




# The chloroalkaloid (–)-acutumine is biosynthesized via a Fe(II)- and 2-oxoglutarate-dependent halogenase in Menispermaceae plants

Colin Y. Kim <sup>1,2</sup>, Andrew J. Mitchell<sup>1</sup>, Christopher M. Glinkerman<sup>1</sup>, Fu-Shuang Li<sup>1</sup>, Tomáš Pluskal <sup>1</sup> & Jing-Ke Weng <sup>1,3</sup>✉

Plant halogenated natural products are rare and harbor various interesting bioactivities, yet the biochemical basis for the involved halogenation chemistry is unknown. While a handful of Fe(II)- and 2-oxoglutarate-dependent halogenases (2ODHs) have been found to catalyze regioselective halogenation of unactivated C–H bonds in bacteria, they remain uncharacterized in the plant kingdom. Here, we report the discovery of dechloroacutumine halogenase (DAH) from Menispermaceae plants known to produce the tetracyclic chloroalkaloid (–)-acutumine. DAH is a 2ODH of plant origin and catalyzes the terminal chlorination step in the biosynthesis of (–)-acutumine. Phylogenetic analyses reveal that DAH evolved independently in Menispermaceae plants and in bacteria, illustrating an exemplary case of parallel evolution in specialized metabolism across domains of life. We show that at the presence of azide anion, DAH also exhibits promiscuous azidation activity against dechloroacutumine. This study opens avenues for expanding plant chemodiversity through halogenation and azidation biochemistry.

<sup>1</sup>Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA. <sup>2</sup>Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>3</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ✉email: [wengj@wi.mit.edu](mailto:wengj@wi.mit.edu)

Halogenated natural products are commonly found in bacteria, marine animals and macroalgae, and harbor a multitude of biological activities ranging from antibiotic to anticancer<sup>1–3</sup>. Despite >5000 characterized halogenated natural products to date, only a few examples have been identified in land plants<sup>1,4,5</sup>. Some known cases include maytansine, a cytotoxic chlorinated ansamacrolide isolated from the Ethiopian shrub *Maytenus serrata*<sup>5,6</sup>, and 4-chloroindole-3-acetic acid, an auxin analog found in the seeds of some leguminous plants<sup>4</sup>. While maytansine was recently confirmed to be produced by endophytic bacteria<sup>5</sup>, the synthesis of 4-chloroindole-3-acetic acid may also rely on symbiotic bacterial metabolism, as known classes of indole halogenases, which are well characterized in bacteria, are absent in sequenced legume genomes<sup>7</sup>.

Another notable example is (–)-acutumine, a tetracyclic chloroalkaloid produced in plants of the Menispermaceae family. (–)-Acutumine exhibits selective cytotoxicity to cultured human T cells<sup>8</sup> and memory-enhancing properties in the Wistar rat model<sup>9</sup>. Due to its unusual structure and bioactivities, total synthesis of (–)-acutumine has been pursued by synthetic chemists in recent years<sup>10–12</sup>. In particular, (–)-acutumine contains an *sp*<sup>3</sup>-carbon-halogen bond, which is difficult to install in a site-specific manner, especially at the late stages of chemical synthesis. Early studies have also explored (–)-acutumine biosynthesis in Menispermaceae plants using isotopic tracing showing that tyrosine is a biosynthetic precursor of (–)-acutumine and chlorination is the terminal step of its biosynthesis<sup>13,14</sup>.

Regio-selective and stereo-selective halogenation is of great interest in medicinal chemistry, as the bioactivities and pharmacokinetics of lead pharmacophores can often be finetuned through halogenation<sup>15</sup>. Using bacterial halogenases, previous studies have introduced halide functional groups to vinca alkaloids in periwinkle plant *Catharanthus roseus*<sup>16</sup>, and have also generated new-to-nature halogenated indigo precursors in *Nicotiana benthamiana*<sup>17</sup>. However, with the scarcity of naturally occurring halogenated plant metabolites, no halogenase of plant origin has been characterized to date. We therefore set out to investigate the chlorination biochemistry in (–)-acutumine biosynthesis in Menispermaceae plants.

Here, we report the discovery of dechloroacutumine halogenase (DAH) from Menispermaceae plants, an Fe(II)-dependent and 2-oxoglutarate-dependent halogenase (2ODH) that catalyzes the terminal chlorination reaction in (–)-acutumine biosynthesis. Phylogenetic analyses suggest that DAH evolved independently from those 2ODHs previously identified from bacteria<sup>18–23</sup>. We show that DAH can also catalyze regioselective C–H azidation at the presence of azide anion, illustrating the promising broader utilities of 2ODH enzymes for versatile natural product derivatization.

