Discovery of homoharringtonine pathway enzymes reveals a whole plant model for coordinated biosynthesis

Yaereen Dho,[1] Kevin Smith,[2] & Elizabeth S. Sattely[2,3]*

[1]Department of Chemistry, Stanford University, Stanford, CA 94305

[2]Department of Chemical Engineering, Stanford University, Stanford, CA 94305

[3]HHMI, Stanford University, Stanford, CA 94305

*corresponding author

ABSTRACT

Plants have evolved to produce diverse molecules that inhibit protein translation. A lead example is homoharringtonine (HHT), both a key tool for ribosomal profiling and an FDA-approved treatment for chronic myeloid leukemia. HHT is commercially produced through semi-synthesis by esterifying the alkaloid core cephalotaxine (CET) extracted from endangered *Cephalotaxus* species. Despite its medicinal significance, a biosynthetic pathway to CET and HHT has not been described. Here, we use paired untargeted metabolomics (stable-isotope labeled precursor feeding) and transcriptomics to elucidate a near-complete biosynthesis to CET without prior knowledge of intermediates and biosynthetic genes. We show that while CET alkaloid core is actively biosynthesized only in growing root tips, both CET and HHT accumulate throughout the plant. We discovered and characterized seven CET pathway intermediates and six novel biosynthetic enzymes that, together, can be used to produce cephalotaxinone, the likely direct precursor of CET. Included are non-canonical cytochrome P450s, an atypical short-chain dehydrogenase, and a 2-oxogluatrate-dependent dioxygenase that together result in carbon excision and the formation of the characteristic pentacyclic backbone of HHT alkaloids. Our data support a model where cephalotaxinone is the last pathway intermediate produced specifically in the root tips, and its distribution throughout the plant is likely the starting point for subsequent elaboration to HHT. This study not only establishes a metabolic route to the core scaffold of HHT—enabling future sustainable, large-scale production of this valuable drug—but also suggests how *Cephalotaxus* species employ a whole plant coordination to regulate the biosynthesis of eukaryotic ribosomal toxins.

INTRODUCTION

Cephalotaxus (plum yew) is a slow-growing coniferous tree from the *Cephalotaxaceae* family that produces over 70 structurally diverse *Cephalotaxus* alkaloids with the most prominent being cephalotaxine (CET)—the core structure of *Cephalotaxus* alkaloids.[1-3] Although no unique bioactivity for CET itself has been reported, its ester derivatives—such as harringtonine, isoharringtonine, deoxyharringtonine, and homoharringtonine (HHT)—have demonstrated potent antileukemic activity *in vivo*, each significantly inhibiting P388 lymphoid leukemia in mice at doses as low as 1 mg/kg.[2,4-6] HHT, in particular, exhibited nanomolar inhibitory activity against myeloid and lymphocytic leukemia cell lines[7] and was approved by the U.S. Food and Drug Administration (FDA) in 2012 for the treatment of chronic myeloid leukemia (CML),[8,9] especially in patients resistant to tyrosine kinase inhibitors (comprising 30% of CML cases).[10] Also known as omacetaxine mepesuccinate (Synribo®), the HHT drug acts as a protein translation inhibitor by binding to the A-site of the eukaryotic large ribosomal subunit, preventing the initial elongation step in protein synthesis.[11-14] The specificity of harringtonine alkaloids has also made it a critical research tool in ribosomal profiling.[15]

Despite its biomedical value, HHT remains difficult to access. Natural levels of HHT in *Cephalotaxus* species are extremely low (~0.06-3.2 mg/g dry weight),[16,17] and its current commercial production primarily relies on a semi-synthetic route involving the extraction of CET from *Cephalotaxus* needles, followed by chemical esterification with a synthetic side chain.[1,18-20] While total syntheses of CET and HHT have been reported,[18,21,22] these approaches are limited by multi-step procedures and low yields, making them commercially less favorable. Furthermore, *Cephalotaxus* species have been classified as endangered due to extensive harvesting, inherently slow growth rate, and long seed development time.[1,2,23] Since late 2024, although still listed in FDA-approved drug, HHT has been discontinued in the U.S. in part due to manufacturing limitations,[24-26] underscoring the need for alternative production platforms. Metabolic engineering of *Cephalotaxus* alkaloid biosynthesis into tractable heterologous hosts could provide a sustainable, scalable production of CET and HHT; however, the development of such solutions has been impeded by a fundamental lack of knowledge of their biosynthetic pathways. Furthermore, knowledge of biosynthetic genes would enable studies to determine the specific biological role of these alkaloids in the producing species.

Early radioisotope labeling studies in actively growing *Cephalotaxus harringtonia* demonstrated that CET is derived from the amino acids tyrosine[27,28] and phenylalanine[28,29] (Figure 1). Based on these findings—and the minor presence of homoerythrina-type alkaloids in *Cephalotaxus* species—it was hypothesized that a phenethylisoquinoline scaffold is an intermediate of CET biosynthesis, formed from tyrosine and phenylalanine via a pathway analogous to those described for *Colchicum* and *Schelhammera* alkaloids.[2,27-29] However, the failure of cinnamic acid to incorporate into CET has questioned the validity of this hypothesis.[2,28] More recently, a phenethylisoquinoline scaffold (**1**) was reported in *C. hainanensis*, along with upstream biosynthetic enzymes capable of converting tyrosine and phenylalanine into **1**,[30] resembling the early steps in colchicine biosynthesis.[31] These findings support a biosynthetic pathway model that hinges on this Pictet-Spengler metabolite, but to date, there is no direct evidence that this compound **1** is a precursor to CET or HHT alkaloids (**Fig. 1**) and no additional pathway intermediates have been identified. Moreover, CET and HHT are found distributed throughout the plant body, and thus a site of active biosynthesis has remained elusive.[2,16] Collectively, these gaps have made elucidating the biosynthetic pathways of *Cephalotaxus* alkaloids particularly challenging.

Here, we address these long-standing challenges by generating paired metabolomic (via isotope-labeled precursor feeding) and transcriptomic datasets across diverse tissues of *C. harringtonia* harvested during its actively growing season. This integrated approach revealed a site of active biosynthesis of CET, uncovered previously unknown key intermediates, and ultimately enabled the discovery of a set of highly co-expressed biosynthetic enzymes responsible for biosynthesis of the CET core.

MAIN TEXT

Labeling studies identify growing root tips as a site of active biosynthesis of CET and specific phenethylisoquinolines as intermediates in the pathway

In studying biosynthetic pathways of metabolites of interest in non-model plants like *Cephalotaxus harringtonia*, where genomic resources and genetic manipulation tools are limited, transcriptomics with co-expression analyses has often been used when the spatial, temporal, or condition-specific production of the compounds is known.[32-35] However, sequencing a plant transcriptome without prior knowledge of when and where the pathway is expressed may not capture relevant biosynthetic genes. Moreover, knowledge of the correct biosynthetic intermediates is also crucial for the discovery of biosynthetic genes. In our prior work using a D_2O labeling study, we proposed that HHT is actively biosynthesized in new growth needles of *Cephalotaxus*. However, it was unclear whether the entire pathway or only a portion (i.e. the final steps) is active in these tissues.[36] Given that our efforts to identify biosynthetic genes expressed in these tissues were unsuccessful,[37] we considered whether parts of the

pathway were expressed in other plant tissues. Thus, we first attempted to identify a site of active biosynthesis for the alkaloid core, CET, specifically and elucidate potential intermediates in the pathway with stable-isotope precursor feeding experiment across different tissues of *C. harringtonia*.

To collect various tissues of *C. harringtonia*, we harvested plant tissues (young inner needles—YIN, young outer needles—YON, old needles—ON, young stem—YSt, old stem—OSt, and root tips—Rt) when the plant appears to be actively growing—this was typically spring[16,38,39] (**Fig. 2a**). We then first compared natural accumulation of CET and HHT in different tissues using liquid chromatography-mass spectrometry (LC-MS). Both CET (**Fig. 2b**, left) and HHT (**Fig. 2b**, right) were detected at high levels in all different tissues collected in May 2022, with lowest in YSt. CET was about 50- to 200-fold higher in amount than HHT throughout the plant (**Supplementary Fig. 1a**). Tissue samples collected in May 2023 were also analyzed and compared (**Supplementary Fig. 1b** and **1c**). In both sample sets, both CET and HHT are generally abundant in all tissues with no consistent pattern of accumulation.

To determine whether CET and HHT are actively made in all different tissues, as detected, or synthesized in a specific tissue and transported throughout the plant, we next conducted a stable isotope precursor feeding experiment with samples from various parts of an actively growing plant. Sectioned tissues of *C. harringtonia* were incubated in a solution of dopamine-D_4 (deuterium-labeled alkyl groups), a predicted early intermediate in CET biosynthesis based on the prior isotope labeling study that showed tyrosine serving as an upstream precursor[27,28] and on the hypothesis that CET biosynthesis involves a phenethylisoquinoline scaffold.[2,27,30] Interestingly, enrichment of CET-D_4 ($[M_0 + H_4 + H]^+ = m/z$ 320.1794) was observed only in the root tips (up to length of approximately 5-8 cm)—Rt, and not in other tissues (**Fig. 2c** and **Supplementary Fig. 2**). These results support a model where CET biosynthesis from dopamine occurs at the root tips of an actively growing plant, contrasting with the widespread accumulation pattern of CET, and that dopamine serves as a key pathway precursor.

We then performed an untargeted metabolite analysis (XCMS) comparing dopamine-D_4-fed root tips with water-fed controls to search for pathway intermediates. A ranking using increasing fold change in abundance between the two conditions (fold change > 3.5 and $P < 0.2$), with *m/*z in between 200 and 600, revealed several deuterium (D_4) labeled metabolites that may serve as potential intermediates in CET biosynthesis. Notably, these mass features each have an *m/z* close to CET and a large fold-change relative to others in the list of 148 unique mass features found (**Fig. 2d** and Supplementary **Fig. 1d**). We searched for both the deuterated and unlabeled isotopes of these molecules using extracted ion chromatograms (EICs) from both dopamine-D_4 and water treated Rt and found that while the unlabeled isotopes were present in both sample sets, the deuterated isotopes accumulated only in dopamine-D_4-treated samples. These results indicate that these molecules not only arise from dopamine-D_4 but also are true metabolites of *C. harringtonia*. Interestingly, the first two mass signatures in the ranked list were found to correspond to specific phenethylisoquinoline scaffolds (**2**-D_4 and **3**-D_4) when compared to authentic synthesized standards (**Supplementary Fig. 3**). Unlike CET that was found throughout the plant body, **2** and **3** accumulated only in root tips (**Supplementary Fig. 1e-f**). Nonetheless, their D_4-enrichment (**2**-D_4 and **3**-D_4) from dopamine-D_4 feeding showed a similar pattern to that of CET-D_4, which occurs exclusively in growing root tips and not in other tissues (**Supplementary Fig. 1g-h**). These results suggested that **2** and **3** could be intermediates of CET biosynthesis and that they could represent sequential metabolism since they differ only by an O-methylation. Meanwhile, untargeted analyses also identified CET-D_4. Its relatively low fold change (3.8) compared to earlier potential intermediates could be attributed to its position further downstream in the pathway. Three additional mass signatures with large fold change and interesting *m/z* that are close to that of CET (**Fig. 2d**, red dots) were subsequently also found to be true intermediates of CET biosynthetic pathway (see **Fig. 4** and **Supplementary Figs. 17-18**), suggesting that feeding of early precursor (dopamine-D_4 in this case) combined with untargeted analyses using LC-MS effectively revealed intermediates of interest.

To determine whether **2** is an intermediate of CET biosynthesis, different tissues of actively growing *C. harringtonia* were sectioned and soaked in a solution of **2**-D_4 and analyzed with LC-MS in a manner similar to the dopamine-D_4 feeding study. Relative isotope enrichment of CET-D_4 in **2**-D_4 feeding compared to water treatment in six different tissues further supports the growing root tips as a site of active biosynthesis of CET and that **2**-D_4 serves as a pathway intermediate (**Fig. 2e**). Furthermore, unlike when dopamine-D_4 was fed, **3**-D_4 was observed in all different tissues when treated with **2**-D_4. Nevertheless, O-methylation of **2**-D_4 to **3**-D_4 predominantly occurred in the growing root tips compared to other tissues (**Fig. 2f** and **Supplementary Fig. 1i**), following a similar pattern of high **3**-D_4 enrichment in root tips from dopamine-D_4 feeding (**Supplementary Fig. 1h**).

Since both dopamine and **2** are incorporated into **3** and appear to be involved in CET biosynthesis in root tips, we next sought to determine whether **3** is a direct intermediate in the CET biosynthetic pathway rather than a side-product or unrelated metabolite. Similar feed-in experiments and analyses done with dopamine-D_4 and **2**-D_4 were conducted with **3**-D_4. Enrichment of CET-D_4 from **3**-D_4 feeding occurred only in growing root tips and not in other tissues sampled (**Fig. 2g** and **Supplementary Fig. 1j**), indicating that **3** serves as a true intermediate in CET biosynthesis and further supporting that a site of active biosynthesis of CET is growing root tips and not other tissues. Furthermore, our identification of **2** and **3** reveals direct evidence for CET biosynthetic intermediates beyond the tyrosine and phenylalanine precursors that had been identified in the 1970s[27-29] and provides experimental evidence for the long-standing hypothesis that phenethylisoquinolines act as intermediates in CET biosynthesis.[2,27]

Because **2** and **3** were both found to be intermediates of CET and phenethylisoquinolines, especially compound **1**, have long hypothesized to be intermediates of CET,[2,27] we also searched for other possible phenethylisoquinolines such as **1**, **11**, **12**, and **13** to examine their biological relevance, and whether they could serve as potential intermediates in the CET pathway (**Supplementary Figs. 4-5** and **SI Table 1**). Although natural accumulation of **1**[30] and **11** were observed in root tips of *C. harringtonia*, no enrichment of **1**-D_4 nor **11**-D_4 were detected from feeding dopamine-D_4 (**Supplementary Fig. 4**). Moreover, neither **12** nor **13** were detected anywhere in any tissues of *C. harringtonia* sampled (**Supplementary Fig. 5**). We also did not detect deuterium enrichment (**12**-D_4 and **13**-D_4) from feeding dopamine-D_4. Feeding of **1**-D_4 and **12**-D_4 in different tissues sampled from *C. harringtonia* also did not result in enrichment of CET-D_4 (**Supplementary Figs. 4-5**). These findings collectively suggested that CET is derived from **2**, which is formed through a Pictet-Spenger-like condensation of dopamine and 3-(4-hydroxy-3-methoxyphenyl)propanal (**Fig. 2h** and **SI Table 1**). Compounds **1** and **11** are likely not on pathway to CET but represent additional alkaloids that accumulate in *Cephalotaxus* roots but were not actively made during the time of sampling.

Additionally, unlike CET, no enrichment of deuterated HHT (HHT-D_4) was detected in any of the sectioned tissues of *C. harringtonia* fed with dopamine-D_4, **2**-D_4, or **3**-D_4. This result may be interpreted in two ways: i) the feed-in assay failed to capture HHT biosynthesis despite root tips being the sole site of HHT production, or ii) the alkaloid core of HHT gets made only in roots and is transported to other parts of the plant, where side-chain biosynthesis and its attachment occur to produce HHT. The widespread availability of both CET and HHT in all different tissues likely support the latter model.

Collectively, the stable-isotope labeled precursor feeding studies with various tissues sectioned from an actively growing *C. harringtonia* revealed that the root tips are the primary site of CET alkaloid core biosynthesis, which does not align with the widespread accumulation pattern of CET and HHT observed throughout the plant. The biosynthetic pathway begins with tyrosine[27,28] and phenylalanine[28,29] and proceeds through the intermediates dopamine, **2**, and **3**. Compound **3** likely undergoes oxidative coupling to form compound **6** (a proposed intermediate with a mass signature identified via XCMS analyses; **Fig. 2d**), followed by a series of oxidations, including carbon excision, and reduction to ultimately produce CET (**Fig. 2h**). Our data suggest that the alkaloid core of HHT is biosynthesized and subsequently transported from root tips and distributed throughout the plant, where side-chains are

185 attached when needed to produce the toxic ester-derivatives such as HHT.[1,2,4] The availability of
186 downstream intermediates (**2** and **3**) exclusive to root tips further supports our model. Having identified **2**
187 and **3** as intermediates in the pathway, we next sought to identify a candidate gene in biosynthesis that
188 could be used as a bait gene to shortlist candidate CET biosynthetic enzymes.

189

190 **Paired transcriptomics and metabolomics analysis identifies *Ch*OMT-1 as the biosynthetic gene and**
191 **reveals co-expressed candidate genes involved in CET biosynthesis**

192 Transcriptomics with co-expression analyses has often been used in elucidating plant biosynthetic
193 pathways when spatial, temporal, or condition-specific production of a compound of interest is known.[32-
194 35,40] With the knowledge that a site of active CET biosynthesis is specific to root tips, we sought to use
195 transcriptional analyses to identify CET biosynthetic enzymes. To do so, we first aimed to obtain a
196 transcriptome of *C. harringtonia* that fully captures CET biosynthetic genes by generating our own *in-*
197 *house* transcriptome paired with stable-isotope labeling studies on diverse tissues of actively growing *C.*
198 *harringtonia*, harvested in May 2023. Each tissue sample was divided for parallel profiling (one for
199 metabolomics and the other for transcriptomics), ensuring that the biosynthetically active root tip tissues
200 were analyzed using RNA-seq. We also utilized PacBio Iso-Seq to avoid technical limitations of de novo
201 transcriptome assembly, such as misassembled or fragmented contigs,[40,41] and obtain an accurate, full-
202 length transcriptome that could be used as a reference for short-read RNA-seq (29,163 transcripts after
203 processing). Illumina sequencing from multiple various plant tissues (a total of 36 biosynthetic and non-
204 biosynthetic samples) was then used to quantify relative gene expression.

