Catalytic Plasticity of Germacrene A Oxidase Underlies Sesquiterpene Lactone Diversification[1][OPEN]

Trinh-Don Nguyen,[a,2,3] Moonhyuk Kwon,[a,b,2] Soo-Un Kim,[c] Conrad Fischer,[d] and Dae-Kyun Ro[a,4,5]

[a]University of Calgary, Department of Biological Sciences, Calgary, AB T2N 1N4, Canada
[b]Division of Applied Life Science (BK21 Plus), Plant Molecular Biology and Biotechnology Research Center, Gyeongsang National University, Jinju 52828, Republic of Korea
[c]Department of Agricultural Biotechnology and Institute of Agricultural Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea
[d]Department of Chemistry, University of Alberta, Edmonton, AB T6G 2G2, Canada

ORCID IDs: 0000-0002-2759-5791 (T.-D.N.); 0000-0003-0862-2149 (M.K.); 0000-0002-2289-010X (C.F.); 0000-0003-1288-5347 (D.-K.R.).

Adaptive evolution of enzymes benefits from catalytic promiscuity. Sesquiterpene lactones (STLs) have diverged extensively in the Asteraceae, and studies of the enzymes for two representative STLs, costunolide and artemisinin, could provide an insight into the adaptive evolution of enzymes. Costunolide appeared early in Asteraceae evolution and is widespread, whereas artemisinin is a unique STL appearing in a single Asteraceae species, *Artemisia annua*. Therefore, costunolide is a ubiquitous STL, while artemisinin is a specialized one. In costunolide biosynthesis, germacrene A oxidase (GAO) synthesizes germacrene A acid from germacrene A. Similarly, in artemisinin biosynthesis, amorphadiene oxidase (AMO) synthesizes artemisinic acid from amorphadiene. GAO promiscuity is suggested to drive the diversification of STLs. To examine the degree of GAO promiscuity, we expressed six sesquiterpene synthases from cotton (*Gossypium arboretum*), goldenrod (*Solidago canadensis*), valerian (*Valeriana officinalis*), agarwood (*Aquilaria crassna*), tobacco (*Nicotiana tabacum*), and orange (*Citrus sinensis*) in yeast to produce seven distinct sesquiterpene substrates (germacrene D, 5-*epi*-aristolochene, valencene, δ-cadinene, α- and δ-guaienes, and valerenadiene). *GAO* or *AMO* was coexpressed in these yeasts to evaluate the promiscuities of GAO and AMO. Remarkably, all sesquiterpenes tested were oxidized to sesquiterpene acids by GAO, but negligible activities were found from AMO. Hence, GAO apparently has catalytic potential to evolve into different enzymes for synthesizing distinct STLs, while the recently specialized AMO demonstrates rigid substrate specificity. Mutant GAOs implanted with active site residues of AMO showed substantially reduced stability, but their per enzyme activities to produce artemisinic acid increased by 9-fold. Collectively, these results suggest promiscuous GAOs can be developed as novel catalysts for synthesizing unique sesquiterpene derivatives.

[1]This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair (CRC) program to D.-K.R and the Bettina Bahlsen Memorial Graduate Scholarship to T.-D.N. This work was also supported by a grant from the Next-Generation BioGreen 21 Program (SSAC, grant no. PJ01326501), Rural Development Administration (RDA), Republic of Korea, and the Basic Science Research Program by the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1A6A3A03003409 to M.K.).

[2]These authors contributed equally to the article.
[3]Present address: Department of Chemistry, University of British Columbia Okanagan, Kelowna, BC V1V 1V7, Canada.
[4]Author for contact: daekyun.ro@ucalgary.ca.
[5]Senior author.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Dae-Kyun Ro (daekyun.ro@ucalgary.ca).

T.-D.N. and M.K. performed cloning, analytical, and biochemical experiments; T.-D.N. and D.-K.R. designed the project and wrote the manuscript; M.K. performed LC-MS and mutant P450 analyses; S.-U.K. performed molecular modeling and structural interpretation; C.F. interpreted NMR data.

[OPEN]Articles can be viewed without a subscription.
www.plantphysiol.org/cgi/doi/10.1104/pp.19.00629

The Asteraceae (or Compositae) is the largest plant family comprised of more than 24,000 species, including some important crop and medicinal plants, such as sunflower (*Helianthus annuus*), lettuce (*Lactuca sativa*), and *Artemisia annua* (Panero and Funk, 2008). Due to the enormous diversity and convergent evolution, the origin and phylogeny of the Asteraceae have been difficult topics in the field of classical morphology-based plant systematics. Molecular data together with fossil evidence, however, have shown that the Asteraceae first appeared in South America ~50 million years ago and adapted successfully in all continents except in Antarctica (Barreda et al., 2010, 2012). Among 13 subfamilies of the Asteraceae, the Barnadesioideae is considered to be a basal lineage of all Asteraceae plants (Jansen and Palmer, 1987). This is supported by the lack of a 22-kb inversion in the plastidic genome of the Barnadesioideae, a shared feature in all other Asteraceae plants. This unique plastidic genome structure has entitled the Barnadesioideae to be a living fossil or "mother-of-all-Asteraceae," to which many other variations by different environmental adaptations can be referenced (Panero and Funk, 2008). Rooting from the Barnadesioideae, other subfamilies of the Asteraceae

are taxonomically well resolved (Fig. 1A), providing a solid taxonomic framework to investigate the chemical evolution associated with plant diversifications.

One characteristic phytochemical class in the Asteraceae is sesquiterpene lactone (STL), defined as a fifteen-carbon terpenoid possessing an α-methylene γ-lactone group. Although the structures of thousands of STLs have been elucidated, their carbon backbones can be traced to about a dozen skeletal types, on which various side chain decorations occur to increase the structural diversity of STLs (Picman, 1986; Padilla-Gonzalez et al., 2016). Costunolide (**3**, see Fig. 1B for structures) is one of the simplest STLs in the Asteraceae. At the entry point of the biosynthesis of **3**, germacrene A synthase (GAS) catalyzes the formation of the germacrene A (**1**) backbone from farnesyl pyrophosphate (FPP) by a carbocation rearrangement (Fig. 1B; Bennett et al., 2002). Then, C12 of **1** is oxidized by germacrene A oxidase (GAO) to produce germacrene A acid (**2**; Nguyen et al., 2010; Cankar et al., 2011; Ramirez et al., 2013; Eljounaidi et al., 2014). Subsequently, a regio- and stereo-selective hydroxylation of C6 of **2** by costunolide synthase (COS), followed by a spontaneous lactonization, completes the biosynthesis of **3** (Ikezawa et al., 2011; Liu et al., 2011, 2014; Eljounaidi et al., 2014). Costunolide (**3**) is believed to be a gateway compound to some C6-C7-fused STLs (e.g. eudesmanolide, elemanolide, and guaianolide), and **3** and its derivatives have been found in many different Asteraceae plants (Picman, 1986). Analogous reactions occur in the

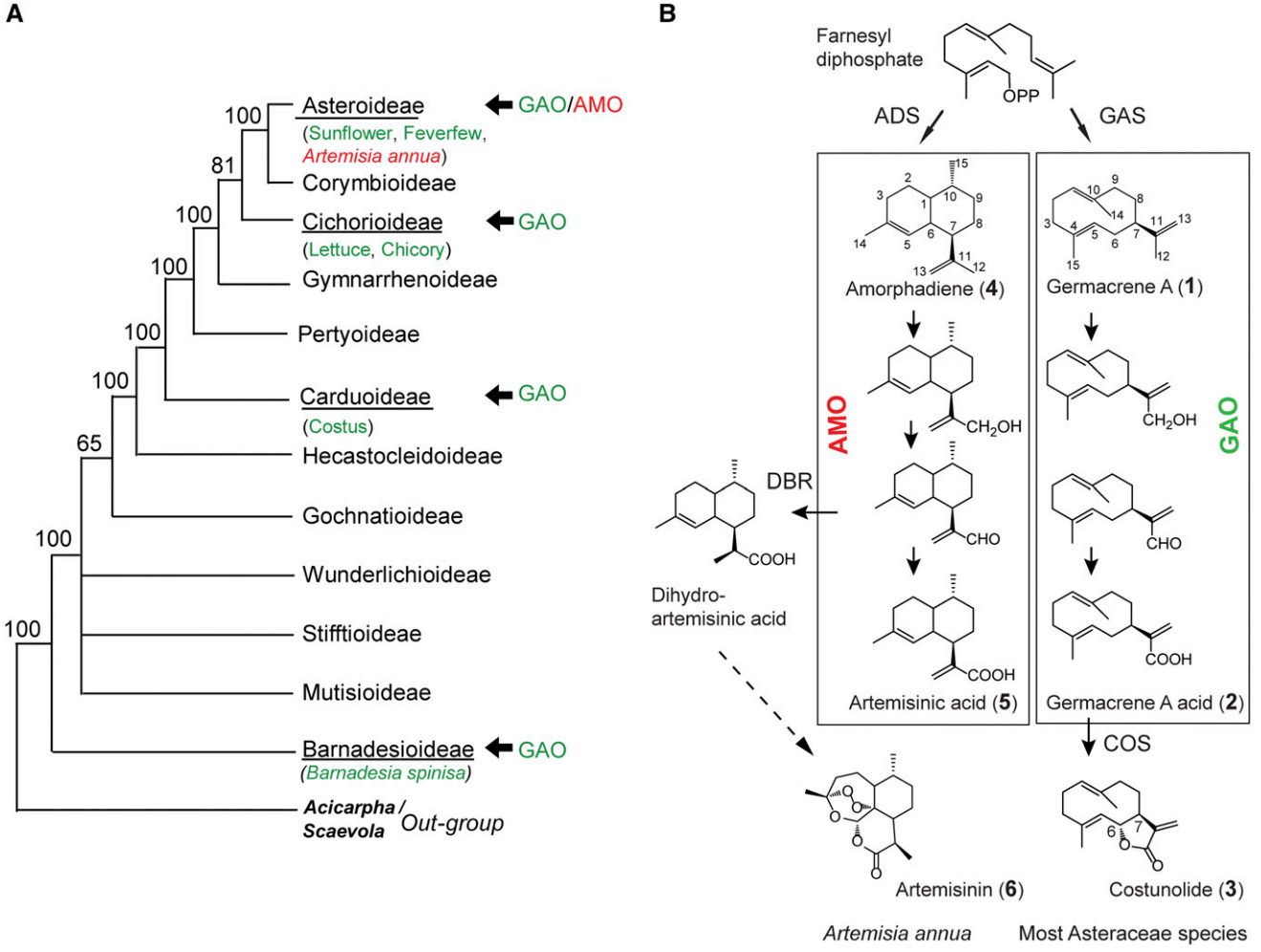

Figure 1. Sesquiterpene lactone metabolism in the Asteraceae family. A, Characterized sesquiterpene oxidases in the biosynthetic pathways of sesquiterpene lactones in the Asteraceae subfamilies. Among these sesquiterpene oxidases, amorphadiene oxidase (AMO) occurs in a single species, *Artemisia annua* of the Asteroideae subfamily, while germacrene A oxidase (GAO) is present in six species in four subfamilies (underlined). Bootstrap values are given at each node. B, Oxidation of sesquiterpenes in the biosynthetic pathways of sesquiterpene lactones. In artemisinin biosynthesis (left), amorphadiene is oxidized by AMO to form artemisinic aldehyde (a biological precursor of artemisinin) and further to artemisinic acid, which can be chemically converted to artemisinin (dashed arrow). Costunolide biosynthesis (right) is considered the general sesquiterpene lactone pathway in the Asteraceae. ADS, Amorphadiene synthase; GAS, germacrene A synthase; DBR, double-bond reductase; COS, costunolide synthase.

biosynthesis of artemisinin (**6**), a well-known potent antimalarial drug only found in a single plant species—*A. annua* (Fig. 1B). Amorphadiene synthase (ADS) and amorphadiene oxidase (AMO or CYP71AV1) catalyze the synthesis of artemisinic aldehyde (Ro et al., 2006; Teoh et al., 2006), which is converted to dihydroartemisinic aldehyde by a double-bond reductase and further to dihydroartemisinic acid. Subsequently, dihydroartemisinic acid further undergoes a photooxidation to produce artemisinin (Zhang et al., 2008; Paddon et al., 2013). Alternatively, artemisinic acid (**5**) could be chemically converted to artemisinin (Fig. 1B) in pharmaceutical industries (Paddon and Keasling, 2014).