## Results and discussion

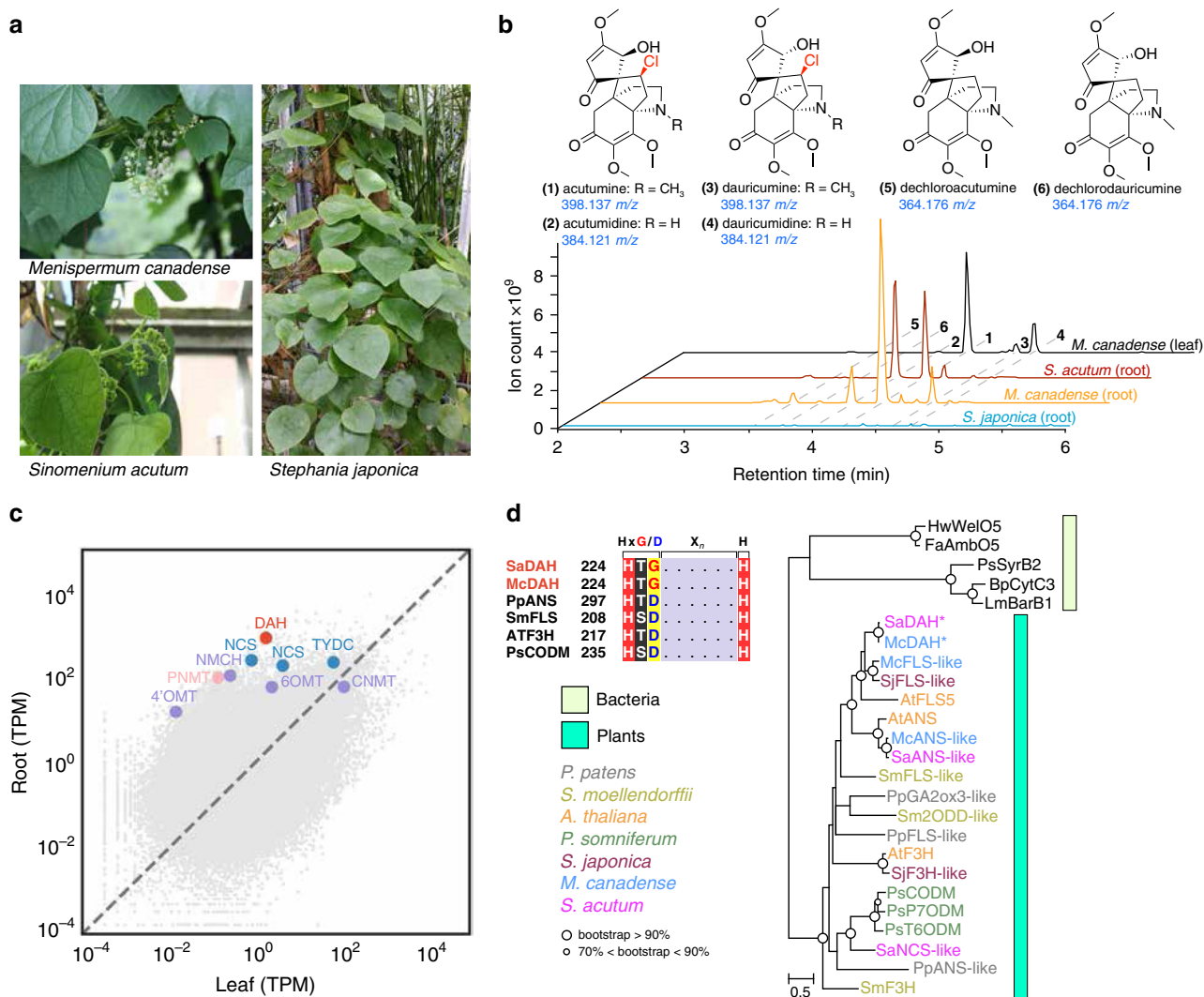
**Discovery of dechloroacutumine halogenase from Menispermaceae plants.** We obtained three species of Menispermaceae plants, two known (–)-acutumine producers *Menispermum canadense* and *Sinomenium acutum*, and one non-producer *Stephania japonica* (Fig. 1a). Using liquid chromatography high-resolution accurate-mass spectrometry (LC-HRAM-MS), we detected several previously reported acutumine-type alkaloids in *M. canadense* and *S. acutum*<sup>13,14,24</sup> (Fig. 1b). The same set of alkaloids were absent in *S. japonica* extracts (Fig. 1b), suggesting that comparative transcriptomic analyses including *S. japonica* as a reference species devoid of the (–)-acutumine chemotype could help identify genes involved in (–)-acutumine biosynthesis. Furthermore, we found that (–)-acutumine accumulation is

relatively more enriched in roots compared to stems and leaves in *M. canadense*.

To enable gene discovery in the target Menispermaceae plants, we performed tissue-specific RNA-Seq experiments, followed by de novo transcriptome assembly for each of the three Menispermaceae species. The resulting transcripts were annotated by BLAST search against the UniProtKB/Swiss-Prot database, and their tissue-specific abundance was quantified by transcripts per million (TPM) values. Using D<sub>2</sub>O-labeling technique, a previous study explored the plausible (–)-acutumine biosynthetic pathway by tracing deuterium-labeled benzyloisoquinoline alkaloids (BIAs) in *M. canadense*<sup>25</sup>. Based on this proposed pathway and prior knowledge about BIA metabolism in other plants, we examined functionally annotated transcripts that are highly and differentially expressed in the root versus the leaf or stem tissue of *M. canadense*, from which numerous candidate genes for the biosynthesis of (–)-acutumine and other related BIAs were identified (Fig. 1c and Supplementary Figs. 1, 2; Supplementary Table 1, 2). We postulated that the candidate enzyme responsible for the terminal chlorination is catalyzed by 2ODH because it is the only characterized class of halogenases capable of targeting unactivated *sp*<sup>3</sup>-hybridized carbon centers for single-step halogenation<sup>26</sup>. We therefore sought highly expressed candidate genes in our transcriptomes that are annotated as Fe(II)-dependent and 2-oxoglutarate-dependent dioxygenases (2ODDs), and manually searched for the potential presence of active-site sequence variation unique to the 2ODH subfamily.

2ODHs were previously identified in bacteria, such as SyrB2 from syringomycin E biosynthesis in *Pseudomonas syringae*<sup>18</sup>, WelO5 from welwitindolinone biosynthesis in *Hapalosiphon welwitschii*<sup>26</sup>, and recently reported amino acid halogenases from *Streptomyces cattleya*<sup>20</sup>. 2ODHs are evolutionarily derived from the 2ODDs, a large enzyme family spanning all kingdoms of life<sup>27</sup>. Canonical 2ODDs activate molecular oxygen using a ferrous cofactor and a 2-oxoglutarate (2OG) cosubstrate to oxidize aliphatic C–H bonds at a highly conserved iron-binding facial triad (HxD/EX<sub>n</sub>H)<sup>28</sup>, typically resulting in a hydroxylation outcome. In contrast, bacterial 2ODHs harbor a mechanistic active-site substitution, replacing the key acidic residue Asp or Glu of the facial triad in 2ODDs with a Gly or Ala. This allows the halide ligand to occupy an iron coordination site, and in turn elicits halogenation chemistry<sup>26</sup>. With this knowledge in mind, we discovered that the most highly expressed 2ODD-family gene in *M. canadense* root transcriptome is likely to encode a candidate 2ODH, because it harbors the unusual active-site Asp226Gly substitution similar to bacterial 2ODHs (Fig. 1c, d and Supplementary Fig. 2, Supplementary Dataset 1). This gene also exhibits a 459-fold higher transcript expression level in roots than in leaves, consistent with the more enriched (–)-acutumine accumulation in roots. This gene is named as *M. canadense* dechloroacutumine halogenase (McDAH) hereafter. Analysis of the *S. acutum* transcriptome identified SaDAH, which exhibits 99.1% sequence identity to McDAH at the protein level, and is likely an ortholog of McDAH (Fig. 1d and Supplementary Fig. 3). No obvious DAH ortholog could be identified in the *S. japonica* transcriptome, and no other annotated 2ODD amongst the three transcriptomes harbor the Asp-to-Gly facial triad mutation. Moreover, no homologs of the other aforementioned classes of halogenases could be identified in the three Menispermaceae plant transcriptomes.