205 Using these paired metabolomic and transcriptomic datasets across diverse tissues of *C.*
206 *harringtonia*, we sought to identify an initial enzyme in CET biosynthetic pathway that could be used as a
207 bait gene for co-expression analyses to discover CET biosynthetic enzymes. With the knowledge that O-
208 methylation of **2** to **3** is involved in CET biosynthesis in root tips, we proposed that the enzyme
209 responsible for this transformation of **2** to **3** would be a good first candidate to be used as the bait gene.
210 Because O-methylation of **2** to **3** mainly occurs in the root tips of the plant and *S*-adenosylmethionine-
211 dependent methyltransferase is a family of enzymes known for performing O-methylation,[31,42] we focused
212 on transcripts that had expression in roots and their Pfam annotated as methyltransferase. We then
213 prioritized our candidate gene list by conducting a Pearson's correlation ($r > 0.85$) between transcript
214 expression levels and **3**-D_4 enrichment across all 15 tissues fed with **2**-D_4. These analyses together
215 resulted in a shortlist of 18 methyltransferase-annotated transcripts, whose expression strongly correlated
216 with *in vivo* metabolic conversion of **2**-D_4 to **3**-D_4 (**Fig. 3a** and **Supplementary Fig. 6**).

217 A batch of 5 out of the 18 that were initially cloned from *C. harringtonia* cDNA was tested for
218 function using *Agrobacterium*-mediated transient expression in *N. benthamiana* leaves with direct co-
219 infiltration of **2** as a substrate.[31,43] LC-MS analysis of infiltrated leaf extracts revealed that one candidate
220 methyltransferase (YD-1-Rt_50321, highlighted in bold in **Fig. 3a**) consumed **2** and produced **3** (**Fig. 3b**
221 and **Supplementary Fig. 7**). This O-methyltransferase responsible for O-methylation of **2** to **3** was
222 referred to as *Ch*OMT-1.

223 *Ch*OMT-1 was then utilized as the bait gene to identify a set of strongly co-expressing CET
224 biosynthetic gene candidates from 29,163 unique transcripts. Initial Pearson correlation analysis across all
225 36 tissue samples using *Ch*OMT-1 as the query gene identified 316 highly co-expressed transcripts ($r >$
226 0.98). As many of these accumulate primarily in root tips and are not detected in other tissues, a second
227 round of correlation analysis ($r > 0.90$) was conducted using Rt samples alone, enabling prioritization of
228 41 top candidates (**Fig. 3c** and **Supplementary Fig. 8**). This top-ranking candidate list is rich in genes
229 that are generally associated with secondary metabolism such as cytochrome P450s (CYPs),[31,33,44]
230 Fe(II)/2-oxoglutarate-dependent dioxygenases (2OGDs),[40,45,46] and dehydrogenase/reductase enzymes,[47,48]
231 suggesting that it contains CET biosynthetic enzymes responsible for oxidations and reductions that are

232 required to form CET from **3**. Furthermore, the list contains early CET biosynthetic genes that encode
233 enzymes proposed to generate dopamine from tyrosine (*Ch*TyDC-1 and *Ch*PPO-1).[30]

234

235 **Discovery and engineering of *Cephalotaxus* alkaloid biosynthesis**

236 With a shortlist of candidate CET biosynthetic genes from co-expression analysis (**Fig. 3c**) and
237 unique masses of the next top two potential intermediates (**6** and **5**) after **2** and **3** identified from
238 untargeted metabolite analysis (XCMS) of stable-isotope precursor feeding study (**Fig. 2d**), we next
239 sought to elucidate the downstream biosynthetic pathway of CET after **3**. To eventually produce CET, we
240 hypothesized that **3** would go through a series of oxidations such as oxidative coupling of two
241 methoxyphenols to form **6** and carbon excision from the lower six-membered ring C to form the five-
242 membered ring C' of CET (**Fig. 2h** and **Fig. 4b**).

243 We speculated that **3** first undergoes oxidative transformation: hydroxylation, oxidative coupling,
244 or both. All the oxidases (10 CYPs and 3 2OGDs)[33,46,49-51] in the shortlist of candidate genes (**Fig. 3c**)
245 were each cloned and tested by Agrobacterium-mediated transient transformation in *N. benthamiana* with
246 **3** co-infiltrated as a substrate. However, LC-MS analysis of leaf extracts revealed that none of the
247 candidates resulted in additional hydroxylation ($[M + H]^+$ = *m/z* 346.1649), oxidative coupling ($[M + H]^+$
248 = *m/z* 328.1543), or both ($[M + H]^+$ = *m/z* 344.1492) of **3**. However, one of the tested enzymes
249 (*Ch*CYP805A11) resulted in partial consumption of **3**. We therefore conducted untargeted metabolite
250 analysis (XCMS, see Methods) comparing *Ch*CYP805A11-expressed leaves to GFP-expressed leaves
251 (negative control), both co-infiltrated with **3**, to examine if any relevant unique mass with a significantly
252 increasing fold change ($\Delta > 5$) in abundance between the two conditions ($P < 0.1$) appeared
253 (**Supplementary Fig. 9**). XCMS analysis reported 27 unique masses, of which 7 are above the signal to
254 noise threshold and can be identified in the raw LC-MS data. These seven (**Supplementary Fig. 9d**, red
255 dots) all represent isotopologues or ions of a single compound ($[M + H]^+$ = *m/z* 348.1805, r.t. = 7.07 min.)
256 that corresponds to a single hydroxylation and reduction of **3**. Based on the MS/MS analyses of both the
257 molecule and its four-deuterated form from **3**-D_4 feeding in root tips of *C. harringtonia* experiment, the
258 compound was assigned the structure of **4** (**Supplementary Fig. 9**). These results suggested that
259 *Ch*CYP805A11 catalyzes oxidative coupling of two methoxyphenols of **3** (rings A and C), forming **4'**
260 (**Fig. 4**). However, due to its unstable imine ring, **4'** likely undergoes imine hydrolysis to form **4''** that
261 contains a very reactive aldehyde moiety, which then gets reduced by a promiscuous reductase in *N.*
262 *benthamiana* to form **4** (**Supplementary Fig. 9h**). The instability of **4'** could explain why this molecule
263 was not detected in LC-MS but instead, a mass corresponding to a single hydroxylation and reduction of
264 **3** (**4**, $[M + H]^+$ = 348.1805 Da) was observed. This observation was consistent with metabolites observed
265 in the native plant *C. harringtonia*, indicating that the promiscuous reductase is not specifically present
266 only in *N. benthamiana* but also in other plants in general.

267 To form the fused bicyclic backbone of CET (rings B' and D), it has been speculated that a
268 precursor similar to **5**, formed through oxidative coupling of phenethylisoquinoline, undergoes an
269 intramolecular cyclization via nucleophilic attack by the nitrogen atom,[2,52,53] indicating that the ring B'' of
270 **4'** needs to be preserved (**Fig. 4b**). We thus hypothesized that there exists a reductase in CET biosynthesis
271 for reducing the imine bond of **4'** to form the stable compound (**5**) and that **4** is not a direct intermediate
272 but a side-product of CET biosynthesis. To evaluate our hypothesis, all five
273 dehydrogenases/reductases[47,48,54,55] in the shortlist of candidate genes (**Fig. 3c**), including an atypical
274 short-chain dehydrogenase (SDR)—nitrogen metabolite repression regulator (NmrA)-like family
275 enzyme[56-60] (*Ch*NmrA-1), were each cloned and co-expressed with *Ch*CYP805A11 in *N. benthamiana*
276 leaves using Agrobacterium-mediated gene delivery, with **3** co-infiltrated as a substrate. LC-MS analysis
277 of the *Ch*NmrA-1 infiltrated leaf extracts revealed partial reduction of **4** and production of a new peak
278 ($[M + H]^+$ = *m/z* 330.1700, r.t. slightly left to that of **3**) (**Supplementary Fig. 10**). This new peak
279 corresponded to the compound with the fourth largest fold-change in earlier XCMS analysis comparing

280 dopamine-D_4-fed root tips with water-fed controls of *C. harringtonia* (**Fig. 2d**), suggesting that it is likely
281 to be an intermediate of CET biosynthesis. MS/MS analyses of both the new molecule and its four-
282 deuterated form were used to assign the structure as **5** (**Supplementary Fig. 10**). Collectively, the results
283 support our hypothesis that **3** goes through a *Ch*CYP805A11 mediated oxidative coupling to form an
284 unstable intermediate **4'**, which gets reduced by *Ch*NmrA-1 to produce **5**.

285 To further verify the functionalities of *Ch*CYP805A11 and *Ch*NmrA-1, *in vitro* assays with yeast
286 microsomal protein fractions of *Ch*CYP805A11 and purified *Ch*NmrA-1 were conducted
287 (**Supplementary Fig. 11**). Unlike in *N. benthamiana*, no production of **4** was detected in *in vitro* assay of
288 purified yeast microsomes of *Ch*CYP805A11 with **3**. However, when both the microsomes of
289 *Ch*CYP805A11 and purified *Ch*NmrA-1 were combined with **3** *in vitro*, production of **5** was observed
290 using LC-MS. These findings collectively confirmed that *Ch*CYP805A11 acts on **3**, producing an
291 unstable intermediate, speculated to be **4'**, which gets reduced to either **4** or **5** in the presence of a
292 promiscuous plant reductase or *Ch*NmrA-1 respectively, and that **4** is not a substrate for *Ch*NmrA-1 to
293 produce **5** but instead a side-product.

294 As discussed earlier, we next speculated that **5** goes through an oxidative intramolecular
295 cyclization via nucleophilic attack by the nitrogen atom to produce the fused bicyclic backbone of
296 CET[2,52,53] (**Fig. 4b**). Therefore, the rest of 9 CYPs and 3 2OGDs in the shortlist of candidate genes (**Fig.**
297 **3c**) were each tested in combination with *Ch*CYP805A11 and *Ch*NmrA-1 in *N. benthamiana* leaves with
298 co-infiltrating **3** as a substrate. Addition of *Ch*CYP805A12 resulted in consumption of **5** and production
299 of a new peak, proposed to be **6** based on its MS/MS fragmentation pattern (**Supplementary Fig. 12**).
300 This was the first intermediate identified containing a tertiary amine and a bicyclo[5.3.0]decane core
301 (rings B' and D) that are crucial to the structural backbone of CET. Compound **6** also corresponded to the
302 molecule with third largest fold-change in earlier XCMS analysis comparing dopamine-D_4-fed root tips
303 with water-fed controls of *C. harringtonia* (**Fig. 2d**), indicating that LC-MS-based untargeted analyses of
304 early precursor feeding studies in dissected tissues of a native plant could be used as an effective tool to
305 identify biosynthetic intermediates.

306 To produce CET from **6**, a carbon excision from ring C and formation of methylenedioxy bridge
307 in ring E are necessary (**Fig. 4b**). Because CYPs and 2OGDs have been known to catalyze such unusual
308 oxidative transformations,[40,49,50,61,62] the remaining 8 CYPs and 3 2OGDs in the shortlist of candidate
309 genes (**Fig. 3c**) were again each tested in the transient expression system, producing **6**, and analyzed using
310 LC-MS/MS. Interestingly, each *Ch*2OGD-1 and *Ch*CYP805A13 resulted in consumption of **6** and
311 formation of different compounds, **7** (**Supplementary Fig. 13**) and **8** as well as **9** (**Supplementary Fig.**
312 **14**) respectively based on their MS/MS analyses. Co-expression of both enzymes with *Ch*CYP805A11,
313 *Ch*NmrA-1, and *Ch*CYP805A12 in *N. benthamiana* leaves co-infiltrated with **3**, led to consumption of **7**
314 and **8** (including **9**) and production of a new peak ($[H+M]^+$ = *m/z* 314.1387) that was confirmed to be
315 cephalotaxinone (**10**), a *Cephalotaxus* alkaloid,[1,17] using LC-MS/MS comparison to a purified
316 cephalotaxinone standard (**Supplementary Fig. 15**). The ability to convert **6** ($[H+M]^+$ = *m/z* 328.1543) to
317 **7** ($[H+M]^+$ = *m/z* 316.1543), and **8** ($[H+M]^+$ = *m/z* 326.1387) to cephalotaxinone ($[H+M]^+$ = *m/z*
318 314.1387)—both involving a 12 Da loss corresponding to one carbon—demonstrates that *Ch*2OGD-1 is
319 responsible for performing the remarkable carbon excision reaction that contracts ring C to C', a key step
320 in the biosynthesis of CET (**Supplementary Figs. 13 and 15**).

321 Moreover, conversion of **6** to **8** and **7** to cephalotaxinone illustrated that *Ch*CYP805A13 catalyzes
322 methylenedioxy bridge formation to produce the ring E of CET (**Supplementary Figs. 14-15**). When
323 *Ch*CYP805A13 acted on **6**, it produced not only **8** but also **9**, a side-product in equilibrium with **8** that
324 likely arose from the spontaneous unfolding of the tertiary amine in **8** to the secondary amine of **9**
325 (**Supplementary Fig. 14**). A similar phenomenon was observed with **5** existing in equilibrium with **6** in
326 *N. benthamiana* leaves co-expressed with *Ch*CYP805A11, *Ch*NmrA-1, and *Ch*CYP805A12 and co-
327 infiltrated with **3**. Although **5** is largely consumed by *Ch*CYP805A12 to form **6**, the presence of **5** in

328 small amounts suggested that **6** can spontaneously open back to **5** (**Fig. 4b-c** and **Supplementary Fig.**
329 **12**). Such spontaneous ring-opening of the tertiary amine was no longer observed upon addition of
330 *Ch*2OGD-1, which locks the tertiary amine structure by preventing dienone-phenol rearrangement
331 through a carbon excision of ring C (**Supplementary Figs. 13** and **15**).

332 The five enzymes in *C. harringtonia* identified through co-expression analysis using *Ch*OMT-1
333 as the bait gene establish a biosynthetic pathway of CET from **3** to cephalotaxinone (**Fig. 4** and
334 **Supplementary Fig. 16**). The order of the pathway was validated through sequential addition of each
335 enzyme (**Supplementary Fig. 16a-b**) and dropout experiments in which each individual enzyme was
336 removed from the engineered pathway to cephalotaxinone (**Supplementary Fig. 16c**). The sequential
337 accumulation of proposed pathway intermediates with no detection of alternative mass signatures also
338 supported that the overall biosynthetic pathway of CET from **3** to cephalotaxinone is linear, with the final
339 steps involving *Ch*2OGD-1 and *Ch*CYP805A13 may occur in bifurcated sequence (**Supplementary Fig.**
340 **16**). Reconstitution of CET biosynthetic pathway from **2** to cephalotaxinone in *N. benthamiana* was also
341 established with co-expression of all the 6 newly discovered biosynthetic genes in this study and co-
342 infiltration of **2** (**Supplementary Fig. 17**). Furthermore, the identities of all the newly discovered
343 intermediates (**3**, **5**, **6**, **7**, **8**, and cephalotaxinone) and side-products (**4** and **9**) produced heterologously in
344 *N. benthamiana* were supported by co-elution with metabolites found in *C. harringtonia* root tips that
345 exhibited identical MS/MS fragmentation patterns (**Supplementary Fig. 18**). This indicated that the
346 compounds produced in the *N. benthamiana* expression system are biologically relevant pathway
347 intermediates. In addition to the authentic synthetic or purified standards used to confirm the structures of
348 **2**, **3**, and cephalotaxinone, structural assignments of other proposed intermediates were all supported by
349 LC-MS/MS analyses of the molecules and comparison of their MS/MS fragmentation patterns to those of
350 the corresponding four-deuterated forms from dopamine-D_4, **2**-D_4, or **3**-D_4 feeding study or to those of the
351 related CET biosynthetic intermediates (**Supplementary Figs. 9-15**).

352 Moreover, all the identified intermediates were only observed or highly detected in growing root
353 tips of *C. harringtonia* relative to other tissues (**Fig. 4a** and **Supplementary Fig. 17c**). The exception is
354 cephalotaxinone, which is found throughout the plant body. This accumulation of CET intermediates in
355 growing root tips aligned with our findings of the core biosynthesis of HHT (CET) occurring at the root
356 tips. However, the wide distribution of not only CET and HHT (**Fig. 2b**) but also cephalotaxinone (**Fig.**
357 **4a**) throughout the native plant as well as reversible conversion between cephalotaxinone and CET from
358 early radioisotope labeling studies[28] indicated that it is cephalotaxinone that gets primarily made in the
359 growing root tips and is transported throughout the plant. Thus, reduction of cephalotaxinone to CET
360 most likely occurs not only in root tips but also in other tissues (**Fig. 5**). Collectively, we have elucidated
361 a biosynthetic pathway to cephalotaxinone, which was barely studied and uncovered. We discovered six
362 novel biosynthetic enzymes that perform unique transformations, seven pathway intermediates, and two
363 side-products in CET biosynthesis from **2** to cephalotaxinone, which is primarily produced in growing
364 root tips of *C. harringtonia* and then gets transported throughout the plant for conversion into CET and
365 HHT.

