



# Ancestral Sequence Reconstruction of a Cytochrome P450 Family Involved in Chemical Defense Reveals the Functional Evolution of a Promiscuous, Xenobiotic-Metabolizing Enzyme in Vertebrates

Kurt L. Harris,<sup>1,†</sup> Raine E.S. Thomson,<sup>1,†</sup> Yosephine Gumulya,<sup>1</sup> Gabriel Foley,<sup>1</sup> Saskya E. Carrera-Pacheco,<sup>2</sup> Parnayan Syed,<sup>1</sup> Tomasz Janosik,<sup>3</sup> Ann-Sofie Sandinge,<sup>4</sup> Shalini Andersson,<sup>5</sup> Ulrik Jurva,<sup>4</sup> Mikael Bodén ,<sup>1</sup> and Elizabeth M.J. Gillam <sup>1,\*</sup>

<sup>1</sup>School of Chemistry and Molecular Biosciences, The University of Queensland, St. Lucia, Brisbane 4072, QLD, Australia

<sup>2</sup>Centro de Investigación Biomédica, Facultad de Ciencias de la Salud Eugenio Espejo, Universidad UTE, Quito 170147, Ecuador

<sup>3</sup>RISE Research Institutes of Sweden, Division Bioeconomy and Health, Chemical Process and Pharmaceutical Development, Södertälje, Sweden

<sup>4</sup>DMPK, Early Cardiovascular, Renal and Metabolism, BioPharmaceuticals R&D, Astrazeneca, Gothenburg, Sweden

<sup>5</sup>Discovery Sciences, BioPharmaceuticals R&D, Astrazeneca, Gothenburg, Sweden

\*Corresponding author: E-mail: e.gillam@uq.edu.au

†These authors contributed equally to the work.

Associate editor: Meredith Yeager

## Abstract

The cytochrome P450 family 1 enzymes (CYP1s) are a diverse family of hemoprotein monooxygenases, which metabolize many xenobiotics including numerous environmental carcinogens. However, their historical function and evolution remain largely unstudied. Here we investigate CYP1 evolution via the reconstruction and characterization of the vertebrate CYP1 ancestors. Younger ancestors and extant forms generally demonstrated higher activity toward typical CYP1 xenobiotic and steroid substrates than older ancestors, suggesting significant diversification away from the original CYP1 function. Caffeine metabolism appears to be a recently evolved trait of the CYP1A subfamily, observed in the mammalian CYP1A lineage, and may parallel the recent evolution of caffeine synthesis in multiple separate plant species. Likewise, the aryl hydrocarbon receptor agonist, 6-formylindolo[3,2-*b*]carbazole (FICZ) was metabolized to a greater extent by certain younger ancestors and extant forms, suggesting that activity toward FICZ increased in specific CYP1 evolutionary branches, a process that may have occurred in parallel to the exploitation of land where UV-exposure was higher than in aquatic environments. As observed with previous reconstructions of P450 enzymes, thermostability correlated with evolutionary age; the oldest ancestor was up to 35 °C more thermostable than the extant forms, with a <sup>10</sup>T<sub>50</sub> (temperature at which 50% of the hemoprotein remains intact after 10 min) of 71 °C. This robustness may have facilitated evolutionary diversification of the CYP1s by buffering the destabilizing effects of mutations that conferred novel functions, a phenomenon which may also be useful in exploiting the catalytic versatility of these ancestral enzymes for commercial application as biocatalysts.

**Key words:** cytochrome P450, drug metabolism, CYP1A2, CYP1B1, ancestral sequence reconstruction, thermostability.

## Introduction

Cytochrome P450 enzymes (P450s) are a superfamily of heme-containing monooxygenases found in all domains of life. Their ability to functionalize unactivated carbon centers has led to their evolutionary diversification to many important and diverse physiological roles including the utilization of carbon sources in microorganisms, production of defensive compounds in plants and microorganisms (e.g., many antibiotics), synthesis of chemical signaling molecules (hormones, eicosanoids), and the

metabolic clearance of lipophilic chemicals. The vertebrate P450 families, CYP1–CYP3, include many important human drug-metabolizing enzymes and are principally responsible for the clearance of xenobiotics including drugs and environmental toxins, but also contribute to the metabolism of steroids, eicosanoids, and fatty acids.

Xenobiotic-metabolizing P450s represent an interesting case study for enzyme evolution since they are involved in protecting animals from a diverse array of potentially toxic environmental chemicals, such as plant secondary metabolites as well as numerous endogenous chemicals. Thus,

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

it is important for the survival of an organism that they show a high degree of promiscuity, either individually or collectively, to ensure adequate defense of the host organism against the accumulation of potentially toxic lipophilic chemicals, especially those ingested via the diet.

A finite set of enzymes in each organism (e.g., ~16 in humans of which five are dominant) deals with an extensive, diverse, and constantly changing array of xenobiotics (Guengerich 2015)—a function that can be described as “chemical defense” (Gillam and Hunter 2007). They use a conserved diflavin redox partner, NADPH-cytochrome P450 reductase (CPR), the redox cofactor, NADPH, and molecular oxygen to insert a single oxygen into a substrate, reducing the other to water (Aigrain et al. 2012). P450-mediated monooxygenation leads to an increase in the water solubility of the metabolite compared with the parent compound, thereby facilitating its elimination directly in urine. Alternatively, monooxygenation may introduce or expose a functional group that enables conjugation with more water-soluble moieties. The extant forms show broad and often overlapping substrate specificity on xenobiotic substrates, but exhibit routinely poor stability and typically feeble catalytic rates (less than  $\sim 10 \text{ min}^{-1}$ ) consistent with not being specialized to any one xenobiotic substrate.

The CYP1 family enzymes are responsible for the metabolism of  $\sim 19\%$  of chemicals subject to metabolic clearance in humans, representing  $\sim 20\%$  of the total contribution by P450s (Rendic and Guengerich 2015). They make a proportionally greater contribution to the bioactivation of environmental carcinogens (32% of the bioactivation of carcinogens; 48% of the total P450 contribution; Rendic and Guengerich 2012). Most xenobiotic substrates of CYP1s are fused ring structures or heterocyclic amines such as biphenyls, caffeine, and benzo[*a*]pyrene (Guengerich 2015) but they also metabolize various endogenous substrates including 17 $\beta$ -estradiol (E2), testosterone, uroporphyrin, and arachidonic acid (Lambrecht et al. 1992; Spink et al. 1992; Hayes et al. 1996). There are three human forms, CYP1A1, CYP1A2, and CYP1B1: CYP1A1 and CYP1A2 are the principal xenobiotic-metabolizing CYP1 forms, with the hepatic form, CYP1A2, contributing  $\sim 9\%$  of the total P450-mediated clearance of drugs (Rendic and Guengerich 2015). CYP1A1 is mostly extrahepatic and may contribute to tissue-specific xenobiotic metabolism, for example, in the lung, especially in smokers. CYP1A1 and CYP1B1 can also activate certain environmental carcinogens, in particular benzo[*a*]pyrene, resulting in DNA-adduct forming products (Penman et al. 1994; Buters et al. 1995; Shimada et al. 1996; Crespi et al. 1997; Shimada et al. 1999; Shimada et al. 2001). Genetic polymorphisms in CYP1B1 have been linked to congenital glaucoma (Stoilov et al. 1997), whereas polymorphisms in CYP1A2 have been associated with the risk of cardiovascular disease (Cornelis et al. 2004) and with dietary coffee intake (Rodenburg et al. 2012).

Mammalian CYP1 gene expression is regulated by the aryl hydrocarbon receptor (AhR), and many known

CYP1 substrates, for example, components of cigarette smoke, are also AhR ligands. One of the most potent endogenous AhR ligands identified to date is 6-formylindolo[3,2-*b*]carbazole (FICZ) (Rannug et al. 1987; Rannug et al. 1995), which is also an exceptionally good substrate for all three human CYP1 enzymes, displaying a catalytic efficiency that widely exceeds those of all other CYP1A1 substrates (Wincent et al. 2009). This naturally occurring compound can be generated by the photo-oxidation of tryptophan and can also be formed systemically (Smirnova et al. 2016) or by certain microorganisms (Schallreuter et al. 2012; Magiatis et al. 2013). FICZ is, therefore, a likely candidate for an ancestral substrate of CYP1 enzymes (reviewed in Rannug and Rannug 2018). However, given their ability to metabolize a wide range of other endogenous substrates and xenobiotic compounds that were not present historically, the true ancestral function of these enzymes and how they evolved to carry out their modern-day activities remains largely unknown.

Although the literature exists on the phylogeny and evolution of CYP1s (Morrison et al. 1998; Goldstone and Stegeman 2006; Goldstone et al. 2007; Kawai et al. 2010; Jönsson et al. 2011) complementary experimental studies on the evolution and ancestral function of this family are lacking. Therefore, the aim of the present study was to infer and resurrect ancestral CYP1 forms in order to explore the evolution of function in this P450 family. In addition, previous reconstructions of P450s and many other enzymes have shown that the thermostability of ancestral enzymes is often (but not always) greatly enhanced over that of extant forms (Risso et al. 2013; Hart et al. 2014; Gumulya et al. 2018; Gumulya et al. 2019). We, therefore, hypothesized that the resurrected CYP1 ancestors would be more thermostable than the extant forms.

Here we show that CYP1 enzymes evolved from a highly thermostable ancestor with relatively poor activity toward typical xenobiotic and endobiotic substrates of the well-characterized extant human CYP1 forms. This study addresses the gap in experimental evidence and provides insights into the evolution of activity and promiscuity in xenobiotic-metabolizing enzymes involved in chemical defense.

## Results

### Ancestral Sequence Reconstruction

A total of 471 sequences from the CYP1 clade (encompassing the vertebrate CYP1A, CYP1B, CYP1C, and CYP1D subfamilies plus the available CYP1E and CYP1F sequences identified in *Ciona* and *Lottia gigantea*, respectively) were aligned using Multiple Alignment using Fast Fourier Transform (MAFFT) and used to generate a tree in MEGA, using the maximum-likelihood method with 100 bootstrap replicates (Alignment and Bootstrapped Tree, Supplementary material online). The tree and alignment were submitted to Graphical Representation of Ancestral

Sequence Predictions (GRASP; Foley 2022) to infer a series of ancestors using a maximum likelihood, joint reconstruction approach. GRASP models insertion and deletion events via the inference of a partial order graph at each ancestor, representing the inclusion/exclusion of sequence by “jumps” (edges) between positions determined by the original input alignment (Foley 2022; Ross et al. 2022). Although content is determined by maximum likelihood, inference of edges for an ancestor is based on maximum parsimony. Each ancestor was denoted by a node number (N#) and the representative evolutionary branch or subfamilies from which it was derived (e.g., N52\_1A\_Actino denotes Node 52, derived from the CYP1A sequences from Actinopterygii). Eleven ancestral nodes were selected for resurrection based on evolutionary position and relevance to characterized extant enzymes (fig. 1), including the vertebrate CYP1 ancestor (N9\_1ABCD) and individual subfamily ancestors (N17\_1A, N20\_1B, N18\_1D). N25\_1C\_Actino was selected over the more ancient CYP1C subfamily ancestor (N19\_1C), due to the paucity of sequence data from the CYP1C subfamily outside the actinopterygian CYP1C group. Ancestors of individual forms of interest were also resurrected (specifically: mammalian CYP1A1, CYP1A2, and CYP1B1; avian CYP1A4 and CYP1A5; and actinopterygian CYP1A).

### Recombinant Expression and Thermostability Analysis of Ancestral Proteins

All ancestral enzymes were expressed in *Escherichia coli* at levels comparable to or exceeding those of the extant forms (table 1). The thermostability of each enzyme was quantified in whole cells (table 1, supplementary figs. S1 and S2, Supplementary Material online). In general, the older ancestors were found to exhibit greater thermostability than the extant forms with the oldest ancestor (N9\_1ABCD) demonstrating a  $^{60}T_{50}$  value (temperature at which 50% of the protein remains folded after heating for 60 min) of  $62.2 \pm 0.7$  °C and a  $^{10}T_{50}$  value (temperature at which 50% of the protein remains folded after heating for 10 min) of  $70.8 \pm 0.4$  °C; increases of 22–29 °C above those of the extant forms tested (human [h]CYP1A1, hCYP1A2, hCYP1B1, and turkey [t]CYP1A5). Each subfamily ancestor (N17\_1A, N18\_1D, N20\_1B, and N25\_1C\_Actino) also demonstrated a significantly higher stability than the corresponding descendant, extant forms ( $P < 0.005$ , unpaired two-tailed Student's *t*-test), the greatest difference being seen in the 1B subfamily, with the ancestor (N20\_1B) showing a thermostability  $>18$  °C higher than hCYP1B1. Younger ancestors, namely those corresponding to a single form, demonstrated smaller improvements in stability, with N98\_1B1\_Mam, N127\_1A2\_Mam, and N223\_1A5\_Aves, showing increases in  $^{60}T_{50}$  of 5–6 °C over the corresponding extant forms ( $P < 0.05$ , unpaired Student's two-tailed *t*-test). The  $^{60}T_{50}$  values of each form correlated with evolutionary branch length, a measure of evolutionary age that indicates the relative

separation of ancestors along evolutionary lineages (fig. 1 inset).