**Biochemical characterization of dechloroacutumine halogenase.** To examine the biochemical function of DAH, we first expressed and purified recombinant SaDAH protein from *Escherichia coli* (Supplementary Fig. 4), and performed in vitro

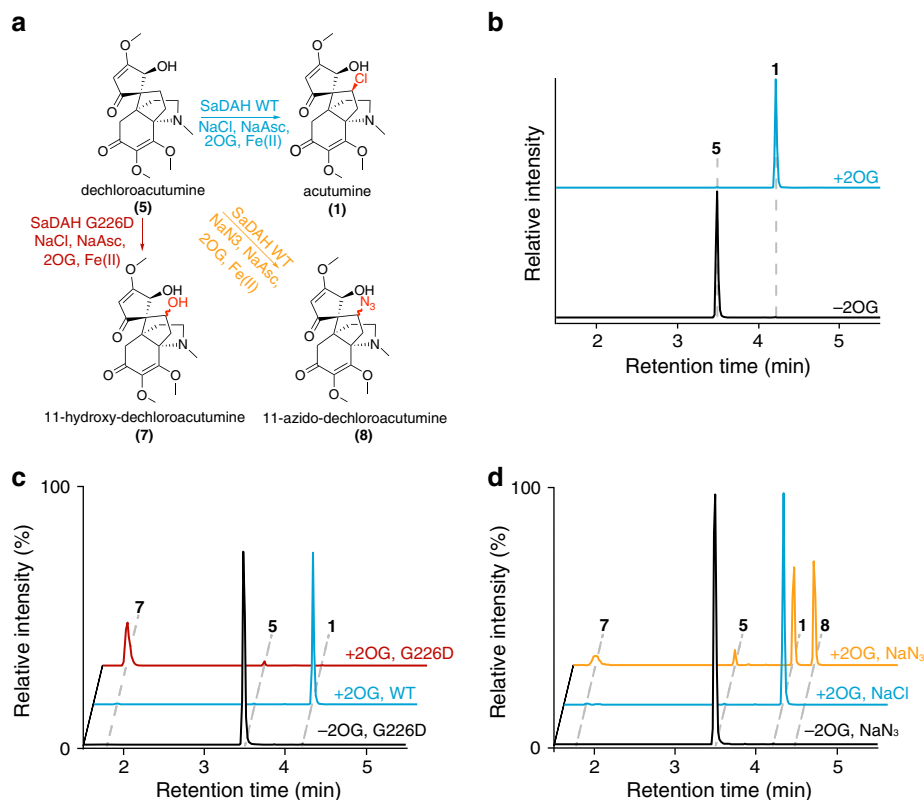


**Fig. 1 Identification of candidate genes encoding DAH in Menispermaceae plants.** **a** Images of *M. canadense*, *S. acutum*, and *S. japonica* plants. **b** Extracted ion chromatograms (XICs) exhibiting the presence and absence of (–)-acutumine and its related compounds, in three Menispermaceae plants. **c** Transcript differential expression analysis in *M. canadense* root vs. leaf tissues. Transcript abundance is quantified by transcripts per million (TPM) values. Candidate (–)-acutumine biosynthetic enzymes are colored based on their postulated roles in the proposed pathway depicted in Supplementary Fig. 1. **d** Sequence and phylogenetic analysis of DAH. The insert at the upper left corner shows multiple sequence alignment of two DAHs and other ODDs highlighting the Hx(D/G)<sub>n</sub>H motif. The full sequence alignment is shown in Supplementary Fig. 3. The phylogenetic tree of DAHs together with select 2ODD-family proteins is inferred using maximum-likelihood method. Bootstrap statistics (200 replicates) are indicated at the tree nodes. The scale measures evolutionary distance in substitutions per amino acid. Asterisks denote the two orthologous DAHs identified in this study. The multiple sequence alignment used for building the phylogenetic tree is in Supplementary File 1. NCS norcoclaurine synthase, TYDC tyrosine decarboxylase, PNMT pavine N-methyltransferase, 6OMT (*RS*)-norcoclaurine 6-*O*-methyltransferase, CNMT (*S*)-coclaurine *N*-methyltransferase, NMCH (*S*)-*N*-methylcoclaurine-3-hydroxylase, 4'OMT (*S*)-3'-hydroxy-*N*-methylcoclaurine 4'-*O*-methyltransferase, ANS anthocyanidin synthase, FLS flavonol synthase, F3H flavanone-3-hydroxylase, CODM codeine *O*-demethylase, GA2ox3 gibberellin 2- $\beta$ -dioxygenase 3, 2ODD iron(II)- and 2-oxoglutarate-dependent dioxygenase, P7ODM papaverine 7-*O*-demethylase, T6ODM thebaine 6-*O*-demethylase.

enzyme assays. SaDAH exhibits specific 2OG-dependent halogenase activity that converts (–)-dechloroacutumine to (–)-acutumine in a reaction buffer containing 1 mM NaCl (Fig. 2a, b and Supplementary Fig. 5), whereas no halogenase activity could be detected when the structurally related BIA sinomenine was tested as a substrate (Supplementary Fig. 6). Recombinant McDAH exhibits identical activity to SaDAH (Supplementary Fig. 7). Since the recombinant protein yield was higher for SaDAH than McDAH, we chose SaDAH as the representative DAH for detailed characterizations hereafter. To examine the potential substrate promiscuity of SaDAH, we tested its halogenase activity against a panel of additional alkaloids

including vinpocetine, codeine, scoulerine, berberine, and boldine. No halogenated products could be detected in these assays, suggesting that SaDAH is likely a specific halogenase towards its native substrate (–)-dechloroacutumine (Supplementary Fig. 8).