366

367 **DISCUSSION**

368 Plants have evolved to produce diverse, complex secondary metabolites for ecological
369 interactions (i.e. to attract pollinators and defend against pests and predators).[63] The biosynthesis of these
370 specialized toxic defense metabolites is strategically regulated and coordinated, often being stress-induced
371 and tissue-specific.[33,64-67] By producing these defense molecules in specific tissues, storing in non-toxic
372 forms or in separate compartments, transporting to target locations, or converting into their active toxic
373 forms when needed, plants effectively defend themselves and avoid autotoxicity.[33,34,65-69]

374 Our study demonstrates careful regulation and coordination of biosynthesis of CET, the core
375 scaffold of the ribosome inhibitor HHT, by the plant. *Cephalotaxus* species are slow-growing conifers
376 that follow a seasonal growth cycle, with bud expansion beginning in late March, shoot emergence in late
377 April, and rapid shoot elongation occurring throughout May, followed by winter dormancy.[16,38,39] Our
378 stable-isotope labeled precursor feeding experiments with sectioned tissues of *C. harringtonia* showed
379 that this seasonal cycle strongly influences CET biosynthesis. In tissues collected during active shoot
380 growth in May, incorporation of dopamine-D_4 and **2**-D_4 into CET-D_4 in root tips was observed, whereas
381 sectioned tissues from dormant plants harvested in November and December showed no enrichment of
382 CET-D_4 (**SI Table 2**). Furthermore, our *in-house* transcriptome generated from tissues of actively
383 growing *C. harringtonia* collected in May 2023 captured all CET biosynthetic transcripts, whereas our *in-*
384 *house* transcriptome generated from tissues harvested after the spring growth period (August 2018) had
385 all the CET biosynthetic transcripts missing (**Supplementary Fig. 8**). Although potential limitations of
386 feeding assays and technical errors in RNA sampling from August 2018 cannot be entirely excluded, the
387 consistent detection of CET-D_4 enrichment in root tips of an actively growing *C. harringtonia* harvested
388 in May 2022, May 2023, and June 2024, together with transcriptome data from May 2023, supports that
389 CET biosynthesis is seasonally activated during active spring growth, particularly in May and June.

390 Our findings further illustrate strategic coordination of *C. harringtonia* to produce *Cephalotaxus*
391 alkaloids through whole plant coordination metabolism. The stable-isotope labeled precursor feeding
392 studies and metabolite profiling of various tissues of actively growing *C. harringtonia* revealed that the
393 root tips are the primary site of CET alkaloid core biosynthesis, contrasting with the widespread
394 accumulation pattern of cephalotaxinone, CET, and HHT observed throughout the plant (**Fig. 2** and **Fig.**
395 **4a**). The non-toxic cephalotaxinone[1,70] gets synthesized in the growing root tips and transported
396 throughout the plant in an inert form, which can later be reduced to CET[28]—an apparent non-toxic
397 precursor[2,4] poised for esterification with different side-chains to generate toxic defense compounds such
398 as HHT[1,2,4] when and where needed (**Fig. 5**). Together, our findings in *C. harringtonia* highlights a model
399 in which a plant employs whole-organism metabolic coordination to carefully produce and store toxic
400 defense metabolites, thereby ensuring effective defense while minimizing self-toxicity and unnecessary
401 energy expenditure.

402 It is notable that HHT is one of the few known plant-derived small molecule NPs that inhibit
403 eukaryotic protein translation, whereas most known plant-derived inhibitors of this process are
404 proteins.[71,72] Compared to prokaryotes, far fewer eukaryotic translation inhibitors are known in plants.[13,71]
405 Notable plant NPs other than HHT that act as eukaryotic protein translation inhibitors by binding to the
406 large ribosomal subunit include ricin, a protein from castor beans,[73] as well as lycorine and narcisclasine,
407 small molecules from *Amaryllidaceae* family.[13,71] Interestingly, *Amaryllidaceae* species have been shown
408 to exhibit spatiotemporal regulation of the biosynthesis of defense molecules like lycorine,[33] which may
409 represent a strategy to mitigate potential autotoxicity—similar to the coordinated CET/HHT biosynthetic
410 pathway observed in *Cephalotaxus*. These examples underscore how plants have evolved to produce a
411 class of eukaryotic translation inhibitors as defense compounds without disrupting their own ribosomal
412 function through sophisticated regulation and coordination of the toxic defense metabolite biosynthesis.

413 Our study elucidates the long-elusive biosynthetic pathway to cephalotaxinone—the core alkaloid
414 scaffold of an FDA-approved anti-cancer drug HHT. Stable-isotope labeled precursor feeding, integrated
415 with tissue-resolved metabolite profiling and transcriptomic co-expression analysis, enabled us to identify
416 growing root tips as a site of active HHT alkaloid core biosynthesis, discover seven CET pathway
417 intermediates, and functionally characterize six novel biosynthetic enzymes. Leveraging this paired, *in-*
418 *house* generated metabolomic and transcriptomic dataset across diverse tissues of actively growing *C.*
419 *harringtonia* was critical in fully capturing the spatiotemporally expressed CET biosynthetic genes and
420 identifying co-expressed genes in biosynthetically active tissues, demonstrating a powerful framework for
421 pathway discovery in genetically intractable, slow-growing non-model plants. Furthermore, our findings
422 suggest that *C. harringtonia* employs a sophisticated whole-plant coordination strategy to produce the

toxic ribosomal inhibitor HHT, offering deeper insights into how certain plants regulate the biosynthesis of eukaryotic ribosomal toxins as a defense mechanism while minimizing autotoxicity. Overall, our work not only establishes a metabolic route to the core scaffold of HHT, enabling future sustainable, large-scale production of this clinically essential drug, but also expands our understanding of whole-plant coordination for biosynthesis of specialized metabolites and the unique biochemical transformations catalyzed by plant enzymes to generate these complex, bioactive molecules.

METHODS

Chemicals and reagents

All biological and chemical reagents were purchased from commercial vendors, unless stated otherwise. These include dopamine (MilliporeSigma), 3-hydroxy-4-methoxyphenethylamine (Combi-Blocks), 3-methoxy-4-hydroxyphenethylamine (Combi-Blocks), 3-(4-hydroxyphenyl)propanal (Chemscene), 3-(3,4-dihydroxyphenyl)propanal (Astatech), 3-(4-hydroxy-3-methoxyphenyl)propanal (Astatech), dopamine-D_4 (1,1,2,2-D_4; Cambridge Isotope Laboratories), cephalotaxine (Selleckchem), and homoharringtonine (MilliporeSigma).

Plant growth

The *Cephalotaxus harringtonia* var. Fastigiata plants used to measure natural abundance of their secondary metabolites and to conduct stable isotope labeled precursor feeding experiments were purchased from Forest Farms (forestfarm.com). They were grown in pots with soil, watered approximately once every week. Growth conditions consisted of ambient laboratory temperature and lighting.

The *N. benthamiana* plants used for heterologous gene expression were sown and grown in PRO MIX HP Mycorrhizae soil (Premier Tech Horticulture) on growth shelves with a 16/8-h light/dark cycle at ambient laboratory temperature. 4-5 weeks old plants after seeding were used for *Agrobacterium*-mediated transformation.

Stable isotope labeled precursor feeding experiments to identify the site of active biosynthesis of CET

Various tissues from *C. harringtonia* var. Fastigiata were excised from the whole plant using ethanol-sterilized scissors. For young inner needles (YIN), actively growing inner young light-green yellowish needles of 0.5-2 cm in length were harvested. Young outer needles (YON) were actively growing outer light-green needles, 1.5-3 cm in length. For young stems (YSt), light-green stems with YIN and YON of 2-3 cm in length from the tip of the stem (meristem) was harvested; all the surrounding needles were removed using a sterile scalpel. Old needles (ON) were mature, dark green needles (≥1-2 years old), 4-7 cm in length. For old stems (OSt), at least 1-2 year-old mature dark green stems (4-5 cm in length) with ON removed were collected. Before harvesting root tips (Rt), they were thoroughly washed with water, quickly rinsed with 30% ethanol, and washed again with water as a final washing step. Then, roots of 5-8 cm in length from the root tip, including the root meristem, were collected. Needles and roots, due to their thin structure, were latitudinally split into equally sized sections for paired metabolomics (feeding experiments) and transcriptomics (RNA-seq). Stems were longitudinally sectioned into equal halves for the same paired analyses.

Tissues used for accumulated metabolite analysis and RNA extraction were immediately flash frozen in liquid nitrogen and stored at -80°C until further processing. For feeding experiments, sterile 12-

466 well plates were filled with 2 mL of an aqueous solution of 100 μM dopamine-D_4, **1**-D_4, **2**-D_4, **3**-D_4, **12**-
467 D_4, or merely deionized water (negative control); see **SI Table 2** for details on when and which
468 experiments were conducted. Each well contained 7-9 needles for YIN, 3 needles for YON, 1 stem for
469 YSt, 3 needles for ON, 1 stem for OSt, and 4-5 roots for Rt, with all tissues cut into smaller pieces to
470 maximize direct contact with the substrate solution. For Rt, the 12-well plates were wrapped in aluminum
471 foil to simulate the darkness experienced by roots growing underground. The assay was conducted for
472 five days, with the solution replaced by freshly prepared substrate solution halfway through (after ~2.5
473 days). Following the remaining ~2.5 days, sectioned tissues were collected into pre-weighed 2 mL
474 Eppendorf Safe-Lock tubes, flash frozen in liquid nitrogen, and lyophilized to dryness. Samples were then
475 processed for analysis using LC-MS (see *Metabolite extraction from plant material and LC-MS analysis*
476 *of metabolites* in Methods). All feeding experiments were performed on tissues from the same plant.

477

478 *RNA extraction*

479 Flash-frozen *C. harringtonia* tissues (from May 2023 paired feeding experiments) were
480 homogenized to a fine powder under liquid nitrogen using a chilled mortar and pestle that were pre-
481 treated with RNAse Zap RNAse Decontamination Solution (Thermo Fisher Scientific). RNA from the
482 powdered tissue samples were extracted using RNeasy Plant Mini Kit (Qiagen) with on-column DNase
483 digestion. RLC buffer was used as the lysis buffer. For root samples, the extracted RNA was further
484 cleaned up using the same kit. RNA concentration and purity were measured using a NanoDrop and Qubit
485 (Qubit® 2.0 Fluorometer), and quality was further assessed using an Agilent 2100 Bioanalyzer.

486

487 *Preparation and sequencing of Illumina NovaSeq 6000 libraries and processing*

488 Library preparation for Illumina NovaSeq 6000 (S4 Flow cell – PE150) sequencing was
489 performed using the NEBNext® Ultra™ II RNA Library Prep Kit for Illumina® (non-directional; New
490 England Biolabs), as per the manufacturer instructions. Individual libraries were barcoded with internal
491 index sets at Novogene. For fragmentation, Diagenode Bioruptor Pico was used, and inserts were size
492 selected with NEBNext beads to have an average insert size of 150-200 bp. The PCR products were
493 purified (AMPure XP system). The quality and size distribution of each library were confirmed with
494 analysis using a High Sensitivity DNA chip on a 2100 Bioanalyzer (Agilent). Each library was quantified
495 using Qubit (Qubit® 2.0 Fluorometer) and real-time PCR (Kapa). Equimolar amounts of libraries for each
496 sample (36 total, from May 2023 feeding experiments) were pooled for paired-end sequencing within a
497 single Novaseq 6000 lane (S4 Flow cell – PE150).

498 The quality of raw Illumina sequencing reads was evaluated using FastQC (Babraham
499 Bioinformatics, Babraham Institute). These reads were processed to remove adapters, barcodes, and low-
500 quality sequences using Trimmomatic,[74] using the following parameters: <ILLUMINACLIP:TruSeq3-
501 PE-2.fa:2:30:10 LEADING:5 TRAILING:5 SLIDINGWINDOW:4:5 MINLEN:50>. The trimmed reads
502 were then re-assessed with FastQC to confirm quality.

503 While the processed Illumina sequencing reads were used to quantify each transcript abundance
504 against PacBio Iso-Seq transcriptome, we also generated a *de novo* transcriptome assembly from Illumina
505 reads using Trinity[75] with default settings to leverage the higher sequencing depth and broader range of
506 tissues provided through Illumina sequencing, resulting in generation of 541,990 contigs. Quality of the
507 newly assembled transcriptome was assessed by mapping Illumina short reads to the transcriptome using
508 bowtie2[76] and through Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis for
509 completeness.[77] Then, peptide sequences, with longest ORF identified for each transcript, were predicted
510 using Transdecoder,[75] which resulted in 144,299 contigs. Redundant contigs were removed using CD-

511 HIT-EST[78] with 99% identity threshold, yielding a final Illumina *de novo* transcriptome of 102,958
512 contigs.

513

514 *Preparation and sequencing of PacBio Iso-Seq libraries and processing*

515 Preparation of PacBio Iso-Seq libraries were performed as previously described.[40] For Iso-Seq,
516 only a subset of tissues from the 36 total samples was sequenced: Rt1 (root tips), Rt6 (root tips), an equal
517 combination of Rt4 and Rt8 (root tips), an equal combination of YIN3 and YIN4 (young inner needles),
518 an equal combination of YSt3 and OSt3 (young and old stems), and an equal combination of ON1 and
519 ON2 (old needles). Inclusion of Rt1, Rt,6 and Rt8—tissues with active CET biosynthesis—ensured that
520 the transcriptome contained CET biosynthetic genes. Additional tissues were included to represent all
521 other major tissue types beyond root tips, enabling capture of complete transcripts from *C. harringtonia*.

522 PacBio SMRT library preparation and sequencing were performed by QB3 Genomics, UC
523 Berkeley, Berkeley, CA (RRID: SCR_022170) on the PacBio Sequel II System using the Sequel II 3.1
524 binding kit with a 30-hour movie time. RNA concentration was adjusted to <5 ng/µL using the Qubit HS
525 RNA Kit, and RNA quality was assessed via capillary electrophoresis on an Agilent Bioanalyzer 2100
526 using the Eukaryotic Total RNA Pico Assay, as per the manufacturer instructions, with a 2 min
527 denaturation at 70°C followed by a 2 min cold shock on ice in a freezer. PacBio SMRT sequencing
528 library was prepared as described in the Iso-Seq™ Template Preparation for Sequel® Systems manual
529 using the NEBNext® Single Cell/Low Input cDNA Synthesis & Amplification Module and the Pacific
530 Biosciences Iso-Seq Express Oligo Kit with the barcoded primers provided in the manufacturer's protocol
531 for first-strand cDNA synthesis (standard purification) and PCR amplification. Barcoded cDNAs from six
532 representative samples (containing 10 unique tissue samples as discussed above) were pooled in
533 equimolar amounts, and a single library was prepared from this pooled cDNA using the SMRTbell Prep
534 Kit 3.0 following the standard protocol. The raw data were then processed using the Iso-Seq 2 pipeline in
535 SMRT® Link software (PacBio) with default settings, resulting in generation of 357,877 high quality
536 transcript isoforms. The quality of the Iso-seq transcriptome was assessed by mapping short read Ilumina
537 short reads to the long read transcriptome assembly using bowtie2[76] and through BUSCO analysis for
538 completeness.[77] Peptide sequences, with longest ORF identified for each transcript, were then predicted
539 using Transdecoder,[75] resulting in 349,712 contigs. Redundant contigs were removed using CD-HIT-
540 EST[78] with a cut-off at 99% identity, yielding a final PacBio Iso-Seq transcriptome of 29,163 unique
541 transcripts.

542 For the old *in-house* PacBio Iso-Seq transcriptome—derived from tissues (young and old needles,
543 old stem, and roots) harvested in August 2018 that lack CET biosynthetic genes and used as a comparison
544 to the newly prepared *in-house* transcriptome from tissues harvested in May 2023 (see **Supplementary**
545 **Fig. 8**)—library preparation and sequencing were performed by the Joint Genome Institute (JGI). The
546 2018 dataset was prepared and processed in the same manner as the May 2023 Iso-Seq transcriptome,
547 except for this difference in sequencing facility. The final processed transcriptome consisted of 48,657
548 unique transcripts.

549

550 *Analysis of RNA-seq data*

551 Both the final processed PacBio Iso-seq and *de novo* Illumina-generated transcriptomes (from
552 tissue samples in May 2023) were annotated with best-hit Pfam terms, followed by all subsequent (lower-
553 ranked) Pfam hits.[79] Annotations also included the best Basic Local Alignment Search Tool (BLAST) hit
554 from *Arabidopsis thaliana* proteomes[80] (via UniProt), *Taxus chinensis* proteomes,[81] *Cephalotaxus*
555 *hainanensis* proteomes,[30] and the *in-house* old Iso-seq transcriptome of *Cephalotaxus harringtonia*
556 (harvested in August 2018), using BLASTP. Quantification of processed Illumina reads against both the

557 PacBio and Illumina-generated transcriptomes was performed using Kallisto[82] with default parameters.
558 For all downstream expression analysis, log_2-transformed, TMM-normalized CPM expression values
559 were used. Ultimately, the PacBio Iso-seq-generated transcriptome (from tissue samples in May 2023
560 paired feeding experiment) was used for all analyses and candidate gene selection to avoid technical
561 limitations of *de novo* Illumina transcriptome assembly (i.e. misassembled or fragmented contigs).[40,41] For
562 example, *Ch*OMT-1, an essential CET biosynthetic gene we discovered, was detected in the PacBio Iso-
563 seq transcriptome but absent from the *de novo* Illumina assembly.

564

565 *Cloning of candidate biosynthetic genes*

566 cDNA was prepared from extracted RNA of *C. harringtonia* tissues harvested in May 2023 for
567 paired metabolomics and transcriptomics using the SuperScript™ IV First-Strand Synthesis System Kit.
568 Because the candidate biosynthetic genes were highly expressed in root tips tissues, they were cloned
569 mostly from cDNA generated from root tips of *C. harringtonia*. Coding sequences were PCR-amplified
570 using Q5® High-Fidelity DNA Polymerase (New England Biolabs) following the manufacturer's
571 protocol. Oligonucleotide primers were designed to anneal to the 5′ and 3′ ends of the coding sequence for
572 each target gene, incorporating flanking sequences homologous to the intended Gibson assembly sites of
573 the appropriate expression plasmid. Following PCR amplification, the PCR products were assessed using
574 gel electrophoresis on a 1% (w/v) agarose gel. PCR products were either purified using the DNA Clean &
575 Concentrator-5 kit (Zymo Research) prior to transient expression construct assembly or, when necessary,
576 were excised from the gel and purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research).

