3-1-2016

Probing the promiscuity of ent-kaurene oxidases via combinatorial biosynthesis

Sibongile Mafu
Iowa State University

Meirong Jia
Iowa State University

Jiachen Zi
Iowa State University

Dana Morrone
Iowa State University

Yisheng Wu
Iowa State University

See next page for additional authors

Probing the promiscuity of ent-kaurene oxidases via combinatorial biosynthesis

Abstract

The substrate specificity of enzymes from natural products' metabolism is a topic of considerable interest, with potential biotechnological use implicit in the discovery of promiscuous enzymes. However, such studies are often limited by the availability of substrates and authentic standards for identification of the resulting products. Here, a modular metabolic engineering system is used in a combinatorial biosynthetic approach toward alleviating this restriction. In particular, for studies of the multiply reactive cytochrome P450, *ent*-kaurene oxidase (KO), which is involved in production of the diterpenoid plant hormone gibberellin. Many, but not all, plants make a variety of related diterpenes, whose structural similarity to *ent*-kaurene makes them potential substrates for KO. Use of combinatorial biosynthesis enabled analysis of more than 20 such potential substrates, as well as structural characterization of 12 resulting unknown products, providing some insight into the underlying structure–function relationships. These results highlight the utility of this approach for investigating the substrate specificity of enzymes from complex natural products' biosynthesis.

Keywords

Enzyme evolution, cytochromes P450, natural products, terpenoids

Disciplines

Biotechnology | Enzymes and Coenzymes | Molecular Biology | Natural Products Chemistry and Pharmacognosy

Comments

This is a manuscript of an article published as Mafu, Sibongile, Meirong Jia, Jiachen Zi, Dana Morrone, Yisheng Wu, Meimei Xu, Matthew L. Hillwig, and Reuben J. Peters. "Probing the promiscuity of ent-kaurene oxidases via combinatorial biosynthesis." *Proceedings of the National Academy of Sciences* 113, no. 9 (2016): 2526-2531. DOI: 10.1073/pnas.1512096113. Posted with permission.

Authors

Sibongile Mafu, Meirong Jia, Jiachen Zi, Dana Morrone, Yisheng Wu, Meimei Xu, Matthew L. Hillwig, and Reuben J. Peters

Classification: Biological Sciences (Biochemistry and Plant Sciences)

Probing the promiscuity of *ent*-kaurene oxidases via combinatorial biosynthesis

Short title: A promiscuous ent-kaurene oxidase

Sibongile Mafu[a,b], Meirong Jia[a], Jiachen Zi[a,c], Dana Morrone[a,d], Yisheng Wu[a,e], Meimei Xu[a], Matthew L. Hillwig[a,f] and Reuben J. Peters[a,#]

[a]Roy J. Carver Department of Biochemistry, Biophysics & Molecular Biology, Iowa State University, Ames, Iowa, 50011, USA; [b]Current address: Department of Plant Biology, University of California, Davis, California, 95616, USA; [c]Current address: Biotechnological Institute of Chinese Materia Medica, College of Pharmacy, Jinan University, Guangzhou, Guangdong, 510632, China; [d]Current address: Department of Basic Sciences, St. Louis College of Pharmacy, Missouri, 63110, USA; [e]Current address: Conagen Inc., Bedford, Massachusetts, 01730, USA; [f]Current address: Department of Chemistry, St. Vincent College, Latrobe, Pennsylvania, 15650, USA.

[#]**Corresponding author:** Reuben J. Peters, Iowa State University, Mol. Biol. Bldg. Rm. 4216, Ames, Iowa, 50011, USA. Phone: 1(515)294-8580; e-mail: rjpeters@iastate.edu

Keywords: Enzyme evolution, cytochromes P450, natural products, terpenoids

Abstract

The substrate specificity of enzymes from natural products metabolism is a topic of considerable interest, with potential biotechnological use implicit in the discovery of promiscuous enzymes. However, such studies are often limited by the availability of substrates and authentic standards for identification of the resulting products. Here a modular metabolic engineering system is employed in a combinatorial biosynthetic approach towards alleviating this restriction. In particular, for studies of the multiply-reactive cytochrome P450, *ent*-kaurene oxidase (KO), which is involved in production of the diterpenoid plant hormone gibberellin. Many, but not all, plants make a variety of related diterpenes, whose structural similarity to *ent*-kaurene makes them potential substrates for KO. Use of combinatorial biosynthesis enabled analysis of more than 20 such potential substrates, as well as structural characterization of 12 resulting novel products, providing some insight into the underlying structure-function relationships. These results highlight the utility of this approach for investigating the substrate specificity of enzymes from complex natural products biosynthesis.