Because GAOs from various species (lettuce [*L. sativa*], sunflower [*H. annuus*], costus [*Saussurea lappa*], and spiny barnadesia [*Barnadesia spinosa*]) and AMO share significant sequence identity (> 79% amino acid identity) and catalyze the formation of structurally similar sesquiterpenoid acids, adaptations of these enzymes for different substrates were expected to occur along the diverse speciation of Asteraceae plants. GAO is involved in the biosynthesis of costunolide and its derivatives in many Asteraceae plants. GAO orthologs have been identified from three major subfamilies that constitute 95% of all Asteraceae plants (Carduoidea, Cichorioideae, and Asteroideae) and the phylogenetically basal subfamily Barnadesioideae (Nguyen et al., 2010; Cankar et al., 2011; Ramirez et al., 2013; Eljounaidi et al., 2014). Furthermore, their authentic GAO activities were biochemically confirmed from sunflower, chicory (*Cichorium intybus*), costus, and *B. spinosa*. On the contrary, AMO is found only in one species, *A. annua*, in the Asteroideae subfamily, suggesting that the speciation of *A. annua* is tightly associated with the restricted occurrence of artemisinin (**6**). Previous biochemical studies showed that GAOs display a substantial cross reactivity for a nonnative substrate, amorphadiene (**4**), to synthesize artemisinic acid (**5**), whereas AMO activity is restricted to its native substrate amorphadiene and cannot oxidize germacrene A (**1**; Nguyen et al., 2010).

Taking the phylogenetic occurrences and biochemical data of GAO and AMO into consideration, we contemplated that GAO is a widespread, generalist enzyme occurring in the majority of species in the Asteraceae, while AMO is a restrictive, specialist enzyme only present in *A. annua*. It was hypothesized that inherent substrate plasticity (or promiscuity) of GAO can drive the structural diversity of STLs in the Asteraceae, and therefore closer analyses of GAO and AMO catalytic activities would allow us to gain insights into the enzyme promiscuity and speciation occurring in nature. However, examining the cross reactivity of GAO and AMO only for oxidizing **1** and **4** (Nguyen et al., 2010) still left uncertainties as to whether GAO is truly promiscuous for structurally different sesquiterpene substrates and whether AMO is a specialized enzyme evolved only for the oxidation of **4**, and hence, here we embarked on a systematic analysis of GAO/AMO catalytic properties using diverse substrates. Such studies of enzyme promiscuity inspired by adaptive evolution can ultimately enable us to identify and improve specific P450 activities capable of catalyzing chemically difficult reactions.

To address these questions, *GAO*s from *B. spinosa* (*CYP71AV7*) and lettuce (*CYP71AV3* and *CYP71AV15*) or *AMO* (*CYP71AV1*) from *A. annua* was coexpressed in yeast with several sesquiterpene synthases, which synthesize diverse skeletons of sesquiterpene substrates for GAO or AMO enzymes. These experiments substantiated that the substrate plasticity is embedded in GAO sequences, but not in AMO sequences. These results also suggest that the latent catalytic plasticity of GAO is an underlying principle in promoting the STL diversity, such as the case with artemisinin, in the Asteraceae.

RESULTS

Isolation and Characterization of LsGAO2 from Lettuce

In silico screening of the updated lettuce transcriptomics data generated by the Illumina sequencing led to the identification of a second *GAO* isoform, and its full-length complementary DNA (cDNA) was isolated from lettuce. This new *GAO* gene in lettuce was named *LsGAO2* (*CYP71AV15*; GenBank: KF981867), while previously characterized lettuce *GAO* (GU198171) was named *LsGAO1* (*CYP71AV3*). *LsGAO2* encodes a polypeptide of 497 amino acids with 76% and 75% identities to LsGAO1 and AMO, respectively. To examine whether *LsGAO2* also encodes a C12 oxidation activity of germacrene A (**1**), it was coexpressed with *GAS* and *cytochrome P450 reductase* (*CPR*) in yeast. For this purpose, the triple expression vector pESC-Leu-2d::*LsGAO2/GAS/CPR* was constructed as previously shown (Nguyen et al., 2012) and transformed into EPY300 yeast strain, which was engineered to increase the endogenous FPP level. The sesquiterpene metabolites synthesized from this transgenic yeast were compared to those from the EPY300 transformed with pESC-Leu-2d::*LsGAO1/GAS/CPR*, known to produce germacrene A acid (**2**). In liquid chromatography mass-spectrometry (LC-MS) analysis, both yeast strains could synthesize **2** as a single dominant product (Fig. 2). This result showed that lettuce has a second isoform of *GAO* that also catalyzes three sequential oxidations on the isopropenyl moiety's terminal carbon (C12) of **1** (Fig. 1B).

We previously reported that LsGAO1 shows a substantial cross reactivity for C12 oxidations of its nonnatural substrate, amorphadiene (**4**; Nguyen et al., 2010). Since LsGAO2 shows a high degree of sequence identity to both LsGAO1 and AMO, the cross reactivity of LsGAO2 on **4** was also examined by coexpressing *LsGAO2* with *ADS/CPR* in yeast. Metabolite analysis by LC-MS showed that the transgenic yeast expressing *LsGAO2/ADS/CPR* could synthesize artemisinic acid (**5**; Fig. 2). These results revealed that

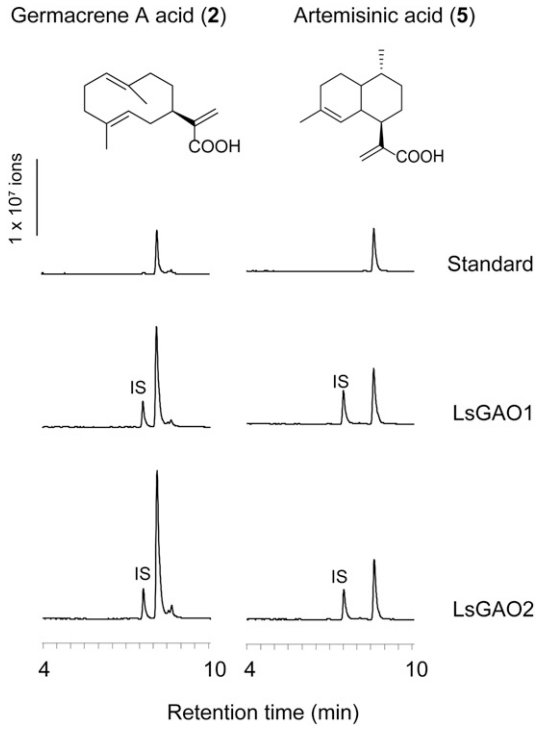

Figure 2. Negative Ion LC-MS Analyses of LsGAO2 Activities in Yeast. Metabolite profiles from EPY300 coexpressing *LsGAO2/CPR/GAS* (left) or *LsGAO2/CPR/ADS* (right). IS indicates internal standard (decanoic acid).

LsGAO2 has a cross reactivity for the oxidation of **4** as reported for LsGAO1. We concluded that two copies of *GAO* are encoded in lettuce, and both showed catalytic activities for **1** and **4**.

Establishing Yeast Strains Producing Diverse Sesquiterpenes

GAOs are widely present and conserved in the Asteraceae family, starting from the phylogenetically basal subfamily Barnadesioideae (Fig. 1A). Intriguingly, GAOs can oxidize a structurally distinct substrate, amorphadiene (**4**), to produce artemisinic acid (**5**), which suggests that a latent catalytic plasticity is encoded in GAOs (Nguyen et al., 2010). However, it remained unknown whether GAOs are truly promiscuous for diverse sesquiterpene substrates.

To examine the substrate plasticity of GAOs beyond **4**, we aimed to test GAO activities for seven other sesquiterpenes with different structural backbones in this work. The structures of these sesquiterpenes (shown in Fig. 3) are one 10-carbon-ring monocyclic structure (germacrene D [**8**] from *Solidago canadensis*), three bicyclic structures with two fused six-carbon rings (δ-cadinene [**7**] from cotton [*Gossypium arboreum*], 5-*epi*-aristolochene [**12**] from tobacco [*Nicotiana tabacum*], and valencene [**13**] from *Citrus sinensis*), two bicyclic structures with fused five- and seven-carbon rings (α- and δ-guaienes [**10** and **11**, respectively] from *Aquilaria crassna*), and a bicyclic sesquiterpene with fused five- and six-carbon rings (valerenadiene [**9**] from *Valeriana officinalis*). The side chain moieties of these sesquiterpenes, to be oxidized by AMO and GAOs as observed in the cases of germacrene A and amorphadiene, are also structurally distinct. Isopropenyl group is present in compounds **10**, **11**, **12**, and **13**; isopropyl group occurs in the compounds **7** and **8**; and isobutenyl group exists in compound **9**. Therefore, these sesquiterpenes represent the diverse hydrocarbon skeletons of sesquiterpenes that are present in the Asteraceae (**7**, **8**, **10**, and **11**) and those not reported to be present in the Asteraceae (**9**, **12**, and **13**).

Since most of these compounds are not commercially available, their published sesquiterpene synthase (*STS*) cDNAs were either isolated from respective plant sources or chemically synthesized (i.e. valencene and guaiene synthase denoted as *SynVLS* and *SynGUS*, respectively; Fig. 3). The acquired cDNA clones were then used to produce sesquiterpenes in yeast. When organic extracts of the transgenic yeast culture were analyzed by gas chromatography-mass spectrometry (GC-MS), the transgenic yeasts expressing each *STS* cDNA produced the desired sesquiterpenes with identical mass-fragmentation patterns previously published with estimated yields around 100 (δ-cadinene 7),

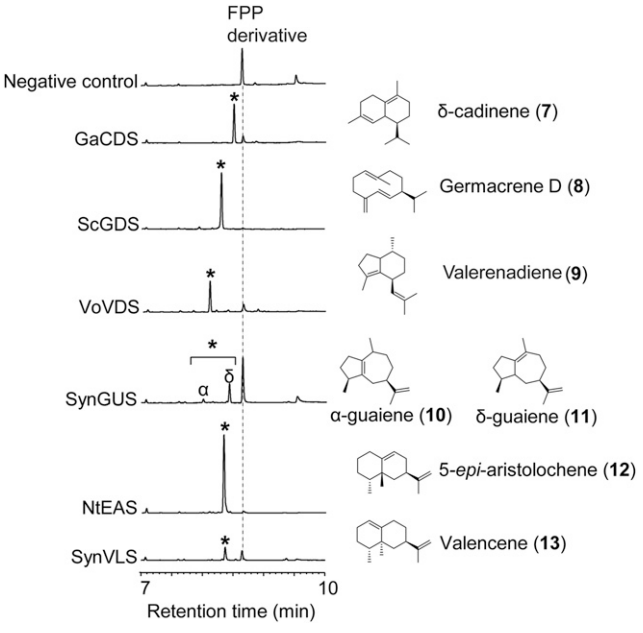

Figure 3. De novo synthesis of plant sesquiterpenes in yeast. GC-MS analysis of metabolites from EPY300 yeast expressing sesquiterpene synthases. Peaks of sesquiterpene products are indicated with asterisks; the corresponding structures are depicted to the right. Negative control was EPY300 transformed with empty *pESC-Leu-2d* vector. GaCDS, cotton (*Gossypium aboreum*) δ-cadinene synthase; ScGDS, Canadian goldenrod (*Solidago canadensis*) germacrene D synthase; VoVDS, valerian (*Valeriana officinalis*) valerenadiene synthase; SynGUS, synthetic guaiene synthase; NtEAS, tobacco (*Nicotiana tabacum*) 5-*epi*-aristolochene synthase; SynVLS, synthetic valencene synthase.