SaDAH exhibits a  $K_M$  of  $18.35 \pm 7.13 \mu\text{M}$  against (–)-dechloroacutumine as the substrate, and an apparent  $k_{\text{cat}}$  of  $63.44 \pm 8.75 \text{ min}^{-1}$  (Supplementary Fig. 9), which are comparable to the kinetic constants measured for previously characterized bacterial 2ODHs<sup>20</sup> and plant 2ODDs<sup>29</sup>. Interestingly, a trace amount of the hydroxylation product 11-hydroxy-dechloroacutumine was also detected in the SaDAH in vitro enzyme assay (Supplementary Fig. 10), suggesting that SaDAH still retains a low level of



**Fig. 2 Functional characterization of DAH and its G226D mutant by in vitro enzyme assays.** **a** Alternative catalytic activities of SaDAH-WT and SaDAH-G226D. The chlorination, hydroxylation and azidation activities are colored in blue, red, and yellow, respectively. The putative regiochemistry of the  $-OH$  and  $-N_3$  in **7** and **8** are highlighted in red, while the stereochemistry of the products remains undetermined. **b** XICs showing the in vitro activity of SaDAH-WT that chlorinates **5** to yield **1** in a 2OG-dependent manner. **c** XICs showing the in vitro activity of SaDAH-G226D that predominantly hydroxylates **5-7** in a 2OG-dependent manner. The chlorination product **1** is also produced as a minor product. **d** XICs showing the 2OG-dependent azidation activity of SaDAH-WT that converts **5-8** in reaction buffer containing 1 mM  $NaN_3$ . The chlorination product **1** and the hydroxylation product **7** are also produced as side products. Mass windows used for displaying the XICs: **5**, 364.176  $m/z$ ; **1**, 398.137  $m/z$ ; **7**, 380.17  $m/z$ ; and **8**, 405.18  $m/z$ .

latent hydroxylase activity, although the hydroxylation product was not detected in any of the Menispermaceae plant extracts. This retention of ancestral hydroxylation activity has also been observed in the previously characterized 2ODHs<sup>26</sup>.

**Independent evolution of 2ODHs in plants and bacteria.** To the best of our knowledge, our discovery of DAH in the context of (–)-acutumine biosynthesis represents the first 2ODH of plant origin. Remarkably, DAH seems to harbor the same active-site Asp226Gly substitution previously shown to be characteristic of bacterial 2ODHs<sup>26</sup>. To discern whether DAH might have evolved from ancestral plant 2ODDs, or, alternatively, through a plausible horizontal gene transfer event from bacteria, we performed maximum-likelihood phylogenetic analysis of DAHs together with bacterial 2ODHs and other 2ODD-family proteins from Menispermaceae plants, *Arabidopsis thaliana*, *Papaver somniferum*, *Selaginella moellendorffii*, and *Physcomitrella patens*. DAHs appear to be immediately sister to the well-characterized plant flavonol synthases (FLSs), whereas bacterial 2ODHs form a distinct clade separate from all plant 2ODD-family proteins (Fig. 1d and Supplementary Data 1). Moreover, an extended phylogenetic analysis including 2ODD homologs identified from the available transcriptomes of Ranunculales plants further supports that DAH recently evolved from an FLS-like progenitor through a gene duplication event that occurred in during the radiation of Ranunculales, contributing to the occurrence of (–)-acutumine biosynthesis in Menispermaceae (Supplementary Fig. 11, Supplementary Data 2).

**Expansion of C–H functionalization using dechloroacutumine halogenase.** To assess the role of the potential mechanistic Asp226Gly substitution found in DAHs relative to canonical 2ODDs, we generated the SaDAH G226D mutant and examined its in vitro enzymatic activity. The SaDAH G226D mutant is entirely devoid of halogenase activity, and instead exhibits exclusive (–)-dechloroacutumine hydroxylase activity to give the hydroxylation product 11-hydroxy-dechloroacutumine (Fig. 2a, c and Supplementary Fig. 12). This result indicates that DAH adopts the same mechanistic substitution as in bacterial 2ODHs, wherein the canonical active-site iron-coordinating Asp is replaced with an Gly or Ala that facilitates  $Cl^-$  binding and the subsequent chlorination chemistry<sup>26</sup>. As predicted by this mechanism, mutating Gly226 of DAH to Asp therefore restores the canonical 2ODD iron-binding facial triad, prevents halide coordination, and consequently reverts the reaction outcome to hydroxylation<sup>26</sup>. Interestingly, after an exhaustive search and examination of the abundant plant 2ODD-family protein sequences identifiable at NCBI, Uniprot, and 1KP databases, the two DAHs identified in this study remain the only plant 2ODD-family proteins that harbor the unusual Asp226Gly substitution, suggesting the occurrence of 2ODH activity through this particular molecular mechanism is a rare serendipity in plant metabolic evolution.

It was previously shown that bacterial halogenase SyrB2 exhibits extended catalytic activities to install alternative anions<sup>30</sup>. We therefore tested SaDAH's ability to derivatize (–)-dechloroacutumine using a panel of alternative anions:  $N_3^-$ ,  $Br^-$ ,  $NO_3^-$ ,

$\text{NO}_2^-$ , and  $\text{OCN}^-$ . Among the anions tested, only  $\text{N}_3^-$  resulted in a derivatized product, which was identified as 11-azido-dechloroacutumine by LC-HRAM-MS analysis (Fig. 2a, d and Supplementary Figs. 12, 13). (–)-Acutumine and 11-hydroxy-dechloroacutumine were also produced in the same assay, likely due to the presence of trace  $\text{Cl}^-$  carried over from the enzyme storage buffer and elevated hydroxylation outcome when enzyme access to native  $\text{Cl}^-$  substrate is limited, respectively (Fig. 2d). The regio-specificity and stereo-specificity of the –OH and – $\text{N}_3$  moieties is not directly determined, but can be reasonably postulated to be at the C11 position of (–)-dechloroacutumine with the same stereochemistry as the –Cl group in (–)-acutumine (Fig. 2a, Supplementary Fig. 12). The fact that only a single –Cl product is observed in SaDAH chlorination assay indicates that the enzyme is not capable of abstracting an alternate C–H bond (Fig. 2b and Supplementary Fig. 5). This postulation is consistent with the nature of substrate binding and C–H bond abstraction previously reported in other bacterial 2ODHs<sup>26</sup>.