Significance statement

Combinatorial biosynthesis is enabled by the use of promiscuous enzymes, but also can be applied to the discovery of such biocatalysts. Here such an approach was employed for investigation of the substrate selectivity of cytochromes P450 from diterpenoid phytohormone biosynthesis, enabling facile characterization of both multiple substrates and structural analysis of the resulting products. The discovery of one such promiscuous cytochrome P450 reported here provides insight into the underlying structure-function relationship and further may have biotechnological implications.

/body

We have co-opted natural products, the small molecules produced by living organisms, for a wide variety of purposes (1, 2), many of which are only tangentially related to their original biological function. For this reason, it is desirable to screen related compounds for the activity of interest. One means to access such chemical diversity is combinatorial biosynthesis (3). However, this is dependent on the use of promiscuous enzymes that can readily react with a variety of substrates. Unfortunately, investigation of substrate specificity is often hindered by the limited availability of structurally related and metabolically accessible compounds. This is particularly true for enzymes involved in the biosynthesis of more complex molecules, such as hormones and related natural products.

Gibberellins (GAs) are hormones essential for vascular plant growth and development, meeting the criteria for primary metabolites. Accordingly, the genes encoding the enzymes for the GA biosynthetic pathway are present in all tracheophytes (4). Thus, these plants all contain *ent*-kaurene oxidases (KOs), which are cytochrome P450 mono-oxygenases (CYPs) from the CYP701 family that catalyze multiple oxidations of the C4α methyl of the tetracyclic olefin intermediate *ent*-kaurene to form *ent*-kauren-19-oic acid (Figure 1), as an early step in GA biosynthesis (5).

Notably, the requirement for GA metabolism provides a reservoir of genes encoding biosynthetic enzymes, which has given rise to the production of many related diterpenoid natural products (6). These are characterized by initial bicyclization of the general diterpenoid precursor (*E,E,E*)-geranylgeranyl diphosphate (GGPP) to a labdadienyl/copalyl diphosphate (CPP) intermediate by class II diterpene cyclases then termed CPP synthases (CPSs). This has led to designation of the derived natural products as labdane-related diterpenoids (LRDs), and the

presence of the corresponding decalin ring structure in almost all of these makes the LRDs easily recognizable (7). In the case of GA biosynthesis, the corresponding intermediate is *ent*-CPP, with 9*R*,10*R* stereochemistry, and this is further cyclized to *ent*-kaurene by *ent*-kaurene synthases (KSs)(5). Given the relatively close phylogenetic relationship between KSs and other terpene synthases that act on CPP (of various stereochemistries), these class I diterpene synthases have been termed KS-like (KSL)(7). Accordingly, the CPS and KS required in all tracheophytes for GA production have served as a genetic reservoir from which more specialized LRD metabolism has repeatedly evolved by gene duplication and neo-functionalization (6).

By contrast, KO appears to have been much less frequently diverted to more specialized LRD metabolism, with the only biochemically distinct CYP701 family members characterized to date found in rice (*Oryza sativa*). In particular, rice contains an expanded family of CYP701A sub-family members (8), at least two of which have distinct biochemical function – i.e., can not act as KOs (9, 10) – while genetic evidence clearly indicates that only one (CYP701A6/OsKO2) actually functions in gibberellin metabolism (11). On the other hand, *Arabidopsis thaliana* is known to have only a single KO, CYP701A3/AtKO, as required for gibberellin biosynthesis (12). Previous investigation has demonstrated some promiscuity for AtKO in that it will react with at least two LRD olefins that are structurally distinct yet related to *ent*-kaurene – i.e., *ent*-atiserene and *ent*-beyerene (13). However, this promiscuity has not been thoroughly investigated with a wider range of potential substrates, nor compared to any other KO. As with many such enzymes involved in the metabolism of more complex natural products, such studies are hindered by the severely limited access to the relevant compounds, both substrates and products. Here we report the use of combinatorial biosynthesis for such investigations, enabling analysis of over 20 potential substrates, revealing significant promiscuity for AtKO, while OsKO2 was

found to be much more specific. Critically, this approach further allowed structural characterization of 12 resulting novel compounds, providing insight into the underlying structure-function relationship.