200 (germacrene D **8**), 100 (valerenadiene **9**), 10 (δ-guaiene **10**), 50 (α-guaiene **11**), 400 (5-*epi*-aristolocheme **12**), and 50 (valenecene **13**) μg mL^{-1} (Fig. 3). These data indicated that all *STS* cDNAs used in this study were properly expressed in yeast, allowing de novo biosynthesis of the expected sesquiterpenes. Therefore, in these yeast backgrounds, substrate specificities of GAO and AMO for diverse sesquiterpenes can be examined.

Selections of GAOs for Catalytic Activity Tests

The conserved germacrene A (**1**)-oxidizing activities of GAOs from various subfamilies of the Asteraceae family were demonstrated previously (Nguyen et al., 2010; Cankar et al., 2011; Ramirez et al., 2013; Eljounaidi et al., 2014) and in this study (LsGAO2). Phylogenetic analysis of the characterized GAOs and AMO showed that LsGAO2 and CiGAO2 clustered in the same clade with BsGAO, while the rest of the AMO/GAO homologs form a distinct clade supported by a significant bootstrap value (99%; Supplemental Fig. S1). Based on these studies, three GAO clones were selected for more refined analyses of their activities in comparison to AMO activity. The first one was BsGAO, representing a sesquiterpene oxidase from the Barnadesioideae subfamily, the phylogenetic base of the Asteraceae family. Two other GAOs selected were LsGAO1 and LsGAO2 in lettuce, representing two GAO clades that diverged in the Cichoriodeae subfamily during adaptive speciation from the Barnadesioideae subfamily (Fig. 1; Supplemental Fig. S1). These three GAOs represent widespread and conserved GAOs in the Asteraceae family. In contrast, AMO is an explicitly specialized P450 only found in a single species, *A. annua*, and it displays a high sequence identity to GAOs. Pairwise combinations of six STS and four P450 cDNAs (three *GAO*s and one *AMO*) resulted in constructing 24 sets of combinatorial expression cassettes with *CPR* as a shared gene. All 24 expression cassettes were generated in a pESC-Leu-2d-based triple expression plasmid (Nguyen et al., 2012). Subsequently, these constructs were transformed to EPY300 yeast, and the metabolites from the transgenic yeasts were analyzed by LC-MS.

Evaluating Promiscuous Activities of GAO and AMO in Yeast

Catalytic promiscuity of GAOs for various sesquiterpenes was first examined by (−)-LC-MS to identify the respective carboxylic acids from sesquiterpene substrates. The major (−)-ions detected were *m/z* (−)-251 from **7** to **13** sesquiterpenes (triangles in Fig. 4A) together with the minor (−) ion displaying *m/z* (−)-233 (diamonds in Fig. 4A). However, when *ADS* and *GAO*s were coexpressed to allow **4** to react with GAOs, only (−)-233 ion was detected without the formation of (−)-251 ion. The *m/z* (−)-233 ions are anticipated sesquiterpene acids by oxidations of **4** and **7** to **13** by AMO or GAOs, whereas the *m/z* (−)-251 ions are hydrated compounds of sesquiterpene acids as represented by $[M\text{-}H+H_2O]^-$. Expected molecular formula of all (−)-233 and (−)-251 molecules were validated by high-resolution (HR)-LC-MS analyses within $\Delta 3\ 10^{-6}$ mass accuracy (Supplemental Table S1).

Biosynthesis of (−)-ion compound with *m/z* 233 and *m/z* 251 showed that each of three GAOs can catalyze the three sequential oxidations on diverse sesquiterpene hydrocarbons to produce respective forms of sesquiterpene acids. Of the three GAOs examined, GAO2 showed the highest level of cross reactivities. On the contrary, when *AMO* was coexpressed with various *STS* cDNAs, the (−)-ion compound with the same mass and retention time as those from *GAO* coexpressing yeasts were not detected in three samples (the oxidized compounds from **7**, **8**, and **13**; Fig. 4A, chromatogram in the top row), and whenever detected, their abundance was 10- to 100-fold lower than *GAO* coexpressers, such as the oxidized products from **9** to **12**. It was therefore evident that GAOs are promiscuous while AMO is not in these experiments.

To examine whether oxidized sesquiterpenes other than sesquiterpene acids can be formed, metabolites from the yeast coexpressing *LsGAO2* with an individual *STS* were further analyzed by (+)-LC-MS (Fig. 4B). Compounds with *m/z* (+)-221 were detected in all strains except for the *ScGDS/LsGAO2*-expressing strain, and this mass corresponds to sesquiterpene alcohol $[M+O+H]^+$, putative intermediates of various sesquiterpene acids (Fig. 4B). Another (+)-ion product of *m/z* (+)-239 was detected from the yeast strains expressing *ScGDS/GAO2* or *NtEAS/LsGAO2* (Fig. 4B). Based on the interpretations of the mass data, these unique compounds are diol compounds $[M+O+H_2O+H]^+$. The third (+)-ion product with *m/z* (+)-219 corresponding to sesquiterpene ketone were also detected from three yeast strains (Fig. 4B). Expected molecular formula of all (+) ions were validated by HR-LC-MS analyses with $\Delta 3.5\ 10^{-6}$ mass accuracy (Supplemental Table S1). Structural analyses of some of these compounds are described below.

To ensure that the negligible promiscuous activities of AMO on various sesquiterpene substrates are not due to the lack of AMO enzyme, immunoblot analyses of AMO enzyme were performed in parallel with three other GAO enzymes in all experimental sample sets. The immunoblot data showed that the protein levels of AMO were not compromised in all experiments (Fig. 4, top). With these data, we concluded that AMO is specific to its native substrate amorphadiene, whereas GAOs can utilize a wide range of substrates with different backbones and side chains.

Structural Analyses of Oxidized Sesquiterpenes

Chemical identities of some oxidized sesquiterpene products identified from (+/−)-LC-MS were

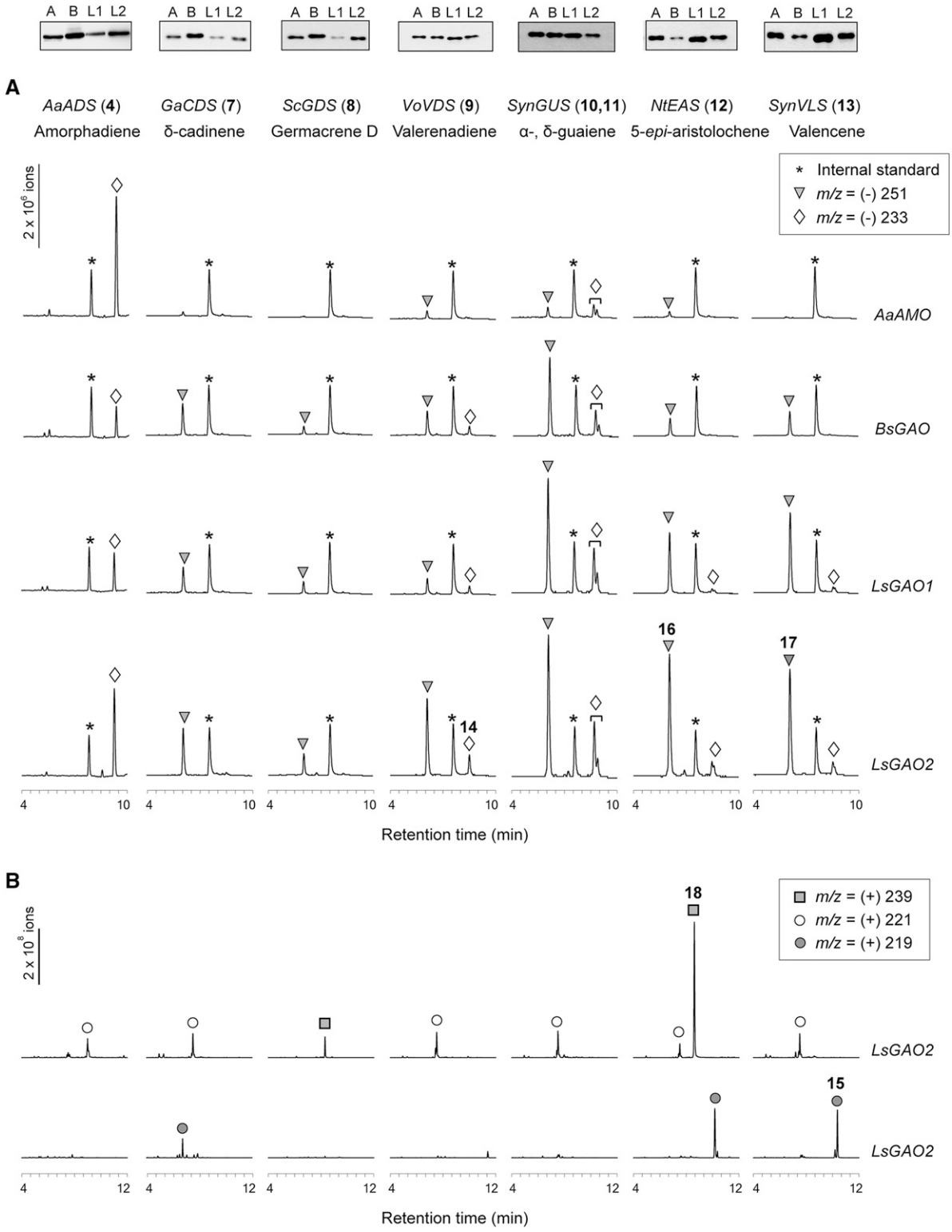

Figure 4. Substrate plasticity of AMO and GAOs against seven sesquiterpene substrates synthesized by six sesquiterpene synthases. Top shows immunoblot analysis for the expression of AMO (A), BsGAO (B), LsGAO1 (L1), and LsGAO2 (L2) in EPY300 coexpressing these P450s, CPR, and each of the seven sesquiterpene synthases. A, (−)-LC-MS metabolite profiling of the transgenic EPY300 yeasts. The coexpressed sesquiterpene synthases and P450s are indicated on top and at the far right, respectively. The extracted ions are *m/z 171* for decanoic acid (asterisk, internal standard), *m/z 233* for sesquiterpene acids (diamond), and *m/z 251* for hydrated sesquiterpene acids (triangle). B, (+)-LC-MS metabolite profiles from transgenic EPY300 yeasts.

confirmed by authentic standards, when available (**14**, **15**), or by nuclear magnetic resonance (NMR) analyses after purifications by silica column and high-performance liquid chromatography (HPLC; **16**, **17**, **18**). The anticipated product from the yeast coexpressing *VoVDS* and *GAO* is valerenic acid (Fig. 5A; **14**). The C12 oxidation of the isobutenyl group in valerenadiene (**9**) to **14** is known to occur naturally in valerian plant (*V. officinalis*), and a valerenic acid standard isolated from valerian plant is available. Using the authentic standard, the oxidized sesquiterpene with *m/z* (−)-233 from the transgenic yeast was identified as valerenic acid (**14**) by (−)-LC-MS and GC-MS (Fig. 5, A–C). In shake-flask cultures, the transgenic yeast could synthesize 3.6 ± 1.6 μg mL^{-1} of compound **14** (n = 3). This result corroborated that GAOs catalyze three sequential oxidations at C12 terminal carbon of the nonnative substrate **9** to yield valerenic acid **14** (Fig. 5, A–C).