In summary, we report the discovery of the first 2ODH of plant origin, DAH, that catalyzes the terminal chlorination step in (–)-acutumine biosynthesis in Menispermaceae plants. The independent occurrences of 2ODHs in plant and bacteria through the same key active-site mechanistic substitution illustrate an exemplary case of parallel evolution in specialized metabolism across domains of life<sup>31</sup>. Although it is unknown why halogenated natural products are so scarce in plants compared to those found in bacteria, fungi, and marine organisms, the lack of access to halogen salts could be one of the explanations. Future experiments to knockdown expression of DAH and other candidate BIA metabolic enzyme-encoding genes in Menispermaceae plants not only will help elucidate the (–)-acutumine biosynthetic pathway, but also will shed light on the functional significance of (–)-acutumine biosynthesis to its native plant hosts. In addition to chlorination, we show that DAH is also capable of catalyzing azidation chemistry, and thus opens up several new avenues for future applications. Azidation of plant natural products could enable rapid metabolite detection using click-chemistry-based fluorogenic probes<sup>32,33</sup>. The same chemical principle can also be applied to pull-down experiments for identifying unknown protein targets of specific bioactive plant natural products amenable to site-specific azidation. Moreover, as a rare 2ODH of plant origin, DAH presents a starting point for future directed evolution experiments to further expand substrate specificity and anion selectivity, which will find use in the production of new-to-nature plant natural products through the means of metabolic engineering<sup>34</sup>.

## Methods

**Plant materials and chemicals.** *S. acutum* and *S. japonica* plant samples were collected at the University of Connecticut's EEB Biodiversity Education & Research Greenhouses, Connecticut. *M. canadense* plant samples were purchased from the Toadshade Wildflower Farm, New Jersey. All chemical reagents were obtained from Sigma-Aldrich or Thermo Fisher Scientific, unless otherwise stated. Ultrapure water was generated by a Milli-Q system (EMD Millipore). (–)-Acutumine (>98%) was obtained from BOC Sciences and as a gift from the Herzon Lab at Yale University. Sinomenine, scoulerine, and vinpocetine were obtained from Santa Cruz Biotechnology.

**Plant metabolite extraction.** Approximately 10 mg of plant tissue was dissected, weighed, transferred into grinding tubes containing approximately six zirconia/silica disruption beads (2 mm diameter; Research Products International), and flash-frozen in liquid nitrogen. The frozen samples were homogenized on a TissueLyser II (QIAGEN) at 25 s<sup>-1</sup> for 2 min. Metabolites were extracted using 100 volumes (w/v) of 80% MeOH at 65 °C for 10 min. Extracts were centrifuged once (13,000×g, 5 min) to pellet plant tissue, and the supernatants were filtered using 0.2 µm filter vials (Thomson Instrument Company) for LC-HRAM-MS analysis.

**Transcriptome sequencing and assembly.** Total RNA was extracted using the RNeasy Mini Kit (Qiagen) from the leaf, stem, and root of *Menispermum*

*canadense* (3–6 biological replicates each), from the root of *Sinomenium acutum* (2 biological replicates), and from the root of *Stephania japonica* (2 biological replicates). RNA quality was assessed by Bioanalyzer (Agilent Technologies). Strand-specific mRNA libraries were prepared using the KAPA Hyper Prep mRNA Library Prep Kit (Kapa Biosystems), and sequenced on a HiSeq 2500 sequencer (Illumina) in paired-end mode (PE100). Total 342 M, 49 M, and 50 M paired sequencing reads were acquired for *M. canadense*, *S. acutum*, and *S. japonica*, respectively. Sequence FASTQ files were trimmed for sequencing adapters and poor-quality reads using Trimmomatic (v. 0.39) and assembled into de novo transcriptomes using Trinity (v. 2.8.5) in strand-specific mode<sup>35,36</sup>. Putative open reading frames in transcripts were identified using Transdecoder (v. 5.5.0)<sup>37</sup>. Gene expression transcripts-per-million (TPM) statistics were determined using kallisto (v. 0.46.0)<sup>38</sup>. Transcripts and predicted protein sequences were annotated with TPM values and closest BLAST hits from the UniProtKB/Swiss-Prot database using in-house scripts. Transcriptome mining was performed on a local BLAST server<sup>39</sup>. To compare the RNA expression profiles of *M. canadense*, *S. acutum*, and *S. japonica*, orthologs were identified by clustering their gene set peptides using the OrthoFinder (v.2.2.7)<sup>40</sup> pipeline with default parameters.

**Organic synthesis of (–)-acutumine.** The reaction schematic is shown in Supplementary Fig. 14. All reagents were used as supplied without further purification with the exception of AIBN, which was recrystallized from acetone and dried over P<sub>2</sub>O<sub>5</sub>. Analytical TLC was conducted using Millipore SiO<sub>2</sub> 60 F254 TLC (0.250 mm) plates. Mass spectrometric analysis was performed on a TSQ Quantum Access Max mass spectrometer (Thermo Fisher Scientific) operated in positive ionization mode with a full scan range of 100–400 m/z. A solution of (–)-acutumine (6.0 mg, 0.015 mmol) in dry toluene (6.0 mL) at 23 °C was treated successively with <sup>10</sup>Bu<sub>3</sub>SnH (0.10 mL, 0.38 mmol) and AIBN (1.3 mg, 0.008 mmol) under a nitrogen atmosphere. The resulting solution was refluxed for 3 h, then cooled to 23 °C and treated with saturated NaHCO<sub>3</sub>(aq) (15 mL) and sat. KF(aq) (1 mL). The organic layer was separated and the aqueous layer was extracted with CHCl<sub>3</sub> (2 × 10 mL). The combined organic extracts were dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated on a rotary evaporator. The resulting solid residue was purified by preparative reverse-phase HPLC (Shimadzu Preparative HPLC with LC-20AP pump and SPD-20A UV-VIS detector) using Kinetex 5 µ C<sub>18</sub> 100 Å, 150 × 21.2 mm column, solvent A-(H<sub>2</sub>O-0.1% TFA), solvent B-(CH<sub>3</sub>CN-0.1% TFA), 0–50 min 40–80% B, 10 mL/min) to provide (–)-dechloroacutumine<sup>8</sup> as an amorphous white solid (30 µg, 1%). Spectrometric data are in agreement with the values published previously<sup>11</sup>.