Results

Screening substrate specificity

We have developed a modular metabolic engineering system that enables facile transformation of *Escherichia coli* to engineer the production of any terpene for which the corresponding synthase(s) are known (14). This has been further coupled to strategies for functional expression of plant CYPs in *E. coli*, to (re)create terpenoid biosynthetic pathways via a synthetic biology approach (15). Given the difficulty of obtaining an extended range of relevant substrates for in vitro reactions – e.g., by (retro)synthesis – here we have adopted this approach to screen AtKO and OsKO2 activity against a range of LRD olefins in *E. coli*. Specifically we co-expressed codon-optimized and N-terminally modified constructs for each of these KOs (9, 13), along with a corresponding cytochrome P450 reductase (CPR) from the same species (i.e., AtCPR1 or OsCPR1)(16, 17), with upstream CPS and KS(L) genes from a variety of sources, not only plants but bacteria as well, that together produce an array of more than 20 LRDs (Figure 2, also see SI Appendix, Table S1). Based on the ability of AtKO to react with a range of LRDs derived from *ent*-CPP, we further screened both AtKO and OsKO2 against LRD olefins produced via the available alternative stereoisomers – i.e., normal (9*S*,10*S*) and *syn* (9*S*,10*R*).

As expected, when either AtKO or OsKO2 were co-expressed in *E. coli* engineered to produce *ent*-kaurene, turnover to *ent*-kauren-19-oic acid was observed (see SI Appendix, Fig. S1). Upon wider screening, AtKO was found to react with a much broader range of the tested LRD olefins than OsKO2 (Figure 3). In particular, OsKO2 only reacted with *ent*-kaurene, its double-bond isomer *ent*-isokaurene, and the closely related *ent*-trachylobane, converting the C4α methyl to a mixture of the alcohol and further oxidized carboxylic acid with each substrate (see SI Appendix, Figs S1 – S3). By contrast, AtKO is more promiscuous, reacting with LRDs derived from *ent*-CPP to transform the C4α methyl to varying ratios of alcohols and derived carboxylic acids (see SI Appendix, Figs S1 – S6). In addition, AtKO will even react with LRDs derived from normal or *syn*- CPP (see SI Appendix, Figs S7 – S14), although largely only to the extent of carrying out a single oxygenation reaction to produce hydroxylated derivatives.

Determination of hydroxylation positions

While we have authentic standards for the C19 oxygenated products mentioned above (e.g., *ent*-kauren-19-oic, *ent*-isokauren-19-oic, *ent*-trachyloban-19-oic acids), as well as *ent*-beyeren-19-ol and *ent*-beyeren-19-oic acid, for many of the observed AtKO products no such standards are available. Accordingly, to determine the regio- and stereo- chemistry of the catalyzed reaction, we carried out structural characterization of these compounds by NMR. To obtain sufficient amounts for this purpose, we increased flux to terpenoid metabolism in the engineered cultures by further including a pMBI vector that enables conversion of mevalonate to isoprenoid precursors (18). The resulting major product(s) were then isolated and subjected to NMR structural analysis, elucidating their composition (Figure 4). Through these studies it was found that AtKO targets the C4α methyl group in *ent*-CPP derived LRDs, including *ent*-