Valencene (**13**) is a common metabolite in the fruits of citrus plants, and nootkatone (**15**) is a known flavor of sesquiterpene ketone derived from **13** in grapefruit (Fig. 5D). We suspected that the putative sesquiterpene ketone detected in Figure 4B could be nootkatone **15**. Both LC-MS and GC-MS analyses confirmed the chemical identity of **15** to be nootkatone in comparison to authentic standard (Fig. 5, D–F). In shake-flask cultures, the titer of de novo synthesis of **15** from our transgenic yeast reached 3.6 ± 1.1 μg mL^{-1} (n = 3).

To our best knowledge, there are no reports describing other putative oxidized sesquiterpenes shown in Figure 4. Therefore, purifications of some oxidized sesquiterpenes were carried out to elucidate their structures by NMR analyses. As discussed previously, (−)-LC-MS analyses showed that the sesquiterpene acids (*m/z* 233) and their hydrated forms (*m/z* 251) are produced from the transgenic yeasts. Since the yields of putative sesquiterpene acids were low, we pursued the purifications of hydrated forms of 5-*epi*-aristolochenoic acid (*m/z* 251; **16**) and a hydrated form of valencenoic acid (*m/z* 251; **17**), instead of sesquiterpene acids. Cultures of the EPY300 coexpressing *LsGAO2/CPR* with *SynVLS* or *NtEAS* were scaled up to 5-L, and the culture media were extracted at 48 h after Gal induction. These two compounds (**16**, **17**) were purified by silica gel column chromatography and reversed phase HPLC, followed by NMR analyses (^{1}H-, ^{13}C-NMR, HMBC, HSQC, and rotating frame nuclear Overhauser effect spectroscopy [ROESY]; Supplemental Figs. S2 and S3).

NMR analyses showed that the hydrocarbon skeletons of **16** and **17** are identical and matched to a known natural product, illicic acid (Abu Irmaileh et al., 2015; Fig. 6A). At first sight, these results were enigmatic; however, considering that 5-*epi*-aristolochene and valencene are allylic positional isomers (C9–C10 versus C1–C10 allylic group; Fig. 6, B and C), it is reasonable to postulate that the protonation of the allylic moiety followed by methyl- and hydride-migrations, keeping the same plane, on either upper or lower plane of **12** and **13** can lead to the formation of illicic acid skeletons in both cases as proposed in Figure 6, B and C. We postulate that the hydroxyl group is stereo-specifically added to C4 of carbocation (**16** and **17**) on the opposite site of Me_{14} to avoid steric crowding. This proposed addition results in two stereo-specific orientations of Me_{14} and Me_{15} supported by 2D ROESY assignments for the two conformers (Supplemental Figs. S4 and S5). To confirm the conformation, structural models of both compounds (Merck molecular force field 94 conformational search) were used for the determination of center of gravity (*Cg*) distances. In the 2D ROESY data of **17**, a strong $Me_{14} \leftrightarrow Me_{15}$ cross peak confirms their relatively close proximity (*Cg*Me_{14} to *Cg*Me_{15}, 3.32 Å), whereas this cross peak is weak in compound **16** (*Cg*Me_{14} to *Cg*Me_{15} 4.42 Å). Additional cross peaks of proximal protons as outlined in the calculated structures support the structure of **16** (i.e. H3(eq)↔H14, H8(ax)↔H14, H9(ax)↔H14) and **17** (i.e. H2↔H14, H8(ax)↔H14, H6(ax)↔H14, H14↔H15, H6(ax)↔H15; see Supplemental Figs. S4 and S5 for details). Regardless of the structural rearrangements, a key message relevant to this work is that the terminal C12 of isopentenyl group of **12** and **13** is oxidized to carboxylic acid, validating again the three-step oxidations of the sesquiterpene backbones by GAO.

Another major compound (*m/z* (+)-239, **18**) biosynthesized from the yeast coexpressing *NtEAS* and *GAO2* was purified from a 1-L culture, and its structure was elucidated by NMR analyses (Fig. 6A; Supplemental Fig. S6). In this compound, alcohol moieties are attached to C11 and C12 on 5-*epi*-aristolochene, and this compound is also considered an acid-rearranged product from the sesquiterpene alcohol precursor as proposed in Figure 6D.

Yeast favors acidic culture conditions, and the acidity of the culture medium typically decreases to pH 3 to 4 over the course of cultivations. In such acidic conditions, structural rearrangements (i.e. allylic rearrangement and hydration) can occur. In an attempt to purify intact sesquiterpene acids by preventing acid-induced rearrangements, we cultured the transgenic yeast strains in HEPES-buffered medium, and the sesquiterpene metabolites from these neutralized culture conditions were analyzed. The buffered culture maintained pH 6.5 at the end of cultivation, whereas the acidity of unbuffered medium dropped to pH 3 to 4. However, (−)-LC-MS analysis of the extracts from the buffered medium (~pH 6.5) still showed the substantial amounts of the hydrated sesquiterpene acids displaying *m/z* (−)-251, although their abundance was decreased (Fig. 7). Additionally, new compounds with *m/z* (−)-233 appeared in three yeast cultures (SynGUS, NtEAS, and SynVLS), indicating neutral pH influences the product profile. However, these compounds are in low abundance and could not be purified for further structural studies. On the other hand, the occurrence of **18** from the yeast coexpressing *NtEAS/LsGAO2*

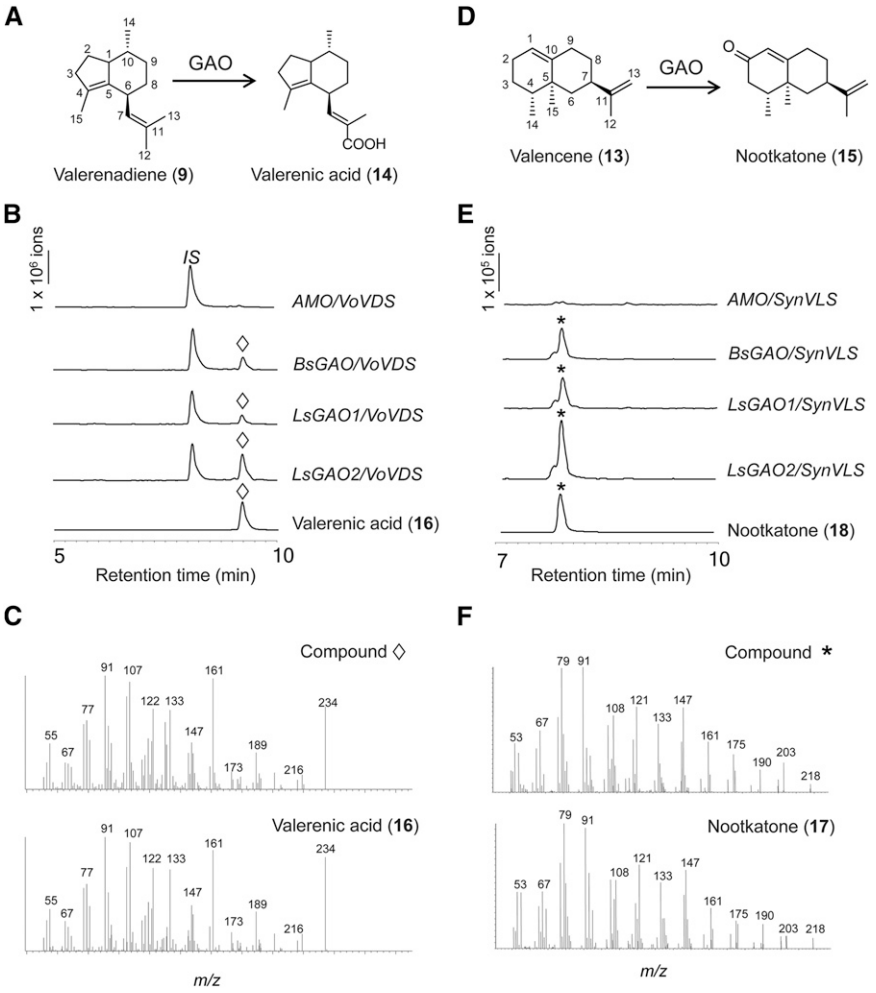

Figure 5. Enzymatic synthesis of nootkatone and valerenic acid by GAOs. A and D, Schematic presentations of the oxidations at C12 and C2 positions of valerenadiene and valencene, respectively. B and E, Extracted ion chromatograms for the oxidation products: *m/z 219* for ketone (asterisk) from (+)-LC-MS and *m/z 233* for sesquiterpene acid (diamond) from (−)-LC-MS. C and F, Comparison of the electron impact (EI) fragmentation patterns of the de novo oxidation products and authentic standards.

completely disappeared under the buffered condition (pH ~6.5).

Structure-Function Analysis of AMO and GAOs

Despite high sequence identity between AMO and several GAOs (>79% amino acid identity), these enzymes display a substantial difference in their substrate specificities. Presumably a few critical residues could determine their substrate specificities. To achieve structural insights into the substrate specificity and promiscuity embedded in AMO and GAOs, homology modeling of AMO, LsGAO1 (highest identity to AMO), and BsGAO (lowest identity to AMO) was thus carried out. The AMO substrate, amorphadiene, was computationally docked into the structural models of these P450s, and the C12 of isopropenyl group was oriented in a close proximity to the heme group in P450. Among the amino acids within 10Å of the substrate, 12 and 17 residues differentiate LsGAO1 and BsGAO, respectively, from AMO. All of these residues position on or are adjacent to the six putative substrate-recognition sites of the P450s (Supplemental Fig. S7; Gotoh, 1992; Denisov et al., 2005), suggesting that they may specify GAO activity.

Using the structural guidance, we examined whether incorporating the AMO-specific residues onto the GAO scaffold can specify and improve the amorphadiene oxidizing activity in GAOs. Since there is only one native form of AMO (the native enzyme from *A. annua*) for semisynthetic artemisinin production at present, generating new AMO isozymes based on the GAO scaffolds can benefit yeast strain development for a higher titer of artemisinin production. As a first step toward this goal, the mutant versions of LsGAO1 and BsGAO were synthesized to include all AMO-specific residues within 10Å of substrate-binding site (Supplemental Fig. S8). The catalytic activities of these two mutant GAOs against both amorphadiene and germacrene A substrates were examined using a coexpression strategy.