577 For transient expression construct assembly, pEAQ-HT plasmid was digested with AgeI *and*
578 XhoI restriction enzymes, and PCR amplicons were inserted into the digested pEAQ-HT vector using
579 NEBuilder® HiFi DNA Assembly Mix (New England Biolabs), as per the manufacturer instructions.
580 These Gibson assembly reactions were directly transformed into NEB 10β *Escherichia coli* cells (New
581 England Biolabs) and grown on Luria–Bertani (LB) agar plates containing kanamycin (50 μg/mL)
582 overnight at 37°C. Positive colonies, with successful insertion of the target gene, were selected by
583 screening with Colony PCR and were used to inoculate 4 mL LB liquid cultures with kanamycin (50
584 μg/mL), which were grown overnight on a rotary drum at 37°C. Plasmid DNA was isolated using the ZR
585 Plasmid Miniprep Kit (Zymo Research), and the sequences of the inserts were verified by either whole-
586 plasmid sequencing (Plasmidsaurus) or Sanger sequencing (ELIM Biopharm Inc.).

587 pEAQ-HT expression constructs were transformed into *Agrobacterium tumefaciens*
588 (GV3101:pMP90) via the freeze-thaw method. The transformed cultures were grown on LB agar plates
589 containing kanamycin (50 μg/mL) and gentamycin (30 μg/mL) at 30°C for two days to select positive
590 transformants, which were confirmed by colony PCR. The verified positive colonies were then used to
591 inoculate 2 mL LB liquid cultures (50 μg/mL kanamycin and 30 μg/mL gentamycin), which were grown
592 for two days at 30°C on a rotary drum and subsequently used to prepare 25% glycerol stocks that were
593 snap-frozen in liquid nitrogen and stored at −80°C for long-term use.

594

595 *Agrobacterium-mediated transient expression of Cephalotaxus genes in N. benthamiana*

596 Transient expression in *Nicotiana benthamiana* was conducted much as previously described.[31,48]
597 Glycerol stocks of *Agrobacterium tumefaciens* strains containing the pEAQ-HT expression construct of
598 interest were streaked onto LB agar plates (kanamycin 50 μg/mL and gentamicin 30 μg/mL), which were
599 then incubated at 30°C for two days until a dense bacterial lawn developed. The bacterial cells were
600 scraped off with a sterile pipette tip and resuspended in 1 mL of *Agrobacterium* induction buffer (10 mM
601 MES, pH 5.6,10 mM $MgCl_2$, 150 μM acetosyringone; Acros Organics), which were then incubated at

602 room temperature for 1-3 hours. The optical density at 600 nm (OD_{600}) was measured for each strain
603 resuspension to determine relative cell density. Individual strain was diluted in induction buffer to an
604 OD_{600} of 0.2-0.3 for infiltration. For transient co-expression, multiple strains of interest were combined in
605 equal concentrations, at a final OD_{600} of 0.2-0.3 for each strain. These individual strains or strain mixtures
606 were infiltrated into the abaxial side of 4-5 week-old *N. benthamiana* leaves using a 1 mL needleless
607 syringe. For each transient expression reaction, three leaves—each from a different plant—were used as
608 biological replicates to minimize any batch effects or biological variations among plants. After 3-4 days
609 of infiltration, 150-250 μL of substrate (25-50 μM of compound **2** or **3** in water) was infiltrated into the
610 abaxial side of previously *Agrobacterium*-infiltrated leaves using a needleless syringe. The area infiltrated
611 with substrate was marked, excised one day after the substrate infiltration, and collected into pre-weighted
612 2 mL Safe-lock tubes (Eppendorf), which were flash frozen in liquid nitrogen and lyophilized to dryness
613 for subsequent metabolite analysis.

614

615 *Heterologous expression of ChCYP805A11 in yeast*

616 Expression of *Ch*CYP805A11 in *Saccharomyces cerevisiae* (yeast) was performed as much
617 previously described.[31,83] The coding sequence of *Ch*CYP805A11 was PCR amplified from its pEAQ-HT
618 construct and annealed into BamHI/EcoRI digested pYeDP60 plasmid ($Carb^R$ for *E. coli* selection; *ADE2*
619 for yeast selection) through Gibson assembly, as discussed above in cloning of candidate biosynthetic
620 genes. Similarly, the assembled plasmid construct was transformed into NEB 10β *E. coli* cells, which
621 were plated for selection on LB agar (100 μg/mL carbenicillin) and grown overnight at 37°C. Positive
622 colonies, verified through colony PCR, were grown on a rotary drum overnight in liquid LB media
623 containing carbenicillin (100 μg/mL) at 37°C. The plasmid DNA was isolated using the Zymo Miniprep
624 Kit and sequenced through whole-plasmid sequencing (Plasmidsaurus) to confirm the sequence of the
625 inserted gene.

626 The assembled plasmid construct was transformed into chemically competent *S. cerevisiae*
627 WAT11 (*ade*2), which contains a chromosomal copy of the *A. thaliana* cytochrome P450 reductase 1
628 gene (*ATR1*),[84] using the Frozen-EZ Yeast Transformation II Kit (Zymo Research). Positive
629 transformants were selected on synthetic drop-out media plates lacking adenine (6.7g/L yeast nitrogen
630 base without amino acids, 20 g/L glucose, 2 g/L drop-out mix minus adenine [−Ade], 20 g/L agar)[85],[86]
631 through growth for 2 days at 30°C, which were confirmed through colony PCR. A single, positive colony
632 was then used to inoculate a 4 mL of liquid cultures of synthetic drop-out medium, grown for 2 days at
633 30°C with shaking at 250 rpm. The liquid cultures were then used to prepare 25% glycerol stocks and
634 stored at -80°C for long-term usage.

635 Freshly streaked colonies from *S. cerevisiae* WAT11 strain of interest (harboring pYeDP60
636 plasmid of *Ch*CYP805A11) were inoculated in 4 mL of liquid cultures of synthetic drop-out media (-Ade)
637 and were grown at 30°C for 2 days with shaking at 250 rpm. Then, 2 mL of the starter culture was used to
638 inoculate 500 mL of the synthetic drop-out media (-Ade) and grown at 28°C with 250 rpm shaking until
639 reaching a cell density of approximately 5×10^7 cells/mL, estimated through OD_{600} measurements. At this
640 density, gene expression was induced by adding 50 mL of a sterile galactose solution (200 g/L) to a final
641 concentration of 10% (v/v). The culture was grown for another 18 h at 28°C and 250 rpm to reach a cell
642 density of 5×10^8 cell/mL, which was immediately used for microsomal protein isolation. Moreover,
643 other than *Ch*CYP805A11, a negative control of empty vector was prepared as well.

644

645 *Isolation of ChCYP805A11-enriched microsomes from yeast*

646 Yeast microsomal proteins were isolated from yeast following a previously described protocol.[83]
647 The galactose-induced yeast cultures were centrifuged at 5,000 x g for 5 min at 4°C to pellet cells. The
648 cells were then resuspended in 1 mL of ice-cold TEK buffer (50 mM Tris-HCl, 1 mM EDTA, 100 mM
649 KCl, pH 7.4) per 0.5 g of wet cell pellet mass and incubated at room temperature for 5 min. The cells
650 were centrifuged again as above and resuspended in 5 mL ice-cold TES B buffer (50 mM Tris-HCl, 1
651 mM EDTA, 600 mM sorbitol, pH 7.4). All subsequent steps were performed at 4 °C and/or on ice.
652 Approximately 5 mL (a volume equal to that of the cell resuspensions) of glass beads (0.5 mm diameter)
653 was added to the cells, which were lysed by vigorous shaking for 10 min (30 s of shaking followed by 30
654 s on ice to avoid heating). Next, 10 mL of ice-cold TES B buffer was added to the lysate, which was then
655 removed from the beads and saved. To extract any residual microsomal proteins on the beads, the beads
656 were washed twice with 10 mL TES B buffer, and all washes were combined with the initial lysate (~30
657 mL total). This pooled lysate was centrifuged at 23,000 x g for 10 min at 4°C to pellet and remove large
658 cell debris, and the resulting supernatant (~30 mL) was diluted two-fold with ice-cold TES B buffer (to a
659 final volume of ~60 mL). Microsomes were precipitated by adding NaCl to a final concentration of 150
660 mM and polyethylene glycol (PEG)-4000 to a final concentration of 0.1 g/mL, followed by incubation on
661 ice for about 1 h with periodic mixing to dissolve the solutes. Microsomal proteins were then pelleted by
662 centrifugation at 10,000 x g for 10 min. at 4°C and were resuspended in 1 mL of ice-cold TEG storage
663 buffer (50 mM Tris-HCl, 1 mM EDTA, 20% (v/v) glycerol, pH 7.4). The protein concentration of the
664 microsomal fraction was measured using Bradford Assay, and 50-100 µL aliquots were snap-frozen in
665 liquid nitrogen and stored at -80°C until usage.

666

667 *Heterologous production and purification of ChNmrA-1*

668 For heterologous production and purification of *Ch*NmrA-1, the gene was expressed in *E. coli*.
669 The coding sequence of *Ch*NmrA-1 was first PCR amplified from its pEAQ-HT construct and inserted
670 into NdeI/XhoI digested pET-24b (Km^R) plasmid such that the resulting construct contained a C-terminal
671 6xHis tag through Gibson assembly, as discussed above in cloning of candidate biosynthetic genes.
672 Similarly, the assembled plasmid construct was transformed into NEB 10β *E. coli* cells, which were
673 plated for selection on LB agar (50 µg/mL kanamycin) and grown overnight at 37°C. Positive colonies,
674 verified through colony PCR, were grown on a rotary drum overnight in liquid LB media containing
675 kanamycin (100 µg/mL) at 37°C. The plasmid DNA was isolated using the Zymo Miniprep Kit and
676 sequenced through whole-plasmid sequencing (Plasmidsaurus) to confirm the sequence of the inserted
677 gene. The verified plasmid DNA was then transformed into *E. coli* BL21 (DE3) cells (NEB) as an
678 expression host for heterologous protein production. A single positive colony confirmed via colony PCR
679 was then inoculated in 3 mL of LB media (50 µg/mL kanamycin) and grown overnight on a rotary drum
680 at 37°C, which was used to prepare 25% glycerol stocks that were snap-frozen in liquid nitrogen and
681 stored at −80°C for future use.

682 Freshly streaked colonies from *E. coli* BL21 strains of interest (harboring pET-24b of *Ch*NmrA-
683 1) were inoculated in 2 mL of liquid LB cultures (50 µg/mL kanamycin) and grown overnight on a rotary
684 drum at 37°C. A 500 µL of the starter culture was inoculated in 200 mL of liquid LB culture containing
685 kanamycin (50 µg/mL), which was grown on a culture shaker at 37°C with 250 rpm until OD_{600} of ~0.6
686 was reached. Once reached, the temperature was reduced to 18°C and shaken for 0.5-1 h, followed by 0.5
687 mM IPTG induction. The cultures were shaken at 250 rpm for 18-20 h at 18°C and then centrifuged at
688 5,000 rpm for 10 min. at 4°C to pellet the cells. All subsequent steps were performed on ice and/or at 4°C.
689 The cell pellets (~1-2 g) were resuspended in 5-8 mL of ice-cold binding buffer (50 mM potassium
690 phosphate, pH 7.8, 100 mM NaCl, 10 mM imidazole) and lysed using EmulsiFlex®-B15 High-Pressure
691 Homogenizer with pressure at 15,000 psi; samples were homogenized three times to ensure full lysis of
692 the cells. Cell lysates were centrifuged at 10,000 x g for 30 min. at 4°C to pellet cell debris.

693 For protein purification, the clarified lysate was added to Ni-NTA resin (Ni-NTA Agarose,
694 Invitrogen; R901-15) and incubated at 4°C for 1 hour with gentle rocking. Bound proteins were eluted
695 sequentially using potassium phosphate buffer (50 mM potassium phosphate, 100 mM NaCl, pH 7.8) with
696 increasing concentrations of imidazole. Each eluted fraction was analyzed for protein content by sodium
697 dodecyl sulfate polyacrylamide gel (SDS-PAGE) with Coomassie Blue staining (see **Supplementary Fig.**
698 **11** for the protein gel). Fractions containing pure enzyme were combined and concentrated using
699 Amicon® Ultra-15 Centrifugal Filter Units (10 kDa molecular weight cut-off) by centrifugation at 4,000
700 x g for 20 min. at 4°C. The unfiltered protein-containing solution (concentrated to <1 mL) was diluted
701 with 10-13 mL of ice-cold potassium phosphate buffer and reconcentrated as above; this dilution-
702 concentration cycle was repeated 6-8 times to dilute out imidazole. The final concentrated protein
703 solution was diluted to approximately 1 mL with ice-cold phosphate buffer, which had glycerol added to a
704 final concentration of 10% (v/v). The purified protein concentration as measured using Bradford Assay,
705 and 30 μL aliquots were snap-frozen in liquid nitrogen and stored at -80°C for future use.

706

707 *In vitro reactions with purified ChNmrA-1 and ChCYP805A11-enriched microsomes*

708 *In vitro* reactions with purified *Ch*NmrA-1 and *Ch*CYP805A11-enriched microsomes were
709 carried out in 100 μL of potassium phosphate buffer (50 mM potassium phosphate, 100 mM NaCl, pH
710 7.8) with a final concentration of 50 μM of initial substrate (compound **3**) and 1 mM NADPH. For each
711 enzyme, final concentration of 0.4 μg/μL was used in 100 μL of phosphate reaction buffer. As negative
712 controls, microsomal fractions containing empty vector (EV) and denatured (boiled) *Ch*NmrA-1 were
713 used. The dependency of the reaction on both the substrate (**3**) and cofactor (NADPH) was also shown
714 through dropping out each component. The reactions were incubated at 30°C for several time points (0.5,
715 1, 2, and 4 h), after which the reactions were quenched by an equal volume (1:1) of acetonitrile with 0.1%
716 formic acid. Quenched reactions were diluted four-fold in 100% MeOH and filtered through 0.45 μM
717 hydrophilic PTFE filter plates (MilliporeSigma). The filtrates were transferred into LC-MS vials for
718 analysis.

719

720 *Metabolite extraction from plant material*

721 The overnight-lyophilized tissues (i.e. various tissues of *C. harringtonia* from feeding
722 experiments and leaves from *N. benthamiana* transient expression experiments) were homogenized on a
723 ball mill homogenizer (Retsch MM 400) using a 5 mm diameter steel bead for each sample tube, with
724 shaking at 25 Hz for 2 min. Steel beads were then removed, and each tissue sample was extracted with 40
725 μL and 20 μL of 80% methanol in water per milligram of dry plant tissue weight for *C. harringtonia*
726 tissues and *N. benthamiana* leaves, respectively. The samples were briefly vortexed and sonicated, then
727 incubated for 30 min. at room temperature for metabolite extraction. For *C. harringtonia* tissues,
728 additional 20 min. of incubation at 4°C was performed. The extracted samples were centrifuged at 16,000
729 x g for 1 min. to pellet plant cell debris, and the supernatant was filtered through hydrophilic PTFE filter
730 plates with 0.45 μM pore size (MilliporeSigma). The filtrates were transferred into LC-MS vials for
731 analysis.

732 For quantification of natural abundance of both CET and HHT, a 1000-fold and 200-fold dilution
733 was performed for tissues harvested in May 2022 and May 2023 respectively to avoid saturation of the
734 parent isotope peaks in LC-MS.

735

736 *LC-MS analysis of metabolites*

All metabolite samples were analyzed on the LC-MS instrument of an Agilent 1290 Infinity II UHPLC coupled to an Agilent 6546 Q-TOF mass spectrometer (6546 LC-MS) . For separation of the metabolites, a reversed-phase liquid chromatography using a ZORBAX RRHD Eclipse Plus 95Å C18 column (Agilent, 2.1 x 50 mm, 1.8 µm) was performed with a mobile phase consisting of two buffers: water with 0.1% formic acid (buffer A) and acetonitrile with 0.1% formic acid (buffer B). An injection volume of 1 µL and a flow rate of 0.6 mL/min were used with the following 29 min. gradient method: 0-0.5 min, 3% B; 0.5-7.5 min, 3-10% B; 7.5-15.5 min, 10-25% B; 15.5-20.5 min, 25-50% B; 20.5-21.5 min, 50-95% B; 21.5-23.0 min, 95% B; 23.0-24.5 min, 95-30% B; 24.5-26.5 min, 30-3% B; 26.5-29.0 min, 3% B. MS data were collected using a dual-inlet Agilent Jet Steam electrospray ionization (Dual AJS ESI) in positive ion mode with a mass range of 50-1,700 *m/z* and a rate of 1.00 spectrum per second (1000 ms per spectrum). The ionization source was set as follows: 325°C gas temperature, 10 L/min drying gas, 35 psi nebulizer, 350°C sheath gas temperature, 12 L/min sheath gas flow, 135 V fragmentor, 45 V skimmer, 750 V Oct 1 RF Vpp, 4000 V VCap, and 0 V nozzle voltage. For tandem mass spectrometry (MS/MS) analysis, the mass ions pertaining to individual metabolites were fragmented with collision energies of 15, 20, 30, or 35 V. All data were collected in centroid mode, but when needed, data were also stored in profile mode.

Metabolomics and MS data analyses

LC-MS data were analyzed using Agilent MassHunter Qualitative Analysis 10.0 and Agilent MassHunter Quantitative Analysis 11.0 software. The ion abundances for a given mass signature were calculated using the automated integration ('Agile' method with default settings) of extracted ion chromatograms (EICs) with a 20 ppm mass tolerance. Peaks integrated by the Quantitative Analysis software were manually inspected to ensure correct peak selection and integration.

Untargeted metabolomic analyses were conducted using XCMS Online[87] with default parameters of UPLC / UHD Q-TOF (155). The output of this analysis contains a tabulated list of identified mass signatures at specific retention times, with associated *m/z* values, peak intensity fold change, *P* value (two-tailed unequal variance Welch's *t*-test), retention times, and extracted peak intensities.