### Activity Towards Typical Marker and Drug Substrates of Extant CYP1 Forms

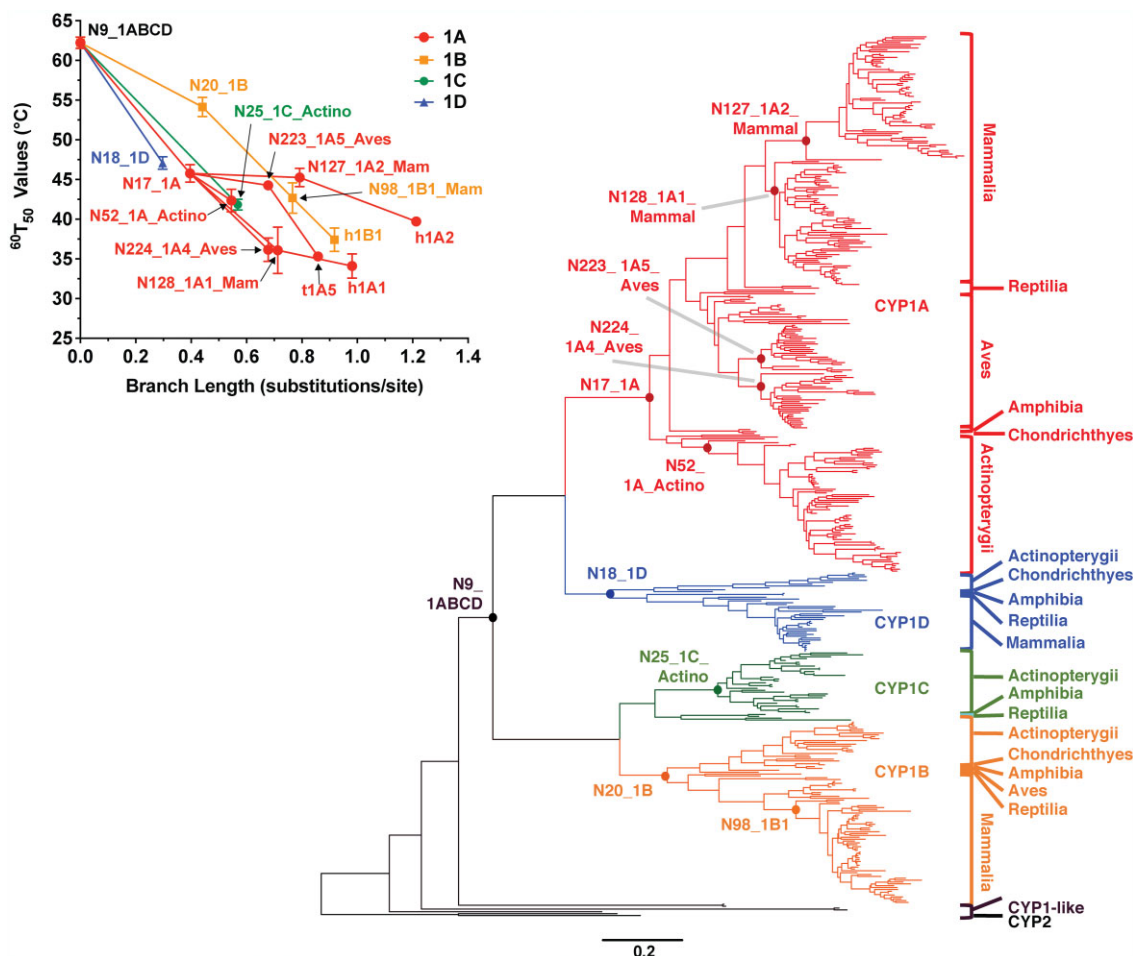
The activity of each ancestral and extant form was tested toward the dealkylation of several high-throughput probe substrates of the extant CYP1 forms (fig. 2, supplementary figs. S3 and S4, Supplementary Material online), namely 7-ethoxyresorufin, 7-methoxyresorufin, 7-benzyloxyresorufin, 7-pentoxeresorufin, and the luminogenic substrates luciferin-CEE and luciferin-ME. The relative activity of mammalian CYP1A forms toward these marker substrates is known to differ between CYP1A1 and CYP1A2 enzymes in a given species. Ratios of activity toward 7-alkoxyresorufins differed markedly between ancestors. Metabolism of the pharmaceutical substrates clozapine, tacrine, imipramine, and aminopyrine was also tested (fig. 3, supplementary figs. S5–S8, Supplementary Material online). Younger ancestors, especially from the CYP1A lineages, tended to demonstrate greater activity toward these substrates than older ancestors with some of the older ancestors (N9\_1ABCD and N18\_1D) showing minimal or no activity toward many substrates.

### E2 Metabolism

Figure 4 and supplementary figures S9 and S10, Supplementary Material online show the ability of each ancestral and extant form to metabolize E2. Most ancestors produced 2-hydroxyestradiol (2-OH-E2) with the exception of N9\_1ABCD, N18\_1D, and N52\_1A\_Actino. Among the ancestral forms, the N128\_1A1\_Mam, N224\_1A4\_Aves, and N127\_1A2\_Mam ancestors produced the highest amounts, whereas older ancestors showed less activity. 4-Hydroxyestradiol (4-OH-E2), the principal metabolite of hCYP1B1, was produced by several ancestral forms including the CYP1B, CYP1A1, CYP1A2, and CYP1A4 ancestors, albeit to a much smaller extent than hCYP1B1. Estriol (16 $\alpha$ -OH-E2) was produced by N25\_1C\_Actino and N20\_1B alone. An unidentified metabolite (denoted M9.6) with a retention time (RT) of 9.6 min was produced by all forms except those along the CYP1D and recent CYP1A2/1A5 lineages, that is, hCYP1A2 and tCYP1A5 and the ancestors N127\_1A2\_Mam, N223\_1A5\_Aves, and N18\_1D. Likewise, unidentified metabolites eluting at 8.5 and 9 min were also produced by hCYP1A1 and hCYP1B1 and in much smaller amounts by the ancestors along the corresponding CYP1A1/4 and CYP1B lineages but were absent from the CYP1A2/5, CYP1C, and CYP1D lineages.

### Testosterone Metabolism

Two main metabolites, 6 $\beta$ -hydroxytestosterone (6 $\beta$ -OH-T) and 16 $\alpha$ -hydroxytestosterone (16 $\alpha$ -OH-T) were formed from testosterone (fig. 4, supplementary figs. S11 and S12, Supplementary Material online). Most forms produced 6 $\beta$ -OH-T, with the highest amounts



**FIG. 1.** CYP1 ancestral nodes reconstructed and correlation of thermostability with evolutionary age. Main figure: The phylogenetic relationship between sequences used to reconstruct CYP1 family ancestors from the CYP1A, CYP1B, CYP1C, and CYP1D subfamilies. Sequences were collected from the NCBI and UniProt databases by BLAST searching for proteins with at least 40% sequence similarity to characterized forms, and aligned with selected outgroup sequences (human CYP2A6 [NP\_000753.3] and mouse CYP2E1 [NP\_067257]) using MAFFT. A maximum-likelihood tree was generated using the JTT evolutionary model, and ancestors were inferred using a joint reconstruction approach in GRASP (Foley 2022). Individual nodes selected for reconstruction are shown as circles, with names based on the node number and sequences from which they are derived. “CYP1-like” sequences are those from the basal chordate *Ciona intestinalis* determined by the NCBI sequence annotation algorithm to likely be CYP1 genes. Inset: The variation in the  $^{60}T_{50}$  values for CYP1 ancestors and extant forms with branch length (a measure of evolutionary age). Each subfamily is colored separately (CYP1A: red, CYP1B: orange, CYP1C: green, CYP1D: blue) and lines connect data points on the same evolutionary branch. Data represent the mean  $\pm$  SD ( $n = 2-3$   $^{60}T_{50}$  determinations). A general trend was observed toward increased thermostability correlating with greater evolutionary age.

produced by hCYP1A1, N20\_1B, N128\_1A1, and N224\_1A4 ancestors. Slightly fewer forms produced significant quantities of  $16\alpha$ -OH-T, with N20 and N98 CYP1B subfamily ancestors, N224\_1A4\_Aves and N25\_1C\_Actino generating the most. No  $16\alpha$ -OH-T was observed with extant hCYP1A2 or most of the CYP1A2/5 lineage, except for small amounts with tCYP1A5 and N17\_CYP1A. A metabolite was observed to elute at 11.5 min (M11.5) that was identified as  $2\alpha$ - and/or  $2\beta$ -hydroxytestosterone ( $2\alpha/\beta$ -OH-T) based on comparison with the RTs of the authentic metabolites. This peak was seen at significant levels with all extant forms but only N25\_1C\_Actino and the CYP1A1/4 lineage among the ancestors (i.e., N17\_CYP1A, N128\_1A1 Mam, and N224\_1A4\_Aves). An additional metabolite with a RT of 6.2 min (M6.2) was produced by N20\_1B,

N98\_1B1\_Mam, and hCYP1B1 that was tentatively identified as  $15\alpha$ -hydroxytestosterone. Trace androstenedione formation was detected at statistically significant levels (exceeding apparent formation in the hCPR control) with all extant forms plus N224\_1A4\_Aves, N25\_1C\_Actino, and N128\_1A1\_Mam. A minor unidentified metabolite peak was also observed at a RT of 4.2 min (M4.2) for N224\_1A4\_Aves.

### Caffeine Metabolism

For the caffeine assay, three negative controls were used to assess apparent P450-independent product formation: minus cofactor (NADPH), time=0, and hCPR-only. Product formation was taken to be significant only if it was statistically higher in complete reactions than in all

**Table 1.** Expression and Thermostability of Ancestral and Extant CYP1 Enzymes.

Protein	Expression (nmol/l culture)	<sup>10</sup> T <sub>50</sub> (°C) <sup>a</sup>	<sup>60</sup> T <sub>50</sub> (°C) <sup>b</sup>
<i>Ancestral forms</i>			
N9_1ABCD	600 ± 400	70.8 ± 0.4	62.2 ± 0.7
N18_1D	620 ± 30	56.5 ± 0.8	47.1 ± 0.8
N17_1A	360 ± 90	52.1 ± 0.1	46 ± 1
N20_1B	150 ± 20	60.1 ± 0.8	54 ± 1
N52_1A_Actino(ptyerygii)	140 ± 10	52 ± 2	42 ± 1
N25_1C_Actino(ptyerygii)	380 ± 20	43 ± 2	41.8 ± 0.7
N224_1A4_Aves	200 ± 30	34.8 ± 0.2	36 ± 2
N223_1A5_Aves	510 ± 60	52.0 ± 0.9	44.3 ± 0.3
N128_1A1_Mam(malia)	300 ± 100	40 ± 5	36 ± 3
N98_1B1_Mam(malia)	1400 ± 600	47 ± 3	43 ± 2
N127_1A2_Mam(malia)	790 ± 20	47.5 ± 0.3	45.4 ± 0.4
<i>Extant forms</i>			
Turkey (t)CYP1A5	120 ± 20	45.5 ± 0.3	35.3 ± 0.4
Human (h)CYP1B1	470 ± 70	42 ± 2	37 ± 2
Human (h)CYP1A1	550 ± 20	46 ± 2	34 ± 2
Human (h)CYP1A2	330 ± 70	44.9 ± 0.2	39.8 ± 0.5

<sup>a</sup>T<sub>50</sub> indicates the temperature at which 50% of the hemoprotein remains intact after an incubation of 10 min.

<sup>b</sup>T<sub>50</sub> indicates the temperature at which 50% of the hemoprotein remains intact after an incubation of an hour.

three types of negative control. Among the ancestors, N128\_1A1\_Mam and N127\_1A2\_Mam form demonstrated the greatest overall activity toward caffeine (fig. 5, supplementary fig. S13, Supplementary Material online) consistent with the highest activity being seen with the extant mammalian 1A forms. Paraxanthine was the dominant metabolite, being produced by all forms except the oldest ancestor (N9\_1ABCD) and the extant tCYP1A5. Levels of the other metabolites were an order of magnitude or more lower. Theophylline formation was only statistically significant for N20\_1B, N98\_1B1\_Mam, N25\_1C\_Actino, N127\_1A2\_Mam, N224\_1A4\_Aves, hCYP1A1, and hCYP1A2. Similarly, trimethyluric acid (TMU) was found at significant levels only with the CYP1A1 and CYP1A2 ancestors and their extant counterparts, N224\_1A4\_Aves and the Actinopterygian ancestors N25\_1C\_Actino and N52\_1A\_Actino. Theobromine production was only statistically significant with hCYP1A1.

### FICZ Metabolism

FICZ was metabolized to various extents by the forms assayed here, resulting in the formation of at least eight metabolites, many of which have been at least partially resolved and identified by Bergander et al.: M1, 2,8-dihydroxy-FICZ (2,8-diOH-FICZ); M2, a combination of 2,10-diOH-FICZ and 4,8-diOH-FICZ; M3, a combination of 2-OH-FICZ and 8-OH-FICZ; indolo(3,2-*b*)carbazole-6-carboxylic acid (CICZ); and MB, an unidentified metabolite (Bergander et al. 2004; Bergander 2005; Wincent et al. 2009; supplementary table S1, fig. S14 and Detailed discussion of FICZ, Supplementary Material online). In addition, at least three other unknown metabolites were observed here, eluting at 3.1 min (M3.1), 3.5 min (M3.5), and 5.0 min

