDNA  GTCATTTTAATATTCACAAATTCTAAAAAATAAAATCCCATCATTTCACCAAACAAGAAAATTCACAAATTCTAGAAAATAAAATCACGTCATTTCTTTATCCAAACATAGTGTAATTATGTTTGACCTT
CMS2gDNA  GTCATTTTAATATTCACAAATTCTAAAAAATAAAATCTCATCATTTCACCAAACAAGAAAATTCACAAATTCTAGAAAATAAAATCACGTCATTTCTTTATCCAAACATAGTGTAATTATGTTTGACCTT
Consensus ..................................................................................................................................

          1301 1310      1320      1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  ----------------------------------------------------------------------------------------------------------------------------------
CMS2mRNA  ----------------------------------------------------------------------------------------------------------------------------------
CMS1gDNA  ACAAATATTTAAAACAAAAATCAAAAGAATAATAGTCCACTATCCACCAAACTAATAGTTAGGCCATGTAAACCTGAGTATTGAAAGTCAACTATATCAGTATACTTCTTGTTCGAACTTTTGAATCCTC
CMS2gDNA  ACAAATATTTAAAACAAAAATCAAAAGAATAATAGTCCACTATCCACCAAACTAATAGTTAGGCCATGTAAACCTGAGTATTGAAAGTCAACTATATCAGTATACTTCTTGTTCGAACTTTTGAATCCTC
Consensus ..................................................................................................................................

          1431 1440      1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  -------------------------------------------------------------------------------------------------------CTGAGAAGGAGTTTCTATATCGAATAT
CMS2mRNA  -------------------------------------------------------------------------------------------------------CTGAGAAGGAGTTTCTATATCGAATAT
CMS1gDNA  TTATATAATTTAGACTAATTTAATGTAAACTGTCCCTCGTAAAAAAGAAGGATCGGGCTTAAAAACTAATTGATTATATGCTTTTAATTAATCTGAATTAGGTCTGAGAAGGAGTTTCTATATCGAATAT
CMS2gDNA  TTATATAATTTAGACTAATTTAATGTAAACTGTCCCTCGTAAAAAAGAAGGATCGGGCTTAAAAACTAATTGATTATATGCTTTTAATTAATCTGAATTAGGTCTGAGAAGGAGTTTCTATATCGAATAT
Consensus .......................................................................................................CTGAGAAGGAGTTTCTATATCGAATAT

          1561 1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680      1690
          |---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
CMS1mRNA  ATGGGGAAGCTATCAAGGCCGGAGCAACAACTTTAACATTCACAGACACTGTTGGCTACAATTTCCCCAGCGAAGTGGAGCAATTTGTTAAGGACTTGAAAGCCAATGTAATAGGGATCGAAAATGCAAT
CMS2mRNA  ATGGGGAAGCTATCAAGGCCGGAGCAACAACTTTAACATTCACAGACACTGTTGGCTACAATTTCCCCAGCGAAGTGGAGCAATTTGTTAAGGACTTGAAAGCCAATGTAATAGGGATCGAAAATGCAAT
CMS1gDNA  ATGGGGAAGCTATCAAGGCCGGAGCAACAACTTTAACATTCACAGACACTGTTGGCTACAATTTCCCCAGCGAAGTGGAGCAATTTGTTAAGGACTTGAAAGCCAATGTAATAGGGATCGAAAATGCAAT
CMS2gDNA  ATGGGGAAGCTATCAAGGCCGGAGCAACAACTTTAACATTCACAGACACTGTTGGCTACAATTTCCCCAGCGAAGTGGAGCAATTTGTTAAGGACTTGAAAGCCAATGTAATAGGGATCGAAAATGCAAT
Consensus ATGGGGAAGCTATCAAGGCCGGAGCAACAACTTTAACATTCACAGACACTGTTGGCTACAATTTCCCCAGCGAAGTGGAGCAATTTGTTAAGGACTTGAAAGCCAATGTAATAGGGATCGAAAATGCAAT
```

SI Appendix, Figure S6. Coding and genomic sequence alignment of *MdCMS_1* and *2*. Genomic sequences of *MdCMS* revealed eight introns within the coding sequence and that the two genes differed in length at the 4th intron. *MdCMS_1* and *2* gDNA differed at only two positions, bp107 and 2488, yielding differences in the amino acid sequence.

```
MdCMS_1     1 .MAFSAENLLFIASTSPVSIIRHSKPTNSN...IDTTLLLKPSSLFPLSQQRAILRTN.......KRCSALSQSP........THPEYIP
MdIPMS1     1 .MAAVCTNPKILPSPAATMSSVNIPNTSQSQLLFRSHLHKPKMPKFLVSRPNLHNTCN.....PHILCSQTD.NPKPT.....PRPDYIP
MdIPMS2     1 .MATVCTHPKILPTPAATVSSINIPKSSQSQLLFRSHLHNPKTPNFLVSPPNLHTTSNPHHLKPHILCSQTD.NPKPT.....RRPDYIP
AtIPMS1     1 MASSLLRNPNL..YSSTTITTTSFLPTFSS........KPTPISSSFRFQPSHRSISLRSQTLRLSCSISDPSPLPPHTPRRPRPEYIP
SlIPMS3     1 .MTSLCANYSFGNYISLHSKNNSVPPIKRN...................FQSSYD.......IRCSNIQ...............R
LiCMS       1 .MTKVETR..................................................................................
BjMAM1-A    1 MASSLLTSSGIIPTTGSNMVVRSFLPFGSV.....RLTRPYNNSSLLISCCSSVSKN.......AETSGTDLKTIVE.....RWPEYIP
BjMAM1-A                                                                                             T

MdCMS_1    72 NRIPNENYVRVVDTTLRDGEQAAGASMTRLEKLAIAH.QLAKLRVDVIEAGFPASSKYDSETVKRI.....AKEVGSRVDEC.GYVPVIS
MdIPMS1    79 NRISDPNYVRIFDTTLRDGEQSPGASLTSKEKLDIAR.QLAKLGVDIIEAGFPASSKDDAEAVKMI.....AKEVGNAVDKD.GYVPVIC
MdIPMS2    84 NRISDPNYVRIFDTTLRDGEQSPGASLTSTEKLDIAR.QLAKLGVDIIEAGFPASSKDDAEAVKMI.....AKEVGNAVDKD.GYVPVIC
AtIPMS1    81 NRISDPNYVRVFDTTLRDGEQSPGATLTSKEKLDIAR.QLAKLGVDIIEAGFPAASKDDFEAVKTI.....AETVGNTVDEN.GYVPVIC
SlIPMS3    44 GKLSDPNYVTIFDTTLRDGEQAPGASMTAKQKMKIAC.QLAKLGVDVIEVGFPAASHAEFDLVKLV.....AQKIGNNIDEE.GYVPMIC
LiCMS       8 .......LEILDVTLRDGEQTRGVSFSTSEKLNIAKFLLQKLNVDRVEIASARVSKGELETVQKIMEWAATEQLTERIEIL.GFV....
BjMAM1-A   73 NKLPDKNYVRVFDTTLRDGEQSPGAALTPPQKIEIAR.QLAKLRVDIMEVGFPVSSEEEFETIQTI.....AKTVGNEVDEETGYIPVIC
BjMAM1-A      T   TT          TT
                     β1   π1            α1                 β2    η1           α2           η2

MdCMS_1   155 AYCRCVRSDVDAAWESVKDATRPRLCIFISTSEIHMKYKLNKTADQVLELAKESVRYARSLGAQDITFVCED...AGRSEKEFLYRIYGE
MdIPMS1   162 GLSRCNRSDIQTAWDAVKYAKRPRIHTFIATSPIHLEYKLRKTKEQVIEIARNMVKFARELGCDDVEFSPED...AGRSEREFLYQILGE
MdIPMS2   167 GLSRCNKNDIQTAWDAVKYAKRPRIHTFIATSPIHLEYKLRKSKEQVIEIARNMVKFARELGCDDVEFSPED...AGRSEREFLYQILGE
AtIPMS1   164 GLSRCNKKDIERAWDAVKYAKRPRIHTFIATSDIHLKYKLKKTKAEVIEIARSMVRFARSLGCEDVEFSPED...AGRSEREYLYEILGE
SlIPMS3   127 GLARSTKEDIDRAWESLKYAKTPMIHMFIATSDMHMKYKLNMSREEVVERARSMVAYATSLGFEHVRFSLED...ATRSDKEFVYHIIEE
LiCMS      85 .........DGNKTVDWIKDSGAKVLNLLTKGSLHHLEKQLGKTPKEFFTDVSFVIEYAIKSGLK.INVYLEDWSNGFRNSPDYVKSLVEH
BjMAM1-A  157 VIARSKERDIKAAWESVKYAKRPRIVIFTSTSDIHLKYKLKMTREEVVDMVASSIRFAKSLGFEDIEFGCED...GGRSDKDYICTVFEE
              β3        α3             β4        α4              α5                β5        η3     α6