cassadiene and *ent*-sandaracopimaradiene, which are converted to a mixture of the alcohol and derived carboxylic acid (see SI Appendix, Figs S15 & S16 and Tables S2 & S3). By contrast, with LRDs derived from normal or *syn*- CPP, AtKO targets the neighboring C3 and C2 positions. Specifically, we found that AtKO largely hydroxylates *syn*-aphidicolene, *syn*-labda-8(17),12*E*,14-triene, *syn*-pimara-7,15-diene and *syn*-stemod-13(17)-ene at the C3β position. But in addition, with *syn*-pimara-7,15-diene AtKO also yields small amounts of the C2α-hydroxyl, and with *syn*-stemodene, converts small amounts of the C4β methyl to a carboxylate (see SI Appendix, Figs S17 – S20 and Tables S4 – S8). Similarly, AtKO hydroxylates isopimara-8,15-diene at the C3β position, while it hydroxylates isopimara-7,15-diene at the C2α position, with a minor amount of further oxidization to the ketone also observed, and sandaracopimaradiene (isopimara-8(14),15-diene) at both the C2α and C3β positions (see SI Appendix, Figs S21 – S23, and Tables S9 – S14).

Relative kinetic analysis

To enable comparison of the reaction rates of the different substrates by AtKO, in vitro assays were carried out with a sub-set of alternative substrates, also obtained via the metabolic engineering system. Given the variability in the relationship between catalytic activity and amount of 'correctly folded' CYP measured by carbon monoxide binding difference spectra observed between enzymatic preparations, we first attempted to carry out competition assays as previously described (19). However, AtKO reacts very efficiently with *ent*-kaurene and invariably preferentially converts this to *ent*-kaurenoic acid before reacting with the alternative (LRD) substrate. Accordingly, to compare catalytic efficiency we normalized our substrate turnover to *ent*-kaurene – i.e., each enzyme preparation was assayed with *ent*-kaurene and a

standard amount of such measured activity used in assays with the alternative substrates. AtKO exhibited much lower affinity for its non-native substrates with K_M ranging from 30 to 50 μM, representing >10-fold increases over the 2 μM K_M for *ent*-kaurene (13). Moreover, V_{max} was generally ~10-fold less than that observed with *ent*-kaurene as well, indicating that AtKO exhibits at least two order of magnitude decreases in catalytic efficiency with these alternative substrates (Table 1).

Molecular phylogenetic analysis

The difference between the activity exhibited by AtKO and OsKO2 may reflect their distinct evolutionary contexts. In particular, while Arabidopsis seems to be representative of the Brassicaceae plant family, which do not produce any more specialized LRDs, only GA, and contain just one CYP701 family member, rice is representative of the Poaceae plant family where it has been indicated that more specialized LRDs arose early, with widespread retention (20), and rice contains a number of CYP701A sub-family members with distinct function (9-11). This presumably exerted differing selective pressures on these genes, which was examined by molecular phylogenetic analysis of all characterized members of the CYP701 family (Figure 5, see also SI Appendix, Table S14). As might be expected, AtKO groups with other dicot KOs, while OsKO2 groups with its paralogs (i.e., the other rice CYP701A sub-family members) and the only other characterized cereal KO, specifically that from barley, CYP701A13/HvKO. To detect adaptive evolution, the ratios of the nonsynonymous and synonymous substitution rates (d_N/d_S or ω) were estimated for this gene family, not only for overall and pairwise comparisons, but also lineage and/or site specific models (21). Across the CYP701 family, purifying selection is dominant ($\omega < 0.2$), and nor is this appreciably different for AtKO relative to the rest of the

family, as indicated by the essentially identical values for ω, as well as log-likelihood score, for this distinct branch model. On the other hand, the paralogs from rice (CYP701A6, 8, 9 & 19) are clearly under different selective pressure than the rest of the family ($p < 0.001$), with $\omega = 0.44$ for these, versus 0.12 for the remainder of the family. Additional analysis further strongly supported variation in selective pressure along the coding sequence, again with $p < 0.001$ from a likelihood ratio test. However, these sites do not generally appear to be under positive selection, as the relevant null test exhibits an identical log-likelihood score. Instead, this branch-site model indicates that a proportion of the sites are released from the usual purifying selection pressure, with ω changing from < 0.1 to 1.0 (i.e., neutral) for ~25% of the residues in the rice CYP701A paralogs. Nevertheless, this analysis does suggest that a few residues may be under positive selection pressure, which may still be possible given the limited number of available sequences. Indeed, the distinct function of two of the rice CYP701A sub-family members, CYP701A8 and CYP701A9 (9, 10), almost certainly required such differentiating positive selection pressure (although this is not evident from the relevant branch model analysis, and comparison of log likelihood scores indicate that selective pressure differs for all the rice paralogs, rather than only these two).