For XCMS analysis comparing dopamine-D_4-fed root tips with water-fed controls of *C. harringtonia* from isotope labeled precursor feeding studies, we used the following filters to look for differential mass signatures: P value < 0.2, fold change > 3.5, and *m/z* range of in between 200 and 600.

For XCMS analysis to determine masses found at differential levels between two samples from *N. benthamiana* transient expression experiments, we used the following filters, unless stated otherwise: P value < 0.1, fold change > 5, average peak intensity > 5×10^4, retention time < 18 min., and 250 < *m/z* < 600.

*Synthesis of different phenethylisoquinolines (**1**, **1**-D_4, **2**, **2**-D_4, **3**, **11**, **12**, **12**-D_4, and **13**)*

The synthesis of compound **1** (1-(4-hydroxyphenethyl)-1,2,3,4-tetrahydroisoquinoline-6,7-diol) was adapted from a previously described protocol.[88] Dopamine (0.42 mmol) and of 3-(4-hydroxyphenyl)propanal (0.50 mmol) were combined in 4 mL of a 1:1 (v/v) mixture of acetonitrile and potassium phosphate buffer (0.1 M, pH 6.0). The reaction mixture was stirred at 50°C for 16 h under N_2 with reflux. The mixture was then frozen in liquid-nitrogen and lyophilized to dryness.

The above procedure was applied to the synthesis of all compounds listed in Table 1 below, substituting the amine and aldehyde starting materials as specified.

781 **Table 1.** Synthesis of phenethylisoquinolines

Compound	Amine (0.42 mmol)	Aldehyde (0.50 mmol)
1	dopamine	3-(4-hydroxyphenyl)propanal
1-D_4	dopamine-D_4	3-(4-hydroxyphenyl)propanal
2	dopamine	3-(4-hydroxy-3-methoxyphenyl)propanal
2-D_4	dopamine-D_4	3-(4-hydroxy-3-methoxyphenyl)propanal
3	3-hydroxy-4-methoxyphenethylamine	3-(4-hydroxy-3-methoxyphenyl)propanal
11	3-hydroxy-4-methoxyphenethylamine	3-(4-hydroxyphenyl)propanal
12	dopamine	3-(3,4-dihydroxyphenyl)propanal
12-D_4	dopamine-D_4	3-(3,4-dihydroxyphenyl)propanal
13	3-methoxy-4-hydroxyphenethylamine	3-(3,4-dihydroxyphenyl)propanal

782 For compounds **1**, **1**-D_4, **11**, **12**, **12**-D_4, and **13**, the lyophilized crude product was resuspended in
783 ~1-2 mL of DMSO and directly loaded onto a Biotage® Sfär C18 D Duo 100 Å 30 µm 6 g column.
784 Purification was performed an automated Biotage Selekt system using water with 0.1% formic acid and
785 acetonitrile with 0.1% formic acid as mobile phases. Eluting fractions containing the compound of
786 interest (identified by LC-MS) were pooled, frozen in liquid nitrogen, and lyophilized to dryness.

787 For compounds **2**, **2**-D_4, and **3**, the lyophilized crude product was resuspended in ~2-3 mL of 80%
788 MeOH and purified by preparative HPLC using an Agilent 1260 Infinity preparative-scale HPLC system
789 with an Agilent 1100 diode array detector and a Clipeus C18 5-µm, 10-mm x 250-mm semi-prep column
790 (Higgins Analytical). Water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B) were
791 used as the mobile phases.

792 For compounds **2** and **2**-D_4, an injection volume of 500 µL and a flow rate of 5 mL/min were
793 used with the following gradient method: 0-5 min, 3% B; 5-8 min, 3-10% B; 8-23 min, 10-20% B; 23-24
794 min, 20-25% B; 24-24.1min, 25-97% B; 24.1-29.1 min, 97% B; 29.1-29.2 min, 97-3% B; 29.2-34 min,
795 3% B. Fractions were collected from 1-34 min, and those containing the compound of interest (verified by
796 LC-MS) were combined, frozen, and lyophilized to dryness.

797 For compound **3,** an injection volume of 500 µL and a flow rate of 5 mL/min were used with the
798 following gradient method: 0-5 min, 3% B; 5-8 min, 3-16% B; 8-18 min, 16-24.5% B; 18-19 min, 24.5-
799 30% B; 19-19.1 min, 30-97% B; 19.1-24.1 min, 97% B; 24.1-24.2 min, 97-3% B; 24.2-29 min, 3% B.
800 Fractions were collected from 1-29 min, and those containing the desired compound (verified by LC-MS)
801 were pooled, frozen, and lyophilized to dryness.

802 All the purified products **1**, **1**-D_4, **2**, **2**-D_4, **3**, **11**, **12**, and **12**-D_4 were analyzed by both LC-MS/MS
803 and NMR. For compound **13**, its structure was verified by LC-MS/MS.

804

805 *Heterologous production and purification of compound 3-D_4 using ChOMT-1*

806 Compound **3**-D_4 was biosynthesized from chemically synthesized **2**-D_4 by heterologous
807 expression of *Ch*OMT-1 in *E. coli* BL21 (DE3). Cloning and expression of *Ch*OMT-1 were performed
808 exactly as described for *Ch*NmrA-1 (see Methods: *Heterologous production and purification of ChNmrA-*
809 *1*), except that the induction culture volume was increased to 900 mL. After 18 h of 0.5 mM IPTG
810 induction of *Ch*OMT-1 at 18°C with shaking at 250 rpm, compound **2**-D_4 was added to a final
811 concentration of 50 µM for O-methylation reaction of **2**-D_4 to **3**-D_4. The biosynthetic reaction was run for
812 5 h with continuous shaking of 250 rpm at 18°C. The bacterial cells were pelleted, resuspended in 900

813 mL of fresh LB media, and subjected to two additional cycles of the O-methylation reaction (total of three
814 900 mL reactions). The combined supernatants containing the biosynthesized **3**-D_4 were extracted with
815 ethyl acetate (1:1, v/v) four times. Organic phases were combined, dried using rotary evaporation, and
816 resuspended in ~3 mL of 80% MeOH. Purification was performed using the same preparative HPLC
817 system, column, and mobile phases as described for compounds **2**, **2**-D_4, and **3**, but with a modified
818 gradient method.

819 For compound **3**-D_4**,** an injection volume of 200 µL and a flow rate of 5 mL/min were used with
820 the following gradient method: 0-5 min, 3% B; 5-8 min, 3-9.5% B; 8-16 min, 9.5% B; 16-21 min, 9.5-
821 11% B; 21-36 min, 11% B; 36-41 min, 11-13% B; 41-44 min, 13-16%; 44-46 min, 16-20 min; 46-47 min,
822 20-30%; 47-47.1 min, 30-97%; 47.1-52.1 min, 97% B; 52.1-52.2 min, 97-3% B; 52.2-57 min, 3% B.
823 Fractions were collected from 1-57 min, and those containing the compound of interest (verified by LC-
824 MS) were combined, frozen, and lyophilized to dryness. The purified product **3**-D_4 was then analyzed by
825 LC-MS/MS and NMR.

826

827 *Extraction and purification of a cephalotaxinone (**10**) standard*

828 Old leaves and stem of *C. harringtonia* were harvested, frozen at -80°C, and lyophilized to
829 dryness. The dried tissues (~10 g) were then ground to a fine powder or small pieces using mortar and
830 pestle, and extracted with 700 mL of methanol for 3-4 days at room temperature with constant stirring.
831 Extracts were filtered using vacuum filtration and dried by rotary evaporation. Because the compound of
832 interest (CET intermediate) is alkaloid, an acid-base extraction was performed to remove water-soluble
833 and non-alkaloid impurities prior to two rounds of chromatography. The residue was resuspended in 450
834 mL of water with HCl (pH 2-3). This fraction was partitioned with 400 mL of diethyl ether or *tert*-butyl
835 methyl ether. The aqueous fraction was collected, and the organic phase was washed twice with 100 mL
836 of water (pH 2-3) to maximize extraction of the desired compound, after which each of these aqueous
837 fractions was combined with the initial 450 mL aqueous fraction. Then, the pH of the aqueous phase was
838 increased to pH 8-9 using saturated sodium bicarbonate. This solution was partitioned with 600 mL of
839 ethyl acetate. The organic layer, which supposedly contains the target compound, was collected,
840 combined with two additional 100 mL ethyl acetate fractions used to wash the aqueous phase, and dried
841 using rotary evaporation.

842 The dried residue was resuspended in ~1-2 mL of methanol and loaded onto a 3.5cm-diameter
843 column packed with 18 g of P60 silica gel (SiliCycle). Hexane (HPLC grade; VWR) and ethyl acetate
844 (ACS reagent grade; J.T. Baker) were used as mobile phases. The following gradient was applied: 100
845 mL each of 50%, 60%, 70%, 80%, 90%, and 100% ethyl acetate. Each step was collected in three
846 fractions, yielding a total of 18 fractions. Fractions containing the compound of interest (identified by LC-
847 MS) were combined and dried by rotary evaporation. The partially purified product was resuspended in
848 ~3-4 mL of 80% MeOH for second round of chromatography.

849 For the second round, purification was performed using the same preparative HPLC system,
850 column, and mobile phases as described for compounds **2**, **2**-D_4, and **3**, but with a modified gradient
851 method. An injection volume of 450 µL and a flow rate of 5 mL/min were used with the following
852 gradient method: 0-5 min, 3% B; 5-8 min, 3-7% B; 8-10 min, 7% B; 10-13 min, 7-10% B; 13-16 min,
853 10% B; 16-20 min, 10-15% B; 20-24 min, 15%; 24-30 min, 15-25 min; 30-37 min, 25-40%; 37-37.1 min,
854 40-97%; 37.1-42.1 min, 97% B; 42.1-42.2 min, 97-3% B; 42.2-47 min, 3% B. Fractions were collected
855 from 1-47 min, and those containing the compound of interest (confirmed by LC-MS) were combined,
856 frozen, and lyophilized to dryness.

857

858 *NMR analysis of purified compound*

859 DMSO-d_6 (ThermoFisher Scientific) was used as the solvent for all NMR samples of
860 phenethylisoquinolines (**1**, **1**-D_4, **2**, **2**-D_4, **3**, **3**-D_4, **11**, **12**, and **12**-D_4). For NMR sample of
861 cephalotaxinone (**10**), $CDCl_3$ (Acros Organics) was used as the solvent. ^1H, ^{13}C, and 2D-NMR spectra
862 were mostly acquired on a Bruker NEO 500-MHz spectrometer, and on a Varian Inova 600-MHz (for
863 ROESY for **10**) at room temperature using VNMRJ 4.2. The data were visualized, analyzed, and
864 processed on MestReNova v.15.1.0. Chemical shifts were reported in ppm downfield from Me_4Si by
865 using the residual solvent peak as an internal standard: 7.26 ppm for ^1H and 77.16 ppm for ^{13}C ppm
866 chemical shift of $CDCl_3$ and 2.50 ppm for ^1H and 39.52 ppm for ^{13}C ppm chemical shift of DMSO-d_6.

867 The NMR spectra for compounds **1**, **1**-D_4, **2**, **2**-D_4, **3**, **3**-D_4, **11**, **12**, **12**-D_4, and cephalotaxinone
868 (**10**) are shown in **Supplementary Figs. 19-69**, and their $^1H/^{13}C$ assignments are detailed in
869 **Supplementary Tables 4-13**.

870

871 *General software use and graph generation*

872 Routine data compilation was performed in Microsoft Excel (version 16.62), including
873 calculation of Pearson's correlation coefficient used for co-expression analysis. General LC-MS data
874 analysis was performed with Agilent MassHunter Qualitative Analysis 10.0. Chromatograms and mass
875 spectra were plotted using IGOR Pro (version 8.04). Bar graphs, lines graphs, and scatterplots were
876 plotted using GraphPad Prism 10, including mean and standard deviations denoted in the graphs. NMR
877 data were processed and visualized on MestReNova v15.1.0. Molecular weight calculation and chemical
878 structural visualization were conducted using ChemDraw Professional v21.0.0.

879

880 *Data Availability*

881 The raw RNA-seq have been deposited in the National Center for Biotechnology Information
882 (NCBI) Sequence Read Archive (SRA) database under the BioProject ID XXX. Cloned sequences have
883 been deposited in GenBank (accession nos. XXX). All other raw data are available upon request.

884

885

886

887

888

889

890

891

892

893

894

895

ACKNOWLEDGEMENTS

We thank Stephen Lynch (Stanford University) for his assistance and helpful discussion related to NMR analysis. We also thank Prof. David Nelson (University of Tennessee) for providing cytochrome P450 nomenclature. We acknowledge Prof. George Lomonossoff for providing us with the pEAQ-HT plasmid. We are grateful to Warren Lau and Catherine Liou for early contributions to this project. The author is also thankful to Diego Wengier and Sarah Niehs for feedback on this manuscript. An initial *C. harringtonia* transcriptome (August 2018) (proposal: 10.46936/10.25585/60001097 led by J. Bohlmann) was performed by the US Department of Energy (DoE) Joint Genome Institute (https://ror.org/04xm1d337), a DoE Office of Science User Facility, and is supported by the Office of Science of the US DoE operated under contract number DE-AC02-05CH11231. The research is supported by an NIH R01 GM121527 to E.S.S.

AUTHOR CONTRIBUTIONS

Y.D. and **E.S.S.** led the project and conceived of experimental procedures. **Y.D.** performed the experiments, generated the new *C. harringtonia* transcriptome (May 2023) used for discovery, and analyzed the data. **K.S.** collected the samples for the initial *C. harringtonia* transcriptome (August 2018) for seasonal comparison and helped develop the hypothesis for a whole plant coordination of HHT biosynthesis. **Y.D.** and **E.S.S.** interpreted the data, and **Y.D.** wrote the manuscript with edits from **E.S.S.**

COMPETING INTERESTS

The authors declare no competing interests related to this work.

MATERIALS & CORRESPONDENCE

E.S.S. (sattely@stanford.edu) should be contacted for any materials or other correspondence.

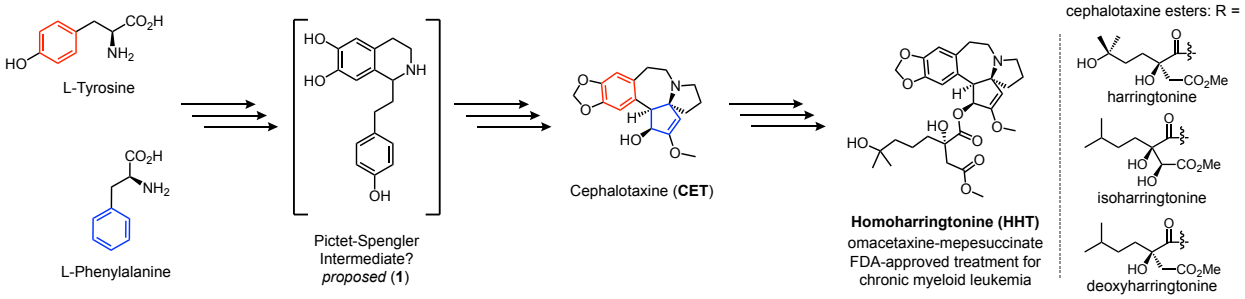


Figure 1. Proposed biosynthetic pathway of CET, the core structure of HHT, and its esterified derivatives. The proposed pathway for CET, the core structure of HHT, biosynthesis is based on radioisotope-labeling studies (shown in red and blue) and structural characterization of alkaloids from *Cephalotaxus* species. Over 20 distinct CET esters have been identified to date, four of which shown here exhibit notable biological activity.