MdCMS_1   242 AIKAGATTLTFTDTVGYNFPSEVEQFVKDLKANVIGIENAILSMHCHNDFGLANANTIAGAYAGARQVEVTINGIGERAGNASLEEFVMA
MdIPMS1   249 VIKAGATTLNIPDTVGYNVPDEYSQLIADIKSNTPGIDNVIISTHCQNDLGLSTANTLAGACAGARQLEVTINGIGERAGNASLEEVVMT
MdIPMS2   254 VIKAGATTLNIPDTVGYNVPEEYRQLIADIKSNTPGIDNVIISTHCQNDLGLSTANTLAGACAGARQLEVTINGIGERAGNASLEEVVLT
AtIPMS1   251 VIKAGATTLNIPDTVGITLPSEFGQLITDLKANTPGIENVVISTHCQNDLGLSTANTLSGAHAGARQMEVTINGIGERAGNASLEEVVMA
SlIPMS3   214 VIKAGATCICVADTVGCNLPNEFAQLIVDIKANTLGIQNVVLAVHCHNDLGLATANTLAGICAGVRQVDVTINGIGERAGNASLEEIVMT
LiCMS     166 LSKEHIERIFLPDTLGVLSPEETFQGVDSLIQKYPDIH...FEFHGHNDYDLSVANSLQAIRAGVKGLHASINGLGERAGNTPLEALVTT
BjMAM1-A  244 AIKAGATTLACPDTVGINMPHEYGKLVRYIKANTPGIDDVIFSAHCHNDLGVATANTIAGICAGARQVEVTINGIGERSGNAPLEEVVMA
                          TT                                    TT                       TT
                     β6            α7             η4   β7           α8          β8   η5

MdCMS_1   332 VKTRGKDILGGLHTGINTKHIIATSKMVEEYSGLSVQPHKAIVGANIFSHASGIHQDGVLKNKSTYEIILAEDIGYVHSNDDGIVLGKHS
MdIPMS1   339 LNCRGEHVLGGLYTGINTKHIYVTSKMVEEYTGLHVQPHKAIVGANAFAHESGIHQDGMLKHKGTYEIISPEDIGYERSNEAGIVLGKLS
MdIPMS2   344 LNCRGEQVLGGLYTGINTKHIYVTSKMVEEYTGLHVQPHKAIVGANAFAHESGIHQDGMLKHKGTYEIISPEDIGYERSNEAGIVLGKLS
AtIPMS1   341 IKCRGDHVLGGLFTGIDTRHIVMTSKMVEEYTGMQTQPHKAIVGANAFAHESGIHQDGMLKHKGTYEIICPEEIGLERSNDAGIVLGKLS
SlIPMS3   304 IKYRGGEVLGGVYTGINTKYIFTTSNMVEEYSGLKLQPNKAIVGANAFAHESGIHQDGMLKNRGTYEFISAEDVGFIRATKHGIKLGKLS
LiCMS     253 IHDKSNS.....KTNINEIAITEASRLVEVFSGKRISANRPIVGEDVFTQTAGVHADGDKKGNLYANPILPERFGRKRS...YALGKLA
BjMAM1-A  334 LKCRGAFVMGGVYTRIDTRQIMATSKMVQEYTGLYVQPHKPIVGANCFVHESGIHQDGILKNRSTYEIISPEDVGVVKSQNSGIVLGKLS
                                                                                                TT
              α9             η6   α10                              α11       η7        η8       η9

MdCMS_1   422 GRHALKSRLLQLGHDLDEKKFHEVFEHFKSLAETKKSLTNEDLESLV...........................................
MdIPMS1   429 GRHALRNRLAELGYELEDDQLATVFEHFKAVAEQKKIITDADLGALVRDEV.FQPEVVWKLHDLQVTCGTLGLSTATVKLIDADGREHVA
MdIPMS2   434 GRHALRNRLAELGYELENDQLATIFERFKAVAEQKKSITDADLGALVRDEV.FQPEVVWKLHDLQVTCGTLGLSTATVKLIDADGREHVA
AtIPMS1   431 GRHALKDRLTKLGYQLDDEQLSTIFWRFKTVAEQKKRVTDADIIALVSDEV.FQPEAVWKLLDIQITCGTLGLSTATVKLADADGKEHVA
SlIPMS3   394 GRHALKNKMLELGYSFEEKQLGDLFWRFKSLAEGKKNITDDDLRALI...........................................
LiCMS     334 GKASISENVKQLGMVLSEVVLQKVLERVIELGDQNKLVTPEDLPFIIADVSGRTGEKVLTIKSCNIHSG.IGIRPHAQIELEYQGKIHKE
BjMAM1-A  424 GRHAVKGRLKELGYEISDEKLNEVFSRFRDLTKQKKRVTDDDLKALV....................TCG....................
                 α12          α13             α14

MdCMS_1       ..........................................................................................
MdIPMS1   518 CSVGTGPVDSAYKAVDLIVKEPVMLVEYSMNAVTEGNDAIATTRVVIRPENRRMVTHAHTGESVQRTFSGVAAGMDIVVSSVKAYIGALN
MdIPMS2   523 CSVGTGPVDSAYKAVDLIVKEPVTLVEYSMNAVTEGIDAIATTRVVIRLENSHTVTHAQTGESVQRTFSGIGAGMDIVVSSVKAYIGALN
AtIPMS1   520 CSIGTGPVDSAYKAVDLIVKEPATLLEYSMNAVTEGIDAIATTRVLIRGSNKYSSTNAITGEEVQRTFSGTGAGMDIVVSSVKAYVGALN
SlIPMS3       ..........................................................................................
LiCMS     423 ISEGDGGYDAFMNALTKITNR....LGISIPKLIDYEVRIPPGGKTDALVETRITWNKSLDLEEDQTFKTMGVHPDQTVAAVHATEKMLN
BjMAM1-A  474 ..................................DEIFSSDKLNGTDDNEINSNGY....................................