(−)-LC-MS analyses of the transgenic yeasts showed that the mutant GAOs can synthesize significantly reduced levels of germacrene A acid (~1000-fold lower), while comparable amounts of artemisinic acid (36% lower in mutant BsGAO and 16% higher in mutant LsGAO1) are synthesized from these mutant enzymes,

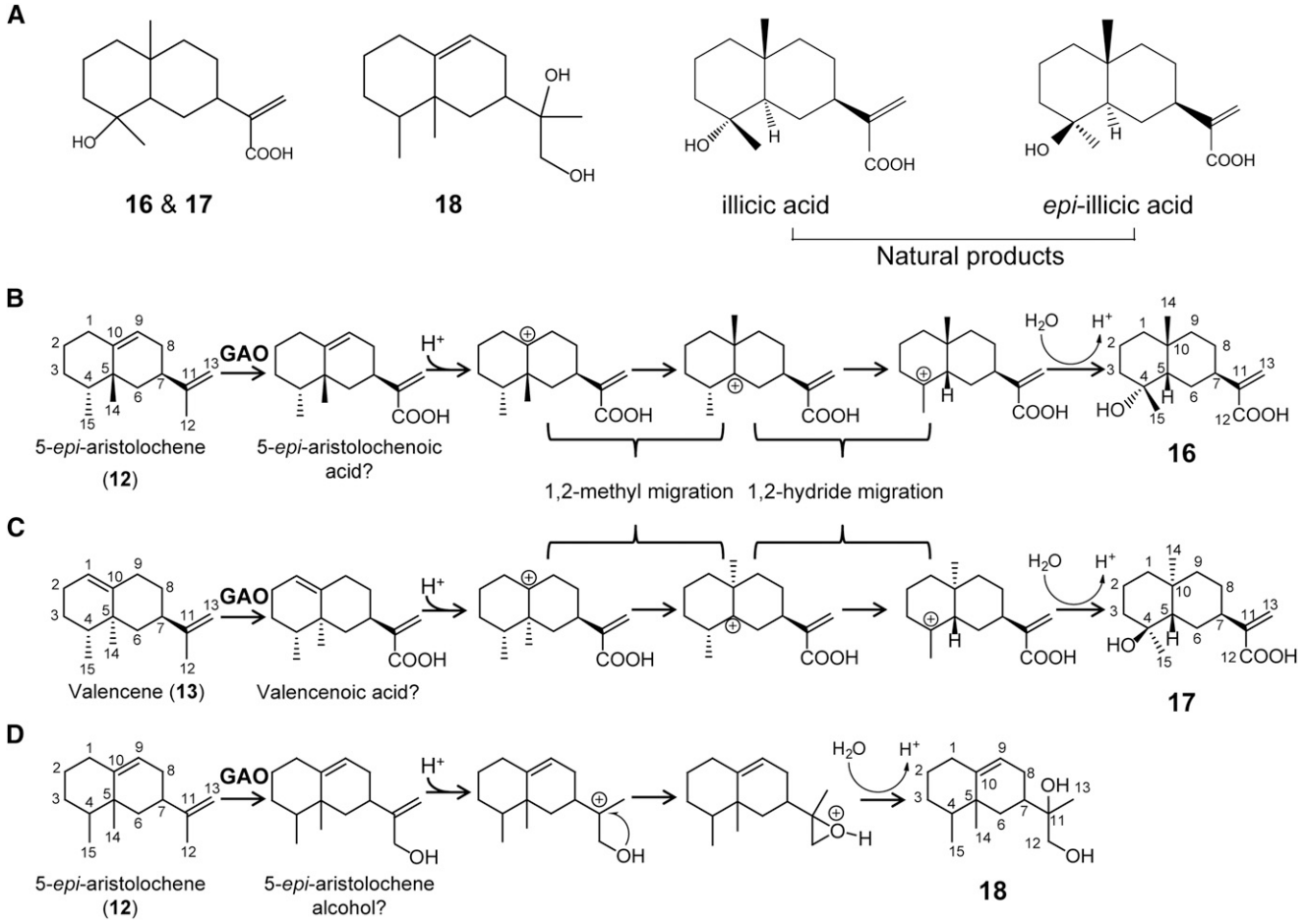

Figure 6. Oxidations and rearrangements of 5-*epi*-aristolochene and valencene in yeast expressing *GAO*s. A, Structures of the compounds synthesized from transgenic yeasts and structures of natural products, illicic acid, and epi-illicic acid. B and C, Oxidations at C12 of 5-*epi*-aristolochene (B) and valencene (C) to the respective sesquiterpene acids. Skeletal rearrangements after protonations by acid are proposed to explain the occurrence of illicic acids. D, A proposed skeletal rearrangement of 5-*epi*-aristolochene alcohol to compound **18**. Question marks indicate hypothetical compounds.

relative to the activities from the native enzymes (Fig. 8A). When enzyme abundance was estimated by immunoblot analysis, mutant versions of BsGAO and LsGAO1 were significantly less abundant than the native ones (Fig. 8B). Mutant BsGAO showed 7% to 10% enzyme abundance, and mutant LsGAO1 showed 13% to 24% enzyme abundance, relative to their respective native enzymes. After normalizing the activities with the enzyme abundance, the mutant BsGAO and LsGAO1 exhibited approximately 9-fold higher activities for amorphadiene while retaining ~0.1% activity for germacrene A, relative to the activities from native BsGAO and LsGAO1 (Table 1). Such lower protein abundance in mutants is putatively caused by decreased protein stability, as the same promoter was used to express both native and mutant genes. These results showed that overall substrate specificity was shifted to amorphadiene in mutant GAOs with improved productivity per enzyme but accompanied by decreased enzyme abundance, which compromised the overall productivity of artemisinic acid in a given culture volume.

DISCUSSION

Evolutionary Insights of GAO Promiscuity

From the perspective of adaptive speciation, the Asteraceae family is the most successful plant family, with >24,000 species on earth, and thus can be an excellent lineage to examine metabolic divergence alongside the extensive speciation. Most Asteraceae plants have retained characteristic STL metabolites, which have structurally diversified over 50 million years while keeping their core carbon 15 α-methylene γ-lactone skeleton. One central oxygenation enzyme in the STL biosynthesis is the multifunctional cytochrome P450 that catalyzes the conversion of sesquiterpene hydrocarbons to sesquiterpene acids. We initiated the studies

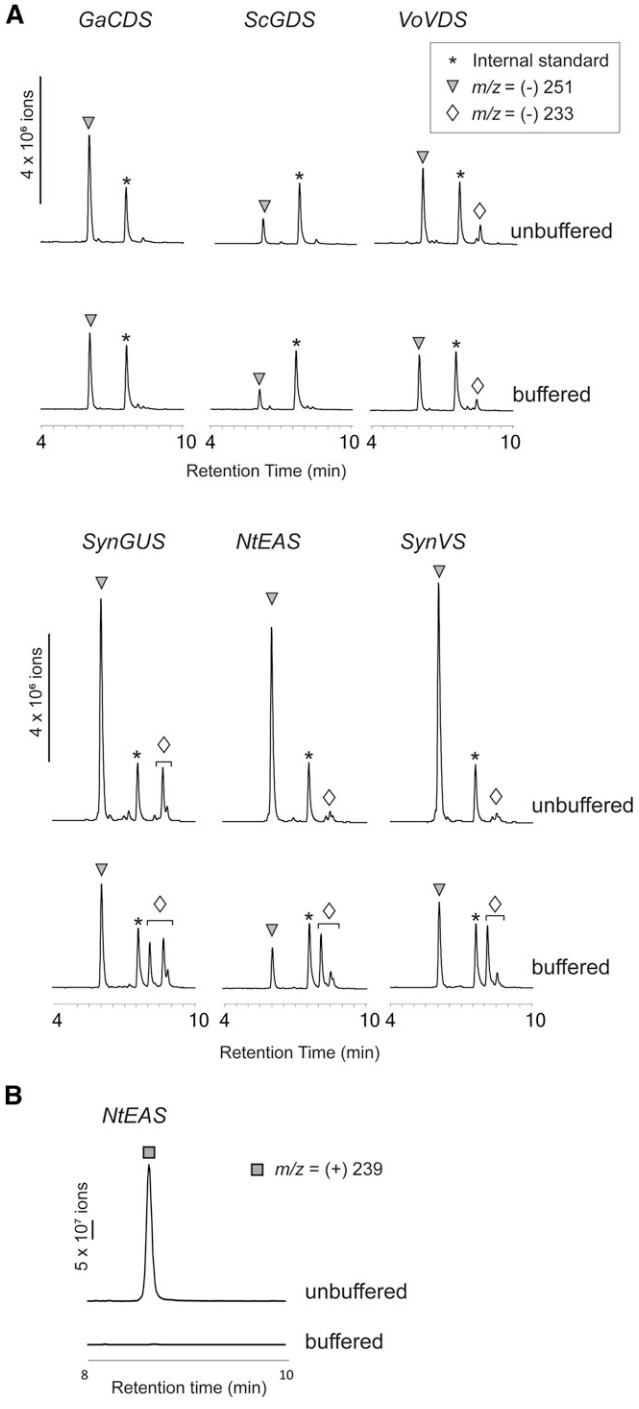

Figure 7. LC-MS analyses of the sesquiterpenoids from transgenic yeasts cultivated in buffered and unbuffered media. A, Yeast strains expressing *LsGAO2/CPR* together with one of the six sesquiterpene synthases were cultivated in unbuffered medium and HEPES-buffered medium, and sesquiterpenoid metabolites were detected by (−)-LC-MS. Triangles indicate compounds with *m/z* (−)-251, $[M+H_2O-H]^-$ ion, and diamonds indicate compounds with *m/z* (−)-233, $[M-H]^-$ ion. B, Disappearance of the compound with *m/z* (+)-239 (rectangle) under buffered medium is shown.

of the core oxygenase enzymes, GAO and AMO, since such comparative studies of catalytic plasticity between specialized AMO and widespread GAO might shed lights on adaptive enzyme evolution that influences the STL diversifications in the Asteraceae. In this study, we demonstrated that GAOs display remarkable catalytic plasticity toward a wide range of sesquiterpene substrates in contrast to the restricted substrate utilization of *A. annua* AMO. CYP71AVs from chicory and *Cynara cardunculus* were reported to oxidize valencene and germacrene D as shown here (Cankar et al., 2011; Eljounaidi et al., 2014), but five additional substrates were further tested in this work to comprehensively demonstrate the inherent catalytic promiscuity of lettuce and *B. spinosa* GAO. The diversity of the accepted structural types of substrates and the occurrence of more than one product (**15** and **17**) from a single substrate (**13**) convincingly demonstrated that GAOs readily accept a wide range of sesquiterpene substrates that GAOs do not meet in nature.

The question of how new enzyme activities have evolved is of fundamental scientific importance, but it remains not fully understood. This is due to the lack of knowledge on the historical events including evolutionary "accidents" and selection pressures (Arnold et al., 2001). Jensen proposed the general hypothesis of enzyme evolution from ancestral promiscuity (Jensen, 1976). Although this hypothesis is well rationalized, it is supported only by a few examples in nature (Khersonsky et al., 2006; Ober, 2010). The differential substrate plasticity between AMO and GAOs presented here provided strong experimental data to support Jensen's hypothesis in the Asteraceae family. The environmental conditions have constantly changed since the Asteraceae started its diversifications 50 million years ago and so have the metabolic pathways of Asteraceae plants. Therefore, the broad substrate ambiguity of GAOs likely exerted selection advantages for the plants, as it readily provides room for selection and optimization of specific catalytic functions. This process resulted in the subfunctionalization of GAO to become a more specialized P450, such as the recently evolved AMO in *A. annua*, and might have resulted in other substrate-specific sesquiterpene oxidases that are yet to be elucidated.