Discussion

While enzymatic promiscuity enables combinatorial biosynthesis, investigation of this has been limited, particularly for more complex natural products metabolism, by the availability of substrates and authentic standards for identification of the resulting products. As demonstrated here, combinatorial biosynthesis also provides a robust approach for discovery and

characterization of promiscuous enzymes. Reflecting the underlying (bio)chemical logic of plant natural product biosynthesis (22), here various diterpene synthases were combined to produce a range of scaffolds that can be acted upon by tailoring enzymes such as plant CYPs like the KOs from gibberellin biosynthesis investigated here. Indeed, this approach enabled investigation of more than 20 potential substrates (Figure 2), along with structural resolution of 12 resulting novel products (Figure 4).

Notably, this approach revealed that, while OsKO2 is quite specific for *ent*-(iso)kaurene, and only also reacts with the structurally closely related *ent*-trachylobane (Figure 5A), AtKO exhibits much broader promiscuity, albeit with restrictions that provide some insight into the limits of its active site to accommodate alternative substrates. For example, with the eight LRD substrates derived from *ent*-CPP, AtKO reacts with all but *ent*-pimara-8(14),15-diene, which differs from *ent*-sandaracopimaradiene only in configuration of the methyl and vinyl groups at C13, and 16α-hydroxy-*ent*-kaurane, which differs from *ent*-(iso)kaurene only in the presence of the hydroxyl group. From this, it appears that differences in LRDs distal from the targeted C4α methyl affect the ability of AtKO to accommodate these as substrates. Nevertheless, with all six reactive LRDs derived from *ent*-CPP, AtKO specifically targets this methyl group.

By contrast, with the reactive LRDs derived from normal and *syn*- CPP, AtKO largely targets C2α and/or C3β, which are roughly apposed but on the opposite 'side' of the 'A' ring from the C4α methyl targeted by AtKO in *ent*-CPP derived LRDs. Indeed, at least in one case (i.e., *syn*-stemodene) the only other identified target is the C4β methyl. This highlights the effect of the configuration of the 'A' ring on the orientation of the LRDs in the active site. In particular, the configuration of the methyl substituent on C10, as the β-configuration of this is the most prominent feature in common between the LRDs derived from normal and *syn*- CPP (Figure 5B).

In addition, while AtKO reacts with all of the tested *syn*-CPP derived LRDs, it exhibits some selectivity with those derived from normal CPP. Specifically, it will only react with isopimaradienes, which have a β-methyl (and α-vinyl) at C13. This further emphasizes the effect that such distal differences in structure have on substrate binding in the AtKO active site. Finally, the preferential yield of alcohols rather than more oxidized products with these alternative substrates may stem from the previously noted reduced affinity of AtKO for even the natural *ent*-kauren-19-ol intermediate relative to the olefin precursor *ent*-kaurene (or more oxidized *ent*-kauren-19-al intermediate)(13).

The contrast between the quite strict specificity of OsKO2 versus the promiscuity exhibited by AtKO is intriguing (Figure 3). Even beyond the question of the basis for their contrasting substrate specificity, given the presence of alternative substrates over substantial evolutionary timescales in rice but not Arabidopsis (i.e., in the Poaceae versus Brassicaceae), and the physiological function of both OsKO2 and AtKO in primary (gibberellin) biosynthesis, this difference, although obviously limited in scope, is consistent with the previously proposed general hypothesis that such evolutionary context (i.e., metabolic background) plays a role in shaping enzymatic specificity (23). Further evidence may be obtained by investigating the substrate specificity of a wider array of KOs from other plant species with distinct ranges of LRD metabolism in the future. In addition, it might then be of interest to examine the effect that substitution of OsKO2 with a more promiscuous KO (such as AtKO) has on rice plants. In any case, to the extent that such promiscuity reflects the activity of KO from plants with limited LRD metabolism, such activity provides a means for immediate production of more elaborated metabolites from novel diterpenes that might arise (e.g., by gene duplication and neo-functionalization of upstream diterpene synthases from gibberellin biosynthesis). These natural

products would be more likely to exhibit biological activity (6), and such latent biosynthetic capacity then may (partially) underlie the broad diversity of LRD natural products observed in plants, consistent with a recently suggested mechanism for the evolution of more specialized metabolism in plants more generally (24).