MdCMS_1   469 .................YQVAI
MdIPMS1   608 KMIGFNERSPTKIPAERTPVSA
MdIPMS2   613 KMIGFKERSPPKFPAERNKVSA
AtIPMS1   610 KMMDFKENSATKIPSQKNRVAA
SlIPMS3   441 .................LDETI
LiCMS     509 QIL..............QPWQI
BjMAM1-A  496 ...........VPAPQISSVV
```

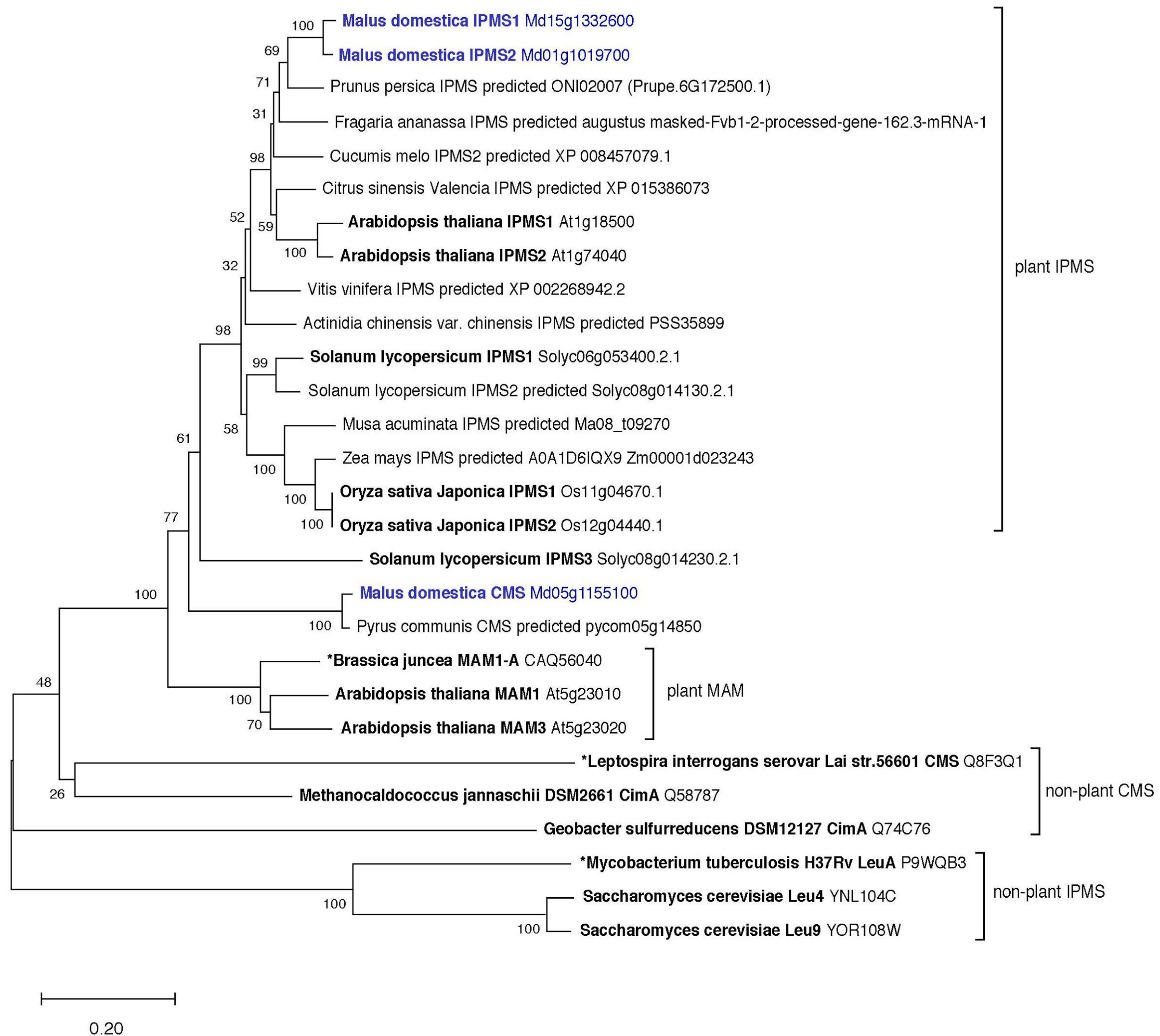
●● α-ketoacid binding sites
○ CoA binding sites
▼ Catalytic sites
★ Metal binding sites
! SNP MdCMS_1/MdCMS_2
↓ Chloroplast target peptide cleavage site

TIM barrel
Subdomain I
Subdomain II
R-region

SI Appendix, Figure S7. Amino acid sequence alignment of citramalate synthase (CMS) and 2-isopropylmalate synthase (IPMS) from apple (*Malus* × *domestica*), tomato (*Solanum lycopersicum*) IPMS3 (Solyc08g014230) (31), *Leptospira interrogans* CMS (UniProtKB-Q8F3Q1) (32), *Arabidopsis thaliana* IPMS1 (At1g18500) (33), and *Brassica juncea* methylthioalkylmalate synthase1-A (BjMAM1-A) (gi: CAQ56040.1) (34). Highlighted dark background shows perfectly conserved sequences across the species. Blocks of bolded residues indicate high conservation. Binding sites, catalytic sites, and secondary structure are derived from BjMAM1-A crystal structure (34). Solid oval symbols depict α-ketoacid binding sites implicated in substrate selectivity; red oval symbols are those shown to impact substrate size discrimination. Yellow highlighted and underlined MdCMS_1 sequences are peptides detected in proteomic analyses of ripening 'Golden Delicious' apple fruit (Table S7). Protein domain designations are derived from *Mycobacterium tuberculosis* IPMS crystal structure (35). Arrow indicates predicted cleavage site of plant chloroplast targeting peptides.

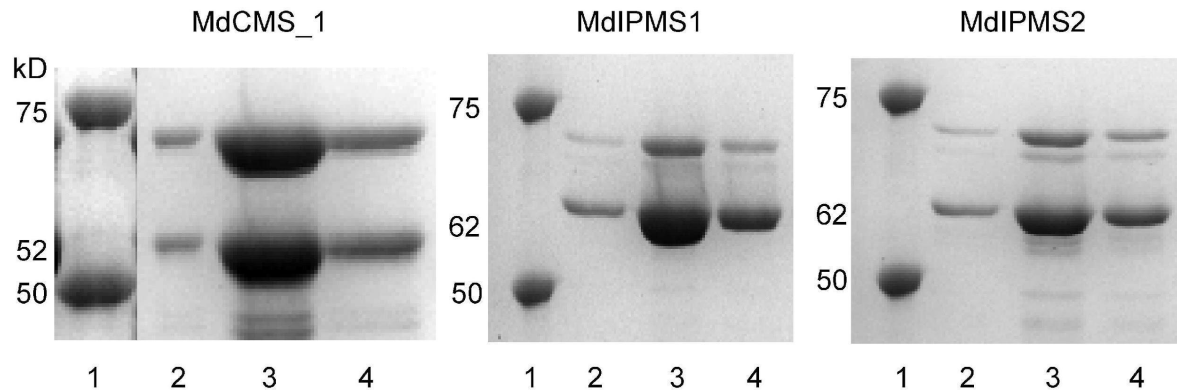
```
          1       10        20        30        40        50        60        70        80        90        100       110       120       130
MdIPMS1   ATGGCGGCTGTCTGCACAAACCCTAAAATTTTACCATCACCGGCAGCTACCATGTCCTCTGTCAACATCCCCAACACCTCCCAATCCCAGCTCCTCTTCCGCTCGCATTTGCACAAACCCAAGATGCCAA
MdIPMS2   ATGGCAACTGTCTGCACGCACCCTAAAATTTTACCAACACCGGCAGCCACCGTTTCCTCCATCAATATCCCCAAGAGCTCCCAATCCCAACTCCTCTTCCGCTCACATTTGCACAACCCCAAGACCCCCA
Consensus ATGGCaaCTGTCTGCACaaACCCTAAAATTTTACCAaCACCGGCAGCcACCaTgTCCTCcaTCAAcATCCCCAAcAcCTCCCAATCCCAaCTCCTCTTCCGCTCaCATTTGCACAAaCCCAAGAccCCaA

          131     140       150       160       170       180       190       200       210       220       230       240       250       260
MdIPMS1   AATTTCTAGTTTCCCGTCCCAATCTTCACAACACTTGTAATCCCCAC--------------ATTCTCTGCTCTCAAACTGACAACCCCAAACCTACTCCCCGACCTGATTACATCCCCAACCGCATCTC
MdIPMS2   ATTTTCTAGTTTCCCCTCCCAATCTCCACACCACCTCTAACCCCCACCACCTCAAGCCCCACATTCTCTGCTCTCAAACTGACAACCCCAAACCTACTCGCCGCCCTGATTACATCCCTAACCGCATCTC
Consensus AaTTTCTAGTTTCCCcTCCCAATCTcCACAaCACcTcTAAcCCCCAC..............ATTCTCTGCTCTCAAACTGACAACCCCAAACCTACTCcCCGaCCTGATTACATCCCcAACCGCATCTC

          261     270       280       290       300       310       320       330       340       350       360       370       380       390
MdIPMS1   TGACCCCAACTATGTCCGCATCTTCGACACCACTCTCCGCGACGGTGAGCAGTCCCCTGGTGCCTCCTTGACCTCAAAAGAAAAACTCGACATCGCCCGGAAGTTAGCAAAGCTTGGGGTTGACATAATT
MdIPMS2   CGACCCTAACTACGTCCGCATCTTCGACACCACTCTCCGCGACGGTGAGCAGTCCCCTGGTGCCTCCTTGACCTCGAAAGAGAAACTCGACATTGCTCGGCAGCTTGCAAAGCTTGGGGTTGACATAATC
Consensus cGACCCcAACTAcGTCCGCATCTTCGACACCACTCTCCGCGACGGTGAGCAGTCCCCTGGTGCCTCCTTGACCTCaAAAGAaAAACTCGACATcGCcCGGaAGcTaGCAAAGCTTGGGGTTGACATAATc