**Recombinant protein expression and purification.** The SaDAH gene was codon-optimized for heterologous overexpression in *Escherichia coli*, and purchased as a synthetic gene (Integrated DNA Technologies). The SaDAH gene was cloned into pHis8-4 expression vector containing an N-terminal 8 × His tag followed by a tobacco etch virus (TEV) cleavage site (Supplementary Table 3). The verified pHis8-4b::SaDAH construct was transformed into *E. coli* BL21 (DE3) for recombinant protein expression. A 1 L culture of terrific broth (TB) medium containing 50 µg/mL kanamycin was inoculated with 30 mL of an overnight starter culture and allowed to grow with shaking at 200 rpm at 37 °C to an OD<sub>600</sub> of 0.6–0.8. Then protein expression was induced by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by cold shock of the medium and subsequent growth with shaking at 200 rpm (18 °C for 18 h).

Cultures were harvested by centrifugation and the resulting cell paste (~10 g/L) was resuspended in lysis buffer (100 mM Tris pH 8.0, 200 mM NaCl, 20 mM imidazole, 10% (vol/vol) glycerol, 1 mM dithiothreitol) containing 1 mg/mL lysozyme and 1 mM phenylmethylsulfonyl fluoride. Cells were lysed via five passes through an M-110L microfluidizer (Microfluidics). The resulting crude protein lysate was clarified by centrifugation (19,000 × g, 45 min) before QIAGEN nickel-nitrilotriacetic acid (Ni-NTA) gravity flow chromatographic purification. After loading the clarified lysate, the Ni-NTA resin was washed with 20 column volumes of lysis buffer and eluted with 1 column volume of elution buffer (100 mM Tris pH 8.0, 200 mM NaCl, 250 mM imidazole, 10% (vol/vol) glycerol, 1 mM dithiothreitol). Then 1 mg of His-tagged TEV protease was added to the eluted protein, followed by dialysis at 4 °C for 16 h in dialysis buffer (25 mM Tris pH 8.0, 200 mM NaCl, 5% (vol/vol) glycerol, 5 mM EDTA, 0.5 mM dithiothreitol). After dialysis, protein solution was passed through Ni-NTA resin to remove uncleaved protein and His-tagged TEV. The recombinant proteins were further purified by gel filtration on an ÄKTA Pure fast protein liquid chromatography (FPLC) system (GE Healthcare Life Sciences). The principal peaks were collected, verified by SDS polyacrylamide gel electrophoresis and dialyzed into a storage buffer (25 mM Tris pH 8.0, 5% (vol/vol) glycerol). Finally, proteins were concentrated to >10 mg/mL using Amicon Ultra-15 Centrifugal Filters (Millipore).

**SaDAH mutagenesis.** SaDAH G226D mutant was generated according to the protocol described in the QuikChange site-directed mutagenesis kit (Agilent Technologies) using plasmid pHis8-4::SaDAH as the template and the following primer sequences:

SaDAH-G226D-5': GTGTAGTCCCTCATACAGACTATCCTGCAATGAC AAT; SaDAH-G226D-3': CACATCAGGGAGTATGTCTGATAGGACGTT ACTGTTA (Supplementary Table 4). The resulting pHis8-4::SaDAH-G226D

construct was verified by sequencing. Recombinant mutant protein production and purification were carried out following the same procedure as described above.

**In vitro enzyme activity assays.** Each enzyme assay for SaDAH-WT and SaDAH-G226D with (–)-dechloroacutumine was carried out in 25 mM Tris buffer, pH = 8.0, on a 50- $\mu$ L scale containing the following components: enzyme (20  $\mu$ M), (–)-dechloroacutumine (100  $\mu$ M), 2OG (500  $\mu$ M), NaCl (1 mM), sodium ascorbate (5 mM), and (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub> (2 mM). In a typical assay, the components were added in the following order: (1) Tris, (2) NaCl, (3) enzyme, (4) sodium ascorbate, (5) 2OG, (6) (–)-dechloroacutumine, and (7) (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub>. The samples were incubated under aerobic conditions at room temperature for 1 h. The assays were quenched by adding methanol to 50% final concentration and centrifugation to spin down enzymes and debris. The supernatants were dried under nitrogen gas and resuspended in 10% acetonitrile in water for injection into the LC-HRAM-MS system for analysis. The azidation activity assay of SaDAH-WT was carried out following the procedure described above, with the replacement of NaCl with NaN<sub>3</sub> (1 mM).

**SaDAH enzyme kinetic assays.** Each enzyme reaction (50  $\mu$ L) contained SaDAH-WT (30 nM), 2OG (500  $\mu$ M), NaCl (1 mM), sodium ascorbate (5 mM), (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub> (2 mM), and (–)-dechloroacutumine (0–105  $\mu$ M). Reactions were initiated by addition of a mixture containing SaDAH-WT and (NH<sub>4</sub>)<sub>2</sub>Fe(SO<sub>4</sub>)<sub>2</sub> in the presence of varying concentrations of (–)-dechloroacutumine (0–105  $\mu$ M; 18 different concentrations). Each reaction was quenched after 10 min by adding methanol to 50% final concentration and centrifugation to spin down enzymes and debris. The supernatants were dried under nitrogen gas and resuspended in 10% acetonitrile in water for injection into the LC-HRAM-MS system for analysis. The reaction velocity for each assay was measured by integrating the peak area corresponding to (–)-acutumine on LC-MS. An internal standard curve of (–)-acutumine was generated using the same set of reactions with known concentrations of (–)-dechloroacutumine (0–105  $\mu$ M), but quenching each reaction at 3 h after the substrate was completely converted to (–)-acutumine.