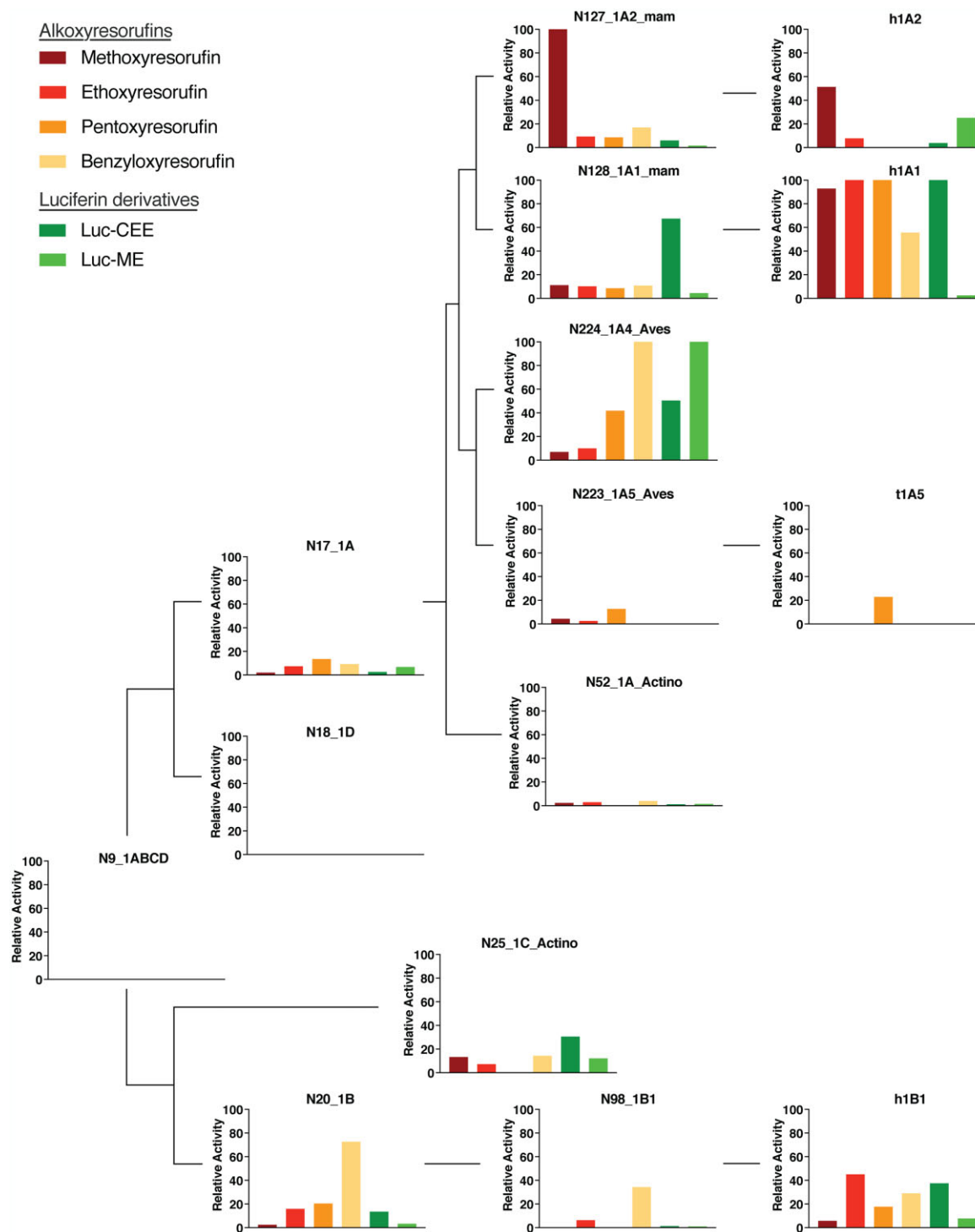
(M5.0). Metabolite identities were inferred by comparing the relative RTs (supplementary fig. S15, Supplementary Material online) to those observed by Bergander et al. (2004, 2005). Although most metabolites eluted within 0.5 min of the RTs Bergander et al. observed, M3.1, M3.5, and M5.0 eluted at least a full minute earlier than the earliest, previously characterized metabolite (2,8-diOH-CICZ; eluting at 6 min). M3.1 appears to consist of two peaks, however, due to the poor resolution of peaks this early in the chromatogram, both peaks were quantified as “M3.1.” The M3.1 peak was seen in all samples, however, it was also present in considerable amounts in the negative controls lacking NADPH or containing the reductase but no P450. M3.1 production was only significantly increased in complete reactions over NADPH- or P450-deficient controls in 10 of the 15 forms, suggesting that the putative metabolite coeluted with a contaminating peak that correlated with the concentration of membranes (supplementary fig. S16, Supplementary Material online).

The extant turkey CYP1A5 (t1A5) showed the highest overall activity of all forms assessed (fig. 6, supplementary fig. S16, Supplementary Material online), predominantly producing the mono-hydroxylated (monoOH) metabolite fraction M3 (where the order of other metabolites was M3 > M2 > CICZ > MB). The closest ancestor along the avian lineage, N223\_1A5\_Aves, was the most active ancestral form overall, showing a similar qualitative metabolite profile as tCYP1A5, with the difference that it produced much less of M3, CICZ, and MB. However, M1 (2,8-diOH-FICZ) and M3.1 were produced in trace amounts by N223\_1A5\_Aves, whereas they were not seen with tCYP1A5.

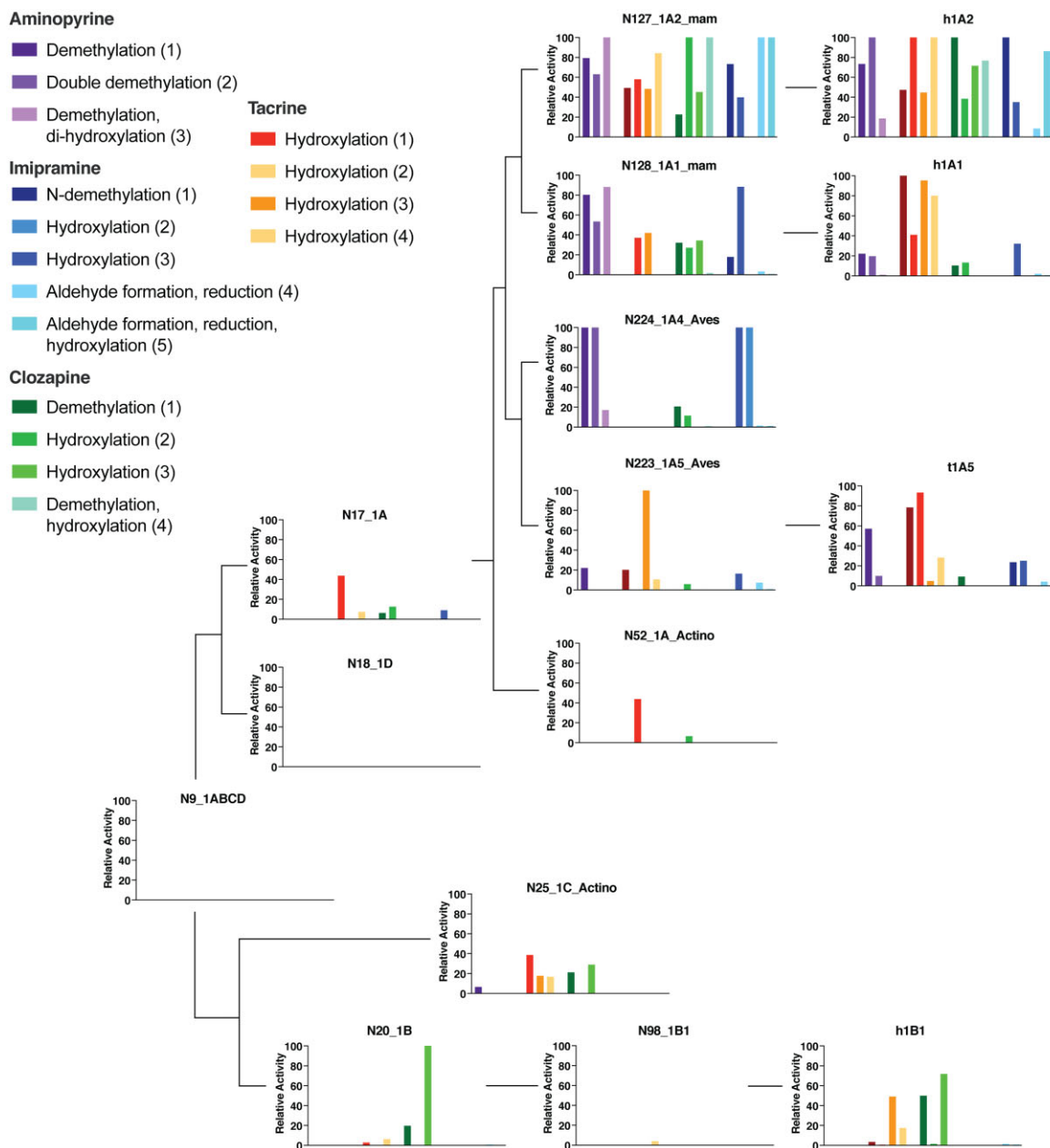
The monoOH M3 fraction, containing 2-OH-FICZ and/or 8-OH-FICZ, was the most frequently seen metabolite fraction, produced in statistically significant amounts by all forms except the mammalian CYP1A1, CYP1A2, and CYP1B1 ancestors and hCYP1A1. This was followed by the diOH M2 and M1 fractions. Assuming no major difference between the absorbance characteristics of the metabolites, for many forms, one or more diOH metabolites were produced in higher amounts than the monoOH metabolites, including some forms that failed to produce significant amounts of M3, suggesting rapid conversion of the primary to the secondary metabolites.

The unidentified metabolite fraction, M3.1, was the next most frequently produced. The early elution time of M3.1, M3.5, and M5.0 suggests increased polarity of these compounds and may reflect further metabolism of the diOH metabolites. The formation of these polar compounds by CYP1B forms is in line with a hypothesis proposed by Bergander et al. (2004) that CYP1B1 is responsible for the further metabolism of one or more diOH-FICZ metabolites. This is based on the observation that inhibition of CYP1B1 by 3-ethynylphenanthrene in microsomes from Aroclor-induced rat liver microsomes resulted in the accumulation of diOH products.

Although many forms overlapped in the metabolites they produced, there were distinct quantitative differences



**FIG. 2.** Relative activity of CYP1 ancestors toward 7-alkoxyresorufin and luciferin derivatives as marker substrates of extant human CYP1 forms. Plots illustrate the relative activity of ancestral and extant forms toward the dealkylation of six CYP1 probe substrates: 7-ethoxyresorufin, 7-methoxyresorufin, 7-benzyloxyresorufin, 7-pentoxyresorufin, luciferin-CEE, and luciferin-ME, normalized to the activity shown toward that substrate by the most active form. The background signal as indicated by the negative controls (hCPR-only in the case of the 7-alkoxyresorufin assay; [supplementary fig. S3, Supplementary Material](#) online) or without NADPH in the case of the luciferin assays ([supplementary fig. S4, Supplementary Material](#) online) was subtracted before calculating relative activity. Only data that are significantly different (two-sided Student's *t*-test;  $P < 0.05$ ) to negative controls are shown. Plots are superimposed on the evolutionary tree of the CYP1 family to show evolutionary relationships; however branch lengths are not to scale. All reactions were carried out at 37 °C. Alkoxyresorufin *O*-dealkylations assays were performed using membranes normalized to a P450 concentration of 5 nM, and a substrate concentration of 5  $\mu$ M in a total volume of 100  $\mu$ l. Luciferin-based assays were performed with bacterial membranes, using a P450 concentration of 10 nM and a substrate concentration of 100  $\mu$ M for luciferin-ME and 30  $\mu$ M for luciferin-CEE in a volume of 50  $\mu$ l. Data represent means of triplicates.

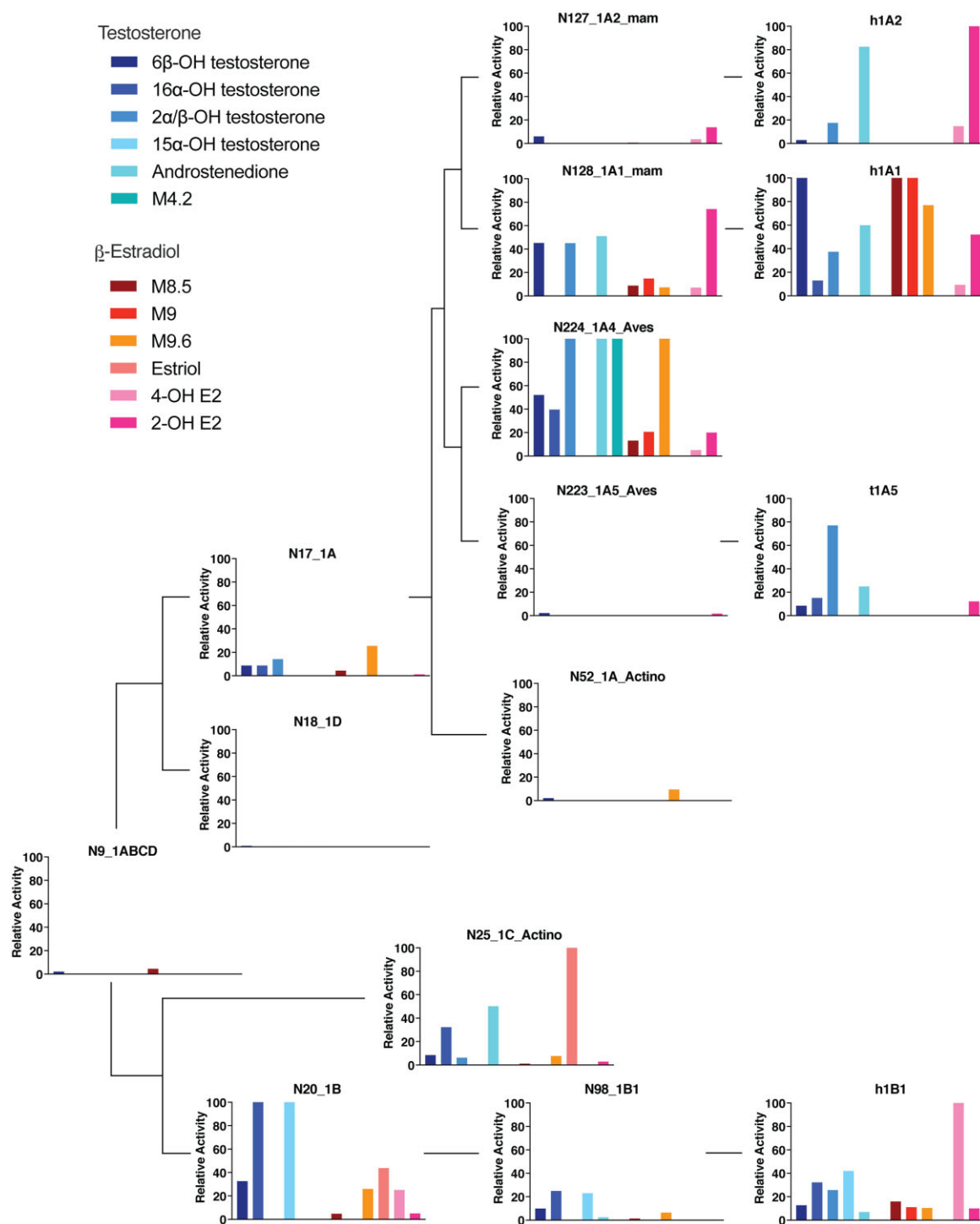


**FIG. 3.** Relative activity of CYP1 ancestors toward drug substrates of extant human CYP1 forms. Plots illustrate the relative activity of ancestral and extant forms toward the metabolism of the pharmaceutical compounds, clozapine, tacrine, imipramine, and aminopyrine normalized to the activity shown toward the substrate and metabolite in question by the most active form. The background signal as indicated by the hCPR-only negative controls ([supplementary figs. S5–S8, Supplementary Material](#) online) was subtracted before calculating relative activity. Only data that are significantly different (two-sided Student's *t*-test;  $P < 0.05$ ) to negative controls are shown. Plots are superimposed on the evolutionary tree of the CYP1 family to show evolutionary relationships; however, branch lengths are not to scale. All reactions were carried out at 37 °C in a total volume of 250  $\mu$ l, with a P450 concentration of 0.1  $\mu$ M and substrate concentrations of 50  $\mu$ M. Data represent means of triplicates. Numbers in the legend reference panel numbers in [supplementary figs. S5–S8, Supplementary Material](#) online.

among the metabolic profiles ([fig. 6](#)). Differences in regioselectivity were seen between the avian CYP1A4 and CYP1A5 lineages, with the CYP1A5 ancestor and extant form producing more M2 and the CYP1A4 ancestor producing more M1. N127\_1A2\_Mam is the major producer of M5.0 and forms in the CYP1B1 lineage produced the most of the very early eluting M3.1 and M3.5.