          391     400       410       420       430       440       450       460       470       480       490       500       510       520
MdIPMS1   GAGGCTGGTTTCCCCGCCTCTTCCAAAGACGATGCCGAGGCTGTGAAGATGATTGCAAAGGAGGTTGGGAATGCAGTTGACAAGGACGGTTATGTTCCTGTCATTTGTGGATTGTCAAGGTGCAATAGAA
MdIPMS2   GAGGCGGGTTTCCCAGCCTCTTCCAAAGACGATGCCGAGGCTGTGAAGATGATTGCAAAGGAGGTTGGGAATGCAGTTGACAAGGACGGTTATGTTCCTGTCATTTGTGGATTGTCACGGTGCAATAAAA
Consensus GAGGCgGGTTTCCCaGCCTCTTCCAAAGACGATGCCGAGGCTGTGAAGATGATTGCAAAGGAGGTTGGGAATGCAGTTGACAAGGACGGTTATGTTCCTGTCATTTGTGGATTGTCAaGGTGCAATAaAA

          521     530       540       550       560       570       580       590       600       610       620       630       640       650
MdIPMS1   ATGATATTCAGACGGCATGGGATGCTGTGAAGTACGCCAAAAGGCCAAGGATTCATACTTTTATTGCGACCAGTCCCATTCATTTGGAGTATAAACTGAGGAAGACCAAGGAGCAGGTAATTGAAATTGC
MdIPMS2   ACGATATTCAGACGGCATGGGATGCTGTGAAGTACGCCAAAAGGCCAAGGATTCATACTTTTATTGCCACCAGTCCCATTCATTTGGAGTACAAACTGAGGAAGAGCAAGGAGCAGGTGATTGAAATTGC
Consensus AcGATATTCAGACGGCATGGGATGCTGTGAAGTACGCCAAAAGGCCAAGGATTCATACTTTTATTGCcACCAGTCCCATTCATTTGGAGTAcAAACTGAGGAAGAcCAAGGAGCAGGTaATTGAAATTGC

          651     660       670       680       690       700       710       720       730       740       750       760       770       780
MdIPMS1   AAGGAACATGGTCAAGTTTGCCAGGGAGTTGGGATGCGATGATGTTGAGTTTAGCCCCGAAGATGCTGGGCGATCCGAAAGGGAGTTTCTATATCAGATTTTGGGTGAAGTTATAAAGGCCGGGGCAACA
MdIPMS2   GAGGAACATGGTCAAGTTTGCCAGGGAGTTGGGATGCGATGATGTTGAGTTTAGCCCCGAAGATGCCGGGCGATCCGAAAGGGAATTTCTATATCAGATTTTGGGTGAAGTTATAAAGGCCGGGGCAACA
Consensus aAGGAACATGGTCAAGTTTGCCAGGGAGTTGGGATGCGATGATGTTGAGTTTAGCCCCGAAGATGCcGGGCGATCCGAAAGGGAaTTTCTATATCAGATTTTGGGTGAAGTTATAAAGGCCGGGGCAACA

          781     790       800       810       820       830       840       850       860       870       880       890       900       910
MdIPMS1   ACCTTGAACATTCCTGACACTGTAGGTTATAATGTGCCAGATGAATATAGTCAGTTGATCGCTGACATAAAATCTAATACCCCTGGAATTGACAACATTATCATTTCTACTCACTGCCAAAATGATCTTG
MdIPMS2   ACTTTGAACATTCCTGACACCGTAGGTTATAATGTGCCAGAAGAATATCGTCAGTTGATTGCTGACATAAAATCTAATACCCCTGGAATTGACAACGTTATCATTTCTACTCACTGCCAAAATGATCTTG
Consensus ACcTTGAACATTCCTGACACcGTAGGTTATAATGTGCCAGAaGAATATaGTCAGTTGATcGCTGACATAAAATCTAATACCCCTGGAATTGACAACaTTATCATTTCTACTCACTGCCAAAATGATCTTG

          911     920       930       940       950       960       970       980       990       1000      1010      1020      1030      1040
MdIPMS1   GACTTTCTACTGCCAACACTTTAGCGGGGGCATGTGCAGGTGCTAGGCAACTGGAAGTAACAATCAATGGCATTGGTGAAAGGGCTGGGAATGCTTCATTAGAGGAGGTTGTCATGACCTTAAATTGTCG
MdIPMS2   GACTTTCTACTGCCAACACTTTAGCGGGGGCATGTGCAGGTGCTCGGCAATTGGAAGTAACAATCAATGGCATTGGTGAAAGGGCTGGGAATGCTTCATTAGAGGAGGTTGTCTTGACCTTAAATTGCCG
Consensus GACTTTCTACTGCCAACACTTTAGCGGGGGCATGTGCAGGTGCTaGGCAAcTGGAAGTAACAATCAATGGCATTGGTGAAAGGGCTGGGAATGCTTCATTAGAGGAGGTTGTCaTGACCTTAAATTGcCG

          1041    1050      1060      1070      1080      1090      1100      1110      1120      1130      1140      1150      1160      1170
MdIPMS1   CGGGGAGCATGTTCTTGGGGGGCTTTATACCGGAATCAATACTAAGCATATCTATGTAACAAGCAAGATGGTGGAAGAGTACACTGGGTTGCATGTGCAGCCACACAAGGCTATTGTTGGAGCTAATGCT
MdIPMS2   CGGGGAGCATGTTCTTGGGGGCCTTTATACCGGAATCAATACTAAGCATATCTATGTAACAAGCAAGATGGTGGAAGAGTACACTGGGTTGCATGTGCAGCCACACAAGGCTATTGTTGGAGCTAATGCT
Consensus CGGGGAGCATGTTCTTGGGGGcCTTTATACCGGAATCAATACTAAGCATATCTATGTAACAAGCAAGATGGTGGAAGAGTACACTGGGTTGCATGTGCAGCCACACAAGGCTATTGTTGGAGCTAATGCT

          1171    1180      1190      1200      1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
MdIPMS1   TTTGCACATGAAAGTGGTATCCATCAGGATGGAATGCTTAAGCACAAAGGTACATATGAAATCATATCTCCTGAAGATATTGGGTATGAACGGTCCAATGAAGCTGGTATTGTTCTTGGGAAACTCAGTG
MdIPMS2   TTTGCACATGAAAGTGGTATCCATCAGGATGGAATGCTTAAGCACAAAGGTACATACGAAATTATATCTCCTGAAGATATTGGGTATGAACGGTCCAATGAAGCTGGTATTGTTCTTGGGAAACTCAGTG
Consensus TTTGCACATGAAAGTGGTATCCATCAGGATGGAATGCTTAAGCACAAAGGTACATAcGAAATcATATCTCCTGAAGATATTGGGTATGAACGGTCCAATGAAGCTGGTATTGTTCTTGGGAAACTCAGTG

          1301    1310      1320      1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430
MdIPMS1   GGCGTCATGCTTTGAGAAATCGACTTGCGGAGCTTGGCTATGAGCTTGAGGATGATCAACTTGCTACTGTGTTCGAGCATTTCAAAGCTGTAGCCGAACAGAAAAAGATTATAACTGATGCAGATCTCGG
MdIPMS2   GACGTCATGCTTTGAGAAACCGACTTGCAGAGCTTGGCTATGAGCTTGAGAATGATCAACTTGCTACTATATTCGAGCGTTTCAAAGCTGTAGCCGAACAGAAAAAGAGTATAACTGATGCAGATCTCGG
Consensus GaCGTCATGCTTTGAGAAAcCGACTTGCaGAGCTTGGCTATGAGCTTGAGaATGATCAACTTGCTACTaTaTTCGAGCaTTTCAAAGCTGTAGCCGAACAGAAAAAGAgTATAACTGATGCAGATCTCGG

          1431    1440      1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560
MdIPMS1   AGCACTGGTGCGAGATGAAGTTTTTCAGCCAGAAGTTGTCTGGAAGCTTCACGATTTACAGGTTACCTGTGGAACTCTTGGTCTTTCTACAGCGACTGTTAAACTAATTGATGCTGATGGGAGAGAGCAT
MdIPMS2   AGCATTGGTGCGAGATGAAATTTTTCAGCCAGAAGTTGTCTGGAAGCTTCATGATCTACAGGTTACCTGTGGAACTCTTGGTCTGTCTACAGCAACTGTTAAACTAATTGATGCTGATGGGAGAGAACAT
Consensus AGCAcTGGTGCGAGATGAAaTTTTTCAGCCAGAAGTTGTCTGGAAGCTTCAcGATcTACAGGTTACCTGTGGAACTCTTGGTCTgTCTACAGCaACTGTTAAACTAATTGATGCTGATGGGAGAGAaCAT