Regardless of evolutionary basis, although the promiscuity of AtKO revealed here is coupled to somewhat reduced catalytic efficiency in vitro, its ability to convert alternative substrates largely to hydroxylated products suggests that the observed activity might be of use in industrial biotechnology approaches. Indeed, the demonstrated incorporation of membrane-associated CYPs such as the KOs into a modular bacterial metabolic engineering system via a synthetic biology approach emphasizes the utility of such microbial host production systems. Moreover, the results reported here further highlight the usefulness of applying combinatorial biosynthesis for investigating the substrate specificity of enzymes from complex natural products metabolism.

Materials and Methods

Combinatorial biosynthesis. The substrate specificity of AtKO and OsKO2 was assessed by use of our previously described modular metabolic engineering system for production of labdane-related diterpenes in *E. coli* (14). The relevant combinations of diterpene synthases are described in the SI Appendix (Table S1). This system is further amendable to incorporation of CYPs. In particular, using synthetic genes, codon-optimized for expression in *E. coli* along with replacement of the N-terminal transmembrane helix sequence with a leader peptide (15), as has been previously reported for AtKO and OsKO2 (9, 13). Accordingly, it was possible to co-

express either KO, along with the requisite CPR from the same species, in *E. coli* also engineered to make any one (or two) of over 20 different LRDs (see the SI Appendix for more details). Organic (hexane) extracts of these recombinant cultures were then analyzed by GC-MS to screen for KO activity. Moreover, in those cases where the resulting products could not be identified from the available authentic standards, it was then possible to produce sufficient quantities for structural analysis via NMR by increasing flux towards isoprenoid metabolism, as well as simply increasing culture volumes, much as previously described (25).

Chemical Structure Identification. Novel products were purified via flash chromatography and HPLC, much as has been previously described (9, 10, 17, 19, 26-28), with further details available in the SI Appendix. The resulting compounds were then resuspended in deuterated chloroform ($CDCl_3$), for structural analysis by NMR, as described in more detail in the SI Appendix. Note that the acquired 1D ^{1}H and ^{13}C spectra can be found in the SI Appendix (Figures S24 – S35). In certain cases, the identity of the observed carboxylic acid products was verified by feeding the structurally characterized hydroxylated precursor (i.e., for *ent*-cassadien-19-oic and *ent*-sandaracopimaradien-19-oic acids, as described in more detail in the SI Appendix).

Analysis of catalytic efficiency. Given the variability between the characteristic CO binding difference spectra usually utilized to quantify CYP concentrations, and the actually observed enzymatic activity in vitro, catalytic efficiency was determined by relative kinetic analyses. This was limited to assays with AtKO, as OsKO2 was not sufficiently stable enough for reliable use in in vitro. These assays were carried out with cell-free extracts, fed the relevant substrate (i.e.,

those shown in Table 1), with comparison of the observed activity to that observed with parallel assays fed the native *ent*-kaurene substrate, enabling normalization of V_{max} (i.e., to the rate observed with *ent*-kaurene, as presented in Table 1). More detailed description of these studies can be found in the SI Appendix.

Molecular phylogenetic analysis. The open reading frames coding for the characterized CYP701 family members were obtained from GenBank (see SI Appendix, Table S14). Alignment, phylogenetic tree construction and overall (as well as pairwise) analyses of evolutionary pressure were carried out using the MEGA 6.06 software package for Mac (29), all of which supported only purifying selection. Lineage specific adaptive evolution was investigated using the PAML 4.8 software package for Mac (30), particularly the CODEML program, employed largely as outlined (31). The fit of various models to the evolutionary history of the CYP701 gene family was assessed by comparison of the calculated log likelihood scores, with statistical significance estimated by likelihood ratio tests where appropriate (i.e., between models with different degrees of freedom). Such comparisons supported a distinct branch model in which the clade of rice paralogs are under less stringent purifying selection than the rest of the CYP701 family (as depicted in Figure 6), with further branch-site analysis indicating this is due to release of such purifying selection on a sub-section of sites/residues, albeit while also suggesting that a few residues may be under positive selection (see the SI Appendix for more details).