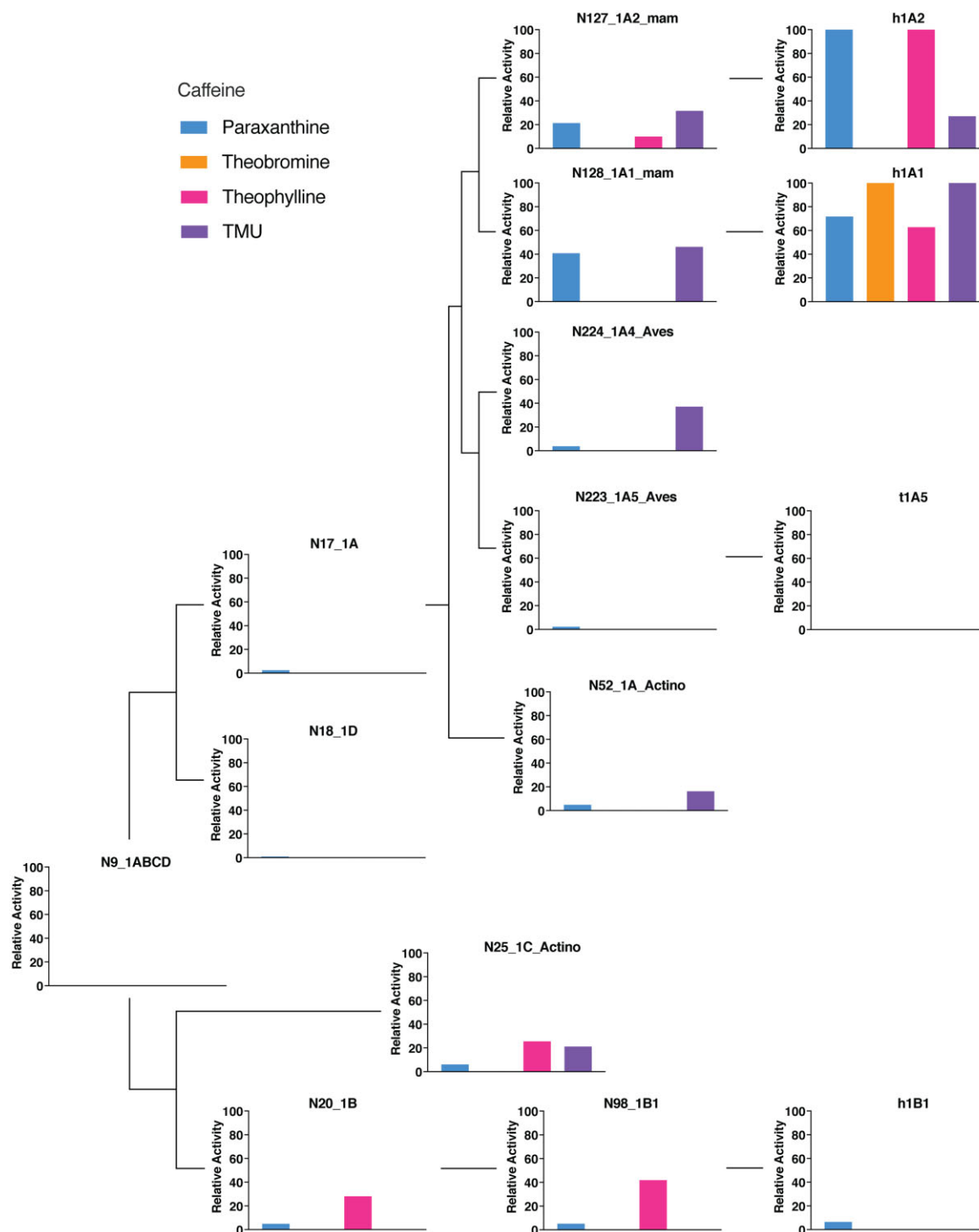
FICZ was also found to be converted to the corresponding carboxylic acid, CICZ. [Wincet et al. \(2009\)](#) have proposed that this reaction is carried out by aldehyde

oxidase (AOX1), however, results from the current study suggest that CICZ can also be produced by CYP1s. CICZ was produced by extant orthologous forms tCYP1A5 and hCYP1A2 (and at levels that failed to reach significance in their closely related ancestors N223\_1A5\_Aves and N127\_1A2\_Mam), as well as ancestors along the fish lineages, namely N52\_1A\_Actino and N25\_1C\_Actino. In general, CICZ makes up 1–3% of total metabolites seen for these forms, except for hCYP1A2, for which it makes up ~10% of metabolites.



**FIG. 4.** Metabolism of E2 and testosterone by ancestral and extant CYP1 forms. Plots illustrate the relative activity of ancestral and extant forms toward the metabolism of the steroids, E2 and testosterone, normalized to the activity shown toward the substrate and metabolite in question by the most active form. The background signal as indicated by the hCPR-only negative controls ([supplementary figs. S10 and S12, Supplementary Material](#) online) was subtracted before calculating relative activity. Only data that are significantly different (two-sided Student's *t*-test;  $P < 0.05$ ) to negative controls are shown. Plots are superimposed on the evolutionary tree of the CYP1 family to show evolutionary relationships; however, branch lengths are not to scale. E2 metabolism was assessed in reactions carried out for 60 min at 37 °C using 300  $\mu$ M E2 and 0.5  $\mu$ M P450 added in bacterial membranes. All metabolites were quantified using a standard curve prepared using the authentic metabolite, with the exception of the unidentified metabolite, which was quantified using the standard curve for 2-OH-E2. Testosterone metabolism was assessed in reactions carried out for 60 min at 37 °C with microsomal membranes in 250  $\mu$ l total volume using 100  $\mu$ M testosterone and 0.1  $\mu$ M P450 added in bacterial membranes. Testosterone was metabolized to five primary metabolites and one unidentified metabolite by the ancestors and extant forms. Metabolites were quantified using a standard curve prepared with authentic standards. Data represent means of triplicates.





**FIG. 5.** Caffeine metabolism by ancestral and extant CYP1 forms. The metabolism of caffeine to four metabolites (paraxanthine, theobromine, theophylline, and TMU) by the ancestors and extant forms is shown. Plots illustrate the relative activity of ancestral and extant forms normalized to the activity shown toward that substrate and product by the most active form, and are superimposed on the evolutionary tree of the CYP1 family to show evolutionary relationships; however branch lengths are not to scale. The background signal inferred from an hCPR-only standard curve ([supplementary fig. S13, Supplementary Material](#) online) were subtracted before calculating relative activity. Only data that are significantly different (two-sided Student's  $t$ -test;  $P < 0.05$ ) to negative controls are shown. Reactions were carried out for 60 min at 37 °C using 2 mM caffeine and 0.5  $\mu$ M P450 added in bacterial membranes. All metabolites were quantified using a standard curve prepared using the authentic metabolite. Data represent means of triplicates.

Lastly, a metabolite observed by Bergander et al. (MB), for which the structure remains to be characterized, was also produced here in minor amounts by tCYP1A5 as

well as its direct ancestor, N223\_1A5\_Aves, at levels that failed to reach statistical significance. This metabolite has been speculated to be an oxidized quinone structure

owing to the shift in its spectral properties at low pH (Bergander et al. 2004; Zhang et al. 2019).

## Discussion

Catalytic and substrate promiscuity is thought to underpin the evolution of new functions in enzymes (Copley 2012): the conventional view is that an enzyme with a potentially useful, promiscuous function can evolve by duplication and genetic drift under the influence of appropriate selection pressure into a novel enzyme with improved properties. However xenobiotic-metabolizing enzymes in animals represent an interesting case study; the survival of an animal requires maintenance of substrate promiscuity in detoxifying enzymes, that is, the versatility to deal with a wide and variable range of potentially toxic xenobiotics, along with the ability to catalyze an effective but general mechanism for detoxification. This is exemplified by animal P450s from the CYP1-3 families: in each animal, a finite set of catalytically versatile P450s with broad and overlapping substrate specificities collectively protect against a myriad of xenobiotics by adding an oxygen to the xenobiotic, thereby increasing its water solubility and facilitating excretion.

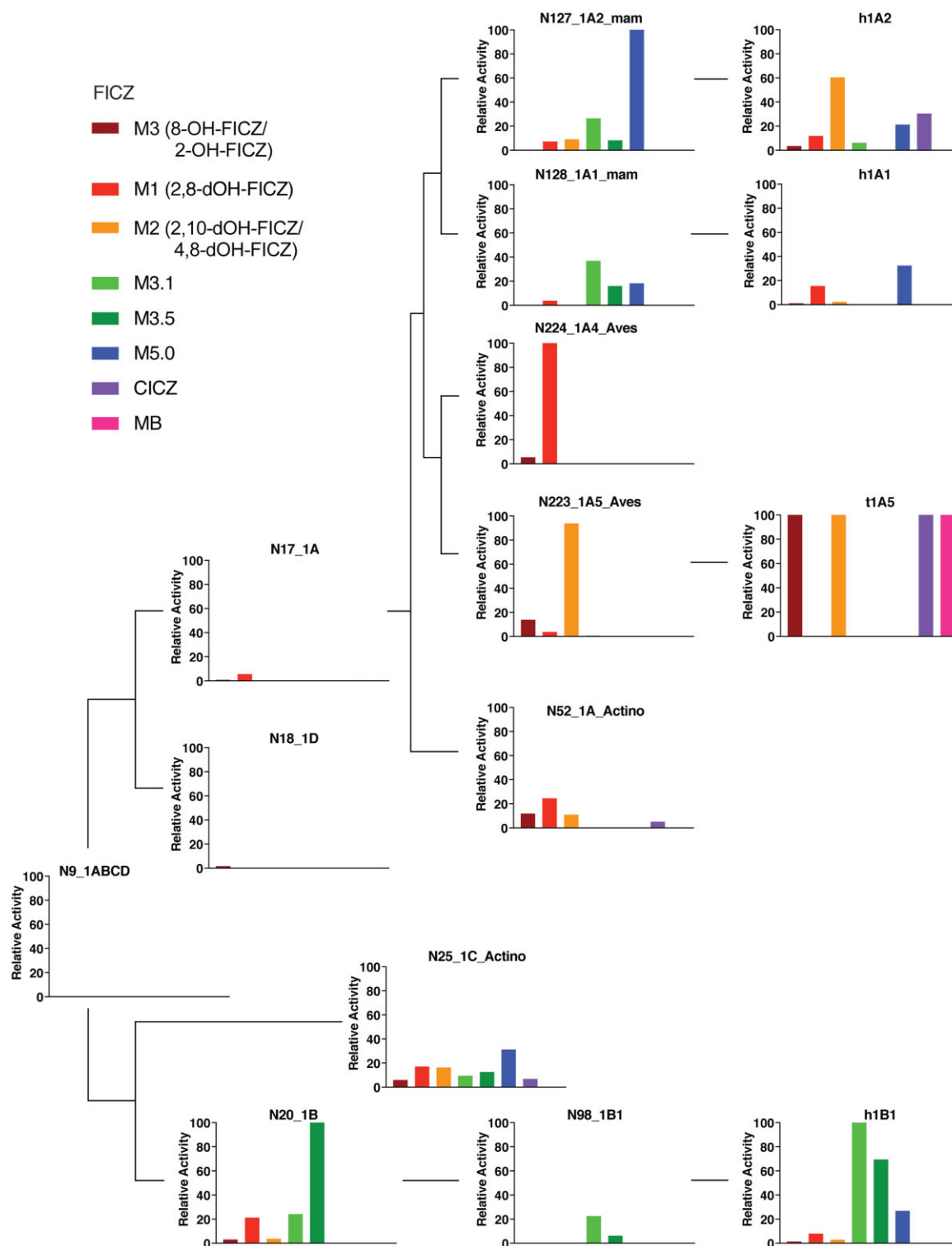
The CYP1 family is evolutionarily complex, littered with gene duplication and deletion events. The presence of CYP1-like genes in tunicates (sea squirts) and other invertebrates suggests the presence of a CYP1 ancestor before the Cambrian explosion (Goldstone et al. 2007). However, due to the limited sequence information available from invertebrate animals, the ancestral protein resurrection performed here was restricted to the vertebrate CYP1 forms, which span four subfamilies (CYP1A, CYP1B, CYP1C, and CYP1D). Two distinct subclades exist in the vertebrate CYP1 phylogenetic tree, CYP1A/D and CYP1B/C having emerged as the result of a gene duplication event. Subsequent gene duplications gave rise to two CYP1A genes in mammals and birds, and two CYP1C genes in fish (Goldstone and Stegeman 2006; Goldstone et al. 2007; Jönsson et al. 2011). Although the CYP1A and CYP1B isoforms appear to be present in all vertebrate species characterized to date, CYP1C genes have been lost in mammals, and CYP1Ds are absent in avian species and some mammals including humans (Kawai et al. 2010; Jönsson et al. 2011). This complex evolutionary history combined with limited knowledge of the non-human forms makes it difficult to predict the ancestral function of the CYP1 clade, motivating the current study in which eleven ancestors from the vertebrate CYP1 family were resurrected and characterized. Ancestral nodes were chosen to reflect major evolutionary groups and different stages of CYP1 evolution, with a bias toward the mammalian CYP1 branches for which more is known concerning the properties of the extant forms.