          1561    1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680      1690
MdIPMS1   GTGGCATGTTCAGTTGGAACGGGTCCAGTAGATTCGGCTTACAAGGCTGTTGATCTCATTGTGAAGGAACCTGTAATGCTCGTTGAGTACTCTATGAATGCGGTCACAGAAGGAAATGATGCAATAGCAA
MdIPMS2   GTGGCATGTTCAGTTGGAACAGGTCCCGTAGATTCAGCTTACAAGGCTGTTGATCTGATTGTGAAGGAACCGGTAAAGCTGGTTGAGTACTCTATGAATGCGGTTACAGAAGGAATTGATGCAATTGCCA
Consensus GTGGCATGTTCAGTTGGAACaGGTCCaGTAGATTCaGCTTACAAGGCTGTTGATCTcATTGTGAAGGAACCgGTAAaGCTcGTTGAGTACTCTATGAATGCGGTcACAGAAGGAAaTGATGCAATaGCaA

          1691    1700      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800      1810      1820
MdIPMS1   CTACTCGTGTTGTAATCCGACCAGAAAACAGACGTATGGTTACTCATGCTCACACCGGAGAATCAGTTCAACGGACATTCAGTGGAGTTGCAGCAGGAATGGATATTGTTGTCTCTAGTGTCAAGGCCTA
MdIPMS2   CTACTCGTGTGGTAATCCGACTGGAAAACAGTCATACGGTTACTCATGCTCAAACTGGAGAATCAGTCCAGCGGACATTCAGTGGAATTGGGGCAGGAATGGATATTGTTGTCTCTAGTGTCAAGGCCTA
Consensus CTACTCGTGTgGTAATCCGACcaGAAAACAGaCaTAcGGTTACTCATGCTCAaACcGGAGAATCAGTcCAaCGGACATTCAGTGGAaTTGcaGCAGGAATGGATATTGTTGTCTCTAGTGTCAAGGCCTA