Acknowledgements

This work was supported by grants to R.J.P. from the NSF (MCB0919735) and NIH (GM109773).

References

1. Newman DJ & Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. *Journal of natural products* 75(3):311-335.
2. Cantrell CL, Dayan FE, & Duke SO (2012) Natural products as sources for new pesticides. *Journal of natural products* 75(6):1231-1242.
3. Zhou H, Xie X, & Tang Y (2008) Engineering natural products using combinatorial biosynthesis and biocatalysis. *Curr Opin Biotechnol* 19(6):590-596.
4. Peters RJ (2013) Gibberellin phytohormone metabolism. *Isoprenoid synthesis in plants and microorganisms: New concepts and experimental approaches*, eds Bach T & Rohmer M (Springer, New York), pp 233-249.
5. Hedden P & Thomas SG (2012) Gibberellin biosynthesis and its regulation. *Biochem J* 444(1):11-25.
6. Zi J, Mafu S, & Peters RJ (2014) To Gibberellins and Beyond! Surveying the Evolution of (Di)Terpenoid Metabolism. *Annu Rev Plant Biol* 65:259-286.
7. Peters RJ (2010) Two rings in them all: The labdane-related diterpenoids. *Nat. Prod. Rep.* 27(11):1521-1530.
8. Itoh H, *et al.* (2004) A rice semi-dwarf gene, *Tan-Ginbozu* (*D35*), encodes the gibberellin biosynthesis enzyme, *ent*-kaurene oxidase. *Plant Mol Biol* 54:533-547.
9. Wang Q, Hillwig ML, Wu Y, & Peters RJ (2012) CYP701A8: A rice *ent*-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism. *Plant Physiol.* 158(3):1418-1425.
10. Kitaoka N, Wu Y, Xu M, & Peters RJ (2015) Optimization of recombinant expression enables discovery of novel cytochrome P450 activity in rice diterpenoid biosynthesis. *Appl. Microbiol. Biotechnol.* 99(18):7549-7558.
11. Sakamoto T, *et al.* (2004) An overview of gibberellin metabolism enzyme genes and their related mutants in rice. *Plant Physiol.* 134:1642-1653.
12. Helliwell CA, *et al.* (1998) Cloning of the *Arabidopsis ent*-kaurene oxidase gene GA3. *Proc. Natl. Acad. Sci., U.S.A.* 95(15):9019-9024.
13. Morrone D, Chen X, Coates RM, & Peters RJ (2010) Characterization of the kaurene oxidase CYP701A3, a multifunctional cytochrome P450 from gibberellin biosynthesis. *Biochem. J.* 431(3):337-344.
14. Cyr A, Wilderman PR, Determan M, & Peters RJ (2007) A Modular Approach for Facile Biosynthesis of Labdane-Related Diterpenes. *J. Am. Chem. Soc.* 129:6684-6685.
15. Kitaoka N, Lu X, Yang B, & Peters RJ (2015) The application of synthetic biology to elucidation of plant mono-, sesqui-, and diterpenoid metabolism. *Mol. Plant* 8(1):6-16.
16. Urban P, Mignotte C, Kazmaier M, Delorme F, & Pompon D (1997) Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. *J. Biol. Chem.* 272(31):19176-19186.