The ancestral enzymes were expressed at relatively high levels in *E. coli* (table 1), with one form (N98\_1B1\_Mam), reaching micromolar concentrations in culture, a significant increase compared with the values routinely obtained

for the coexpression of animal P450s with CPR in *E. coli* (~50–300 nmol/l culture). Although high expression levels may result from optimization of the nucleotide sequence for expression in *E. coli* (e.g., codon optimization combined with mRNA optimization and truncation of the N-terminal membrane anchor), the increased stability of the ancestors may also have potentiated their expression. Most ancestors were found to be significantly more thermostable than their extant counterparts, with the oldest ancestor demonstrating an increase in  $^{60}T_{50}$  of ~28 °C above the most labile extant form, an increase that is comparable to that seen previously for ancestors of the CYP3 (Gumulya et al. 2018), CYP2D (Gumulya et al. 2019), and other CYP2 enzymes (Thomson 2021).

It has been argued that ML methods are biased toward selecting consensus residues (Williams et al. 2006), and that since the consensus choice at a given position is likely to be the most stable, this may cause an artifactual overestimation of the stability of ancestral proteins. We, therefore sought to quantify to what extent the inferred ancestors overlapped with consensus sequences. The inferred ancestral sequences are quite distinct from the consensus sequences, at around only ~64–78% identical in the older ancestors which are the most thermostable, for example, the consensus equivalent of N9\_1ABCD differs by 169 residues throughout, with a sequence similarity of ~75% and an identity of 65% (supplementary table S2, Supplementary Material online), approximately the same degree of identity as seen between the ancestral and extant sequences. Moreover, the extent of the difference in thermostability is more marked than expected from a bias in the inference (Wheeler et al. 2016). We cannot discount the possibility that a single residue or a small set of residues, that happen to be identical between the ancestral and consensus forms, exerts an overwhelming effect on the relative thermostability. However the effect of sequence elements on thermostability has been shown to be predominantly additive (Heinzelman et al. 2009) and the gradual trend we see in the lines on the age vs thermostability plot (see below) supports this conclusion: a general correlation was seen between increased thermostability and evolutionary age, a trend we have also observed for the CYP2 ancestors (Thomson 2021).

The oldest ancestors reconstructed here would have existed hypothetically in the earliest vertebrates, ~450–500 Ma, before the invasion of land by animals and at a time when ocean temperatures probably differed by no more than ~10–20 °C to those today. It is unlikely that such ambient temperatures per se would have exerted sufficient selection pressure to lead to such substantial thermostability as that seen in the oldest ancestors. However, P450s are found throughout all domains of life, and therefore are likely to have existed in both the last universal common ancestor (LUCA) and last eukaryotic common ancestor, organisms that may have existed in primordial hot marine environments. We hypothesize that the primordial forms of P450s in LUCA or its primitive descendants would have required a robust structure to remain



**FIG. 6.** FICZ metabolism by ancestral and extant CYP1 forms. The metabolism of FICZ to five previously characterized and three unknown metabolites by the ancestors and extant forms is shown. Plots illustrate the relative activity of ancestral and extant forms normalized to the maximal activity shown toward that substrate by any form, and are superimposed on the evolutionary tree of the CYP1 family to show evolutionary relationships; however, branch lengths are not to scale. The background signal as indicated by reactions without added NADPH or quenched at 0 min ([supplementary fig. S16, Supplementary Material](#) online) was subtracted before calculating relative activity. Only data that are significantly different (two-sided Student's *t*-test;  $P < 0.05$ ) to negative controls are shown. Reactions were prepared and carried out in the dark for 60 min at 37 °C using 25  $\mu$ M FICZ and 0.5  $\mu$ M P450 added in bacterial membranes. Data represent means of triplicates.

functional at the elevated temperatures present at various time points and/or locations at which the ancestors of bacteria, archaea, and eukaryotes first evolved. However, with

the cooling of the oceans ([Tartèse et al. 2017](#)), this selective pressure would have been alleviated over time. As P450s have evolved, neutral drift and selection for new functions

may have resulted in the acquisition of destabilizing mutations that were tolerated under the prevailing, cooler conditions. Accumulation of mutations is more likely to lead to the loss of a property than its maintenance, if it is not under sufficient selection pressure. Therefore, in the absence of sufficiently strong selection pressure, genetic drift would lead to a gradual erosion of this property (thermostability). It is not necessary to invoke any deeper biological driver of stability or instability going forward in time to explain this trend, rather just an inevitable loss of this property (stability) by neutral drift in the absence of selection pressure for its maintenance. This alleviation of selection pressure for thermostability may not have been “linear” or indeed consistent over different phylogenetic groups, depending on particular environments.

Extant CYP1 forms, especially those from the CYP1A subfamily, show activity toward often broad and overlapping ranges of xenobiotic substrates as well as certain endobiotics such as steroids (reviewed by [Guengerich 2015](#)). The substrate specificity of CYP1B forms appears to be more restricted, including the important endogenous substrates, E2 and retinoids, but also certain procarcinogens including polycyclic aromatic hydrocarbons in smoke and heterocyclic amines formed from protein pyrolysis ([Shimada et al. 1997, 2001](#)). The fact that a single CYP1B form exists in most mammals also hints at a possible important conserved physiological role, and CYP1B1 mutations have been associated with congenital glaucoma, although the mechanism is not yet understood ([Stoilov et al. 1997, 1998](#)). The substrate specificities of the CYP1C and CYP1D forms, mostly studied in fish, are less clear. Zebrafish (z)CYP1C1 and zCYP1C2 have both shown activity toward marker substrates including alkoxyresorufins and coumarins, at similar or lower levels to the zCYP1A and zCYP1B1 genes ([Scornaienchi et al. 2010](#)). zCYP1D1 showed activity toward alkoxyresorufins but not alkoxycoumarins, and poorer affinity toward benzo(a)pyrene than the CYP1C forms; both zebrafish CYP1C and CYP1D forms metabolized testosterone but with different regioselectivity ([Stegeman et al. 2015](#)). The results obtained here with the ancestors are in general agreement with the previous studies on extant forms in that the CYP1A forms showed activity toward the broadest range of substrates. CYP1B and CYP1C forms were active toward a more restricted set of substrates and the CYP1D ancestor showed little or no activity toward the substrates examined.

Animal xenobiotic-metabolizing P450s, such as the CYP1A subfamily forms, have been hypothesized to have diversified in response to animal–plant warfare ([Gonzalez and Nebert 1990](#)). Thus, it was of interest to determine how the activity of the resurrected CYP1 forms varied with evolutionary age. Was the ultimate CYP1 ancestor a form with a defined physiological function that was then co-opted by evolution to metabolize the increasingly diverse range of xenobiotics generated by secondary metabolism in plants, as well as, most recently, combustion products resulting from the use of fire by humans?

Alternatively, did the original CYP1 forms show broad activities toward a diverse range of substrates, with specific subfamilies evolving more defined roles toward endogenous compounds, such as E2 in the case of the CYP1B subfamily?

The comparisons made here can only be regarded as semiquantitative, since rates were measured at a single substrate concentration and a detailed analysis of the kinetics of each enzyme–substrate combination was not performed. Nevertheless a general trend was evident in that one or more CYP1A lineages consistently showed an increase in activity compared with the oldest forms toward a wider range of the xenobiotic substrates studied here (i.e., alkoxyresorufins, substituted luciferins and drugs) of the extant CYP1 enzymes ([figs. 2 and 3](#); [supplementary figs. S3–S8, Supplementary Material online](#)). Typically the mammalian CYP1A extant forms showed the highest activity toward xenobiotics followed by the younger CYP1A ancestors, the CYP1B and CYP1C lineage forms and finally the oldest forms. This may reflect some degree of bias in the choice of substrates, as the younger mammalian CYP1A ancestors are closer evolutionarily to the extant forms that are known to carry out these reactions. Analysis of a wider range of substrates is needed to draw definitive conclusions about the relative substrate promiscuity of ancestral and extant forms. Alternatively, we cannot exclude the possibility that errors in the ancestral inference related to the tree or alignment have led to a relative loss of activity in the older forms, especially N9\_1ABCD, since the uncertainty in the prediction increases going back in time. (An analysis of the effects of differences in the tree topology, outgroup, and evolutionary model is presented in the [Supplementary Material online](#).) However when this was assessed in a previous study by creation of pool of alternative ancestors ([Gumulya et al. 2018](#)), minor quantitative differences were observed rather than all-or-none alterations in activity.

E2 is an important physiological CYP1 substrate in mammals, with human CYP1 forms, along with CYP3A4, producing the catechol estrogens, 2-OH-E2 and 4-OH-E2 ([Aoyama et al. 1990](#); [Waxman et al. 1991](#); [Kerlan et al. 1992](#); [Spink et al. 1992](#); [Lee et al. 2003](#)). E2 was metabolized to a greater extent by the extant CYP1As and younger, mammalian and avian ancestors than older forms ([fig. 4](#); [supplementary figs. S9 and S10, Supplementary Material online](#)). Although hCYP1A1 and hCYP1A2 are known to produce mainly 2-OH-E2, hCYP1B1 preferentially produces 4-OH-E2. hCYP1B1 is implicated in cancer, as 4-OH-E2 reacts readily with oxygen to form the DNA-reactive *o*-quinone as well as redox cycling to generate reactive oxygen species ([Liehr et al. 1986](#)). In contrast to the strong regioselectivity of hCYP1B1, the CYP1B ancestors each produced low levels of both 2- and 4-OH products. [Nishida et al. \(2013\)](#) reported that the human, dog, and monkey CYP1B1 enzymes showed preference for 4-OH, whereas mouse and rat isoforms produced mainly 2-OH-E2. The authors suggested that a single residue

(Val395 in hCYP1B1) is primarily responsible for regioselectivity for the 4-OH product, with mutation to Leu inverting the preference of the enzyme. However, the determinants of 2-OH versus 4-OH selectivity may be more complex, as both N20\_1B and N98\_1B1\_Mam contain Val at this position but fail to show such a strong preference for 4-hydroxylation. These results suggest that the relative regioselectivity of certain extant CYP1B1 isoforms is a recently evolved phenomenon, and that the effect of the residue at the position equivalent to 395 in the human form is dependent on the sequence and/or structural context presented by other parts of the CYP1B protein.

The major metabolite for the CYP1A extant forms, 2-OH-E2, was produced primarily by the younger CYP1A ancestors (but much less by N223\_1A5\_Aves and N52\_1A\_Actino) but in smaller amounts by several other forms. Estriol (16 $\alpha$ -OH-E2) was produced exclusively by N25\_1C\_Actino and in smaller amounts by N20\_CYP1B, aligning with previous findings that zCYP1C1 and zCYP1C2 produce this metabolite (Scornaienchi et al. 2010). An unidentified metabolite (E9.6) was produced by several ancestors and extant forms, particularly N224\_1A4\_Aves and hCYP1A1. This may be the 15 $\alpha$ -hydroxy derivative, known to be produced by hCYP1A1 (of which CYP1A4 is the avian homolog) (Lee et al. 2003). This assignment could not be confirmed since the pure metabolite is not commercially available.

Testosterone is a canonical P450 substrate metabolized to some degree by numerous forms (Waxman et al. 1991). It has been shown to be metabolized by human CYP1A1 and CYP1B1, as well as fish CYP1Cs and CYP1D1 (Crespi et al. 1997; Schwarz et al. 2000; Stegeman et al. 2015). Most of the ancestral and extant forms tested here produced 6 $\beta$ -OH-T as the main metabolite (fig. 4, supplementary figs. S11 and S12, Supplementary Material online). This metabolite is generally considered to be a marker of hepatic CYP3A4 activity in humans (Waxman et al. 1991) but the 6 $\beta$ -position is relatively sterically unhindered and chemically favored due to the neighboring allylic double bond, which may make it more easily targeted by a promiscuous monooxygenase. Notably, N9\_1ABCD, which had not demonstrated statistically significant activity toward canonical CYP1 substrates, metabolized testosterone to 6 $\beta$ -OH-T. Although this activity is relatively low, it confirms that the oldest N9\_1ABCD ancestor is functional, but suggesting that the ancestral function of CYP1 may have differed substantially to that of the modern-day forms.

Overall, the most active forms toward testosterone were from the CYP1B, CYP1A1/1A4, and CYP1C lineages, producing both 6 $\beta$ -OHT and 16 $\alpha$ -OHT. The CYP1B and CYP1C forms share a common ancestor suggesting an enhancement in activity toward testosterone may have occurred early in the common CYP1B/CYP1C branch. Interestingly, the unidentified metabolite eluting at 4.2 min was generated exclusively by the avian CYP1A4 ancestor; it is possible that this metabolite matches the unidentified metabolite produced by zebrafish CYP1D1 but further work is needed to confirm the identities of these

metabolites (fig. 4, supplementary figs. S11 and S12, Supplementary Material online).

Metabolism of the CYP1A xenobiotic substrate, caffeine, showed a similar trend to E2, with the extant mammalian CYP1A forms showing the greatest levels of activity (fig. 5, supplementary fig. S13, Supplementary Material online) followed by the CYP1A1, CYP1A2, CYP1A4, CYP1B, and CYP1C subfamily ancestors; the N9\_1ABCD, N18\_1D, and oldest CYP1A forms were markedly less active. Caffeine biosynthesis only recently evolved in flowering plants through convergent evolution across three different species (coffee, tea, and cacao; Denoeud et al. 2014). Although these plants are not necessarily common components in the diets of most mammals (i.e., other than modern humans), the ability to metabolize caffeine may reflect activity toward structurally related, plant, secondary metabolites. We hypothesize that the CYP1A progenitor possessed some inherent ability to clear caffeine and related compounds to a modest degree, and that this activity may have been subject to positive selection upon diversification of related plant secondary metabolism pathways.