The retention of GAO's substrate plasticity throughout the Asteraceae family could also be attributable to the structural dynamics of its native substrate. Germacrene A (**1**) is known to have four distinct conformations (Faraldos et al., 2007) and might have emerged very early in the evolution of the Asteraceae (Nguyen et al., 2016). A flexible active site may be a necessary characteristic of ancestral and extant GAOs to prevent conflict against different conformers. However, such flexibility of active site could be lost during the evolution of AMO from the ancestral GAO, and the resulting AMO has become to only accommodate the rigid structure of amorphadiene. In concert with the occurrence of various sesquiterpene synthases, the substrate plasticity of ancestral GAO

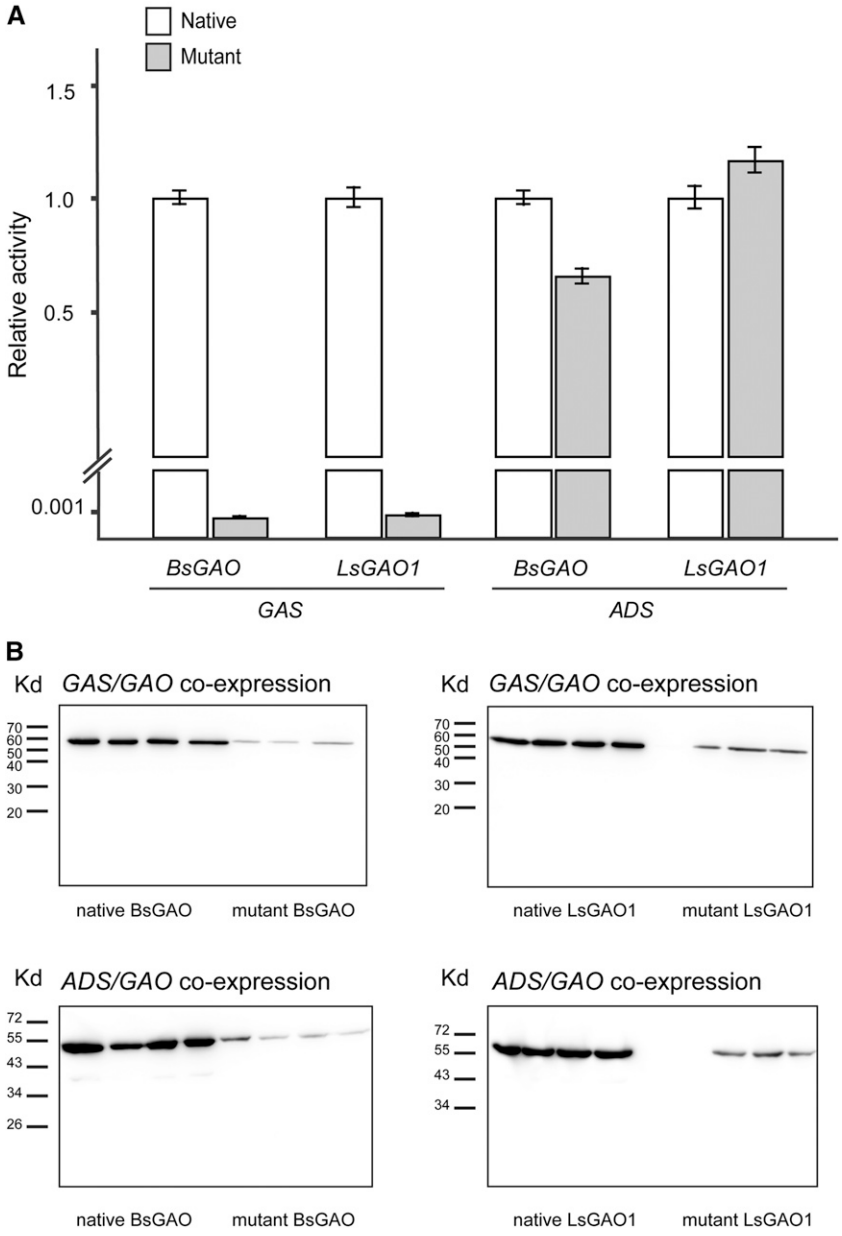

Figure 8. Assessing catalytic activities of native and mutant GAOs. A, Relative catalytic activities to convert germacrene A and amorphadiene to respective sesquiterpene acids by native and mutant forms of BsGAO and LsGAO1. Catalytic activities of the native enzymes were set as 1, which are defined as *GAS/BsGAO* = 27 μg mL^{-1}, *GAS/LsGAO1* = 22 μg mL^{-1}, *ADS/BsGAO* = 20 μg mL^{-1}, *ADS/LsGAO1* = 27 μg mL^{-1}. Data are means ± SD (n = 3–4; biological replicates). B, Immunoblot analyses of native and mutant forms of BsGAO and LsGAO1. FLAG-epitope was tagged to the C terminus of BsGAO and LsGAO1, and monoclonal antibodies were used to detect the recombinant enzymes.

and the emergence of its more specialized homologs may be the key driver for the chemical diversity of STLs in the Asteraceae.

It is highly likely that the specialized AMO activity in *A. annua* has emerged from one of many promiscuous GAO activities as observed in this study, although we do not know the nature of selection pressure that has refined the crude GAO activity to a specialized AMO activity in *A. annua*. There are a number of unique STL backbones in Asteraceae, such as xanthanolide and ambrosanolide (Picman, 1986; Padilla-Gonzalez et al., 2016), and therefore independent adaptive evolutions of GAO may have occurred in different branches of STLs to contribute to the chemical diversity of STLs. Our biochemical knowledge for other STLs in other Asteraceae species is still scarce, but further metabolic elucidation of distinct STL pathways will help us examine the role of promiscuous GAO activities in creating new STL natural products in the Asteraceae.

Potential to Engineer GAO for Improved Activities

In principle, the promiscuous activity of enzymes for nonnative substrates can be selected and improved to a desirable activity by protein engineering. Here, we demonstrated that the latent catalytic potential can be developed to a new AMO enzyme from two GAO scaffolds by transferring a set of residues surrounding the active site of AMO to that of GAOs. Intriguingly, their productivity per enzyme has apparently increased, but the altered and improved new activity was

Table 1. *Catalytic activities of mutant BsGAO and mutant LsGAO1, relative to the native enzymes*

Native BsGAO and LsGAO1 catalytic activities for amorphadiene or germacrene A were set as 100%. The production levels of artemisinic acid or germacrene A acid were normalized by relative cytochrome P450 abundance estimated by immunoblot analyses. Data are means ± SD (n = 3–4).

Substrate (cDNA)	Relative Activity (%)	
	Mutant BsGAO	Mutant LsGAO1
Amorphadiene (*ADS*)	913 ± 159	904 ± 218
Germacrene A (*GAS*)	0.07 ± 0.01	0.08 ± 0.02

severely compromised by low protein abundance likely due to the loss of stability in both GAO mutants. Protein stability is a net balance between natively folded proteins and unfolded proteins dictated by their free energy differences (Shoichet et al., 1995; Bloom et al., 2006). Unfolded proteins undergo irreversible modifications, such as aggregation, proteolysis, disulphide rupture, and chemical degradations, resulting in the removal of active proteins in cells (Fágáin, 1995). To harness the potential of the desired activity, it is critically important to identify mutant GAOs that maintain stability with an improved new activity. However, such a task can be difficult, although not impossible, because the acquisition of new function is often accompanied by reduced enzyme stability, leading to the notion of a tradeoff between new function and protein stability (Tokuriki et al., 2008; Soskine and Tawfik, 2010). One such experimental data set was obtained from the studies of clinical mutants of β-lactamases, which acquired a new catalytic function to degrade "β-lactamase resistant antibiotic," cephalosporin (Wang et al., 2002). Crystal structures of the mutants showed that the geometric change in the active site of β-lactamase mutants allowed them to accept cephalosporin as a new substrate but is accompanied by a severe loss of thermodynamic stability of the mutant enzymes. From additional clinical β-lactamase mutants, a secondary mutation far from the active site was found to effectively restore the stability of the original mutants. Through the computational simulations, the analogous pattern (i.e. the reduced stability with occurrences of new functions) was also observed among hundreds of mutants generated by directed evolutions (Tokuriki et al., 2008). The loss of stability observed in our work could be a common consequence in the course of enzyme evolution. To overcome the stability problem, we project it is necessary to identify and implement compensatory silent mutations, as well as mutations in the active site, by sophisticated computational design of enzymes (Borgo and Havranek, 2012).

Interpretation of Acid-Catalyzed Addition of Water in Sesquiterpene Acid

Formations of sesquiterpene acids by promiscuous GAO activities were evident from the presence of valerenic acid (**14**), and the two compounds **16** and **17** as confirmed by NMR and HR MS analyses. However, formation of hydrated products from oxidation of the sesquiterpenes by yeast-expressed P450 poses two questions. The first question is why the hydration occurs only after oxidation of the sesquiterpenes to sesquiterpene carboxylic acids, whereas the unoxidized hydrocarbon sesquiterpenes do not undergo such hydrations. Oxidation-hydration of valencene and 5-*epi*-aristolochene in buffered medium indicates that a lower pH environment favors the formation of hydrated products, suggesting hydration by acid catalysis. Isolation of illicic acids from incubation of valencene and 5-*epi*-aristolochene supports acid-catalyzed hydration mechanism, as illustrated in Figure 6, B and C. It is well known that the hydration of an olefin is catalyzed by strong acids. However, such a strong acidic condition is not achievable in the yeast environment. Efficient hydration of the sesquiterpene carboxylic acid only after introduction of carboxylic group thus suggests intramolecular acid catalysis by -COOH group. Although carboxylic acid usually is not strong enough to catalyze hydrations, catalysis by intramolecular reaction would enhance an effective concentration of acid near the reaction site. If such is the case, conformational flexibility of the sesquiterpene carboxylic acid must allow the approach of -COOH group to double bond to be hydrated. We performed molecular mechanics calculation to estimate the energy difference between the most relaxed conformation of sesquiterpene acids and their strained form with -COOH group within 2Å distance from the sp^2 carbon where proton is to be transferred. The energy difference between the two conformations is 5 to 15 kcal mol^{-1} except valerenic acid to allow such conformations for proton transfer (Supplemental Table S2). If the distance was relaxed to 3Å, the strain energies of all molecules were less than 9 kcal mol^{-1}. Therefore, conformations for intramolecular general acid catalysis are possible with energy input attainable from physiological condition. As a reference, activation energy for cyclohexane ring flip is known to be 10 kcal mol^{-1}.

The second question is why artemisinic acid is recalcitrant toward hydration. To answer this question, we qualitatively estimated stability of carbocation formed by protonation of the sesquiterpene acid. We reasoned that the initial protonation site must be the one that would give the most stable carbocation (Supplemental Fig. S9). In the case of artemisinic acid, the initial tertiary carbocation cannot be further stabilized because a less-stable secondary carbocation intervenes in proceeding toward another tertiary carbocation. However, in the other sesquiterpene acids, direct formation of tertiary allylic carbocation (germacrene D acid), isomerization into allylic carbocation through hydride and methyl shifts (valerenic acid) would result in stabilization of carbocations from the relatively unstable initial tertiary carbocations. In the case of acids derived from valencene, 5-*epi*-aristolochene, and α- and δ-guaienes, where only straight tertiary carbocations are involved, these carbocations

form extra hyperconjugation with adjacent C-H or C-C bonds, which are stereo-electrically aligned with empty *p*-orbital of cationic carbon (Tantillo, 2010). In short, extra stabilization of initial carbocation, either by allylic interaction or geometrically imposed hyperconjugation, would be possible for the acids derived from the aforementioned sesquiterpene acids except for artemisinic acid. This extra stabilization would lead to extended lifetime of carbocations for quenching by water molecule in sesquiterpene acids except for artemisinic acid.

Insight into *GAO/AMO* Evolution in *A. annua*

Although the precise evolutionary path leading to artemisinin biosynthesis in *A. annua* may not be fully reconstructed, *A. annua* trichome transcriptome data and recently published genome sequences provide clues as to how *GAS/GAO* might have diverged to *ADS/AMO* for artemisinin biosynthesis (Bertea et al., 2006; Shen et al., 2018). Functional *GAS* is encoded in genome (annotated CDS: PWA48097) and expressed in trichomes (GenBank: DQ447636), whereas no sequence signatures for *GAO* could be found in *A. annua* genome and trichome transcriptome. Interestingly, two copies of *AMO* are encoded in the *A. annua* genome (annotated CDS: PWA40082 and PWA47004), of which only one *AMO* copy is expressed in trichome, with the other copy not expressed. From these data, we speculate that after duplication of *GAO*, one copy of *GAO* has evolved to *AMO* followed by duplication of *AMO* itself, while the other *GAO* copy disappeared from the *A. annua* genome. Even though a catalytically active GAS is still present in *A. annua* trichome (Bertea et al., 2006), absence of GAO and lack of cross activity of AMO for germacrene A render *A. annua* unable to biosynthesize germacrene A acid and other STLs derived from germacrene A acid.