          1821    1830      1840      1850      1860      1870      1880      1890      1900 1905
MdIPMS1   CATCGGCGCATTGAATAAGATTATAGGTTTCAACAAAAGGTCGCCGACAAAGATTCCAGCGGAACTAACGCCGGTGTCTGCATGA
MdIPMS2   CATCGGTGCATTGAATAAGATGATAGGTTTCAAGGAAAGGTCGCCGCCAAAGTTTCCAGCTGAACGAAACAAAGTGTCTGCATGA
Consensus CATCGGcGCATTGAATAAGATgATAGGTTTCAAcaAAAGGTCGCCGaCAAAGaTTCCAGCgGAACgAAacaaaGTGTCTGCATGA
```

SI Appendix, Figure S8. Coding sequence alignment of *MdIPMS1* and *2*. *MdIPMS1* and *MdIPMS2*, had coding sequences of 1890 bp and 1905 bp, respectively.

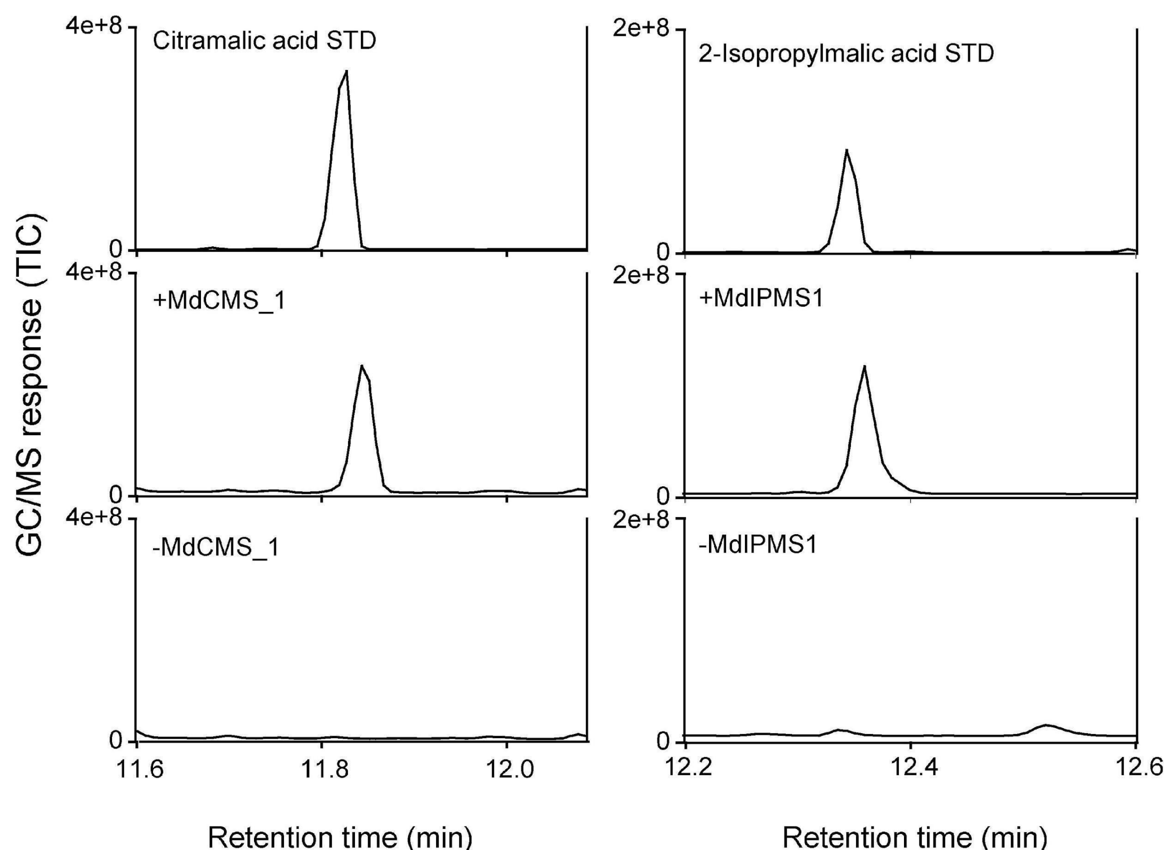


SI Appendix, Figure S9. Phylogenetic tree for 2-isopropylmalate synthase (IPMS), methylthioalkylmalate synthase (MAM), and citramalate synthase (CMS) sequences from selected bacteria, yeast, and plants. Bold print indicates characterized proteins and an asterisk indicates the protein was subjected to crystallographic analysis. Blue print indicates this paper. Where available, the NCBI (plants), Uniprot (bacteria and yeasts), Genbank accession numbers, and annotation are indicated along with species. Sequences were aligned using MUSCLE (25), and the tree was using the Neighbor-Joining method (27). The percentage of replicate trees in

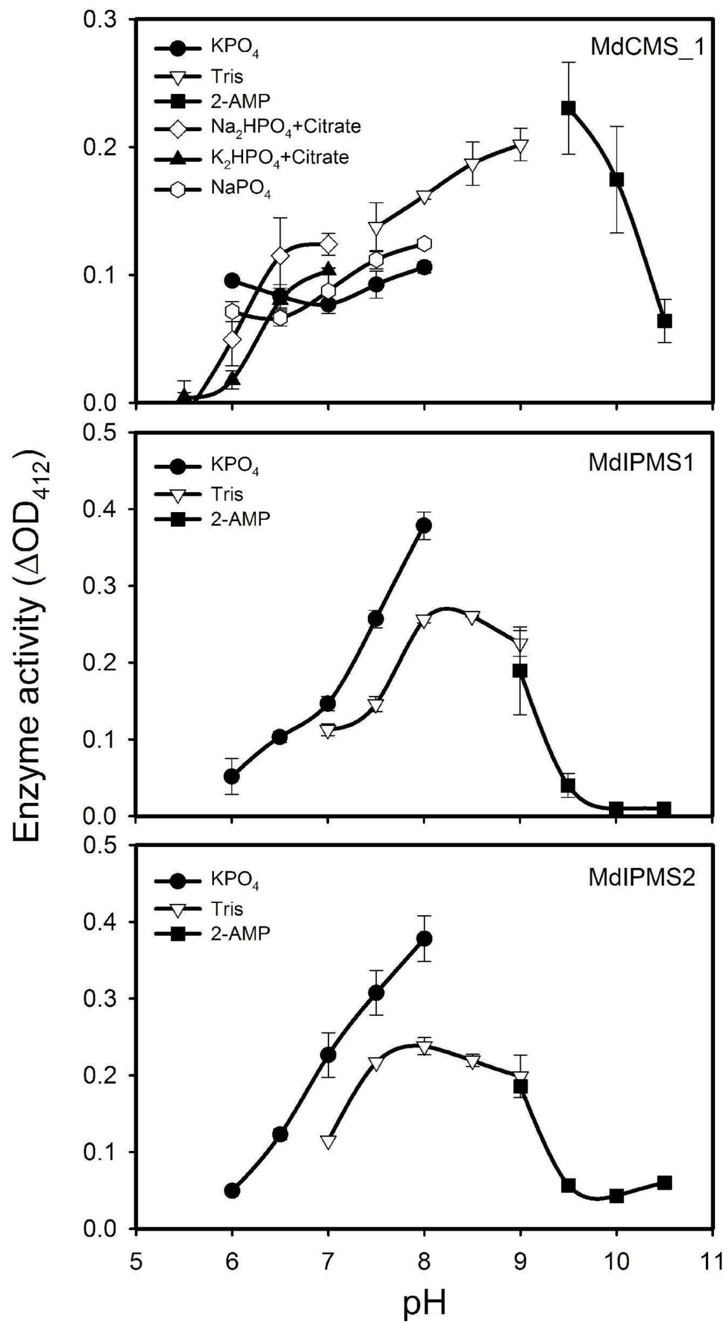
which the associated taxa clustered together in the bootstrap test (500 replicates) is shown next to the branches (28). The evolutionary distances were computed using the Poisson correction method (29) in MEGA X software (30) and indicate the number of amino acid substitutions per site. Signal sequences, the R-region, alignment gaps, and missing data were eliminated.



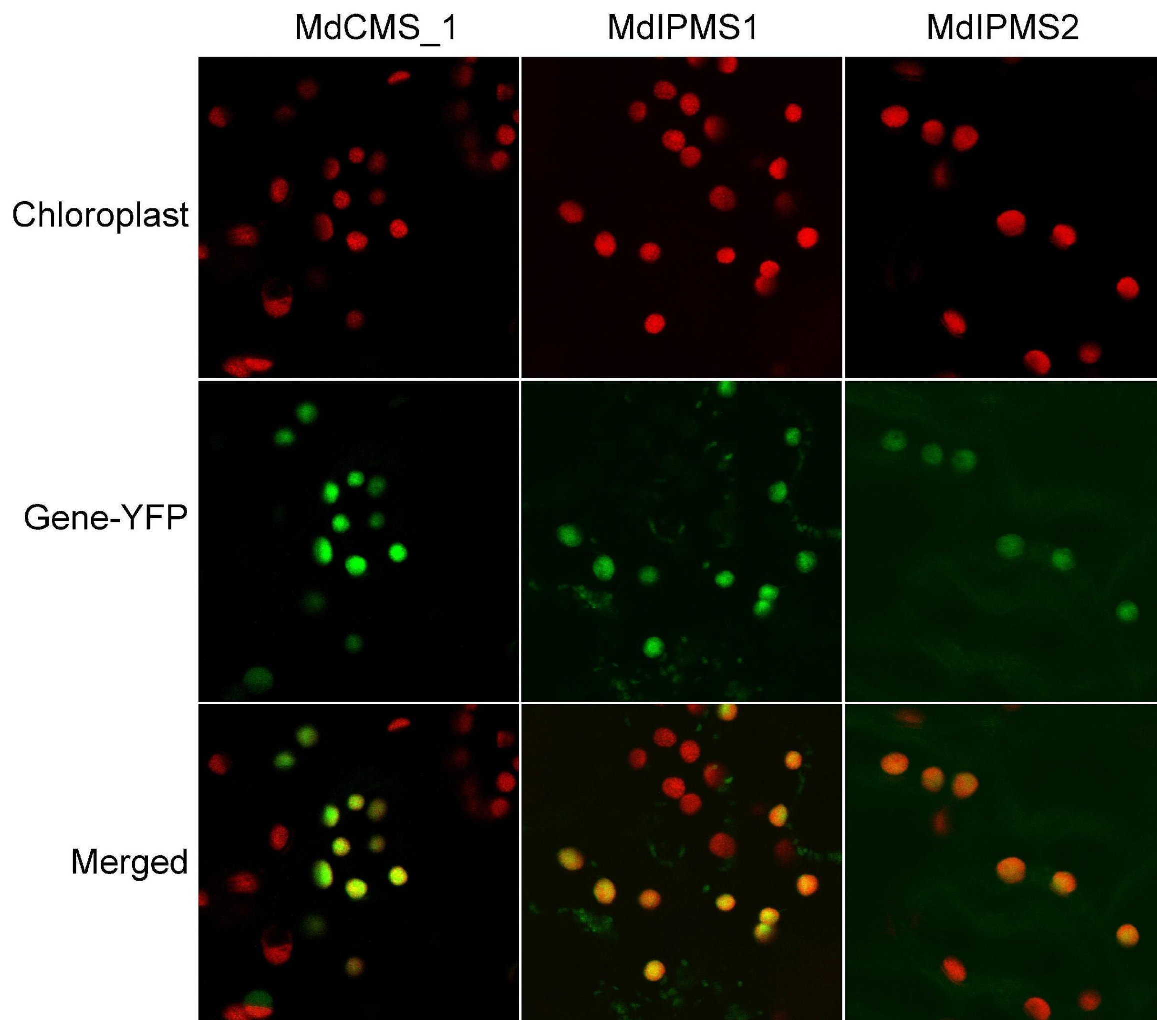
SI Appendix, Figure S10. Protein size of citramalate synthase (MdCMS_1) and 2-isopropylmalate synthase (MdIPMS1 and 2) as determined by electrophoresis on a 10% SDS-PAGE gel and stained using Coomassie Brilliant Blue R-250. MdCMS_1 was expressed as full length ORF and MdIPMS1 and 2 were truncated, being expressed without the chloroplast targeting sequence. The predicted protein size excluding 6x His-tag (0.8 kD) was around 52 kD for full length MdCMS_1 and 62 kD for truncated MdIPMS1 and 2. Lane 1 on each gel contains 50 and 75 kD protein standards for size estimation. Lanes 2 to 4 are replications of purified proteins differing in concentration during the protein purification process. The upper bands on the gel are unknown non-specific proteins from the BL21(DE3) *E. coli* preparation.



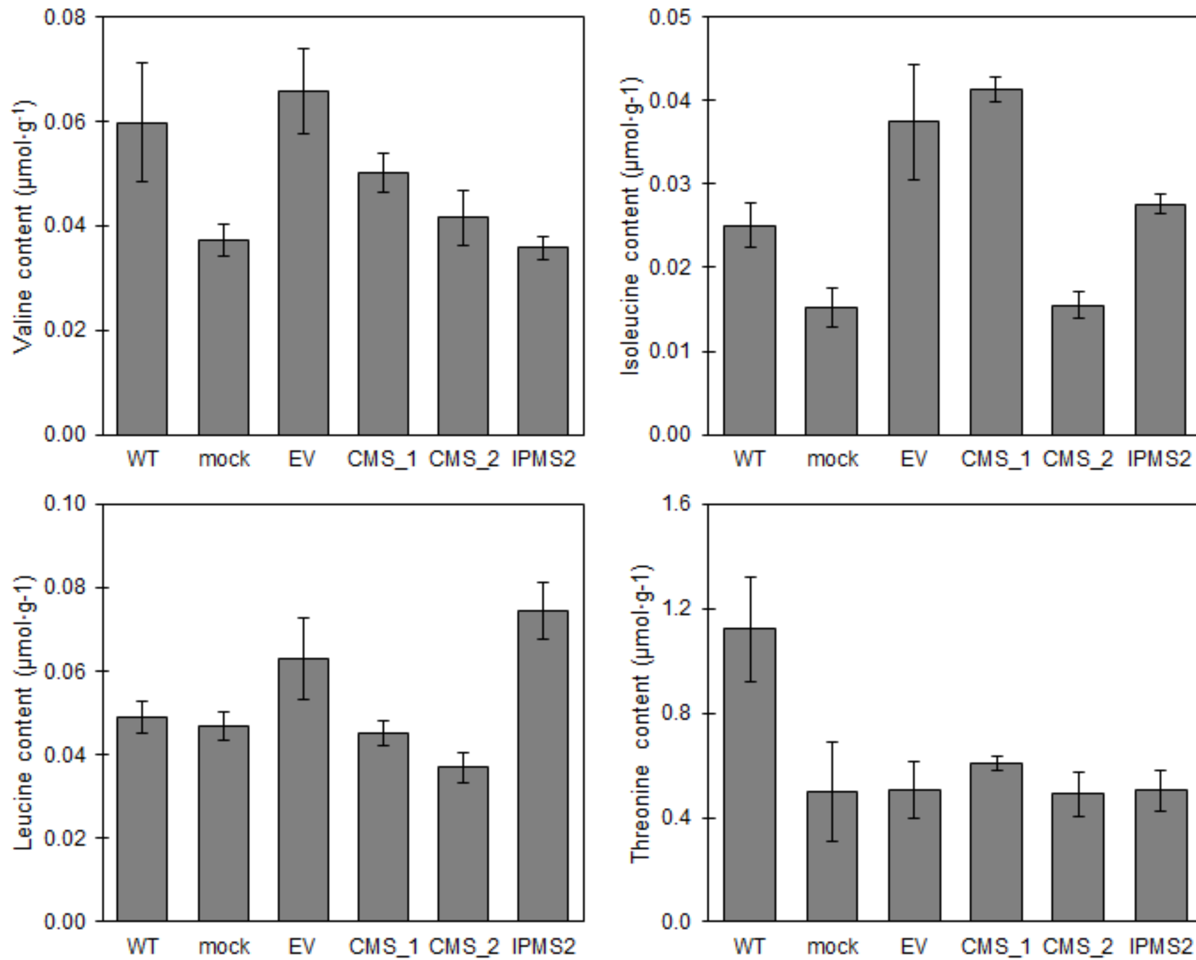
SI Appendix, Figure S11. Gas chromatography coupled with mass spectrometry (GC/MS) analysis of the reaction products of citramalate synthase (MdCMS_1, left) using acetyl-CoA (250 µM) in combination with pyruvate (10 mM) as substrates and 2-isopropylmalate synthase (MdIPMS1, right) with acetyl-CoA (250 µM) and α-ketoisovalerate (10 mM) as substrates following incubation at 22 °C for 120 min. Data are expressed as total ion current (TIC) chromatograms. Assay mixtures and authenticated standards were derivatized and analyzed by GC/MS and identity was verified by matching retention time and mass spectra. Upper graphs are the chromatograms of a standard for the expected assay product, the middle graphs are the chromatograms of the actual enzyme assay product, and the bottom graphs are chromatograms of the assay mixture without added enzymes.



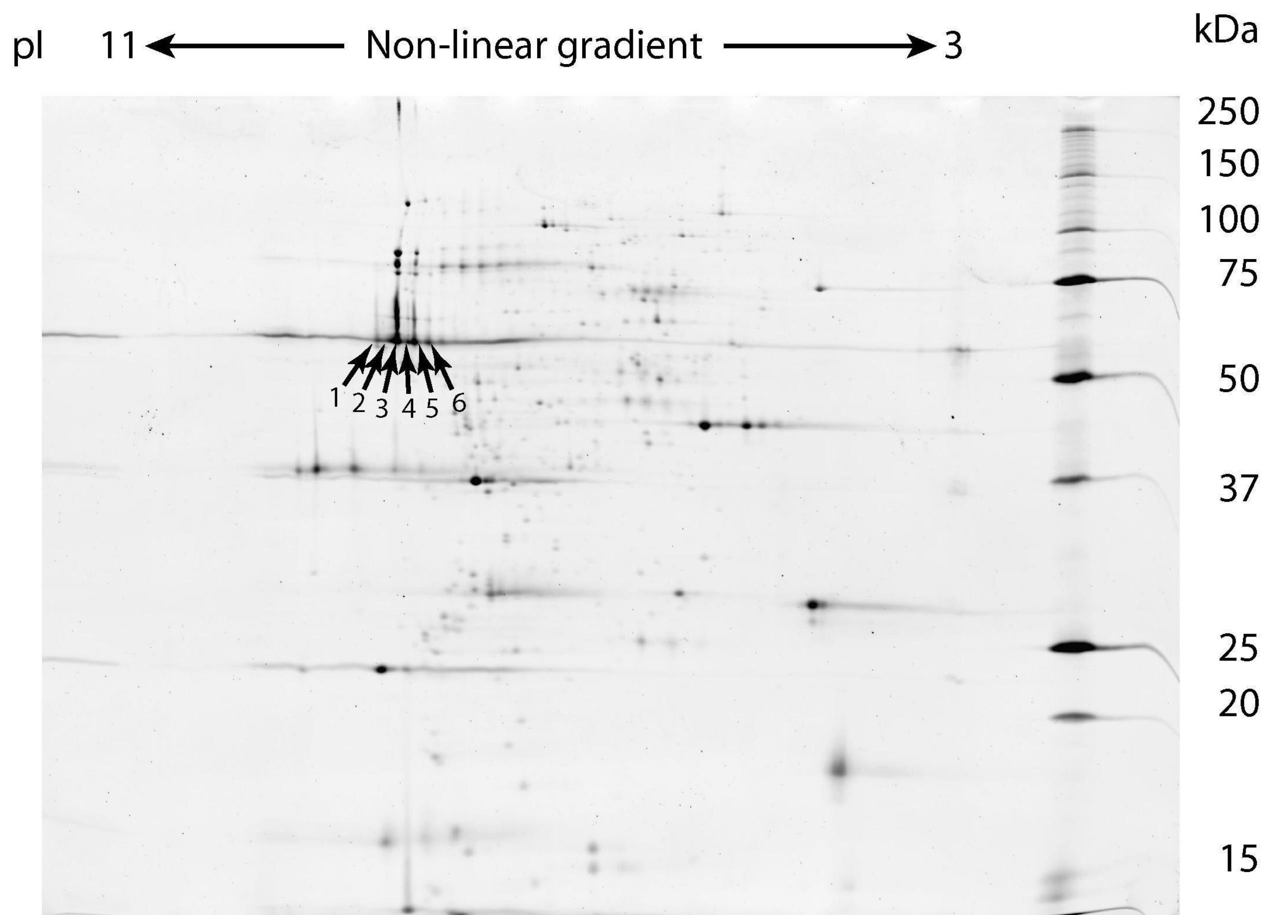
SI Appendix, Figure S12. Activity of citramalate synthase (MdCMS_1) and 2-isopropylmalate synthase (MdIPMS1 and 2) as affected by pH. Pyruvate (10 mM) was used as a substrate for MdCMS_1 and α-ketoisovalerate (10 mM) was used for the two MdIPMS proteins. Acetyl-CoA was kept constant at 250 µM. Assays were repeated twice for each data point. Vertical bar represents ± SD of the mean.



SI Appendix, Figure S13. Transient expression of MdCMS_1 and MdIPMS1 and 2 in tobacco (*Nicotiana tabacum* cv. Samson). Proteins were fused with YFP and infiltrated into tobacco leaves; the infiltrated areas were analyzed after three days by confocal microscopy. Top, middle and lower panels are chloroplast, YFP-protein fusion product, and overlay images.