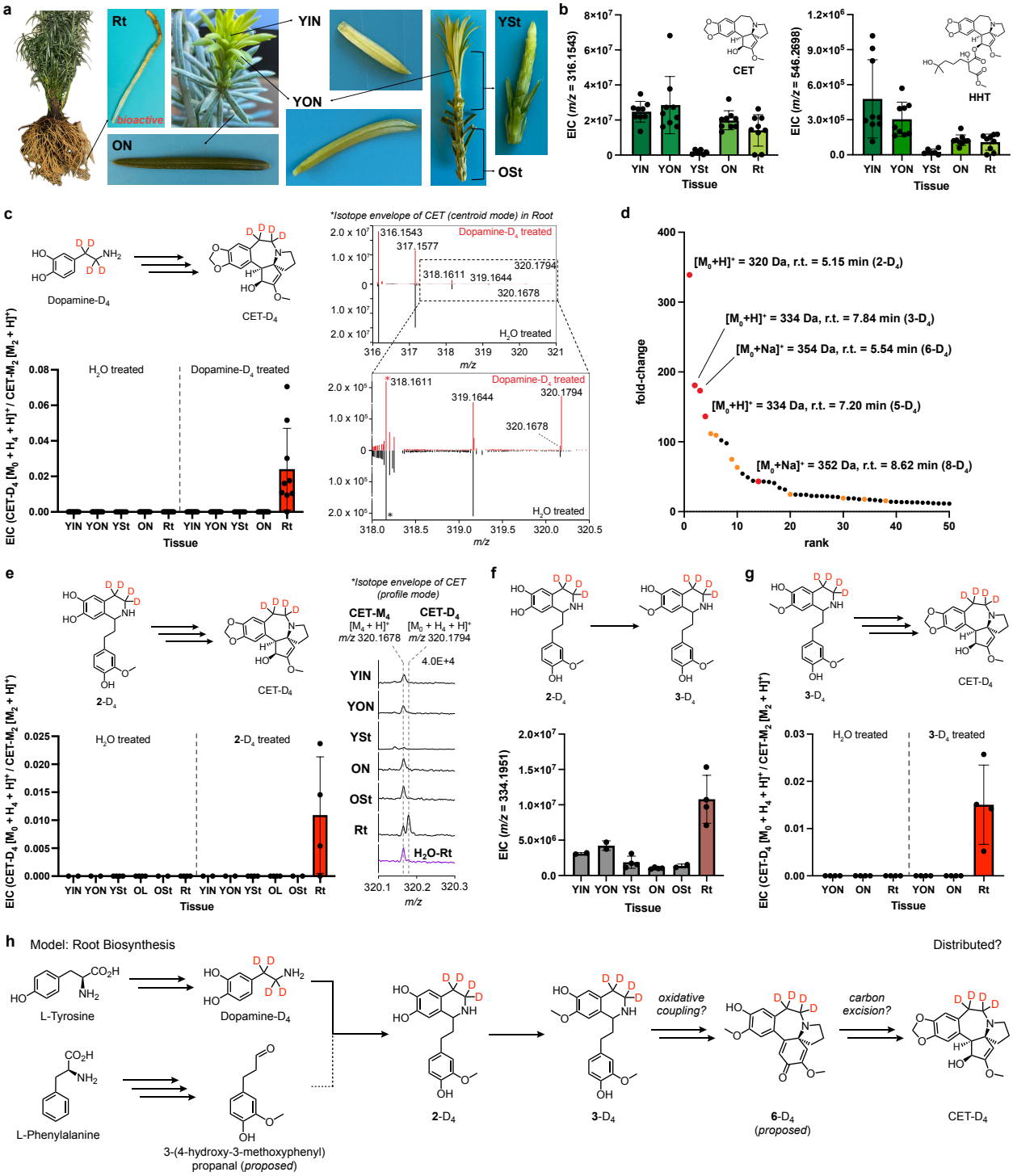


Figure 2. Stable-isotope precursor feeding study reveals that root tips (up to length of approximately 5-8 cm) of actively growing plant are the sites of active biosynthesis of cephalotaxine and that specific phenethylisoquinolines serve as intermediates in this process. (a) Different tissues of *C. harringtonia* (young inner needles—YIN, young outer needles—YON, old needles—ON, young stem—YSt, old stem—OSt, and root tips—Rt) harvested during actively growing season of the plant (May or June) and dissected for paired feed-in experiments (metabolomics) and RNA-seq (transcriptomics). **(b)** Natural accumulation of CET ($[M + H]^+ = m/z$ 316.1543, r.t. = 2 min.) and HHT ($[M + H]^+ = m/z$ 546.2698, r.t. = 10.8 min.) throughout the native plant, *C. harringtonia* harvested in May 2022. **(c)** Relative isotope enrichment of CET-D_4 in deuterated precursor dopamine-D_4 feeding (right half of bar graph) compared to water treatment (left half of bar graph) in five different tissues

945 indicates active biosynthesis of CET in the growing root tips (up to length of approximately 5-8 cm) of the native
946 plant. On the right, isotope envelope of CET in centroid mode, comparing dopamine-D_4 treated Rt (red) vs H_2O
947 treated Rt (black), is shown—with dopamine-D_4 treated Rt having an extra peak of *m/z* 320.1794 corresponding to
948 CET-D_4 ($[M_0 + H_4 + H]^+$). **(d)** Untargeted metabolite analysis (XCMS) comparing dopamine-D_4 fed roots tips to
949 water-fed root tips of *C. harringtonia* ($n = 9$ independent replicates for each experimental condition). The unique
950 mass features ($P < 0.2$; fold change > 3.5; $200 < m/z < 600$) are ranked based on their increasing fold change in
951 abundance between the two conditions, with top 50 shown. Features confirmed to be above signal to noise and
952 represent true mass signatures corresponding to potential intermediates (with four deuteriums, resulting from
953 incorporation of dopamine-D4) in CET biosynthesis are shown in red and their isotopologues and/or adducts in
954 orange. r.t., retention time. While not shown here, CET-D_4 ranked 127th out of 148 unique features identified by
955 XCMS with increasing fold change. **(e)** Relative isotope enrichment of CET-D_4 in **2**-D_4 feeding (right half of bar
956 graph) compared to water treatment (left half of bar graph) in six different tissues indicates active biosynthesis of
957 CET in the growing root tips (up to length of approximately 5-8 cm) of the native plant and **2**-D_4 as an intermediate
958 in CET biosynthesis. On the right, isotope envelope of CET in **2**-D_4 fed tissues (YIN, YON, YSt, ON, OSt, and Rt)
959 and in H_2O fed Rt in profile mode, focusing on the fourth isotope peak, M_4 (M + 4), is shown. In the **2**-D_4 fed Rt
960 sample, an additional fourth isotope peak appears at *m/z* 320.1794, corresponding to CET-D_4 ($[M_0 + H_4 + H]^+$) due
961 to four deuterium incorporation, which is distinct from the natural carbon isotope peak at *m/z* 320.1678,
962 corresponding to CET-M_4 ($[M_4 + H]^+$) observed in all different tissues fed with **2**-D_4 or water. **(f)** Observation of **3**-
963 D_4 ($[M+H]^+$ = *m/z* 334.1951, r.t. = 7.8 min.) in all six different tissues treated with **2**-D_4. O-methylation of **2**-D_4 to **3**-
964 D_4 predominantly occurs in the actively growing root tips. **(g)** Relative isotope enrichment of CET-D_4 in **3**-D_4
965 treatment (right half of bar graph) compared to water treatment (left half of bar graph) in three different tissues
966 indicates active biosynthesis of CET in the growing root tips (up to length of approximately 5-8 cm) of the native
967 plant and **3**-D_4 as an intermediate in CET biosynthesis. **(h)** Summary of updated proposed biosynthetic pathway for
968 CET with stable-isotope precursor feeding experiments. Dotted line represents proposed aldehyde incorporation not
969 yet supported by labeling studies. For all quantification, EIC of LC-MS was used. For CET-D_4 enrichment studies
970 **(c, e, g)**, CET-D_4 was normalized by M_2 (M + 2) isotopologue to avoid detector saturation effects, since CET
971 ionized well in LC-MS and was highly abundant in our measurement conditions. Experiments in **(b-d), (e-f),** and **(g)**
972 were conducted in May 2022 ($n = 9$ biological replicates for each tissue type, except YSt, $n = 6$), in May 2023 ($n = 4$
973 biological replicates for YSt, ON, and Rt, and $n = 2$ for YIN, YON, and OSt), and in June 2024 ($n = 4$ biological
974 replicates for each tissue type) respectively. The values and error bars are mean and ± standard deviation.
975
976
977
978
979

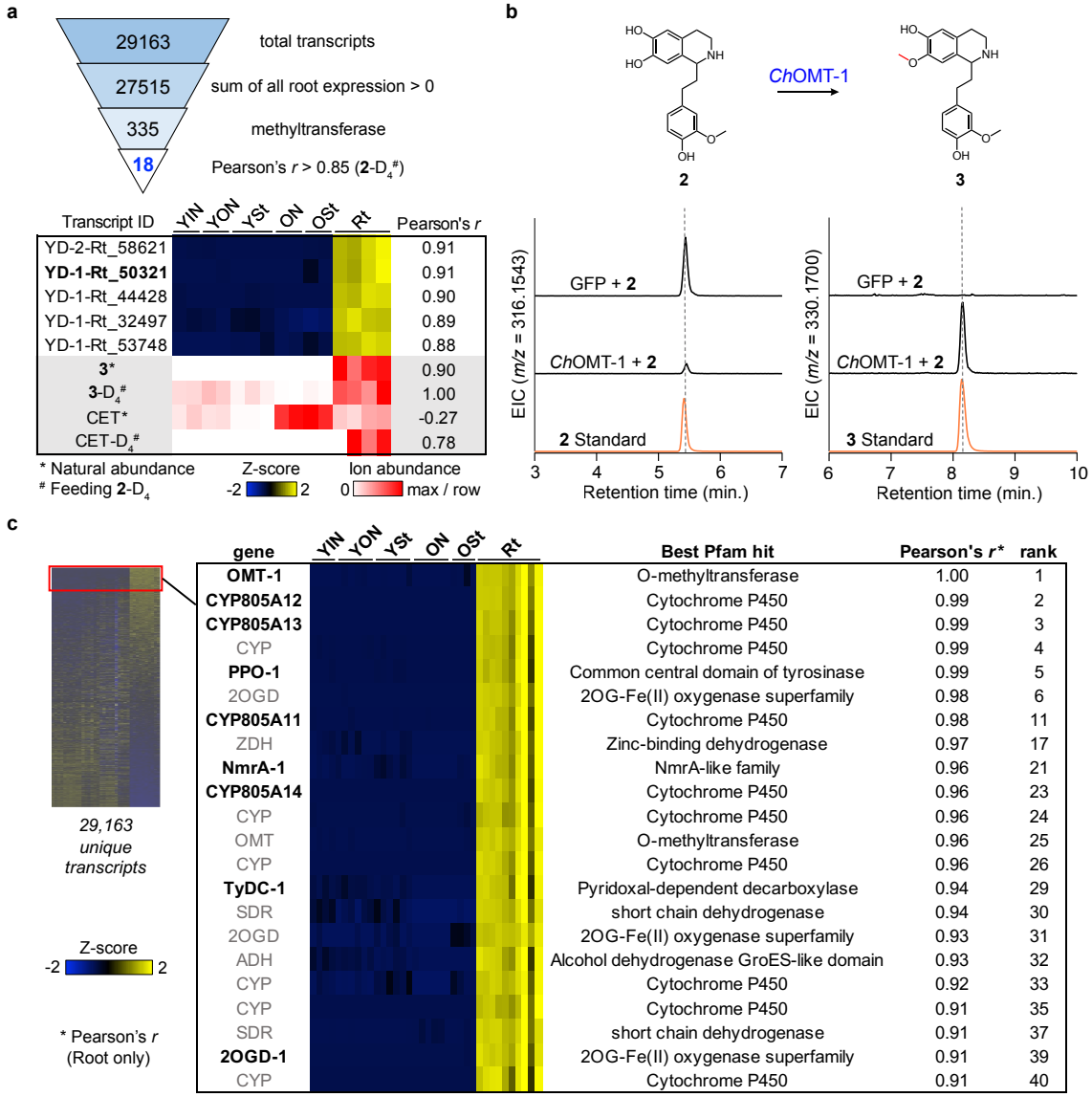


Figure 3. Combined transcriptomics and metabolomics analysis identifies *Ch*OMT-1 as the biosynthetic gene and reveals co-expressed candidate genes involved in CET biosynthesis. (a) The funnel plot shows the transcript selection process from 29,163 total transcripts to 18 candidate methyltransferases. The heatmap shows expression profiles (Z-score scaled) of the five (out of 18) tested methyltransferase-annotated transcripts in 15 samples of six different tissues of *C. harringtonia* from an independent experiment A conducted in May 2023, selected based on expression in root tips (site of active biosynthesis) and strong correlation with **3**-D_4 enrichment following **2**-D_4 feeding (Pearson's $r > 0.85$). Transcripts are ordered by decreasing correlation, with YD-1-Rt_50321 (in bold) identified as *Ch*OMT-1. Below, a metabolomic heatmap (red scale) shows LC-MS ion abundances of **3** and CET at natural isotope abundance (*) and their four-deuterated analogs (#) after **2**-D_4 feeding, measured in the same *C. harringtonia* tissues used for transcriptomic analysis (each tissue was divided for parallel profiling). Six different tissues (15 samples): 2 young inner needles (YIN), 2 young outer needles (YON), 3 young stems (YSt), 2 old needles (ON), 2 old stems (OSt), and 4 root tips (Rt). **(b)** Expression of *Ch*OMT-1 in *N. benthamiana* with co-infiltrated **2** leads to consumption of **2** (*m/z* 316.1543, r.t. of 5.4 min) and production of an O-methylated compound **3** (*m/z* 330.1700, r.t. of 8.1 min), as shown by the LC-MS chromatograms. This experiment was repeated more than three times, with similar results observed each time. Retention times of authentic standards (orange traces) are shown for comparison. Note: minimal retention time shifts (~0.2-0.3 min) are often observed in LC-MS runs conducted at different times. **(c)** RNA-seq analyses across various tissue types—YIN, YON, YSt, ON, OSt, and Rt—combined with Pearson's correlation using *Ch*OMT-1 as the bait gene, identify a set of strongly co-expressing biosynthetic gene candidates from 29,163 unique transcripts. Initial Pearson correlation analysis across all tissues

using *Ch*OMT-1 as a query gene identified 316 transcripts (r > 0.98). As many of these showed expression only in root tips, a second round of correlation analysis was conducted using Rt samples alone, yielding 41 top candidates (r > 0.90). Out of 41, genes of biosynthetically relevant enzyme families such as oxidases, reductases, O-methyltransferases, and pyridoxal phosphate (PLP)-dependent enzymes are shown here, and those involved in CET biosynthesis are highlighted in bold. Expression values within the heat map are log_2-transformed values of each transcript calculated as Trimmed mean of M (TMM)-normalized counts per million (CPM) in Z-score scale. Pearson's *r* shown here are from the second round of correlation analysis done using only Rt samples. The whole heatmap is composed of 36 unique samples—YIN (4), YON (6), YSt (6), ON (6), OSt (4), and Rt (10)—harvested from the same plant but at two different times (one set, A, harvested in early May 2023 and the other set, B, two weeks after).

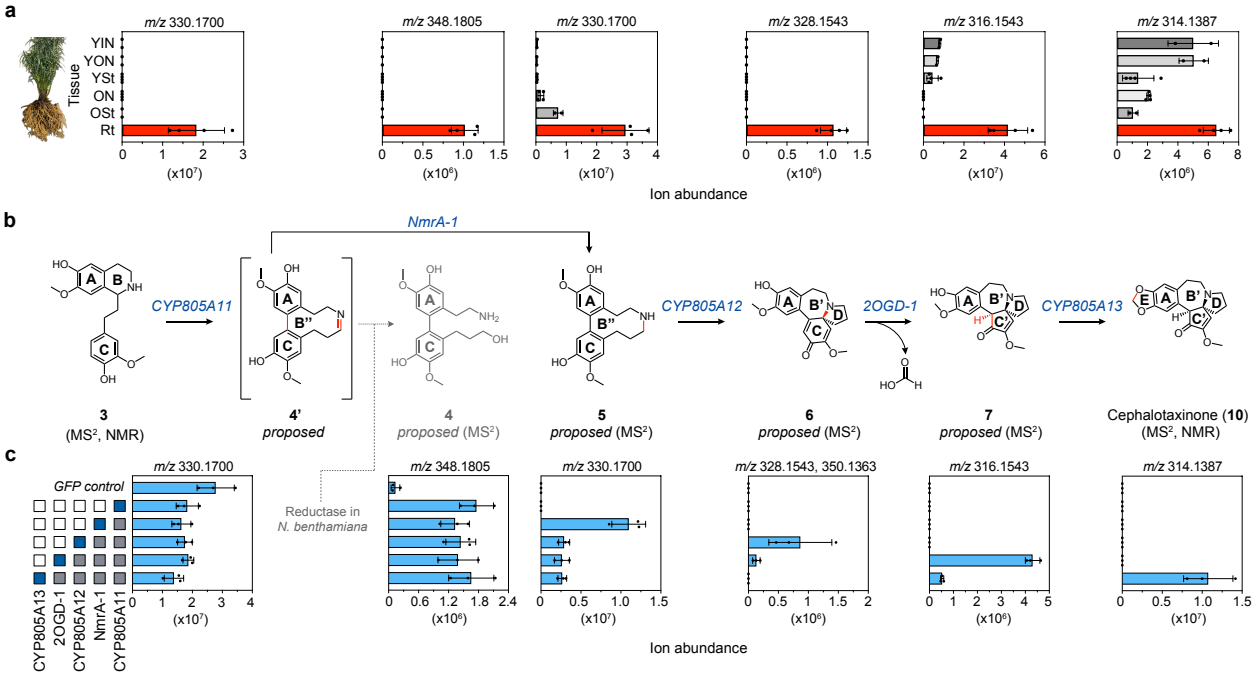


Figure 4. Discovery of a pathway for CET alkaloid biosynthesis, from phenethylisoquinoline (3) to cephalotaxinone. (a) Natural accumulation of CET biosynthetic intermediates in the native plant, *C. harringtonia*, harvested in May 2023. The data are reported as mean ± SD of the extracted ion abundance ($n = 4$ biological replicates for YSt, ON, and Rt, and $n = 2$ for YIN, YON, and OSt) for the exact ion mass $[M + H]^+$ corresponding to each compound. For quantification of cephalotaxinone, a 200-fold dilution was performed to the extracted samples to avoid saturation of the parent isotope peaks in LC-MS. Young inner needles (YIN), young outer needles (YON), Young stems (YSt), old needles (ON), old stems (OSt), and root tips (Rt). **(b)** Proposed biosynthetic pathway of CET reconstituted in *N. benthamiana* through transient co-expression of five identified biosynthetic genes from *C. harringtonia*, illustrating the stepwise conversion from compound **3** to cephalotaxinone (**10**). All the structures are supported by LC-MS/MS analyses, with the first and final intermediates (**3** and cephalotaxinone) further supported by authentic synthesized and purified standards, including NMR analyses. **(c)** Gray boxes to the left of the graphs indicate biosynthetic genes that are co-expressed in *N. benthamiana* leaves; blue box represents the final acting enzyme. For each intermediate, the data are reported as mean ± SD of the extracted ion abundance ($n = 3$) for the exact ion mass $[M + H]^+$ (for **6**, both $[M + H]^+$ and $[M + Na]^+$) corresponding to each compound. While the overall biosynthetic pathway of CET is linear as shown, the final steps involving *Ch*CYP805A13 and *Ch*2OGD-1 may occur in bifurcated sequence (Supplementary Fig. 16).