The tryptophan photoproduct FICZ is the most potent AhR agonist discovered and is also a possible ancestral CYP1 substrate (Bergander et al. 2004; Wincent et al. 2009; Rannug and Rannug 2018). The AhR mediates the up-regulation of a battery of genes including CYP1s. The induction of CYP1 enzymes has been reported in many species including mammals, fish, and birds, and many high-affinity xenobiotic ligands of this receptor, such as certain polyaromatic hydrocarbons as well as the high-affinity endogenous ligands, FICZ, and indolo[3,2-b]carbazole, have been identified as CYP1 substrates, thereby up-regulating their own clearance (Bergander et al. 2004; Uno et al. 2004). The toxicity of FICZ (Farmahin et al. 2016; Jönsson et al. 2016; Wincent et al. 2016) and its photo-oxidation products (Park et al. 2015; Zhang et al. 2019) may have acted as a selection pressure that facilitated the evolution of CYP1 forms with enhanced activity toward FICZ clearance. Consistent with this theory, all the ancestors exhibited activity toward FICZ to various degrees, producing a wide range of metabolic profiles (fig. 6, supplementary figs. S15 and S16, Supplementary Material online), including putative novel metabolites not identified in several previous studies (Bergander et al. 2003, 2004, 2005; Wincent et al. 2009).

The oldest ancestor exhibited only trace activity toward M3 production, as did the individual subfamily ancestors N18\_1D and N17\_1A (supplementary fig. S16, Supplementary Material online). However, N20\_1B and multiple younger ancestors (e.g., N223\_1A5\_Aves, N224\_1A4\_Aves, N127\_1A2\_Mam) showed specialization toward the production of particular monoOH or diOH metabolites (fig. 6, supplementary figs. S15 and S16, Supplementary Material online). This indicates that the common ancestors possessed some inherent ability to metabolize FICZ, and the evolution of new CYP1 genes and species resulted in a diversification of FICZ metabolism,

leading to an array of different regioselectivities in the younger ancestors and extant forms, exemplified in the CYP1A and CYP1B lineages (for further discussion of the FICZ results, please refer to the [Supplementary Material](#) online). Overall, the trends identified here, based on a necessarily constrained set of CYP1 substrates, suggest that the primordial CYP1 ancestor had a restricted substrate specificity compared with that of the extant forms, and that evolutionary pressures, especially on the mammalian CYP1A lineage, have led to an increase in the overall turnover and range of substrates. Further studies are underway toward profiling these forms against large compound libraries to define a pharmacophore for each form. However, the picture that is emerging for the evolution of CYP1 family differs from that reported for the CYP3 ([Gumulya et al. 2018](#)) and more recent branches of the CYP2 family ([Thomson 2021](#)), in which the most ancient ancestors show comparable activity to the extant forms toward a diverse set of substrates, suggesting a role in xenobiotic metabolism that has been conserved throughout their evolution. Rather, we hypothesize that the original function of the CYP1 ancestors was more likely to be the metabolism of a particular, endobiotic substrate, rather than structurally diverse xenobiotics. This putative endogenous substrate is yet to be identified but may have been FICZ or a sterol (based on the low but significant activity of the oldest forms toward FICZ and testosterone). Both targeted and unbiased metabolomics studies are underway to investigate possible substrates of the oldest CYP1 ancestors.

We cannot exclude the possibility that the CYP1 ancestors (or indeed the nonhuman extant tCYP1A5) couple poorly with the extant human reductase used here to support catalytic activity. An ancestral CPR from a closer evolutionary stage to the older ancestors failed to support higher levels of turnover by any of the older ancestors. Rather, in two of three cases, turnover of testosterone instead decreased compared with hCPR, and with the other (CYP1ABCD\_N9), the regioselectivity was shifted slightly toward more minor metabolites ([supplementary fig. S17, Supplementary Material](#) online). No difference was seen in the turnover of alkoxyresorufins between reactions containing hCPR and the ancestral CPR (data not shown). Further work is in progress to identify the cognate ancestral reductases from various points in the CYP1 evolution and study their interactions with the ancestral P450s reported here. We also cannot exclude the possibility that the activity observed may change with different amounts of CPR; however such an effect is more likely to cause a minor quantitative effect rather than reveal latent activity in the oldest ancestors ([Yamazaki et al. 2002](#)).

Using cumene hydroperoxide as an oxygen surrogate led to increased activity toward testosterone by the oldest ancestors suggesting that they could be supported by lipid peroxides ([supplementary fig. S17, Supplementary Material](#) online). However still no activity could be detected toward the typical xenobiotic substrates, 7-ethoxy- and 7-pentoxoresorufin in the presence of cumene

hydroperoxide. This observation further supports the contention that the oldest ancestors are functional on at least some substrates in the right circumstances, but raises the intriguing possibility that they relied on lipid peroxides rather than a redox partner.

In general, there appeared to be an inverse correlation between thermostability and substrate promiscuity; the older, more stable ancestors were less active toward typical modern CYP1 substrates than the younger and extant forms. This could indicate a trade-off between enzyme stability and promiscuity. The acquisition of mutations to modulate enzymatic function is thought to be generally destabilizing ([Bloom et al. 2006](#); [Modarres et al. 2016](#)) since amino acid changes have more chance of compromising than stabilizing a structure ([Guo et al. 2004](#)). Although the younger forms have acquired mutations allowing them to move into previously inaccessible reaction space, this versatility may come at the cost of reduced stability, a correlation that is evident from the study of homologous mesophilic versus thermophilic proteins ([Siddiqui 2017](#)).

It could be hypothesized that older forms are more rigid, which impairs their activity. However previous work with a parallel family of xenobiotic-metabolizing P450s has revealed an ancestor that, whereas much more thermostable than extant forms, was equally promiscuous with comparable kinetic properties to the extant form. P450s conform to the model of a highly innovable enzyme advanced by [Toth-Petroczy and Tawfik \(2014\)](#), that is, the core of the fold holding the heme can be rigid, whereas the structural elements responsible for interacting with substrates are flexible. It is well known that elements such as the F and G helices and intervening loop and the B–C loop that open and close as the structure “breathes” alter conformation in a single P450 binding different substrates ([Ekroos and Sjögren 2006](#)). Therefore it is more likely that the failure to observe a wide range of activities by the older forms is instead due to the substrates selected for analysis being biased toward substrates of extant forms. Confirmation of this hypothesis awaits NMR structures or other experimental evidence of the relative flexibility of different parts of the structure.

Apart from elucidating the functional evolution of this important P450 family, this study has implications in the use of P450s for potential commercial applications. Industrial conditions often require enzymes to have high stability and longevity, so that reactions can be performed at higher temperatures and/or for longer periods of time ([Vieille and Zeikus 2001](#)). CYP1 enzymes show useful activities toward xenobiotics including drugs and environmental toxins making them appropriate starting materials for development of biocatalysts for applications such as the production of pharmaceutical metabolites and the bioremediation of environmental contaminants ([Sakaki et al. 2013](#); [U.S. Food and Drug Administration: Center for Drug Evaluation and Research \(CDER\) 2016](#)). Many of the ancestral enzymes characterized herein show both useful substrate promiscuity and enhanced thermostability suggesting they may have potential applications in

producing high amounts of CYP1A metabolites for pharmaceutical testing (U.S. Food and Drug Administration: Center for Drug Evaluation and Research (CDER) 2016). Current studies are directed toward the further development of these ancestors for industrial applications.

## Materials and Methods

### Materials

DNA sequences and oligonucleotides were synthesised by Thermo Fisher Scientific (Scoresby, VIC, Australia) and Integrated DNA Technologies (Baulkham Hills, NSW, Australia), respectively. Cloning reagents (restriction endonucleases, ligases, buffers, and Gibson Assembly kits) were from New England Biolabs Inc. (Ipswich, MA, USA). Media components (bactopeptone, bactotryptone, yeast extract) were obtained from Becton Dickinson Pty Ltd. (North Ryde, NSW, Australia). Luciferin substrates were sourced from Promega (Sydney, NSW, Australia). Testosterone and E2 metabolites were purchased from Steraloids Inc. (Newport, MI, USA). FICZ was sourced from Syntastic AB (Stockholm, Sweden). Other chemicals, including substrates and standards, were purchased from Sigma-Aldrich (Castle Hill, NSW, Australia) or from local suppliers at the highest grade commercially available.

Expression vectors for hCYP1A1, hCYP1A2, and hCPR were constructed in previous studies (Guo et al. 1994; Sandhu et al. 1994; Parikh et al. 1997) and made available by Professor F.P. Guengerich (Vanderbilt University, Nashville, USA). The expression vector for tCYP1A5 was kindly donated by Dr Roger Coulombe (Univ. Utah) (Yip and Coulombe 2006) and that for hCYP1B1 was from a previous study (Shimada et al. 1998). pCW/pGro7 was obtained from Professor K. Nishihara (HSP Research Institute, Kyoto, Japan).

### Methods

#### *Ancestral Sequence Reconstruction*

Extant CYP1 amino acid sequences were obtained from the NCBI and UniProt databases by BLAST searching and a sequence identity cut-off of 40%. CD-HIT (Li and Godzik 2006) was used to remove redundant sequences from the combined BLAST results. The remaining sequences were aligned using MAFFT (Kato et al. 2002) and manually curated to remove those containing apparent anomalies such as sequencing errors, miscalled exon/intron boundaries, partial ORFs, and internal gaps/insertions. The remaining 471 sequences were then aligned with selected outgroup sequences (Human CYP2A6, NCBI NP\_000753.3, and Mouse CYP2E1, NCBI NP\_067257) and the resulting alignment was used to generate a maximum-likelihood tree in MEGA (Kumar et al. 1994) using the Jones-Taylor-Thornton (JTT) evolutionary model (Supplementary Material online—phylogenetic tree). GRASP (Foley 2022) was used to infer the maximum-likelihood joint and marginal ancestors for this tree and

alignment using the JTT evolutionary model, and ancestral sequences from the joint reconstruction at the nodes shown in figure 1 were selected for synthesis. The robustness of the inferences to choice of outgroup, evolutionary model, and phylogeny was tested as described in the Supplementary Material online.

#### *Optimization and Expression of Ancestral Sequences*

N-terminal sequence modifications were applied to improve recombinant expression based on the MAKKTSS modification originally derived from the truncation of CYP2C3 to improve recombinant expression (supplementary table S3 and supplementary alignment, Supplementary Material online) (von Wachenfeldt et al. 1997). Between 23 and 51 N-terminal residues were removed from each sequence, before the KGK motif preceding the conserved PPGP region, and replaced with a sequence encoding MAKKTSS. Sequences were reverse translated and codon optimized for expression in *E. coli* by the GeneArt (Thermo Fisher, Scoresby, VIC, Australia) codon optimization algorithm. The N-terminal nucleotide sequence (−21 to +96 nucleotides with respect to the start site) was optimized to reduce the formation of mRNA secondary structure elements, using mRNA optimizer (Gaspar et al. 2013) and then by further manual fine-tuning of sequences. Potential secondary structure formation was visualized using NUPACK (Zadeh et al. 2011). Optimized sequences were synthesised (GeneArt, Thermo Fisher, Scoresby, VIC, Australia) with an additional 60 nucleotides upstream and downstream of flanking restriction sites from the target vector, then assembled (Gibson et al. 2009) into the pCW vector containing the ORF for hCPR via NdeI and XbaI sites. The three oldest ancestors, CYP1ABCD\_N9, CYP1D\_N18, and CYP1A\_N17 were also subcloned in bicistronic format with an alternative ancestral CPR (Harris 2020). Constructs were verified by Sanger sequencing at the Australian Genome Research Facility (Brisbane node) then transformed (Inoue et al. 1990) into DH5 $\alpha$  *E. coli* cells pretransformed with the pCW/pGro7 chaperone expression plasmid (Nishihara et al. 1998; Johnston et al. 2008). Expression, harvest, and preparation of membranes were carried out as described (Gillam et al. 1993). P450 content was measured using Fe<sup>2+</sup>.CO versus Fe<sup>2+</sup> difference spectroscopy as described (Johnston et al. 2008). CPR was quantified using the initial rate of cytochrome *c* reduction as described (Guengerich 1994).

#### *Thermostability Analysis*

A Bio-Rad MyCycler thermal cycler was used to heat samples to various temperatures (25–80 °C) for either 10 or 60 min, followed by incubation at 4 °C and room temperature for 5 min each. Residual P450 content was measured as described above, and compared with an unheated sample. Data were fitted to the equation for sigmoidal melting in GraphPad Prism to interpolate the  $T_{50}$  (Muslin et al. 2002).

### Catalytic Assays

All substrates were incubated with bacterial membranes containing coexpressed P450s and human CPR; additional incubations were undertaken with testosterone and CYP1ABCD\_N9, CYP1D\_N18, and CYP1A\_N17 coexpressed with the ancestral CPR. Alkoxyresorufin O-dealkylation assays were carried out using a P450 concentration of 5 nM. Reactions were carried out at 37 °C in 96-well plates in a total volume of 200 µl 100 mM potassium phosphate buffer (pH 7.4) with 5 µM substrate and initiated with an NADPH generation system (NGS, final concentrations: 250 µM NADP<sup>+</sup>, 10 mM glucose-6-phosphate, 0.5 U/ml glucose-6-phosphate dehydrogenase). The linear initial rate of product formation was determined by fluorescence on a CLARIOstar microplate reader with an excitation wavelength of 530 nm and an emission wavelength of 582 nm. The dealkylation of luciferin-CEE and luciferin-ME was carried out according to the manufacturer's protocol (Promega Corporation) using P450 concentrations as specified in the results for individual experiments and luminescence was detected using a MicroBeta microplate counter.