17. Swaminathan S, Morrone D, Wang Q, Fulton DB, & Peters RJ (2009) CYP76M7 is an *ent*-cassadiene C11α-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell* 21:3315-3325.
18. Martin VJJ, Pitera DJ, Withers ST, Newman JD, & Keasling JD (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature biotechnology* 21(7):796-802.
19. Wu Y, Wang Q, Hillwig ML, & Peters RJ (2013) Picking sides: Distinct roles for CYP76M6 and -8 in rice oryzalexin biosynthesis. *Biochem. J.* 454(2):209-216.
20. Schmelz EA, *et al.* (2014) Biosynthesis, elicitation and roles of monocot terpenoid phytoalexins. *Plant J* 79(4):659-678.
21. Anisimova M & Liberles D (2012) Dectecting and understanding natural selection. *Codon Evolution: Mechanisms and Models*, eds Cannarozzi G & Schneider A (Oxford University Press, New York).
22. Anarat-Cappillino G & Sattely ES (2014) The chemical logic of plant natural product biosynthesis. *Current opinion in plant biology* 19:51-58.
23. Tawfik DS (2014) Accuracy-rate tradeoffs: how do enzymes meet demands of selectivity and catalytic efficiency? *Curr Opin Chem Biol* 21:73-80.
24. Weng JK, Philippe RN, & Noel JP (2012) The rise of chemodiversity in plants. *Science* 336(6089):1667-1670.
25. Morrone D, *et al.* (2010) Increasing diterpene yield with a modular metabolic engineering system in E. coli: comparison of MEV and MEP isoprenoid precursor pathway engineering. *Appl. Microbiol. Biotechnol.* 85:1893-1906.
26. Wang Q, *et al.* (2012) Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster. *J. Biol. Chem.* 287(9):6159-6168.
27. Wang Q, Hillwig ML, & Peters RJ (2011) CYP99A3: Functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice. *Plant J.* 65(1):87-95.
28. Wu Y, Hillwig ML, Wang Q, & Peters RJ (2011) Parsing a multifunctional biosynthetic gene cluster from rice: Biochemical characterization of CYP71Z6 & 7. *FEBS Lett.* 585:3446-3451.
29. Tamura K, Stecher G, Peterson D, Filipski A, & Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30(12):2725-2729.
30. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586-1591.
31. Bielawski JP (2013) Detecting the signatures of adaptive evolution in protein-coding genes. *Curr Protoc Mol Biol* Chapter 19:Unit 19 11.

Figure Legends

Figure 1: Conversion of *ent*-kaurene to *ent*-kaurenoic acid catalyzed by *ent*-kaurene oxidases (13).

Figure 2: Labdane-related diterpenes screened in this study, with origin of the underlying stereochemistry (i.e., CPP stereoisomer) also shown and indicated by color – i.e., CPP and derived diterpenes are depicted in blue, *ent*-CPP and derived diterpenes in purple, with *syn*-CPP and derived diterpenes in red.

Figure 3: Relative activity of AtKO versus OsKO2 with screened labdane-related diterpenes (LRDs). Shown is turnover observed with recombinant AtKO (and AtCPR1) or OsKO2 (and OsCPR1) in *E. coli* also metabolically engineered to produce the noted LRD.

Figure 4: AtKO catalyzes hydroxylations of multiple labdane-related diterpenes. Shown are the characterized novel reactions, as elucidated by structural analysis of the resulting major product(s).

Figure 5: Structure-function relationship for the activity exhibited by A) OsKO2 and B) AtKO. A) Three-dimensional rendering illustrating the close structural relationship between *ent*-(iso)kaurene and *ent*-trachylobane, the only substrates accepted by OsKO2. B) Configuration of the 'A' ring from the three underlying stereoisomers (i.e., of CPP) illustrating targeting of the same two roughly apposed sites with normal and *syn*- CPP derived labdane-related diterpenes that may be related to their similar configuration in this region. Previous work has demonstrated that other non-KO CYP701A sub-family members (i.e., from rice) target the C3α position with *ent*- and *syn*- CPP derived labdane-related diterpenes (9, 10).

Figure 6: Molecular phylogenetic analysis of the CYP701 family, based on full-length open reading frames for the functionally characterized members (see SI Appendix, Table S14). Shown is a rooted phylogenetic tree, with the corresponding bootstrap values (from 100 replicates). Also indicated are the branches that delinate the relevant classes of plants from which these CYP701 family members originate. The thicker lines delinate the clade of rice paralogs that lineage specific adaptive evolutionary analysis strongly indicates is under distinct selective pressure relative to the rest of the family. More detailed branch-site analysis further indicates that this is due to release of purifying selection on a sub-set (~25%) of the sites/residues in the rice CYP701 family members, two of which are functionally distinct (these are indicated by *).