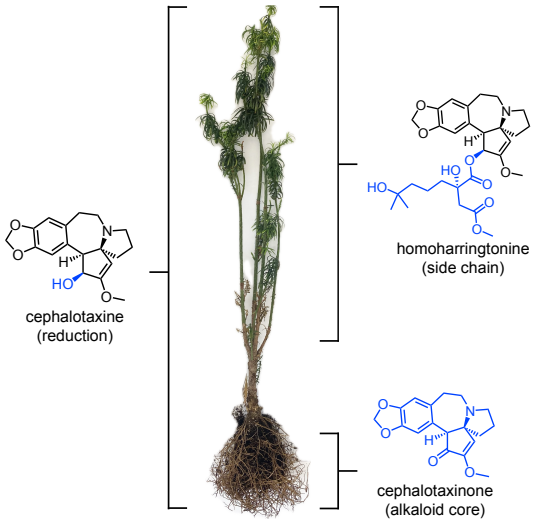


Figure 5. Proposed model illustrating sites of active metabolism for whole plant coordination of *Cephalotaxus* alkaloid biosynthesis. Proposed tissue specific metabolism indicated in the parenthesis. Molecule components that are proposed to be actively produced are highlighted in blue.

REFERENCES

1 Pérard-Viret, J., Quteishat, L., Alsalim, R., Royer, J. & Dumas, F. Cephalotaxus Alkaloids. *Alkaloids Chem Biol* **78**, 205-352 (2017).

2 Abdelkafi, H. & Nay, B. Natural products from Cephalotaxus sp.: chemical diversity and synthetic aspects. *Natural Product Reports* **29**, 845-869 (2012).

3 Takano, I. *et al.* Ester-type cephalotaxus alkaloids from Cephalotaxus harringtonia var. drupacea. *Phytochemistry* **44**, 735-738 (1997).

4 Powell, R. G., Weisleder, D. & Smith, C. R., Jr. Antitumor alkaloids for Cephalataxus harringtonia: structure and activity. *J Pharm Sci* **61**, 1227-1230 (1972).

5 Powell, R. G., Weisleder, D., Smith, C. R. & Rohwedder, W. K. Structures of harringtonine, isoharringtonine, and homoharringtonine. *Tetrahedron Letters* **11**, 815-818 (1970).

6 Mikolajczak, K. L., Powell, R. G. & Smith, C. R. Deoxyharringtonine, a new antitumor alkaloid from Cephalotaxus: Structure and synthetic studies. *Tetrahedron* **28**, 1995-2001 (1972).

7 Zhou, J. Y., Chen, D. L., Shen, Z. S. & Koeffler, H. P. Effect of homoharringtonine on proliferation and differentiation of human leukemic cells in vitro. *Cancer Res* **50**, 2031-2035 (1990).

8 Kantarjian, H. M., O'Brien, S. & Cortes, J. Homoharringtonine/omacetaxine mepesuccinate: the long and winding road to food and drug administration approval. *Clin Lymphoma Myeloma Leuk* **13**, 530-533 (2013).

9 Lü, S. & Wang, J. Homoharringtonine and omacetaxine for myeloid hematological malignancies. *J Hematol Oncol* **7**, 2 (2014).

10 Chen, Y.-f. & Fu, L.-w. Mechanisms of acquired resistance to tyrosine kinase inhibitors. *Acta Pharmaceutica Sinica B* **1**, 197-207 (2011).

11 Gürel, G., Blaha, G., Moore, P. B. & Steitz, T. A. U2504 determines the species specificity of the A-site cleft antibiotics: the structures of tiamulin, homoharringtonine, and bruceantin bound to the ribosome. *J Mol Biol* **389**, 146-156 (2009).

12 Huang, M. T. Harringtonine, an inhibitor of initiation of protein biosynthesis. *Mol Pharmacol* **11**, 511-519 (1975).

13 Garreau de Loubresse, N. *et al.* Structural basis for the inhibition of the eukaryotic ribosome. *Nature* **513**, 517-522 (2014).

14 Tujebajeva, R. M., Graifer, D. M., Karpova, G. G. & Ajtkhozhina, N. A. Alkaloid homoharringtonine inhibits polypeptide chain elongation on human ribosomes on the step of peptide bond formation. *FEBS Letters* **257**, 254-256 (1989).

15 Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M. & Weissman, J. S. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nature Protocols* **7**, 1534-1550 (2012).

16 Delfel, N. E. Alkaloid distribution and catabolism in Cephalotaxus harringtonia. *Phytochemistry* **19**, 403-408 (1980).

17 Powell, R. G. Structures of homoerythrina alkaloids from cephalotaxus harringtonia. *Phytochemistry* **11**, 1467-1472 (1972).

18 Ju, X. & Beaudry, C. M. Total Synthesis of (-)-Cephalotaxine and (-)-Homoharringtonine via Furan Oxidation-Transannular Mannich Cyclization. *Angew Chem Int Ed Engl* **58**, 6752-6755 (2019).

19 Robin, J.-P. *et al.* The first semi-synthesis of enantiopure homoharringtonine via anhydrohomoharringtonine from a preformed chiral acyl moiety. *Tetrahedron Letters* **40**, 2931-2934 (1999).

20 Hiranuma, S. & Hudlicky, T. Synthesis of homoharringtonine and its derivative by partial esterification of cephalotaxine. *Tetrahedron Letters* **23**, 3431-3434 (1982).

21 Isono, N. & Mori, M. Total Synthesis of (-)-Cephalotaxine. *The Journal of Organic Chemistry* **60**, 115-119 (1995).

22 Li, W. D. & Wang, Y. Q. A novel and efficient total synthesis of cephalotaxine. *Org Lett* **5**, 2931-2934 (2003).

23 Yang, Y. L., D.; Katsuki, T. Cephalotaxus harringtonii. The IUCN Red List of Threatened Species 2013. e.T39589A2929537 (2013).

24 Dakota, B. C. B. S. o. N. *Omacetaxine mepesuccinate (Synribo) (ARCHIVED)*, <https://www.bcbsnd.com/providers/policies-precertification/medical-policy/o/omacetaxine-mepesuccinate-synribo> (2024).

25 Administration, U. S. F. a. D. *Drugs@FDA: FDA-approved drugs – New Drug Application (NDA) 203585 (SYNRIBO)*, <https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=203585> (n.d.).

26 American Society of Health-System Pharmacists, I. *Omacetaxine Injection*, <https://medlineplus.gov/druginfo/meds/a613012.html> (2024).

27 Parry, R. J. & Schwab, J. M. Biosynthesis of Cephalotaxus alkaloids. I. Novel mode of tyrosine incorporation into cephalotaxine. *Journal of the American Chemical Society* **97**, 2555-2557 (1975).

28 Gitterman, A., Parry, R. J., Dufresne, R. F., Sternbach, D. D. & Cabelli, M. D. Biosynthesis of the Cephalotaxus alkaloids. Investigations of the biosynthesis of deoxyharringtonine, isoharringtonine, and harringtonine. *Journal of the American Chemical Society* **102**, 2074-2081 (1980).

29 Schwab, J. M., Chang, M. N. & Parry, R. J. Biosynthesis of Cephalotaxus alkaloids. 3. Specific incorporation of phenylalanine into cephalotaxine. *Journal of the American Chemical Society* **99**, 2368-2370 (1977).

30 Qiao, F. *et al.* Elucidation of the 1-phenethylisoquinoline pathway from an endemic conifer Cephalotaxus hainanensis. *Proceedings of the National Academy of Sciences* **120**, e2209339120 (2023).

31 Nett, R. S., Lau, W. & Sattely, E. S. Discovery and engineering of colchicine alkaloid biosynthesis. *Nature* **584**, 148-153 (2020).

32 Jacobowitz, J. R. & Weng, J.-K. Exploring uncharted territories of plant specialized metabolism in the postgenomic era. *Annual review of plant biology* **71**, 631-658 (2020).

33 Mehta, N., Meng, Y., Zare, R., Kamenetsky-Goldstein, R. & Sattely, E. A developmental gradient reveals biosynthetic pathways to eukaryotic toxins in monocot geophytes. *Cell* **187**, 5620-5637. e5610 (2024).

34 Lau, W. & Sattely, E. S. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* **349**, 1224-1228 (2015).

35 Caputi, L. *et al.* Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science* **360**, 1235-1239 (2018).

36 Nett, R. S. *et al.* D(2)O Labeling to Measure Active Biosynthesis of Natural Products in Medicinal Plants. *AIChE J* **64**, 4319-4330 (2018).

37 Lau, W., Sattely, E., Smolke, C. D. & Spormann, A. M. *Pathway discovery and engineering of pharmaceuticals from medicinal plants [electronic resource]*. (2017).

38 in *Evolution of the Arborescent Gymnosperms: Pattern, Process and Diversity: Volume 1: Northern Hemisphere Focus* Vol. 1 (ed Christopher N. Page) 338-351 (Cambridge University Press, 2024).

39 Tomlinson, P. B. & Zacharias, E. H. Phyllotaxis, phenology and architecture in Cephalotaxus, Torreya and Amentotaxus (Coniferales). *Botanical Journal of the Linnean Society* **135**, 215-228 (2001).

1163 40 Nett, R. S., Dho, Y., Low, Y.-Y. & Sattely, E. S. A metabolic regulon reveals early and late
1164 acting enzymes in neuroactive Lycopodium alkaloid biosynthesis. *Proceedings of the National*
1165 *Academy of Sciences* **118**, e2102949118 (2021).
1166 41 Conesa, A. *et al.* A survey of best practices for RNA-seq data analysis. *Genome biology* **17**, 13
1167 (2016).
1168 42 Morris, J. S. & Facchini, P. J. Molecular origins of functional diversity in benzylisoquinoline
1169 alkaloid methyltransferases. *Frontiers in plant science* **10**, 1058 (2019).
1170 43 Sainsbury, F., Thuenemann, E. C. & Lomonossoff, G. P. pEAQ: versatile expression vectors for
1171 easy and quick transient expression of heterologous proteins in plants. *Plant biotechnology*
1172 *journal* **7**, 682-693 (2009).
1173 44 Nguyen, T.-D. & Dang, T.-T. T. Cytochrome P450 enzymes as key drivers of alkaloid chemical
1174 diversification in plants. *Frontiers in Plant Science* **12**, 682181 (2021).
1175 45 Kakizaki, T. *et al.* A 2-Oxoglutarate-Dependent Dioxygenase Mediates the Biosynthesis of
1176 Glucoraphasatin in Radish. *Plant Physiology* **173**, 1583-1593 (2017).
1177 46 Kawai, Y., Ono, E. & Mizutani, M. Evolution and diversity of the 2–oxoglutarate-dependent
1178 dioxygenase superfamily in plants. *The Plant Journal* **78**, 328-343 (2014).
1179 47 Hong, B. *et al.* Biosynthesis of strychnine. *Nature* **607**, 617-622 (2022).
1180 48 Nett, R. S. *et al.* Plant carbonic anhydrase-like enzymes in neuroactive alkaloid biosynthesis.
1181 *Nature* **624**, 182-191 (2023).
1182 49 Martinez, S. & Hausinger, R. P. Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent
1183 Oxygenases. *J Biol Chem* **290**, 20702-20711 (2015).
1184 50 Podust, L. M. & Sherman, D. H. Diversity of P450 enzymes in the biosynthesis of natural
1185 products. *Natural Product Reports* **29**, 1251-1266 (2012).
1186 51 Mizutani, M. & Ohta, D. Diversification of P450 genes during land plant evolution. *Annual*
1187 *review of plant biology* **61**, 291-315 (2010).
1188 52 Kupchan, S. M., Dhingra, O. P. & Kim, C.-K. New biogenetic-type approach to Cephalotaxus
1189 alkaloids and the mechanism of Schelhammera-type homoerythrinadienone formation in vitro.
1190 *The Journal of Organic Chemistry* **43**, 4464-4468 (1978).
1191 53 Marino, J. P. & Samanen, J. M. Biogenetic-type approach to homoerythrina alkaloids. *The*
1192 *Journal of Organic Chemistry* **41**, 179-180 (1976).
1193 54 Meloni, M. *et al.* Structural and biochemical characterization of Arabidopsis alcohol
1194 dehydrogenases reveals distinct functional properties but similar redox sensitivity. *Plant J* **118**,
1195 1054-1070 (2024).
1196 55 Zhang, Y. *et al.* Bioinformatic analysis of short-chain dehydrogenase/reductase proteins in plant
1197 peroxisomes. *Frontiers in Plant Science* **14**, 1180647 (2023).
1198 56 Stammers, D. K. *et al.* The structure of the negative transcriptional regulator NmrA reveals a
1199 structural superfamily which includes the short-chain dehydrogenase/reductases. *The EMBO*
1200 *Journal* **20**, 6619-6626-6626 (2001).
1201 57 Kotaka, M. *et al.* Structural analysis of the recognition of the negative regulator NmrA and DNA
1202 by the zinc finger from the GATA-type transcription factor AreA. *Journal of molecular biology*
1203 **381**, 373-382 (2008).
1204 58 Mao, T. *et al.* Genome-wide characterization of NmrA-like proteins and the regulatory function
1205 of soybean GmNmrA6 in response to salt and oxidative stresses. *Environmental and*
1206 *Experimental Botany* **213**, 105447 (2023).
1207 59 Persson, B. *et al.* The SDR (short-chain dehydrogenase/reductase and related enzymes)
1208 nomenclature initiative. *Chemico-Biological Interactions* **178**, 94-98 (2009).
1209 60 Min, T. *et al.* Crystal Structures of Pinoresinol-Lariciresinol and Phenylcoumaran Benzylic Ether
1210 Reductases and Their Relationship to Isoflavone Reductases*. *Journal of Biological Chemistry*
1211 **278**, 50714-50723 (2003).
1212 61 Guengerich, F. P. & Yoshimoto, F. K. Formation and Cleavage of C–C Bonds by Enzymatic
1213 Oxidation–Reduction Reactions. *Chemical Reviews* **118**, 6573-6655 (2018).

62 Díaz Chávez, M. L., Rolf, M., Gesell, A. & Kutchan, T. M. Characterization of two methylenedioxy bridge-forming cytochrome P450-dependent enzymes of alkaloid formation in the Mexican prickly poppy Argemone mexicana. *Archives of Biochemistry and Biophysics* **507**, 186-193 (2011).
63 Wink, M. Plant Secondary Metabolites Modulate Insect Behavior-Steps Toward Addiction? *Front Physiol* **9**, 364 (2018).
64 Jeon, J. E. *et al.* A Pathogen-Responsive Gene Cluster for Highly Modified Fatty Acids in Tomato. *Cell* **180**, 176-187.e119 (2020).
65 Li, J. *et al.* Controlled hydroxylations of diterpenoids allow for plant chemical defense without autotoxicity. *Science* **371**, 255-260 (2021).
66 Jones, P. & Vogt, T. Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. *Planta* **213**, 164-174 (2001).
67 Morant, A. V. *et al.* β-Glucosidases as detonators of plant chemical defense. *Phytochemistry* **69**, 1795-1813 (2008).
68 Yamada, K. & Mine, A. Sugar coordinates plant defense signaling. *Science Advances* **10**, eadk4131
69 Samanani, N., Alcantara, J., Bourgault, R., Zulak, K. G. & Facchini, P. J. The role of phloem sieve elements and laticifers in the biosynthesis and accumulation of alkaloids in opium poppy. *The Plant Journal* **47**, 547-563 (2006).
70 He, Y. R. *et al.* Alkaloids from Cephalotaxus lanceolata and their cytotoxicities. *Chemistry & Biodiversity* **10**, 584-595 (2013).
71 Brönstrup, M. & Sasse, F. Natural products targeting the elongation phase of eukaryotic protein biosynthesis. *Natural Product Reports* **37**, 752-762 (2020).
72 Zhu, F., Zhou, Y.-K., Ji, Z.-L. & Chen, X.-R. The plant ribosome-inactivating proteins play important roles in defense against pathogens and insect pest attacks. *Frontiers in Plant Science* **9**, 146 (2018).
73 Grela, P., Szajwaj, M., Horbowicz-Drożdżal, P. & Tchórzewski, M. How Ricin Damages the Ribosome. *Toxins (Basel)* **11** (2019).
74 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
75 Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* **8**, 1494-1512 (2013).
76 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359 (2012).
77 Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
78 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).
79 Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic acids research* **44**, D279-D285 (2016).
80 Berardini, T. Z. *et al.* The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *genesis* **53**, 474-485 (2015).
81 Xiong, X. *et al.* The Taxus genome provides insights into paclitaxel biosynthesis. *Nature Plants* **7**, 1026-1036 (2021).
82 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**, 525-527 (2016).
83 Pompon, D., Louerat, B., Bronine, A. & Urban, P. in *Methods in Enzymology* Vol. 272 (eds Eric F. Johnson & Michael R. Waterman) 51-64 (Academic Press, 1996).
84 Urban, P., Mignotte, C., Kazmaier, M., Delorme, F. & Pompon, D. Cloning, Yeast Expression, and Characterization of the Coupling of Two Distantly Related Arabidopsis thalianaNADPH-

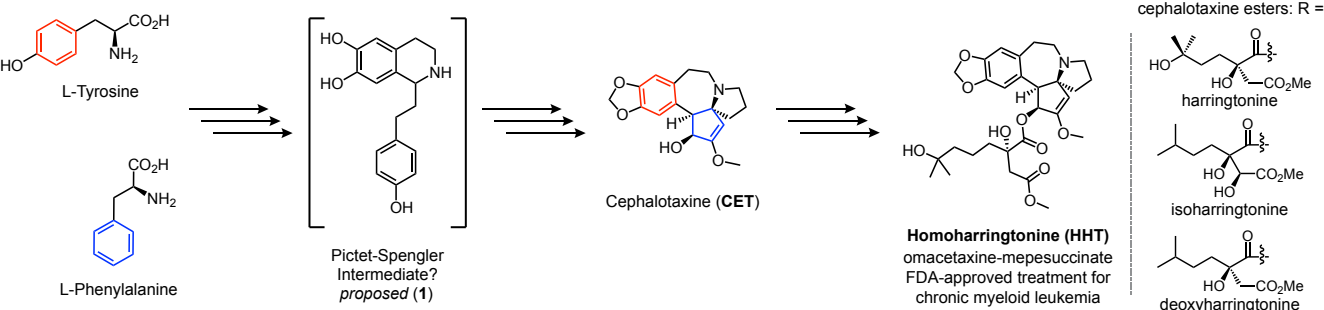
Cytochrome P450 Reductases with P450 CYP73A5*. *Journal of Biological Chemistry* **272**, 19176-19186 (1997).