**LC-MS analysis.** LC was conducted on a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific) using water with 0.1% formic acid as solvent A and acetonitrile with 0.1% formic acid as solvent B. Reverse phase separation of analytes was performed on a Kinetex C18 column, 150  $\times$  3 mm<sup>2</sup>, 2.6  $\mu$ m particle size (Phenomenex). The column oven was held at 30 °C. Most injections were eluted with 5% B for 0.5 min, a gradient of 5–36% B for 5 min, 95% B for 2 min and 5% B for 2.5 min, with a flow rate of 0.5 mL/min. The analysis of Menispermaceae species tissue extracts and enzyme substrate promiscuity tests was performed with 5% B for 0.5 min, a gradient of 5–95% for 14.5 min, 95% B for 2 min and 5% B for 3 min, with a flow rate of 0.5 mL/min. Most MS analyses were performed on a high-resolution Q-Exactive benchtop Orbitrap mass spectrometer (Thermo Fisher Scientific) operated in positive ionization mode with full scan range of 100–600 *m/z* and top five data-dependent MS/MS scans. Initial enzyme activity assays were analyzed with a TSQ Quantum Access Max mass spectrometer (Thermo Fisher Scientific) using multiple reaction monitoring at *m/z* transitions corresponding to (–)-acutumine (398.14–273.08, 305.10, 341.08) and (–)-dechloroacutumine (364.18–257.08, 289.11, 307.12) with scan width of 0.5 and collision energy of 20 V. For detection of (–)-acutumine derivatives, single ion monitoring was used to scan the following *m/z* values: 364.20, 380.17, 398.1, 405.17, 409.16, 425.16, and 442.08 with scan width of 0.5 and collision energy of 20 V. The MS analysis for DAH substrate promiscuity tests was conducted on the TSQ Quantum Access Max mass spectrometer (Thermo Fisher Scientific) operated in positive ionization mode with a full scan range of 250–500 *m/z*. Raw LC-MS data were analysed using XCalibur (Thermo Fisher Scientific).

**Sequence alignment and phylogenetic tree analysis.** The sequences for sequence alignment and phylogenetic tree construction were obtained from the NCBI database, 1KP database, Phytozome v12.1.6. pBLAST with DAH as a query sequence was conducted and top 20 full length protein sequences were collected for each species: *P. patens*, *S. moellendorffii*, *A. thaliana*, all available species in Ranunculales plant family, *S. japonica*, *M. canadense*, and *S. acutum*. In-house scripts were used to extract sequences from the assembled transcriptomes discussed in this manuscript. Sequence alignments were performed using the MUSCLE<sup>41</sup> algorithm in MEGA7<sup>42</sup> and visualized using ESPrnt 3<sup>43</sup>. Evolutionary histories were inferred by using the maximum-likelihood method on the basis of the JTT matrix-based model. Bootstrap statistics were calculated using 200 replicates. All phylogenetic analyses were conducted in MEGA7<sup>42</sup>. All alignment files can be found under Associated Data.

## Data availability

The sequences of the genes reported in this article have been deposited in NCBI GenBank (accessions [MT040708](#), [MT040709](#)). Protein expression plasmids used in this study are available from Addgene (accessions [140128-140130](#)). The raw sequencing reads have been submitted to NCBI SRA (accessions [SRR10947794-SRR10947810](#)) and the de novo assembled transcriptome to NCBI TSA (accessions [GIU000000000.1](#),

[GIU000000000.1](#), [GHIS000000000.1](#)). The raw data underlying Supplementary Figs. 4, 9 are provided as a Source Data file. Additional data that support the findings of this study are available from the corresponding authors upon reasonable request.

## Code availability

Codes used to assemble the transcriptomes and extract sequences discussed in this manuscript are available on request.

Received: 27 November 2019; Accepted: 25 March 2020;