Reactions with clozapine, tacrine, imipramine, aminopyrine, testosterone, caffeine, E2, and FICZ were carried out in 250 µl incubations in 100 mM potassium phosphate buffer (pH 7.4) using a P450 concentration of 0.1 µM (clozapine, tacrine, imipramine, aminopyrine, testosterone) or 0.5 µM (caffeine, E2, FICZ). Substrates were used at a final concentration of 50 µM, except for testosterone (100 µM), caffeine (2 mM), E2 (300 µM), and FICZ (25 µM). Reactions with all substrates except FICZ were initiated by the addition of 25 µl of a 10-fold concentrated NGS. After incubation at 37 °C for 1 h with gentle shaking, 25 µl internal standard was added to incubations with testosterone, caffeine, and E2 (100 µM progesterone, 100 µM acetaminophen, 100 µM hydrocortisone, respectively), and reactions were immediately quenched by the addition of 1 ml ethyl acetate (or 500 µl acetonitrile for clozapine, tacrine, imipramine, and aminopyrine). Quenched samples were then centrifuged at 14,000× *g* for 15 min and the organic phase was extracted and desiccated before resuspension in mobile phase.

FICZ incubations were prepared and incubated in the dark due to the photolability of FICZ. FICZ reactions were initiated with the substrate and quenched after 1 h with 250 µl ice-cold acetone, then extracted twice with ethyl acetate as described above. Dried samples were resuspended in 60 µl methanol followed by an equal volume (60 µl) of water.

Testosterone (Hunter et al. 2011) and E2 (Suchar et al. 1995) metabolites were separated using established methods at 20 and 25 °C, respectively, with detection by absorbance at 247 nm (testosterone) and 280 nm (E2). Caffeine analysis was performed at 30 °C using a modified version of the protocol of Begas et al. (2015). The mobile phase consisted of 0.1% acetic acid:methanol, 88:12 v/v, delivered isocratically for 25 min at a flow rate of 1.3 ml/min, followed by a wash with 100% methanol for 10 min and recalibration to initial conditions for 20 min after each sample.

FICZ samples were analyzed by reverse phase high pressure liquid chromatography as previously described (Bergander et al. 2003) measuring UV absorbance at 386 nm.

Clozapine, tacrine, imipramine, aminopyrine metabolites were analyzed using liquid chromatography–mass spectrometry (LC–MS/MS). Samples (10 µl) were injected on a UPLC system coupled to a Sciex X500B QTOF mass spectrometer with an electrospray ionization source. Metabolites were separated at 45 °C on a Kinetix XB-C18 column (1.7 µm, 2.1 mm × 100 mm). Mobile phase was 0.1% formic acid in ultrapure water (A) and 0.1% formic acid in 90% acetonitrile, (B) at a flow rate of 0.5 ml/min. The following mobile phase gradients were used: 0–6.0 min (10–70% B), 6.0–6.7 min (70–90% B), 6.7–6.8 min (90–10% B) for imipramine and clozapine and 0–5.0 min (5–40% B), 5.0–6.7 min (40–90% B), 6.7–6.8 (90–5% B) for aminopyrine and tacrine. MS source parameters included curtain gas = 30 psi, Gas 1 = 50 psi, Gas 2 = 50 psi, temp = 500 °C, and ionspray voltage 5500 V with DP = 40 V, and CE = 5 V. MS was acquired across 100–1000 *m/z*, for 0.1 s. An Information Dependent Analysis experiment was performed on the top 10 candidates, with intensity >200 cps, across 50–1000 *m/z* for 0.025 s, with CE set to 35 V ± 15 V. MS instrument calibration was performed every six injections using the Sciex ESI Positive Calibration mix and autocalibration function in OS software. Data were processed in MetabolitePilot (Sciex). Proposed metabolites were reviewed manually.

### Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

### Acknowledgments

This work was supported by Australian Research Council Discovery Project Grants DP120101772 and DP160100865 and by AstraZeneca Innovative Medicines and Early Development, Cardiovascular, Renal and Metabolism, Gothenburg. The authors are indebted to Drs Jan Bergman, Agneta Rannug, and Ulf Rannug for advice regarding FICZ biology and metabolism, to Dr Jonas Hellberg (Chemtronica, Sweden) for supplying FICZ for these studies, and to Jong Min (Joseph) Baek and Dr Amanda Nouwens for timely assistance with metabolite identification by LC–MS. Thanks are also extended to Dr Roger Coulombe for donating the CYP1A5 plasmid and to Sabar Budiman and Dr Ian Ross for assistance with sequence validation. K.L.H., R.E.S.T., and G.F. were supported by Australian Postgraduate Research Training Awards.

### Author Contributions

E.M.J.G. conceived the project with input from Y.G. and K.L.H. K.L.H., R.E.S.T., Y.G., P.S., S.E.C.-P., and A.-S.S. performed the experiments and/or metabolite analysis.



K.L.H., R.E.S.T., Y.G., A.-S.S., U.J., and E.M.J.G. analyzed the results. G.F., T.J., U.J., S.A., and M.B. contributed specialist expertise and resources. K.L.H., R.E.S.T., and E.M.J.G. wrote the paper with input from all other authors.

## Data Availability

The data that support the findings of this study are available in the supporting information or from the corresponding authors upon reasonable request.

## References

- U.S. Food and Drug Administration: Center for Drug Evaluation and Research (CDER). 2016. *Safety testing of drug metabolites: guidance for industry*. Silver Spring (MD): Author.
- Aigrain L, Fatemi F, Frances O, Lescop E, Truan G. 2012. Dynamic control of electron transfers in diflavin reductases. *Int J Mol Sci*. **13**:15012–15041.
- Aoyama T, Korzekwa K, Nagata K, Gillette J, Gelboin HV, Gonzalez FJ. 1990. Estradiol metabolism by complementary deoxyribonucleic acid-expressed human cytochrome-P450s. *Endocrinology* **126**: 3101–3106.
- Begas E, Kouvaras E, Tsakalof AK, Bounitsi M, Asproдини EK. 2015. Development and validation of a reversed-phase HPLC method for CYP1A2 phenotyping by use of a caffeine metabolite ratio in saliva. *Biomed Chromatogr*. **29**:1657–1663.
- Bergander L. 2005. *Formation and metabolism of the tryptophan-derived 6-formylindolo[3,2-b]carbazole – a light-induced Ah-receptor ligand*. Stockholm: Stockholm University.
- Bergander L, Wahlstrom N, Alsberg T, Bergman J, Rannug A, Rannug U. 2003. Characterization of in vitro metabolites of the aryl hydrocarbon receptor ligand 6-formylindolo[3,2-b]carbazole by liquid chromatography–mass spectrometry and NMR. *Drug Metab Dispos*. **31**:233–241.
- Bergander L, Wincent E, Rannug A, Foroozesh M, Alworth W, Rannug U. 2004. Metabolic fate of the Ah receptor ligand 6-formylindolo[3,2-b]carbazole. *Chem Biol Interact*. **149**:151–164.
- Bloom J, Labthavikul S, Otey C, Arnold F. 2006. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A*. **103**:5869–5874.
- Buters JTM, Shou MG, Hardwick JP, Korzekwa KR, Gonzalez FJ. 1995. cDNA-Directed expression of human cytochrome-P450 CYP1A1 using baculovirus – purification, dependency on NADPH-P450 oxidoreductase, and reconstitution of catalytic properties without purification. *Drug Metab Dispos*. **23**:696–701.
- Copley SD. 2012. Toward a systems biology perspective on enzyme evolution. *J Biol Chem*. **287**:3–10.
- Cornelis MC, El-Sohemy A, Campos H. 2004. Genetic polymorphism of CYP1A2 increases the risk of myocardial infarction. *J Med Genet*. **41**:758–762.
- Crespi CL, Penman BW, Steimel TD, Smith T, Yang CS, Sutter TR. 1997. Development of a human lymphoblastoid cell line constitutively expressing human CYP1B1 cDNA: substrate specificity with model substrates and promutagens. *Mutagenesis* **12**:83–89.
- Denoeud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng C, Alberti A, Anthony F, Aprea G, et al. 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* **345**:1181–1184.
- Ekroos M, Sjögren T. 2006. Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc Natl Acad Sci U S A*. **103**: 13682–13687.
- Farmahin R, Crump D, O'Brien JM, Jones SP, Kennedy SW. 2016. Time-dependent transcriptomic and biochemical responses of 6-formylindolo[3,2-b]carbazole (FICZ) and 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) are explained by AHR activation time. *Biochem Pharmacol*. **115**:134–143.
- Foley GF. 2022. *Methods for ancestral sequence reconstruction of large and complex protein families* [PhD thesis]. Australia: The University of Queensland.
- Gaspar P, Moura G, Santos MA, Oliveira JL. 2013. mRNA secondary structure optimization using a correlated stem-loop prediction. *Nucleic Acids Res*. **41**:e73.
- Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**:343–345.
- Gillam EMJ, Baba T, Kim B-R, Ohmori S, Guengerich FP. 1993. Expression of modified human cytochrome P450 3A4 in *Escherichia coli* and purification and reconstitution of the enzyme. *Arch Biochem Biophys*. **305**:123–131.
- Gillam EMJ, Hunter DJB. 2007. Chemical defence and exploitation: biotransformation of xenobiotics by cytochrome P450 enzymes. In: Sigel A, Sigel H and Sigel RKO, editors. *Metal ions in life sciences: the ubiquitous roles of cytochrome P450 proteins*. John Wiley & Sons. p. 477–560.
- Goldstone JV, Goldstone HMH, Morrison AM, Tarrant A, Kern SE, Woodin BR, Stegeman JJ. 2007. Cytochrome P450 1 genes in early deuterostomes (tunicates and sea urchins) and vertebrates (chicken and frog): origin and diversification of the CYP1 gene family. *Mol Biol Evol*. **24**:2619–2631.
- Goldstone HMH, Stegeman JJ. 2006. A revised evolutionary history of the CYP1A subfamily: gene duplication, gene conversion, and positive selection. *J Mol Evol*. **62**:708–717.
- Gonzalez FJ, Nebert DW. 1990. Evolution of the P450 gene superfamily: animal-plant “warfare”, molecular drive and human genetic differences in drug oxidation. *Trends Genet*. **66**:164–168.
- Guengerich FP. 1994. Analysis and characterization of enzymes. In: Hayes AW, editors. *Principles and methods of toxicology*. New York: Raven Press, Ltd. p. 1259–1313.
- Guengerich FP. 2015. Human cytochrome P450 enzymes. In: Ortiz de Montellano PR, editors. *Cytochrome p450: structure, mechanism, and biochemistry*. Cham: Springer International Publishing. p. 523–785.
- Gumulya Y, Baek J-M, Wun S-J, Thomson RES, Harris KL, Hunter DJB, Behrendorff JBYH, Kulig J, Zheng S, Wu X, et al. 2018. Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat Catal*. **1**:878–888.
- Gumulya Y, Huang W, D’Cunha SA, Richards KE, Thomson RES, Hunter DJB, Baek J-M, Harris KL, Boden M, De Voss JJ, et al. 2019. Engineering thermostable CYP2D enzymes for biocatalysis using combinatorial libraries of ancestors for directed evolution (CLADE). *Chemcatchem* **11**:841–850.
- Guo HH, Choe J, Loeb LA. 2004. Protein tolerance to random amino acid change. *Proc Natl Acad Sci USA*. **101**:9205–9210.
- Guo Z, Gillam EMJ, Ohmori S, Tukey RH, Guengerich FP. 1994. Expression of modified human cytochrome P450 1A1 in *Escherichia coli*: effects of 5' substitution, stabilization, purification, spectral characterization, and catalytic properties. *Arch Biochem Biophys*. **312**:436–446.
- Harris KL. 2020. *Ancestral reconstruction of cytochrome P450 family 1, 4 and cytochrome P450 reductase: Insights into evolution and applications in biocatalysis* [PhD thesis]. Australia: The University of Queensland.
- Hart KM, Harms MJ, Schmidt BH, Elya C, Thornton JW, Marqusee S. 2014. Thermodynamic system drift in protein evolution. *PLoS Biol*. **12**:e1002091.
- Hayes CL, Spink DC, Spink BC, Cao JQ, Walker NJ, Sutter TR. 1996. 17 $\beta$ -estradiol hydroxylation catalysed by human cytochrome P450 1B1. *Proc Natl Acad Sci U S A*. **93**:9776–9781.
- Heinzelman P, Snow CD, Smith MA, Yu XL, Kannan A, Boulware K, Villalobos A, Govindarajan S, Minshull J, Arnold FH. 2009. Schema recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability. *J Biol Chem*. **284**:26229–26233.
- Hunter DJB, Behrendorff JBYH, Johnston WA, Hayes PY, Huang W, Bonn B, Hayes MA, De Voss JJ, Gillam EMJ. 2011. Facile