CONCLUSION

Regio- and stereo-selective oxygenations on terpene olefins have considerable biotechnological applications, as subtle oxidative modifications can dramatically alter terpene values as shown in nootkatone in grapefruit (*Citrus paradisi*) and santalol in sandalwood (*Santalum album*; Cankar et al., 2014; Celedon et al., 2016). We showed the catalytic potential of GAOs to accept seven sesquiterpene substrates to synthesize distinct STLs, as opposed to the narrow substrate specificity of the recently specialized AMO. The degree of GAO promiscuity and enzyme-product profiles needs to be further examined for other sesquiterpene substrates under different conditions, such as pH variance, to determine the use of GAO for expanding sesquiterpenoid diversity beyond the seven substrates tested here. Nonetheless, the combinatorial expression approach in yeast presented here and a similar method in *Escherichia coli* reported for promiscuous P450 (*ent*-kaurene oxidase) activities on diterpenes (Mafu et al., 2016) showcase effective uses of microbial systems to unveil hidden P450 activities. In principle, P450 with weak initial activity can be molecularly bred to be a specific and potent catalyst for increasing the productivity of specialty chemicals.

MATERIALS AND METHODS

Cloning and Expression of *cytochrome P450s* and *STSs* in Yeast

All primer sequences used in this study are listed in Supplemental Table S3. The full-length cDNA of *LsGAO2* (GenBank: KF981867) was PCR-amplified from lettuce cDNA by primers 1 and 2 and cloned into MCS1 of the pESC-Leu-2d vector. *AMO*, *LsGAO1*, and *BsGAO* were individually cloned into MCS1, and *Artemisia annua CPR* was cloned into MCS2 of these pESC-Leu-2d vectors as previously described (Ro et al., 2008; Nguyen et al., 2010, 2012). All *STS* cDNAs used in this study were PCR-amplified and cloned into MCS2 of empty pESC-Leu-2d vector. Canadian goldenrod (*Solidago canadensis*) *germacrene D synthase* (*ScGDS*; GenBank: AJ583447) was isolated with primers 3 and 4 from Canadian goldenrod cDNA (Prosser et al., 2004). *Guaiene synthase* (*SynGUS*) was synthesized by GenScript for yeast codon optimization based on agarwood (*Aquilaria crassna*) guaiene synthase sequence (GenBank: GU083698; Kumeta and Ito, 2010), including restriction enzyme sequences, and was subsequently subcloned directly to yeast expression vector. These three genes were cloned into the *Apa*I and *Xho*I sites of MSC2 on the pESC-Leu-2d vector. Lettuce (*Lactuca sativa*) *germacrene A synthase* (*LsGAS*), *A. annua amorphadiene synthase* (*AaADS*), valerian (*Valeriana officinalis*) *valerenadiene synthase* (*VoVDS*; GenBank: JQ437840), synthetic *valencene synthase* (*SynVLS*), and thioredoxin-fused tobacco (*Nicotiana tabacum*) 5-epi-*aristolochene synthase* (*NtEAS*) were available in pESC-Leu-2d plasmid from our previous works (Ro et al., 2008; Nguyen et al., 2010, 2012; Pyle et al., 2012). Cotton (*Gossypium arboreum*) δ-*cadinene synthase* (*GaCDS*; GenBank: U23206; Chen et al., 1995) was cloned to MCS2 of the pESC-Leu-2d vector by the homologous recombination-based method using primers 5 and 6 with the In-Fusion HD cloning kit (Clontech Laboratories). Six individual expression cassettes harboring *STSs* were amplified using primers 7 and 8, digested with *Dra*III and *Nae*I, and cloned into the corresponding positions of the aforementioned four pESC-Leu-2d::*P450/CPR* vectors, generating 24 triple-expression constructs. The pESC-Leu-2d-based triple-expression constructs were transformed to EPY300 yeast (Nguyen et al., 2012). All yeast cultivations were performed at 30°C and 200 rpm. Cultures started with overnight inoculations (15–20 h) in appropriate synthetic complete dropout media with 2% Glc. Inoculations were diluted 50- to 100-fold to the same media with 0.2% Glc, 1.8% Gal, and 1 mM Met and cultured for 48 h. When only sesquiterpenes were to be analyzed, a layer of dodecane equivalent to 10% of total culture volume was overlaid to sequester the volatile hydrocarbons. Twenty- to one hundred-milliliter cultures were carried out for metabolite analysis and microsomal preparation, while 1-L or 5-L cultures were used for metabolite purification.

Immunoblot analysis was performed to confirm the presence of the P450 proteins in yeast. Yeast microsome was prepared according to the published protocol (Pompon et al., 1996; Kwon et al., 2016) and some modifications. Yeast cells were collected from cultures by centrifugation and broken open in with glass beads (0.05-mm diameter) at 4°C in Microbead beater (Biospec Products) for 2 min (repeated three times) in TES-B buffer (50 mM Tris-HCl, 1 mM EDTA, 0.6 M sorbitol, pH 7.4). Microsome was isolated by sequential centrifugations at 10,000*g* and 100,000*g*. Pelleted microsome was dissolved in TEG buffer (50 mM Tris-HCl, 1 mM EDTA, 20% [w/v] glycerol, pH 7.4) and stored at −80°C. Microsomal proteins were separated on a 10% SDS-PAGE and transferred onto a polyvinylidene fluoride membrane in transfer buffer (25 mM Tris, 192 mM Gly, 20% methanol). The membrane was blocked in TBS plus Tween 20 (TBST) buffer (25 mM Tris-HCl, 150 mM NaCl, 0.05% Tween 20, pH 7.5) with 5% (w/v) skimmed milk for at least 1 h. The membrane was then incubated with anti-FLAG M2 mouse antibodies (Sigma-Aldrich) in a 1:5000 dilution in 3% (w/v) skimmed milk to detect P450s fused to FLAG epitope on pESC vectors, washed three times with TBST, and incubated with goat anti-mouse antibodies (GE Healthcare) in a 1:10,000 dilution in 3% (w/v) skimmed milk. The excessive anti-mouse antibodies were then washed three times with TBST, and the bound

antibodies were visualized with Luminate Forte western horseradish peroxidase substrate (Millipore). The amount of proteins was quantified by measuring chemiluminescent signal strength using Amersham Imager 600 analyzer.

Metabolite Profile Analysis by GC-MS and LC-MS

Detailed methods (yeast cultivation, medium preparation, and metabolite extraction) can be found in our previous methodology paper Nguyen et al. (2012). To analyze the profiles of sesquiterpenes produced by transgenic yeast, the aforementioned dodecane overlay was separated from the culture by centrifugation at 3,000*g* for 5 min. In other cases, the cultures were extracted with ethyl acetate for nonbuffered cultures or adjusted to pH 6.0 before extraction with ethyl acetate for HEPES-buffered cultures.

For GC-MS analysis, metabolites from transgenic yeast cultures were analyzed on a 6890N gas chromatography coupled with a 5975B mass spectrometer (Agilent). Samples (1–5 μL) were injected onto a DB5-MS column (30 m length $\times$ 250 μm internal diameter $\times$ 0.25-μm film thickness) for separation with a gradient temperature program. Standard inlet temperature was set at 250°C. Carrier gas was helium with a constant flow of 1 mL min^{-1}. For LC-MS analysis, ethyl acetate extracts of yeast metabolites were evaporated completely under gentle nitrogen flows or using a rotary evaporator and replaced with 20% methanol. Five to twenty microliters of these mixtures were injected to an Agilent 1200 Rapid Resolution LC system coupled with an Agilent 6410 triple quadrupole MS or an Agilent 6540 qTOF MS (for high resolution). After separation on a reverse phase Eclipse plus C18 Zorbax column (2.1-mm diameter $\times$ 50-mm length, 1.8-μm particle size) or Poroshell SB-C18 (100 $\times$ 3 mm), the metabolites went through electrospray ionization and were subjected to mass spectroscopy. For most of the LC-MS analyses in this research, metabolites were analyzed with a solvent system of water plus 0.1% acetic acid (A) and acetonitrile (B), starting at the ratio of 80:20 (A:B) to 0:100 (A:B) over 14 min. Both negative and positive modes were used for total ion scans as well as selected ion modes.

To purify compounds of interest, the concentrated ethyl acetate extracts were separated by liquid chromatography on silica gel 60 F_{254} (Merck) with a solvent system of hexane and ethyl acetate. This preliminary purification step for compound **16** and **17** was followed by further separation on HPLC with a gradient solvent system of water and acetonitrile (adjusted to pH 4.0 with 0.05% acetic acid). HPLC separation was achieved on a Waters SunFire C18 column (4.6 mm $\times$ 150 mm, 5 μm) using a Waters 279 separation module. The next purification for compound **18** was performed using flash column made with silica gel 60 F_{254} (Merck) by elution of dichloromethane and ethyl acetate (6:5). Structures of the purified metabolites were elucidated with HR mass spectrometry in a 6500 quadrupole time-of-flight mass spectrometry (Agilent), followed by NMR spectroscopy. NMR spectra were recorded at an ambient temperature on Varian Innova 500 MHz, Varian 500 MHz (equipped with cryoprobe), Varian 600 MHz, Varian 700 MHz, Bruker 400 MHz, and Bruker 600 MHz spectrometers. Chemical shifts of ^{1}H and ^{13}C spectra were referenced to solvent signals at δH/C 7.24/77.0 ($CDCl_3$), 3.30/49.0 (CD_3OD), or 5.33/54.2 (CD_2Cl_2). Signal assignments for **16**, **17**, and **18** were achieved with ^{1}H, ^{13}C, HSQC, and HMBC NMR data. Additional experiments for structural elucidation included 2D ROESY (for **16** and **17**).

Structure-Function Analysis of AMO and GAOs

Structural models of AMO and GAOs were constructed using the modeling software MODELER (Eswar et al., 2008) based on crystal structures of CYP2C9 (PDB: 1R9O), CYP19A1 (PDB: 3EQM), and CYP1B1 (PDB: 3PM0) as suggested by the homology detection and structure prediction server HHpred (Hildebrand et al., 2009). The model was energy-minimized using the Crystallography and NMR System (Brünger et al., 1998). Models of substrates (amorphadiene and germacrene A) were generated and energy-minimized using ChemBioDraw Ultra 12.0 (CambridgeSoft). In silico docking of substrates into the enzymes' models was performed using GOLD Docking (Cambridge Crystallographic Data Centre). Structural data were analyzed using the molecular visualization system PyMOL (Schrödinger). Docking solutions that showed close proximity of substrate's C12 to the protein's heme and no clash between the substrate and side chains of the protein's amino acids were selected for structural analysis. All residues within 10-Å radius of the substrates were selected and compared between the P450s (Supplemental Fig. S4). Amino acids positioned within the 10-Å proximity of the docked substrates and different between AMO and GAOs were selected for mutagenesis. The codons encoding for these selected amino acids in LsGAO1 and BsGAO were changed to the corresponding codons in AMO by gene synthesis. The two new mutant genes were synthesized by GenScript (Supplemental Fig. S4) and cloned into pESC-Leu-2d-based triple-expression vector with *cytochrome P450 reductase* (*CPR*), and *amorphadiene synthase* (*ADS*) or *germacrene A synthase* (*GAS*) using the aforementioned method. The vector was transformed into the EPY300 yeast strain and grown in synthetic complete media. After 48 h of culture, 10 mL was extracted with 2 $\times$ 1.5 mL of ethyl acetate, and analyzed on LC-MS for sesquiterpene acid production. The rest of the culture (20 mL) was used for microsomal preparation.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under the following accession numbers: GU198171 (*LsGAO1*), KF981867 (*LsGAO2*), MN457916 (synthetic valencene synthase, *SynVLS*), MN457917 (synthetic guaiene synthase, *SynGUS*).

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Phylogenetic analysis of characterized AMO and GAOs.

Supplemental Figure S2. NMR spectral data of **16**.

Supplemental Figure S3. NMR spectral data of **17**.

Supplemental Figure S4. Elucidation of stereochemical configuration of compound **16**.

Supplemental Figure S5. Elucidation of stereochemical configuration of compound **17**.

Supplemental Figure S6. NMR spectral data of **18**.