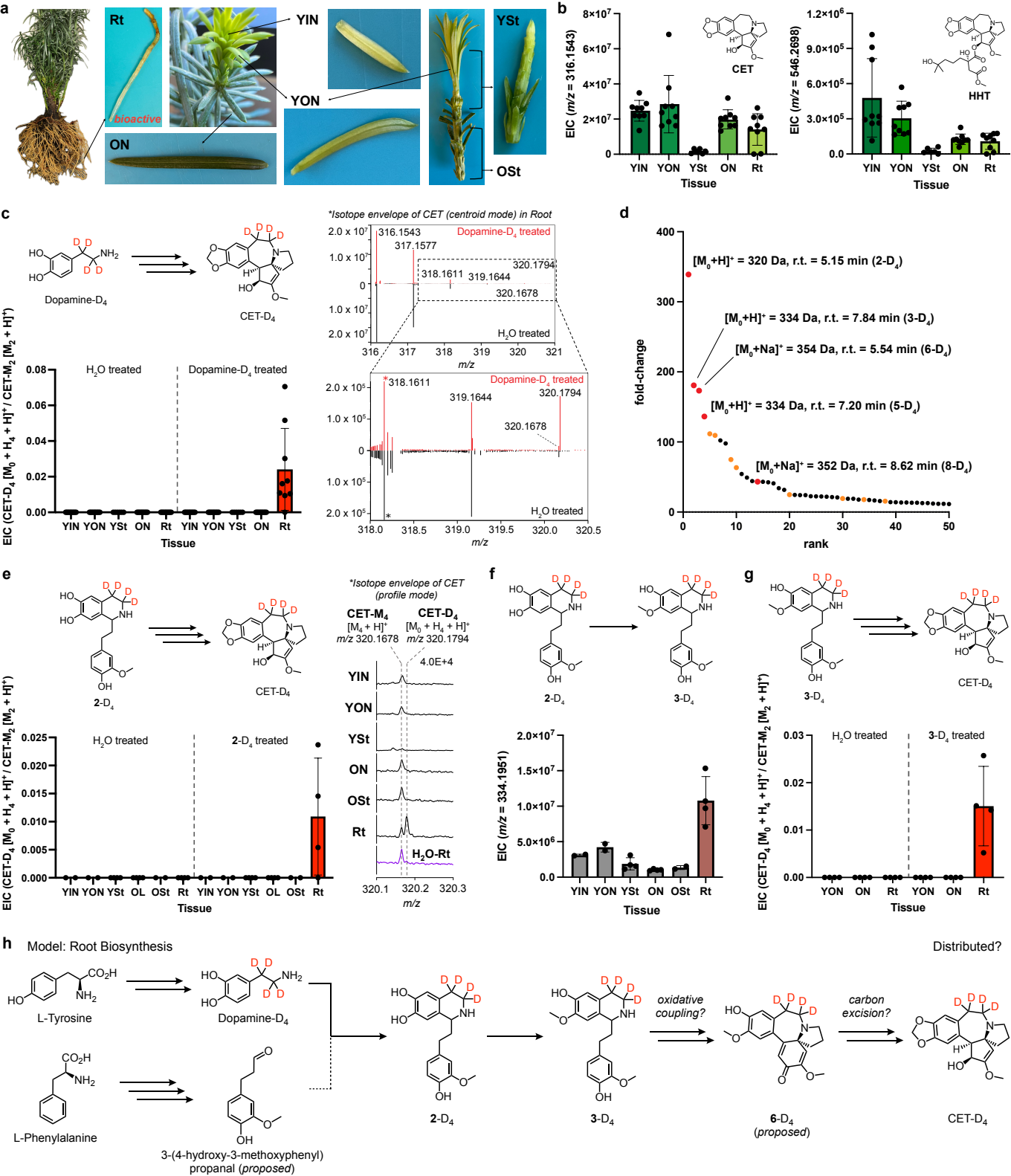
85 Recipe: Complete Minimal (CM) or Synthetic Complete (SC) and Drop-out Media. *Cold Spring Harbor Protocols* (2006). https://doi.org/https://doi.org/10.1101/pdb.rec8190

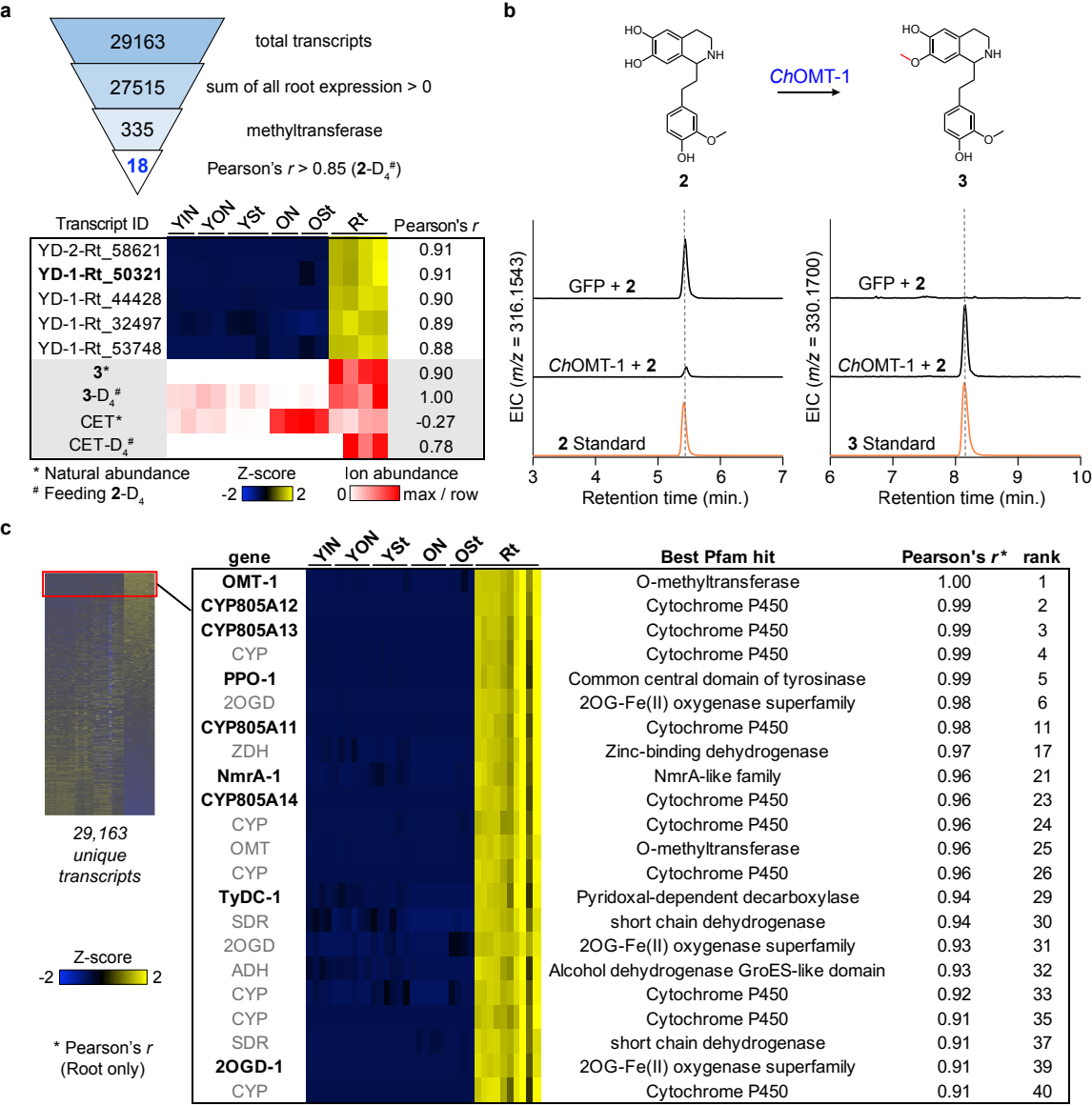
86 Recipe: Dropout Mix. *Cold Spring Harbor Protocols* (2006). https://doi.org/doi:10.1101/pdb.rec8585

87 Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**, 779-787 (2006).

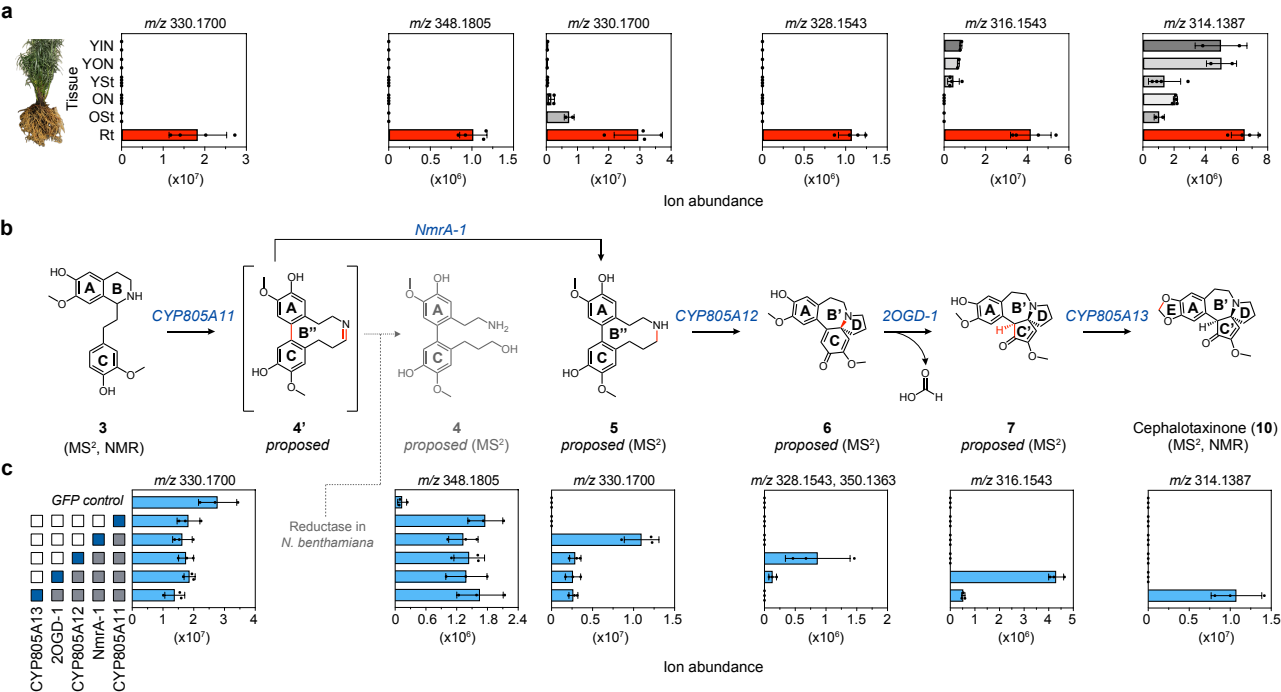
88 Pesnot, T., Gershater, M. C., Ward, J. M. & Hailes, H. C. Phosphate mediated biomimetic synthesis of tetrahydroisoquinoline alkaloids. *Chemical Communications* **47**, 3242-3244 (2011).

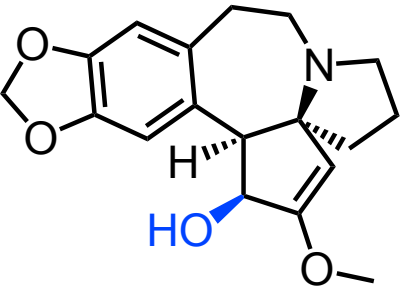

L-Tyrosine
L-Phenylalanine
Pictet-Spengler
Intermediate?
proposed (**1**)
Cephalotaxine (**CET**)
Homoharringtonine (HHT)
omacetaxine-mepesuccinate
FDA-approved treatment for
chronic myeloid leukemia
cephalotaxine esters: R =
harringtonine
isoharringtonine
deoxyharringtonine


a
Rt
bioactive
YIN
YON
ON
YSt
OSt
b
EIC (m/z = 316.1543)
CET
EIC (m/z = 546.2698)
HHT
YIN YON YSt ON Rt
Tissue
c
Dopamine-D4
CET-D4
*Isotope envelope of CET (centroid mode) in Root
Dopamine-D4 treated
H2O treated
316.1543
317.1577
318.1611
319.1644
320.1794
320.1678
m/z
EIC (CET-D4 [M0 + H4 + H]+ / CET-M2 [M2 + H]+)
H2O treated
Dopamine-D4 treated
YIN YON YSt ON Rt YIN YON YSt ON Rt
Tissue
d
[M0+H]+ = 320 Da, r.t. = 5.15 min (2-D4)
[M0+H]+ = 334 Da, r.t. = 7.84 min (3-D4)
[M0+Na]+ = 354 Da, r.t. = 5.54 min (6-D4)
[M0+H]+ = 334 Da, r.t. = 7.20 min (5-D4)
[M0+Na]+ = 352 Da, r.t. = 8.62 min (8-D4)
fold-change
rank
e
2-D4
CET-D4
*Isotope envelope of CET (profile mode)
CET-M4
[M4 + H]+
m/z 320.1678
CET-D4
[M0 + H4 + H]+
m/z 320.1794
4.0E+4
YIN
YON
YSt
ON
OSt
Rt
H2O-Rt
m/z
EIC (CET-D4 [M0 + H4 + H]+ / CET-M2 [M2 + H]+)
H2O treated
2-D4 treated
YIN YON YSt OL OSt Rt YIN YON YSt OL OSt Rt
Tissue
f
2-D4
3-D4
EIC (m/z = 334.1951)
YIN YON YSt ON OSt Rt
Tissue
g
3-D4
CET-D4
EIC (CET-D4 [M0 + H4 + H]+ / CET-M2 [M2 + H]+)
H2O treated
3-D4 treated
YON ON Rt YON ON Rt
Tissue
h
Model: Root Biosynthesis
Distributed?
L-Tyrosine
Dopamine-D4
L-Phenylalanine
3-(4-hydroxy-3-methoxyphenyl) propanal (proposed)
2-D4
3-D4
oxidative coupling?
6-D4 (proposed)
carbon excision?
CET-D4

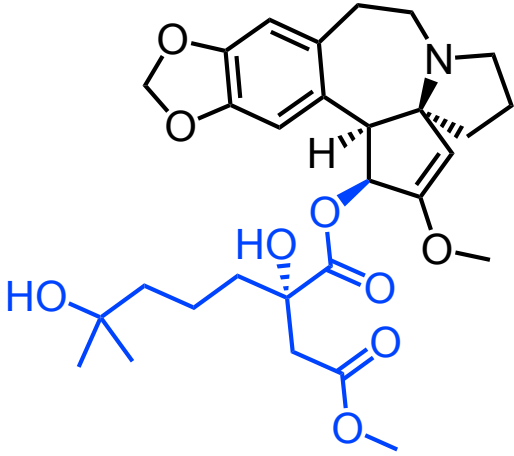


gene	Best Pfam hit	Pearson's r*	rank
OMT-1	O-methyltransferase	1.00	1
CYP805A12	Cytochrome P450	0.99	2
CYP805A13	Cytochrome P450	0.99	3
CYP	Cytochrome P450	0.99	4
PPO-1	Common central domain of tyrosinase	0.99	5
2OGD	2OG-Fe(II) oxygenase superfamily	0.98	6
CYP805A11	Cytochrome P450	0.98	11
ZDH	Zinc-binding dehydrogenase	0.97	17
NmrA-1	NmrA-like family	0.96	21
CYP805A14	Cytochrome P450	0.96	23
CYP	Cytochrome P450	0.96	24
OMT	O-methyltransferase	0.96	25
CYP	Cytochrome P450	0.96	26
TyDC-1	Pyridoxal-dependent decarboxylase	0.94	29
SDR	short chain dehydrogenase	0.94	30
2OGD	2OG-Fe(II) oxygenase superfamily	0.93	31
ADH	Alcohol dehydrogenase GroES-like domain	0.93	32
CYP	Cytochrome P450	0.92	33
CYP	Cytochrome P450	0.91	35
SDR	short chain dehydrogenase	0.91	37
2OGD-1	2OG-Fe(II) oxygenase superfamily	0.91	39
CYP	Cytochrome P450	0.91	40

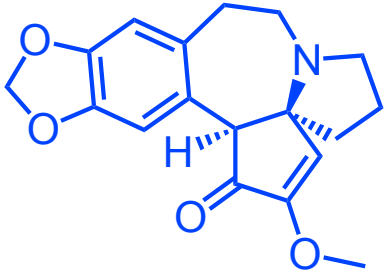

a
m/z 330.1700
m/z 348.1805
m/z 330.1700
m/z 328.1543
m/z 316.1543
m/z 314.1387
Tissue
YIN
YON
YSt
ON
OSt
Rt
(x10^7)
(x10^6)
Ion abundance
b
NmrA-1
CYP805A11
CYP805A12
2OGD-1
CYP805A13
3
(MS², NMR)
4'
proposed
4
proposed (MS²)
5
proposed (MS²)
6
proposed (MS²)
7
proposed (MS²)
Cephalotaxinone (10)
(MS², NMR)
c
GFP control
CYP805A13
2OGD-1
CYP805A12
NmrA-1
CYP805A11
Reductase in
N. benthamiana
m/z 330.1700
m/z 348.1805
m/z 330.1700
m/z 328.1543, 350.1363
m/z 316.1543
m/z 314.1387
(x10^7)
(x10^6)
Ion abundance



cephalotaxine
(reduction)



homoharringtonine
(side chain)



cephalotaxinone
(alkaloid core)