- production of minor metabolites for drug development using a CYP3A shuffled library. *Metab Eng*. **13**:682–693.
- Inoue H, Nojima H, Okayama H. 1990. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**:23–28.
- Johnston WA, Huang W, De Voss JJ, Hayes MA, Gillam EM. 2008. Quantitative whole-cell cytochrome P450 measurement suitable for high-throughput application. *J Biomol Screen*. **13**:135–141.
- Jönsson ME, Mattsson A, Shaik S, Brunstrom B. 2016. Toxicity and cytochrome P450 1A mRNA induction by 6-formylindolo [3,2-*b*]carbazole (FICZ) in chicken and Japanese quail embryos. *Comp Biochem Physiol C Toxicol Pharmacol*. **179**:125–136.
- Jönsson ME, Woodin BR, Stegeman JJ, Brunstrom B. 2011. Cytochrome P450 1 genes in birds: evolutionary relationships and transcription profiles in chicken and Japanese quail embryos. *PLoS One* **6**:e28257.
- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. **30**:3059–3066.
- Kawai YK, Ikenaka Y, Fujita S, Ishizuka M. 2010. The CYP1D subfamily of genes in mammals and other vertebrates. *Mamm Genome* **21**:320–329.
- Kerlan V, Dreano Y, Bercovici JP, Beaune PH, Floch HH, Berthou F. 1992. Nature of cytochromes P450 involved in the 2--/4-hydroxylations of estradiol in human liver-microsomes. *Biochem Pharmacol*. **44**:1745–1756.
- Kumar S, Tamura K, Nei M. 1994. Mega – molecular evolutionary genetics analysis software for microcomputers. *Comput Appl Biosci*. **10**:189–191.
- Lambrecht RW, Sinclair PR, Gorman N, Sinclair JF. 1992. Uroporphyrinogen oxidation catalyzed by reconstituted cytochrome P450IA2. *Arch Biochem Biophys*. **294**:504–510.
- Lee AJ, Cai MXX, Thomas PE, Conney AH, Zhu BT. 2003. Characterization of the oxidative metabolites of 17-beta-estradiol and estrone formed by 15 selectively expressed human cytochrome P450 isoforms. *Endocrinology* **144**:3382–3398.
- Li WZ, Godzik A. 2006. CD-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**:1658–1659.
- Liehr JG, Ulubelen AA, Strobel HW. 1986. Cytochrome P-450-mediated redox cycling of estrogens. *J Biol Chem*. **261**:16865–16870.
- Magiatis P, Pappas P, Gaitanis G, Mexia N, Melliou E, Galanou M, Vlachos C, Stathopoulou K, Skaltsounis AL, Marselos M, et al. 2013. Malassezia yeasts produce a collection of exceptionally potent activators of the Ah (dioxin) receptor detected in diseased human skin. *J Invest Dermatol*. **133**:2023–2030.
- Modarres HP, Mofrad MR, Sanati-Nezhad A. 2016. Protein thermostability engineering. *RSC Adv*. **6**:115252–115270.
- Morrison HG, Weil EJ, Karchner SI, Sogin ML, Stegeman JJ. 1998. Molecular cloning of CYP1A from the estuarine fish *Fundulus heteroclitus* and phylogenetic analysis of CYP1 genes: update with new sequences. *Comp Biochem Physiol C Toxicol Pharmacol*. **121**:231–240.
- Muslin EH, Clark SE, Henson CA. 2002. The effect of proline insertions on the thermostability of a barley alpha-glucosidase. *Protein Eng*. **15**:29–33.
- Nishida CR, Everett S, Ortiz de Montellano PR. 2013. Specificity determinants of CYP1B1 estradiol hydroxylation. *Mol Pharmacol*. **84**:451–458.
- Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. 1998. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese Cedar pollen, Cryj2, in *Escherichia coli*. *Appl Environ Microbiol*. **64**:1694–1699.
- Parikh A, Gillam EMJ, Guengerich FP. 1997. Drug metabolism by *Escherichia coli* expressing human cytochromes P450. *Nat Biotechnol*. **15**:784–788.
- Park SL, Justiniano R, Williams JD, Cabello CM, Qiao SX, Wondrak GT. 2015. The tryptophan-derived endogenous aryl hydrocarbon receptor ligand 6-formylindolo[3,2-*b*]carbazole is a nanomolar UVA photosensitizer in epidermal keratinocytes. *J Invest Dermatol*. **135**:1649–1658.
- Penman BW, Chen LP, Gelboin HV, Gonzalez FJ, Crespi CL. 1994. Development of a human lymphoblastoid cell-line constitutively expressing human CYP1A1 cDNA – substrate-specificity with model substrates and promutagens. *Carcinogenesis* **15**:1931–1937.
- Rannug A, Rannug U. 2018. The tryptophan derivative 6--formylindolo[3,2-*b*]carbazole, FICZ, a dynamic mediator of endogenous aryl hydrocarbon receptor signaling, balances cell growth and differentiation. *Crit Rev Toxicol*. **48**:555–574.
- Rannug A, Rannug U, Rosenkranz HS, Winqvist L, Westerholm R, Agurell E, Grafstrom AK. 1987. Certain photooxidized derivatives of tryptophan bind with very high-affinity to the Ah receptor and are likely to be endogenous signal substances. *J Biol Chem*. **262**:15422–15427.
- Rannug U, Rannug A, Sjoberg U, Li H, Westerholm R, Bergman J. 1995. Structure elucidation of two tryptophan-derived, high affinity Ah receptor ligands. *Chem Biol*. **2**:841–845.
- Rendic S, Guengerich FP. 2012. Contributions of human enzymes in carcinogen metabolism. *Chem Res Toxicol*. **25**:1316–1383.
- Rendic S, Guengerich FP. 2015. Survey of human oxidoreductases and cytochrome P450 enzymes involved in the metabolism of xenobiotic and natural chemicals. *Chem Res Toxicol*. **28**:38–42.
- Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. 2013. Hyperstability and substrate promiscuity in laboratory resurrections of precambrian  $\beta$ -lactamases. *J Am Chem Soc*. **135**:2899–2902.
- Rodenburg EM, Eijgelsheim M, Geleijnse JM, Amin N, van Duijn CM, Hofman A, Uitterlinden AG, Stricker BH, Visser LE. 2012. CYP1A2 and coffee intake and the modifying effect of sex, age, and smoking. *Am J Clin Nutr*. **96**:182–187.
- Ross CM, Foley G, Bodén M, Gillam EMJ. 2022. Using the evolutionary history of proteins to engineer insertion-deletion mutants from robust, ancestral templates using graphical representation of ancestral sequence predictions (GRASP). In: Magnani F, Marabelli C, Paradisi F, editors. *Enzyme engineering: methods and protocols*. New York (NY): Springer US. p. 85–110
- Sakaki T, Yamamoto K, Ikushiro S. 2013. Possibility of application of cytochrome P450 to bioremediation of dioxins. *Biotechnol Appl Biochem*. **60**:65–70.
- Sandhu P, Guo Z, Baba T, Martin MV, Tukey RH, Guengerich FP. 1994. Expression of modified human cytochrome P450 1A2 in *Escherichia coli*: stabilization, purification, spectral characterization, and catalytic activities of the enzyme. *Arch Biochem Biophys*. **309**:168–177.
- Schallreuter KU, Salem M, Gibbons NCJ, Maitland DJ, Marsch E, Elwary SMA, Healey AR. 2012. Blunted epidermal L-tryptophan metabolism in vitiligo affects immune response and ROS scavenging by Fenton chemistry, part 2: epidermal H2O2/ONOO-mediated stress in vitiligo hampers indoleamine 2,3-dioxygenase and aryl hydrocarbon receptor-mediated immune response signaling. *FASEB J*. **26**:2471–2485.
- Schwarz D, Kisselev P, Schunck WH, Chernogolov A, Boidol W, Cascorbi I, Roots I. 2000. Allelic variants of human cytochrome P450 1A1 (CYP1A1): effect of T461N and I462V substitutions on steroid hydroxylase specificity. *Pharmacogenetics* **10**:519–530.
- Scornaienchi ML, Thornton C, Willett KL, Wilson JY. 2010. Cytochrome P450-mediated 17 beta-estradiol metabolism in zebrafish (*Danio rerio*). *J Endocrinol*. **206**:317–325.
- Shimada T, Gillam EMJ, Oda Y, Tsumura F, Sutter TR, Guengerich FP, Inoue K. 1999. Metabolism of benzo[a]pyrene to trans-7,8-dihydroxy-7,8-dihydrobenzo[a]pyrene by recombinant human cytochrome P4501B1 and purified liver epoxide hydroxylase. *Chem Res Toxicol*. **12**:623–629.

- Shimada T, Gillam EMJ, Sutter TR, Strickland PT, Guengerich FP, Yamazaki H. 1997. Oxidation of xenobiotics by recombinant human cytochrome P450 1B1. *Drug Metab Dispos.* **25**:617–622.
- Shimada T, Hayes CL, Yamazaki H, Amin S, Hecht SS, Guengerich FP, Sutter TR. 1996. Activation of chemically diverse procarcinogens by human cytochrome P-450 1B1. *Cancer Res.* **56**:2979–2784.
- Shimada T, Oda Y, Gillam EMJ, Guengerich FP, Inoue K. 2001. Metabolic activation of polycyclic aromatic hydrocarbons and other procarcinogens by cytochromes P450 1A1 and P4501B1 allelic variants and other human cytochromes P450 in *Salmonella typhimurium* NM2009. *Drug Metab Dispos.* **29**:1176–1182.
- Shimada T, Wunsch RM, Hanna IH, Sutter TR, Guengerich FP, Gillam EMJ. 1998. Recombinant human cytochrome P450 1B1 expression in *Escherichia coli*. *Arch Biochem Biophys.* **357**:111–120.
- Siddiqui KS. 2017. Defying the activity-stability trade-off in enzymes: taking advantage of entropy to enhance activity and thermostability. *Crit Rev Biotechnol.* **37**:309–322.
- Smirnova A, Wincent E, Bergander LV, Alsberg T, Bergman J, Rannug A, Rannug U. 2016. Evidence for new light-independent pathways for generation of the endogenous aryl hydrocarbon receptor agonist FICZ. *Chem Res Toxicol.* **29**:75–86.
- Spink DC, Eugster HP, Lincoln DW, Schuetz JD, Schuetz EG, Johnson JA, Kaminsky LS, Gierthy JF. 1992. 17 $\beta$ -Estradiol hydroxylation catalyzed by human cytochrome P450 1A1 – a comparison of the activities induced by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in MCF-7 cells with those from heterologous expression of the cDNA. *Arch Biochem Biophys.* **293**:342–348.
- Stegeman JJ, Behrendt L, Woodin BR, Kubota A, Lemaire B, Pompon D, Goldstone JV, Urban P. 2015. Functional characterization of zebrafish cytochrome P450 1 family proteins expressed in yeast. *Biochim Biophys Acta.* **1850**:2340–2352.
- Stoilov I, Akarsu AN, Alozie I, Child A, Barsoum-Homsy M, Turacli ME, Or M, Lewis RA, Ozdemir N, Brice G, et al. 1998. Sequence analysis and homology modeling suggest that primary congenital glaucoma on 2p21 results from mutations disrupting either the hinge region or the conserved core structures of cytochrome P4501B1. *Am J Hum Genet.* **62**:573–584.
- Stoilov I, Akarsu AN, Sarfarazi M. 1997. Identification of three different truncating mutations in cytochrome P4501B1 (CYP1B1) as the principal cause of primary congenital glaucoma (Buphthalmos) in families linked to the GLC3A locus on chromosome 2p21. *Hum Mol Genet.* **6**:641–647.
- Suchar LA, Chang RL, Rosen RT, Lech J, Conney AH. 1995. High-performance liquid chromatography separation of hydroxylated estradiol metabolites: formation of estradiol metabolites by liver microsomes from male and female rats. *J Pharmacol Exp Therap.* **272**:197–206.
- Tartèse R, Chaussidon M, Gurenko A, Delarue F, Robert F. 2017. Warm Archean oceans reconstructed from oxygen isotope composition of early-life remnants. *Geochem Perspect Lett.* **3**: 55–65.
- Thomson RES. 2021. *Structural and functional characterisation of ancestral cytochromes P450 from family 2 in tetrapods* [PhD thesis]. Australia: The University of Queensland.
- Toth-Petroczy A, Tawfik DS. 2014. The robustness and innovability of protein folds. *Curr Opin Struct Biol.* **26**:131–138.
- Uno S, Dalton TP, Derkenne S, Curran CP, Miller ML, Shertzer HG, Nebert DW. 2004. Oral exposure to benzo a pyrene in the mouse: detoxication by inducible cytochrome P450 is more important than metabolic activation. *Mol Pharmacol.* **65**: 1225–1237.
- Vieille C, Zeikus GJ. 2001. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev.* **65**:1–43.
- von Wachenfeldt C, Richardson TH, Cosme J, Johnson EF. 1997. Microsomal P450 2C3 is expressed as a soluble dimer in *Escherichia coli* following modifications of its N-terminus. *Arch Biochem Biophys.* **339**:107–114.
- Waxman DJ, Lapenson DP, Aoyama T, Gelboin HV, Gonzalez FJ, Korzekwa K. 1991. Steroid-hormone hydroxylase specificities of 11 cDNA-expressed human cytochrome-P450s. *Arch Biochem Biophys.* **290**:160–166.
- Wheeler LC, Lim SA, Marqusee S, Harms MJ. 2016. The thermostability and specificity of ancient proteins. *Curr Opin Struct Biol.* **38**: 37–43.
- Williams PD, Pollock DD, Blackburne BP, Goldstein RA. 2006. Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput Biol.* **2**:e69.
- Wincent E, Amini N, Luecke S, Glatt H, Bergman J, Crescenzi C, Rannug A, Rannug U. 2009. The suggested physiologic aryl hydrocarbon receptor activator and cytochrome P4501 substrate 6-formylindolo[3,2-*b*]carbazole is present in humans. *J Biol Chem.* **284**:2690–2696.
- Wincent E, Kubota A, Timme-Laragy A, Jönsson ME, Hahn ME, Stegeman JJ. 2016. Biological effects of 6-formylindolo[3,2-*b*]carbazole (FICZ) in vivo are enhanced by loss of CYP1A function in an Ahr2-dependent manner. *Biochem Pharmacol.* **110–111**:117–129.
- Yamazaki H, Nakamura M, Komatsu T, Ohyama K, Hatanaka N, Asahi S, Shimada N, Guengerich FP, Shimada T, Nakajima M, et al. 2002. Roles of NADPH-P450 reductase and apo- and holo-cytochrome b5 on xenobiotic oxidations catalyzed by 12 recombinant human cytochrome P450s expressed in membranes of *Escherichia coli*. *Protein Exp Purificat.* **24**:329–337.
- Yip SSM, Coulombe RA. 2006. Molecular cloning and expression of a novel cytochrome P450 from turkey liver with aflatoxin B<sub>1</sub> oxidizing activity. *Chem Res Toxicol.* **19**:30–37.
- Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, Dirks RM, Pierce NA. 2011. Nupack: analysis and design of nucleic acid systems. *J Comput Chem.* **32**:170–173.
- Zhang CY, Creech KL, Zuercher WJ, Willson TM. 2019. Gram-scale synthesis of FICZ, a photoreactive endogenous ligand of the aryl hydrocarbon receptor. *Sci Rep.* **9**:9.