Published online: 20 April 2020

## References

- Chung, W.-J. & Vanderwal, C. D. Stereoselective halogenation in natural product synthesis. *Angew. Chem. Int. Ed. Engl.* **55**, 4396–4434 (2016).
- Li, C. & Shi, D. Structural and bioactive studies of halogenated constituents from sponges. *Curr. Med. Chem.* <https://doi.org/10.2174/0929867325666181112092159> (2018).
- Wagner, C., El Omari, M. & König, G. M. Biohalogenation: nature's way to synthesize halogenated metabolites. *J. Nat. Prod.* **72**, 540–553 (2009).
- Tivendale, N. D. et al. Biosynthesis of the halogenated auxin, 4-chloroindole-3-acetic acid. *Plant Physiol.* **159**, 1055–1063 (2012).
- Kusari, P. et al. Cross-species biosynthesis of maytansine in *Maytenus serrata*. *RSC Adv.* **6**, 10011–10016 (2016).
- Lam, H. K., McAdam, S. A. M., McAdam, E. L. & Ross, J. J. Evidence that chlorinated auxin is restricted to the fabaceae but not to the fabae. *Plant Physiol.* **168**, 798–803 (2015).
- van Pée, K.-H. & Patallo, E. P. Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Appl. Microbiol. Biotechnol.* **70**, 631–641 (2006).
- Yu, B.-W. et al. Alkaloids from *Menispermum dauricum*. *Phytochemistry* **61**, 439–442 (2002).
- Qin, G. W., Tang, X. C., Lestage, P., Caignard, D. H. & Renard, P. Acutumine and acutumine compounds, synthesis and use. Patent WO2004000815A1 (2003).
- Li, F., Tartakoff, S. S. & Castle, S. L. Enantioselective total synthesis of (–)-acutumine. *J. Org. Chem.* **74**, 9082–9093 (2009).
- King, S. M., Calandra, N. A. & Herzon, S. B. Total syntheses of (–)-acutumine and (–)-dechloroacutumine. *Angew. Chem. Int. Ed. Engl.* **52**, 3642–3645 (2013).
- King, S. M. & Herzon, S. B. The hasubanan and acutumine alkaloids. *Alkaloids Chem. Biol.* **73**, 161–222 (2014).
- Sugimoto, Y., Matsui, M. & Babiker, H. A. A. Conversion of dechloroauricumine into chlorinated alkaloids in *Menispermum dauricum* root culture. *Phytochemistry* **68**, 493–498 (2007).
- Babiker, H. A. et al. Biosynthetic relationship between acutumine and dechloroacutumine in *Menispermum dauricum* root cultures. *Biosci. Biotechnol. Biochem.* **63**, 515–518 (1999).
- Agarwal, V. et al. Enzymatic halogenation and dehalogenation reactions: pervasive and mechanistically diverse. *Chem. Rev.* **117**, 5619–5674 (2017).
- Runguphan, W., Qu, X. & O'Connor, S. E. Integrating carbon–halogen bond formation into medicinal plant metabolism. *Nature* **468**, 461–464 (2010).
- Fräbel, S. et al. Engineering of new-to-nature halogenated indigo precursors in plants. *Metab. Eng.* **46**, 20–27 (2018).
- Vaillancourt, F. H., Yin, J. & Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O<sub>2</sub>-dependent halogenase. *Proc. Natl Acad. Sci. USA* **102**, 10111–10116 (2005).
- Hillwig, M. L., Zhu, Q., Ittiarnornkul, K. & Liu, X. Discovery of a promiscuous non-heme iron halogenase in ambigua alkaloid biogenesis: implication for an evolvable enzyme family for late-stage halogenation of aliphatic carbons in small molecules. *Angew. Chem. Int. Ed. Engl.* **55**, 5780–5784 (2016).
- Neugebauer, M. E. et al. A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **15**, 1009–1016 (2019).
- Zhang, Y. et al. An iron(II)- and  $\alpha$ -ketoglutarate-dependent halogenase acts on nucleotide substrates. *Angew. Chem.* <https://doi.org/10.1002/ange.201914994> (2020).
- Galončić, D. P., Barr, E. W., Walsh, C. T., Bollinger, J. M. Jr. & Krebs, C. Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. *Nat. Chem. Biol.* **3**, 113–116 (2007).
- Schofield, C. & Hausinger, R. *2-Oxoglutarate-Dependent Oxygenases* (Royal Society of Chemistry, 2015).
- Sugimoto, Y. et al. Chlorinated alkaloids in *Menispermum dauricum* DC: root culture. *J. Org. Chem.* **66**, 3299–3302 (2001).
- Nett, R. S. et al. DO labeling to measure active biosynthesis of natural products in medicinal plants. *AIChE J.* **64**, 4319–4330 (2018).

26. Mitchell, A. J. et al. Structural basis for halogenation by iron- and 2-oxoglutarate-dependent enzyme WelO5. *Nat. Chem. Biol.* **12**, 636–640 (2016).
27. Mitchell, A. J. & Weng, J.-K. Unleashing the synthetic power of plant oxygenases: from mechanism to application. *Plant Physiol.* **179**, 813–829 (2019).
28. Hagel, J. M. & Facchini, P. J. Expanding the roles for 2-oxoglutarate-dependent oxygenases in plant metabolism. *Nat. Prod. Rep.* **35**, 721–734 (2018).
29. Hagel, J. M. & Facchini, P. J. Dioxygenases catalyze the O-demethylation steps of morphine biosynthesis in opium poppy. *Nat. Chem. Biol.* **6**, 273–275 (2010).
30. Matthews, M. L. et al. Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. *Nat. Chem. Biol.* **10**, 209–215 (2014).
31. Weng, J.-K. The evolutionary paths towards complexity: a metabolic perspective. *N. Phytol.* **201**, 1141–1149 (2014).
32. Thirumurugan, P., Matosiuk, D. & Jozwiak, K. Click chemistry for drug development and diverse chemical–biology applications. *Chem. Rev.* **113**, 4905–4979 (2013).
33. Sawa, M. et al. Glycoproteomic probes for fluorescent imaging of fucosylated glycans in vivo. *Proc. Natl Acad. Sci. USA* **103**, 12371–12376 (2006).
34. Hayashi, T. et al. Evolved aliphatic halogenases enable regiocomplementary C–H functionalization of an added-value chemical. *Angew. Chem. Int. Ed. Engl.* <https://doi.org/10.1002/anie.201907245> (2019).
35. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
36. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
37. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
38. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
39. Priyam, A. et al. Sequenceserver: a modern graphical user interface for custom BLAST databases. *Mol. Biol. Evol.* <https://doi.org/10.1093/molbev/msz185> (2019).
40. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
41. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
42. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
43. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).

## Acknowledgements

This work was supported by grants from the Beckman Foundation (J.K.W.), the Keck Foundation (J.K.W.), and the National Science Foundation (CHE-1709616). We thank

the Herzon Lab at Yale University for providing (–)-acutumine, and the EEB Biodiversity Education & Research Greenhouses at University of Connecticut for providing the *S. acutum* and *S. japonica* plants used in this study, along with their images. We also thank Lady Bird Johnson Wildflower Center for providing the *M. canadense* image.

## Author contributions

C.Y.K. performed most experiments. A.J.M., and T.P. performed the RNA-seq experiment and helped assemble the Menispermaceae plant transcriptomes. F.L. and C.M.G. helped synthesize and purify (–)-dechloroacutumine used for enzyme assays. C.Y.K. and J.K.W. designed experiments, analyzed data, and wrote the manuscript. All authors reviewed the manuscript.

## Competing interests

J.K.W. is a co-founder, a member of the Scientific Advisory Board, and a shareholder of DoubleRainbow Biosciences, which develops biotechnologies related to natural products. All other authors have no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41467-020-15777-w>.

**Correspondence** and requests for materials should be addressed to J.-K.W.

**Peer review information** *Nature Communications* thanks Peter Facchini, Shoji Tsubasa, and the other, anonymous, reviewer(s), for their contribution to the peer review of this work. Peer review reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020