SI Appendix, Figure S14. Accumulation of valine, leucine, isoleucine, and threonine in response to transient expression of *MdCMS_1* (CMS_1), *MdCMS_*2 (CMS_2), and *MdIPMS*2 (IPMS2) in transfected leaves of *N. benthamiana*. Controls include transient expression of an empty vector (EV), mock infiltration with buffer (mock), and wild type (WT). Whole transfected, mock, and WT leaves for three individual plants were sampled. Vertical bars represents ± SD of the mean. There were no statistically significant differences in means ($P < 0.05$, LSD).



SI Appendix, Figure S15. 2-DE analysis of purified MdCMS protein extracts: 100 micrograms of protein was dissolved in DeStreak solution. The 2-DE indicated that MdCMS has an isoelectric point between 6.75 to 7.1 and a molecular mass of 58.3 kDa; these are in agreement with the predicted isoelectric point and mass based on nucleotide sequence (6.79 IEP and 52.3 kDa, respectively). After isoelectrofocusing, proteins were further separated on SDS-PAGE (12.5%) polyacrylamide gels and visualized by Sypro Ruby staining. Arrows indicate six positions from which protein was extracted, digested, and subjected to LC-MS/MS protein fragment analysis.

SI Appendix, Citations

1. J. Song, B.D. Gardner, J.F. Holland, R.M. Beaudry, Rapid analysis of volatile flavor compounds in apple fruit using SPME and GC/time-of-flight mass spectrometry. *J. Agric. Food Chem*. 45, 1801-1807 (1997).

2. K. Biemann, "The mass spectra of isotopically labeled molecules" in Mass Spectrometry: organic chemical applications. (New York: McGraw-Hill, 1962), pp. 204-250.

3. N. Sugimoto, S. Park, S. van Nocker, R. Beaudry, Gene expression associated with apple aroma biosynthesis. *Acta Hortic*. 768, 57-64 (2008).

4. N.A. Mir, E. Curell, N. Khan, M. Whitaker, R.M. Beaudry, Harvest maturity, storage temperature, and 1-MCP application frequency alter firmness retention and chlorophyll fluorescence of 'Redchief Delicious' apples. *J. Am. Soc. Hortic. Sci*. 126, 618-624 (2001).

5. R. López-Gómez, M.A. Gómez-Lim, A method for extracting intact RNA from fruits rich in polysaccharides using ripe mango mesocarp. *HortScience*. 27, 440-442 (1992).

6. Z. Gao, S. Jayanty, R. Beaudry, W. Loescher, Sorbitol transporter expression in apple sink tissues: Implications for fruit sugar accumulation and watercore development. *J. Am. Soc. Hortic. Sci*. 130, 261-268 (2005).

7. E. Scotto-Lavino, G. Du, M.A. Frohman, 5' End cDNA amplification using classic RACE. *Nat. Protoc*. 1, 2555-2562 (2006).

8. N. Sugimoto, P. Forsline, R. Beaudry, Volatile profiles of members of the USDA Geneva *Malus* core collection: Utility in evaluation of a hypothesized biosynthetic pathway for esters derived from 2-methylbutanoate and 2-methylbutanol. *J. Agric. Food Chem*. 63, 2106-2116 (2015).

9. O. Emanuelsson, H. Nielsen, G. von Heijne, ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci*. 8, 978-984 (1999).

10. J.-W. de Kraker, K. Luck, S. Textor, J.G. Tokuhisa, J. Gershenzon, J. Two Arabidopsis genes (*IPMS1* and *IPMS2*) encode isopropylmalate synthase, the branchpoint step in the biosynthesis of leucine. *Plant Physiol*. 143, 970-986 (2007).

11. G.B. Kohlhaw, α-Isopropylmalate synthase from yeast. *Methods Enzymol*. 166, 414-429 (1988).

12. K.W. Earley, J.R. Haag, O. Pontes, K. Opper, T. Juehne, K. Song, C.S. Pikaard, Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J*. 45, 616-629 (2006).

13. S. Reumann, S. Quan, K. Aung, P. Yang, K. Manandhar-Shrestha, D. Holbrook, N. Linka, R. Switzenberg, C.G. Wilkerson, A.P.M. Weber, L.J. Olsen, J. Hu, In-depth proteome analysis of Arabidopsis leaf peroxisomes combined with in vivo subcellular targeting verification indicates novel metabolic and regulatory functions of peroxisomes. *Plant Physiol*. 150, 125-143 (2009).