Supplemental Figure S7. Models of amorphadiene docked in AMO, LsGAO1, and BsGAO structures.

Supplemental Figure S8. Putative plasticity residues of AMO, LsGAO1, and BsGAO.

Supplemental Figure S9. Proposed carbocations formed by protonation of the sesquiterpene acids.

Supplemental Table S1. Mass accuracy of the observed ions measured by LC-HR (high resolution) quadrupole time-of-flight MS analyses.

Supplemental Table S2. Total energy calculations of relaxed sesquiterpene acids and constrained sesquiterpene acids.

Supplemental Table S3. Sequences of primers used for cloning and generating the triple expression vectors.

ACKNOWLEDGMENTS

We thank Drs. David Dietrich and John Vederas (University of Alberta, Canada) for their assistance in NMR data collection and analyses, Dr. David Nelson (University of Tennessee) for assignment of CYP numbers, and Dr. Christopher Keeling (Simon Fraser University, Canada) for valuable discussion on analytical data. We also thank Dr. Ian Lewis and the Calgary Metabolomics Research Facility (CMRF) at University of Calgary for mass spectrometry data.

Received May 28, 2019; accepted September 5, 2019; published September 18, 2019.

LITERATURE CITED

Abu Irmaileh BE, Al-Aboudi AM, Abu Zarga MH, Awwadi F, Haddad SF (2015) Selective phytotoxic activity of 2,3,11β,13-tetrahydroaromaticin and ilicic acid isolated from *Inula graveolens*. Nat Prod Res **29:** 893–898

Arnold FH, Wintrode PL, Miyazaki K, Gershenson A (2001) How enzymes adapt: Lessons from directed evolution. Trends Biochem Sci **26:** 100–106

Barreda VD, Palazzesi L, Katinas L, Crisci JV, Tellería MC, Bremer K, Passalia MG, Bechis F, Corsolini R (2012) An extinct Eocene taxon of the daisy family (Asteraceae): Evolutionary, ecological and biogeographical implications. Ann Bot **109:** 127–134

Barreda VD, Palazzesi L, Tellería MC, Katinas L, Crisci JV, Bremer K, Passalia MG, Corsolini R, Rodríguez Brizuela R, et al (2010) Eocene Patagonia fossils of the daisy family. Science **329:** 1621

Bennett MH, Mansfield JW, Lewis MJ, Beale MH (2002) Cloning and expression of sesquiterpene synthase genes from lettuce (*Lactuca sativa* L.). Phytochemistry **60:** 255–261

Bertea CM, Voster A, Verstappen FWA, Maffei M, Beekwilder J, Bouwmeester HJ (2006) Isoprenoid biosynthesis in *Artemisia annua*: Cloning and heterologous expression of a germacrene A synthase from a glandular trichome cDNA library. Arch Biochem Biophys **448:** 3–12

Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. Proc Natl Acad Sci USA **103:** 5869–5874

Borgo B, Havranek JJ (2012) Automated selection of stabilizing mutations in designed and natural proteins. Proc Natl Acad Sci USA **109:** 1494–1499

Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr **54:** 905–921

Cankar K, van Houwelingen A, Bosch D, Sonke T, Bouwmeester H, Beekwilder J (2011) A chicory cytochrome P450 mono-oxygenase CYP71AV8 for the oxidation of (+)-valencene. FEBS Lett **585:** 178–182

Cankar K, van Houwelingen A, Goedbloed M, Renirie R, de Jong RM, Bouwmeester H, Bosch D, Sonke T, Beekwilder J (2014) Valencene oxidase CYP706M1 from Alaska cedar (*Callitropsis nootkatensis*). FEBS Lett **588:** 1001–1007

Celedon JM, Chiang A, Yuen MM, Diaz-Chavez ML, Madilao LL, Finnegan PM, Barbour EL, Bohlmann J (2016) Heartwood-specific transcriptome and metabolite signatures of tropical sandalwood (*Santalum album*) reveal the final step of (*Z*)-santalol fragrance biosynthesis. Plant J **86:** 289–299

Chen XY, Chen Y, Heinstein P, Davisson VJ (1995) Cloning, expression, and characterization of (+)-δ-cadinene synthase: A catalyst for cotton phytoalexin biosynthesis. Arch Biochem Biophys **324:** 255–266

Denisov IG, Makris TM, Sligar SG, Schlichting I (2005) Structure and chemistry of cytochrome P450. Chem Rev **105:** 2253–2277

Eswar N, Eramian D, Webb B, Shen MY, Sali A (2008) Protein structure modeling with MODELLER. Methods Mol Biol **426:** 145–159

Eljounaidi K, Cankar K, Comino C, Moglia A, Hehn A, Bourgaud F, Bouwmeester H, Menin B, Lanteri S, Beekwilder J (2014) Cytochrome P450s from *Cynara cardunculus L.* CYP71AV9 and CYP71BL5, catalyze distinct hydroxylations in the sesquiterpene lactone biosynthetic pathway. Plant Sci **223:** 59–68

Fágáin C (1995) Understanding and increasing protein stability. Biochim Biophys Acta **1252:** 1–14

Faraldos JA, Wu S, Chappell J, Coates RM (2007) Conformational analysis of (+)-germacrene A by variable temperature NMR and NOE spectroscopy. Tetrahedron **63:** 7733–7742

Gotoh O (1992) Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J Biol Chem **267:** 83–90

Hildebrand A, Remmert M, Biegert A, Söding J (2009) Fast and accurate automatic structure prediction with HHpred. Proteins **77**(Suppl 9): 128–132

Ikezawa N, Göpfert JC, Nguyen DT, Kim SU, O'Maille PE, Spring O, Ro DK (2011) Lettuce costunolide synthase (CYP71BL2) and its homolog (CYP71BL1) from sunflower catalyze distinct regio- and stereoselective hydroxylations in sesquiterpene lactone metabolism. J Biol Chem **286:** 21601–21611

Jansen RK, Palmer JD (1987) A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). Proc Natl Acad Sci USA **84:** 5818–5822

Jensen RA (1976) Enzyme recruitment in evolution of new function. Annu Rev Microbiol **30:** 409–425

Khersonsky O, Roodveldt C, Tawfik DS (2006) Enzyme promiscuity: Evolutionary and mechanistic aspects. Curr Opin Chem Biol **10:** 498–508

Kumeta Y, Ito M (2010) Characterization of δ-guaiene synthases from cultured cells of Aquilaria, responsible for the formation of the sesquiterpenes in agarwood. Plant Physiol **154:** 1998–2007

Kwon M, Kwon EJG, Ro DK (2016) cis-prenyltransferase and polymer analysis from a natural rubber perspective. Methods Enzymol **576:** 121–145

Liu Q, Majdi M, Cankar K, Goedbloed M, Charnikhova T, Verstappen FW, de Vos RC, Beekwilder J, van der Krol S, Bouwmeester HJ (2011) Reconstitution of the costunolide biosynthetic pathway in yeast and *Nicotiana benthamiana*. PLoS One **6:** e23255

Liu Q, Manzano D, Tanić N, Pesic M, Bankovic J, Pateraki I, Ricard L, Ferrer A, de Vos R, van de Krol S, et al (2014) Elucidation and in planta reconstitution of the parthenolide biosynthetic pathway. Metab Eng **23:** 145–153

Mafu S, Jia M, Zi J, Morrone D, Wu Y, Xu M, Hillwig ML, Peters RJ (2016) Probing the promiscuity of ent-kaurene oxidases via combinatorial biosynthesis. Proc Natl Acad Sci USA **113:** 2526–2531

Nguyen DT, Göpfert JC, Ikezawa N, Macnevin G, Kathiresan M, Conrad J, Spring O, Ro DK (2010) Biochemical conservation and evolution of germacrene A oxidase in asteraceae. J Biol Chem **285:** 16588–16598

Nguyen TD, Faraldos JA, Vardakou M, Salmon M, O'Maille PE, Ro DK (2016) Discovery of germacrene A synthases in *Barnadesia spinosa*: The first committed step in sesquiterpene lactone biosynthesis in the basal member of the Asteraceae. Biochem Biophys Res Commun **479:** 622–627

Nguyen TD, MacNevin G, Ro DK (2012) De novo synthesis of high-value plant sesquiterpenoids in yeast. Methods Enzymol **517:** 261–278

Ober D (2010) Gene duplications and the time thereafter - examples from plant secondary metabolism. Plant Biol (Stuttg) **12:** 570–577

Paddon CJ, Keasling JD (2014) Semi-synthetic artemisinin: A model for the use of synthetic biology in pharmaceutical development. Nat Rev Microbiol **12:** 355–367

Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, et al (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. Nature **496:** 528–532

Padilla-Gonzalez GF, Dos Santos FA, Da Costa FB (2016) Sesquiterpene lactones: More than protective plant compounds with high toxicity. CRC Crit Rev Plant Sci **35:** 18–37

Panero JL, Funk VA (2008) The value of sampling anomalous taxa in phylogenetic studies: Major clades of the Asteraceae revealed. Mol Phylogenet Evol **47:** 757–782

Picman AK (1986) Biological activities of sesquiterpene lactone. Biochem Syst Ecol **14:** 255–281

Pompon D, Louerat B, Bronine A, Urban P (1996) Yeast expression of animal and plant P450s in optimized redox environments. Methods Enzymol **272:** 51–64

Prosser I, Altug IG, Phillips AL, König WA, Bouwmeester HJ, Beale MH (2004) Enantiospecific (+)- and (−)-germacrene D synthases, cloned from goldenrod, reveal a functionally active variant of the universal isoprenoid-biosynthesis aspartate-rich motif. Arch Biochem Biophys **432:** 136–144

Pyle BW, Tran HT, Pickel B, Haslam TM, Gao Z, MacNevin G, Vederas JC, Kim SU, Ro DK (2012) Enzymatic synthesis of valerena-4,7(11)-diene by a unique sesquiterpene synthase from the valerian plant (*Valeriana officinalis*). FEBS J **279:** 3136–3146

Ramirez AM, Saillard N, Yang T, Franssen MC, Bouwmeester HJ, Jongsma MA (2013) Biosynthesis of sesquiterpene lactones in pyrethrum (*Tanacetum cinerariifolium*). PLoS One **8:** e65030

Ro DK, Ouellet M, Paradise EM, Burd H, Eng D, Paddon CJ, Newman JD, Keasling JD (2008) Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinic acid. BMC Biotechnol **8:** 83

Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, et al (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature **440:** 940–943

Shen Q, Zhang L, Liao Z, Wang S, Yan T, Shi P, Liu M, Fu X, Pan Q, Wang Y, et al (2018) The genome of *Artemisia* annua provides insight into the evolution of Asteraceae family and artemisinin biosynthesis. Mol Plant **11:** 776–788

Shoichet BK, Baase WA, Kuroki R, Matthews BW (1995) A relationship between protein stability and protein function. Proc Natl Acad Sci USA **92:** 452–456

Soskine M, Tawfik DS (2010) Mutational effects and the evolution of new protein functions. Nat Rev Genet **11:** 572–582

Tantillo DJ (2010) The carbocation continuum in terpene biosynthesis—where are the secondary cations? Chem Soc Rev **39:** 2847–2854

Teoh KH, Polichuk DR, Reed DW, Nowak G, Covello PS (2006) *Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. FEBS Lett **580:** 1411–1416

Tokuriki N, Stricher F, Serrano L, Tawfik DS (2008) How protein stability and new functions trade off. PLOS Comput Biol **4:** e1000002

Wang X, Minasov G, Shoichet BK (2002) Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. J Mol Biol **320:** 85–95

Zhang Y, Teoh KH, Reed DW, Maes L, Goossens A, Olson DJ, Ross AR, Covello PS (2008) The molecular cloning of artemisinic aldehyde Δ11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in *Artemisia annua*. J Biol Chem **283:** 21501–21508