14. J. Luo, X. Deng, C. Buehl, X. Xu and M.H. Kuo, Identification of tension sensing motif of histone H3 in Saccharomyces cerevisiae and its regulation by histone modifying enzymes. *Genetics* 204, 1029-1043 (2016).

15. E. Casalone, C. Barberio, D. Cavalieri, M. Polsinelli, Identification by functional analysis of the gene encoding alpha-isopropylmalate synthase II (LEU9) in Saccharomyces cerevisiae. *Yeast* 16, 539-45 (2000).

16. F. Sherman, Getting started with yeast. *Methods Enzymol*. 194, 3–21 (1991).

17. D. Gietz, A. St Jean, R.A. Woods, R.H. Schiestl, Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res*. 20, 1425 (1992).

18. F. Sainsbury, E.C. Thuenemann, G.P. Lomonossoff, pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J*. 7, 682-693 (2009).

19. M.A. Bedewitz, A.D. Jones, J.C. D'Auria, C.S. Barry, Tropinone synthesis via an atypical polyketide synthase and P450-mediated cyclization. *Nat. Commun*. 9, 5281 (2018).

20. Q. Zheng, J. Song, K. Doncaster, E. Rowland, D.M. Byers, Qualitative and quantitative evaluation of protein extraction protocols for apple and strawberry fruit suitable for two-dimensional electrophoresis and mass spectrometry analysis. *J. Agric. Food Chem*. 55, 1663-1673 (2007).

21. J. Song. F. Bangerth, The effect of harvest date on aroma compound production from 'Golden Delicious' apple fruit and relationship to respiration and ethylene production. *Postharvest Biol. Technol*. 8, 259-269 (1996).

22. Q.F. Zheng, J. Song, L. Campbell-Palmer, K. Thompson, L. Li, B. Walker, Y.S. Cui, X.H. Li, A proteomic investigation of apple fruit during ripening and in response to ethylene treatment. *J. Proteomics*. 93, 276-294 (2013).

23. Zheng Q., Song J., Doncaster K., Rowland E., Byers D.M. Qualitative and quantitative evaluation of protein extraction protocols for apple and strawberry fruit suitable for two-dimensional electrophoresis and mass spectrometry analysis. *J. Ag. Food Chem*. 55:1663-73 (2007).

24. L. Li, J. Song, W. Kalt, C. Forney, R. Tsao, D. Pinto, K. Chisholm, L. Campbell, S. Fillmore, X. Li, Quantitative proteomic investigation employing stable isotope labeling by peptide

dimethylation on proteins of strawberry fruit at different ripening stages. *J. Proteomics* 94, 219–239 (2013).

25. R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. 5, 113 (2004).

26. P. Gouet, E. Courcelle, D.I. Stuart, F. Métoz, ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*. 15, 305-308 (1999).

27. N. Saitou, M. Nei, The neighbor-joining method - A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. 4, 406-425 (1987).

28. J. Felsenstein, Confidence limits on phylogenies: An approach using the bootstrap. *Evolution.* 39, 783-791 (1985).

29. E. Zuckerkandl, L. Pauling, "Evolutionary divergence and convergence in proteins" in Evolving Genes and Proteins, V. Bryson, H.J. Vogel, Eds. (New York: Academic Press, 1965), pp. 97-166.

30. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol*. 35, 1547-1549 (2018).

31. J, Ning, G.D. Moghe, B. Leong, J. Kim, I. Ofner, Z. Wang, C. Adams, A.D. Jones, D. Zamir, R.L. Last, A feedback-insensitive isopropylmalate synthase affects acylsugar composition in cultivated and wild tomato. *Plant Physiol*. 169, 1821-1835 (2015).

32. H. Xu, Y. Zhang, X. Guo, S. Ren, A.A. Staempfli, J. Chiao, W. Jiang, G. Zhao, Isoleucine biosynthesis in *Leptospira interrogans* Serotype Lai strain 56601 proceeds via a threonine-independent pathway. *J. Bacteriol*. 186, 5400-5409 (2004).

33. J.-W. de Kraker, K. Luck, S. Textor, J.G. Tokuhisa, J. Gershenzon, J. Two Arabidopsis genes (*IPMS1* and *IPMS2*) encode isopropylmalate synthase, the branchpoint step in the biosynthesis of leucine. *Plant Physiol*. 143, 970-986 (2007).

34. R. Kumar, S.G. Lee, R. Augustine, M. Reichelt, D.G. Vassão, M.H. Palavalli, A. Allen, J. Gersehnzon, J.M. Jez, N.C. Bisht, Molecular basis of the evolution of methylthioalkylmalate synthase and the diversity of methionine-derived glucosinolates. *Plant Cell*. 31, 1633-1647 (2019).

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